A rapid food chain approach for authenticity screening: The development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices

McVey, C., McGrath, T. F., Haughey, S. A., & Elliott, C. T. (2021). A rapid food chain approach for authenticity screening: The development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices. *Talanta*, *222*, Article 121533. https://doi.org/10.1016/j.talanta.2020.121533

Published in:
Talanta

Document Version:
Peer reviewed version

A rapid food chain approach for authenticity screening: the development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices

Claire McVey*, Terry F. McGrath, Simon A. Haughey, Christopher T. Elliott

Institute for Global Food Security, ASSET Technology Centre, School of Biological Sciences, Queen's University Belfast, Northern Ireland, United Kingdom

*Institute for Global Food Security, Queen's University Belfast, 19 Chlorine Gardens, Belfast, BT9 5DL, Northern Ireland; Phone: +442890974288; Email: c.mcvey@qub.ac.uk

ABSTRACT

This study assesses the application of a handheld, near infrared spectroscopy (NIRS) device, namely the NeoSpectra Micro, for the determination of oregano authenticity. Utilising a large sample set of oregano (n = 295) and potential adulterants of oregano (n = 109), models were developed and validated using SIMCA 15 software. The models demonstrated excellent predictability for the determination of authentic oregano and adulterant samples. The optimal model resulted in a 93.0% and 97.5% correct prediction for oregano and adulterants, respectively. Different standardisation approaches were assessed to determine model transferability to a second NIRS device. In the case of the second device, the best predictions were achieved with data that had not undergone any spectral standardisation (raw). Subsequently, the optimal model was able to correctly predict 90% of authentic oregano samples and 100% of the adulterant samples on the second device. This study demonstrates the potential of the device to be used as a simple, cost effective, reliable and handheld screening tool for the determination of oregano authenticity, at various stages of the food supply chain. It is believed that such forms of monitoring could be highly beneficial in other areas of food authenticity analysis to help combat the negative economical and health implications of food fraud.

Keywords: Near-infrared spectroscopy, food fraud, handheld, chemometrics, model transferability

1. Introduction

The complex and global nature of food supply chains has highlighted a need to provide cost effective and rapid analytical approaches to ensure food authenticity [1]. This has become more evident with the current coronavirus (COVID-19) pandemic that has led to the major disruption of food supply chains, increased demand for food, diminished levels of surveillance and, potentially may lead to increased food prices in the future [2]. This volatility within the food system provides the perfect environment for fraudulent activity, and this has already been evidenced by the seizure of two shipments of horsemeat destined for the EU markets, reminiscent of the 2013 horsemeat scandal [3]. This volatility is likely to intensify with new trading relationships developing and a lack of robust system in place to gather intelligence on potential food fraud risks or incidences [4]. Both the COVID-19 crisis and continuing issues caused by climate change have the potential to leave the food supply chain even more vulnerable to fraudulent activity. To help mitigate against these risks, a testing approach that can be conducted easily and effectively at various stages of the supply chain, to ensure the quality, safety and authenticity of the food we consume, is needed.

Herbs and spices have been identified as one of the most widely adulterated food commodities globally [5]. This is, in part, due to the high intrinsic value of herbs and spices, which is estimated to reach USD $25 billion by

2025 [6]. The complex herb and spice supply chain incorporates small-scale farmers/ growers and intertwines many processors and traders worldwide [7]. Most herbs and spices are grown and processed in the country of origin and consequently, it is this highly susceptible ground or chopped form that makes its way along the food chain [8]. At each point of this long and complex chain, various fraudulent activities can be perpetrated, such as artificial enhancement, dilution and substitution, removal of authentic constituents and mislabelling [5].

This chain can become even more complex and elongated depending on the end product [9]; herbs and spices can be used as a commodity in the domestic kitchen (usually in ground form) or further processed to produce prepared meals and added-value meats [10]. Thus, the product can potentially go through many processing steps, with various other ingredients added before it reaches the consumer. For a consumer purchasing herbs or spices to use in the home, it is very likely that adulteration will be detected. Apart from the economic and quality implications for the consumer, there may also be negative effects on health, such as allergic reactions or stomach problems, if an adulterated product is consumed [5]. Secondary processors are also impacted by non-authentic products, leading to devastating economic implications for legitimate food business operators within the food chain. The potential implications of herb and spice fraud are poor organoleptic properties, which can have a secondary impact on product recipes in terms of product quality and consistency (Oakes, D – personal communication), and public health threat, for e.g. allergens, resulting in the loss of reputation and consumer trust derived from a fraud incident [5]. Adulterated products can travel through the food chain making it difficult to control and mitigate. Many food business operators will have evaluated their own food chain by assessing the vulnerabilities within it [11] and subsequently designed and implemented a mitigation strategy [10]. However, the difficulty associated with validating this mitigation strategy via routine analytical testing, is a serious limitation within quality management procedures [12].

Widely used techniques such as mass spectrometry [13,14] and DNA analysis [15] are well-established, confirmatory methods used to determine food authenticity. These methods are highly sensitive and specific, but due to their high cost and labour-intensive methodology, they are not conducive to implementation in routine monitoring systems [16]. These limitations have initiated an interest in Near Infrared spectroscopy (NIRS) approaches for the determination of food authenticity. This is due to its non-destructive nature, miniaturization of optical components making it portable and/ handheld and low cost and robust components which make it conducive as a screening tool at various points of the food supply chain [17]. Recent research has focused on the application of portable NIRS devices for the determination of authenticity of cod [18], South African honey [19], veal adulterated with pork [20] and the determination of melamine in soya-bean meal [21]. To the best of the author's knowledge, only one publication has outlined the application of a portable NIRS device for the determination of authenticity of herbs and spices [17]. No methods have addressed the need to assess several handheld devices for the determination of herb and spice authenticity at various points of the food supply chain. It is evident that there is a need for robust screening methods that are cost effective and easily used on-site (or handheld), to ensure the authenticity of herbs and spices and products thereof, at several stages of the food supply chain.

A major challenge in developing a NIRS method is the need for the development of classification models to interpret the spectra derived from complex vibrational overtone and combination bands. Previously, when multiple devices were required for use at various stages of the food chain, the development of classification models for each individual device were required to compensate for the inherent instrument-instrument variation. Due to the costly and time-consuming limitations associated with developing models for many devices, model transfer has become very desirable. There are two approaches frequently undertaken: (1) univariate slope and bias correction of the calibration set, and (2) correction of the secondary spectra to mimic that derived from the initial device where the model was developed [22-24]. The second approach is more frequently undertaken and uses spectral standardisation to correct spectra derived from a secondary device. The aim is to transform the spectra of samples assessed on a secondary device, so that they correspond to the spectra derived from a primary device, which was used in the development of the model [25]. Various spectral standardisation approaches have been used, including direct standardisation (DS) [26], piecewise direct standardisation (PDS) [27], orthogonal signal correction (OSC) [28] and wavelet hybrid direct standardisation (WHDS) [29], etc. Each standardisation

approach requires the measurement of a group of standard samples (transfer samples) by both the primary and secondary device, under the same experimental conditions. A mathematical correction is calculated from the data, and this can then be applied to samples analysed on the secondary device [30]. The optimal standardisation approach will ultimately depend on the complexity of the instrument differences.

Herein, the aim of this work was to assess the capabilities of a hand-held NIRS device for the determination of authentic oregano from oregano that has been adulterated. Oregano was chosen to assess the applicability of the method due to its global use and thus inherent vulnerability throughout the food chain [13,14]. Chemometric models were developed and externally validated. A second NeoSpectra device was assessed to determine if the models developed using spectra from the primary device could be used to determine the authenticity of samples scanned on the secondary device. To assess the transferability of the models, various standardisation approaches were undertaken and a correction factor applied to the spectra obtained from a second device.

2. Materials and methods

2.1. Samples

Authentic oregano samples were obtained from different sources and origins, with full provenance and traceability (n = 295). A number of potential adulterants were identified, including sumac, cistus, myrtle and olive (n = 109). All samples were in dried leaf, chopped form and stored in sealed bags at ambient temperature, away from direct light. No further preparation was required prior to analysis.

2.2. Spectra acquisition

A NeoSpectra Micro development kit (SW-001; SWS62231) was obtained from Si-ware Systems (Cairo, Egypt). The NeoSpectra Micro consists of a monolithic micro-electromechanical system (MEMS) Michelson interferometer, a single element photodetector and an electronic driver board. For the purposes of this evaluation, the SpectraMOST software was controlled via connection to a PC. However, the device has the potential to act as a host on its own through the Raspberry Pi board and Bluetooth connectivity, enabling wireless connection to mobile phone / tablet application. Thus, the device has the potential to be utilised on-site and is portable due to its small dimensions (6 (L) x 3 (W) x 4 (H) cm inclusive of Raspberry Pi board). All spectra were collected in diffuse reflectance, over a spectral range of 1350 nm – 2500 nm. The optimal conditions were as follows: 2 second scan time and a default optical gain, which was derived from a background scan. The device was left to warm up for 10 minutes prior to analysis by running in continuous mode. Prior to the first measurement, a background measurement was collected using a Spectralon (99% reflectance) and this was subsequently updated every 10 minutes thereafter. Samples were analysed by continuously rotating a filled petri dish over the collection window. A second NeoSpectra Micro device (SW-002) was used to assess multivariate model transfer under the same parameters.

2.3. Data pre-processing and analysis

Three spectra were collected for each sample and averaged prior to data analysis. The data set was split into a reference set that consisted of 209 samples and a validation set of 195 samples, which were not used in the creation of the model [31]. The samples represented oregano (87%), olive (5%), myrtle (3%), cistus (2.5%) and sumac (1.5%) in the reference set and oregano (58%) or adulterants (42%) in the validation set. The adulterant samples in the validation set were previously tested using a confirmatory mass spectrometry method and determined to be up to 80% adulterated. Pareto scaling was used throughout, which uses the square root of the standard deviation as the scaling factor, allowing for a reduction in the relative significance of large values, whilst at the same time maintaining the structure of the data [32]. Different pre-processing strategies were assessed; data was pre-treated with standard normal variate (SNV), first order derivatives (1DER), Savitzky-Golay (SG) or a

mixture thereof, to remove irrelevant light scatter from the spectra.

Chemometric analysis was carried out using SIMCA 15 software (SIMCA, Sartorius, Sweden). Classification models were built in SIMCA using Partial Least Squares Discriminant Analysis (PLS-DA) or orthogonal PLS-DA (OPLS-DA). The models were validated internally via cross-validation, with 1/7th removed and used as the test set and remaining 6/7th used to build the model. The prediction error was calculated for each validation set and this procedure was repeated for all subsets. The prediction accuracy was also validated using the external validation set (data not used in model development).

Multivariate model transfer was assessed to determine if the model developed on SW-001 (primary) was capable of successfully predicting samples analysed on a second device (SW-002). Several mathematical procedures were assessed using data obtained from both devices. To determine the optimal data standardisation required, twenty scans of four different samples (two oregano and two adulterant samples) were chosen at random. Direct standardisation (DS) and Piecewise direct standardisation (PDS) were applied to determine the mathematical relationship between the data obtained from the two devices, and thus, assess its suitability to the model transfer process [33]. DS uses the whole spectrum measured on the primary device to fit each spectral point of the secondary spectra. In contrast, PDS utilises several spectral measurements in a small window to reconstruct the spectral point on the secondary instrument [34,35]. The optimal number of spectral points (window) when constructing the transformation matrix was assessed. DS and PDS was applied using MATLAB (R2019a, MathWorks Inc.) with PLS Toolbox (Version 8.7.1, Eigenvector Research Inc.). Using this information, a factor was determined and applied to a validation set of 40 samples analysed on SW-002 (secondary device) and the data was uploaded to SIMCA 15 software. A simple baseline correction, which used subtraction or division, was also assessed using the combined spectral differences of the four randomly chosen samples, scanned twenty times on SW-001 and SW-002, and the data was inputted into SIMCA 15 software to determine the compatibility of the model transfer. **Fig. 1** presents an overview of how model development and validation was performed on SW-001, and the process of determining the optimal standardisation approach for model transfer to SW-002.

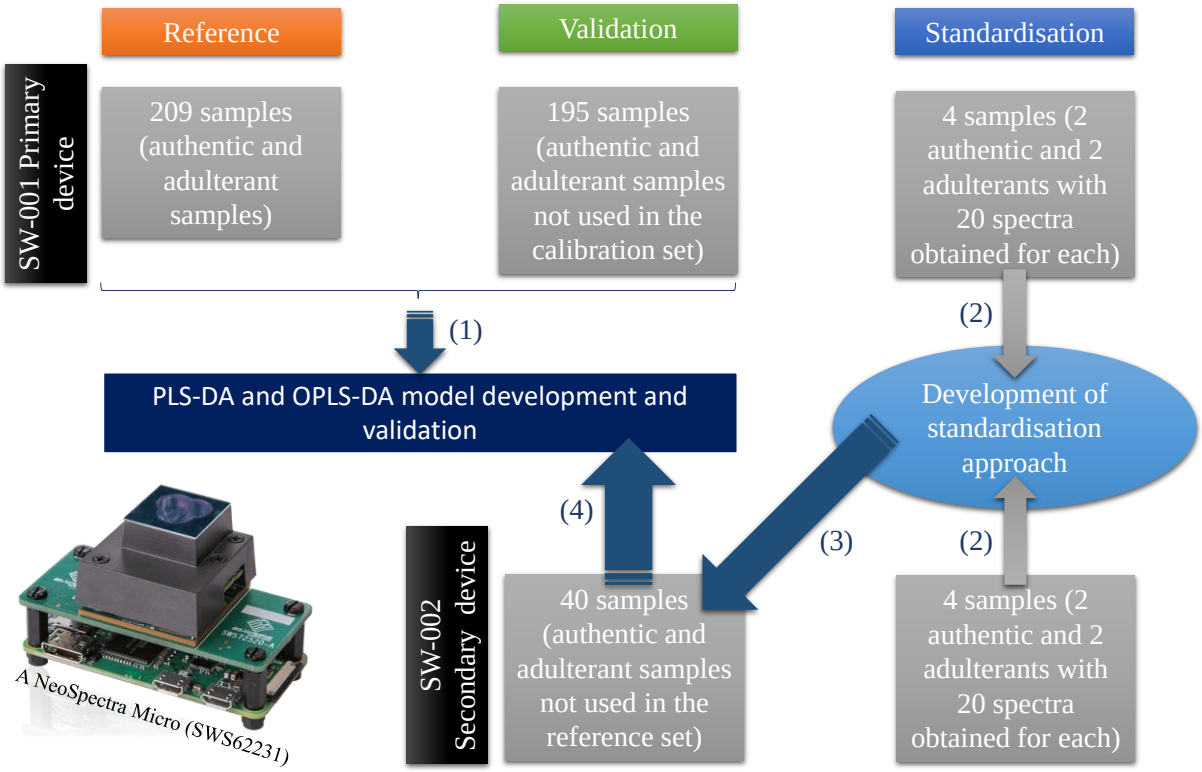

Fig. 1. The workflow for model development and transfer. (1) PLS-DA and OPLS-DA model development and validation using the spectra from the NeoSpectra primary device (SW-001) to predict the authenticity of oregano and adulterant samples; (2) Various standardisation approaches were developed using the spectra of four transfer samples (20 spectra for each sample) analysed on both the primary and secondary device. (3) The standardisation methods were applied to 40 validation set samples that were analysed on the secondary device (SW-002). (4) The standardised spectra derived from the secondary device (SW-002) was assessed against models developed on SW-001 (primary). ***In color for print***

3. Results and discussion

3.1. NIR spectra of oregano and adulterants

The average raw spectra of oregano and adulterants (cistus, myrtle, olive and sumac) collected using the NeoSpectra Micro NIRS (SW-001; primary) are shown in **Fig. 2**. Overtone and combination bands associated with the absorption of NIR radiation by organic molecules are primarily of O-H, C-H, N-H and C=O groups, of which their fundamental molecular stretching are weak in intensity and usually overlapping [36]. Thus, NIR spectra are complex, limiting their usefulness for direct identification purposes. This is evident from the spectral information shown in **Fig. 2**. In general, the collected spectra were characterised by absorption bands at 1650 nm, 1850 nm, 2000 nm, 2200 nm and 2400 nm. Due to the negligible differences between the spectra of authentic oregano and the adulterants, the development of a chemometric approach was necessary to determine the

feasibility of the device for authentication purposes.

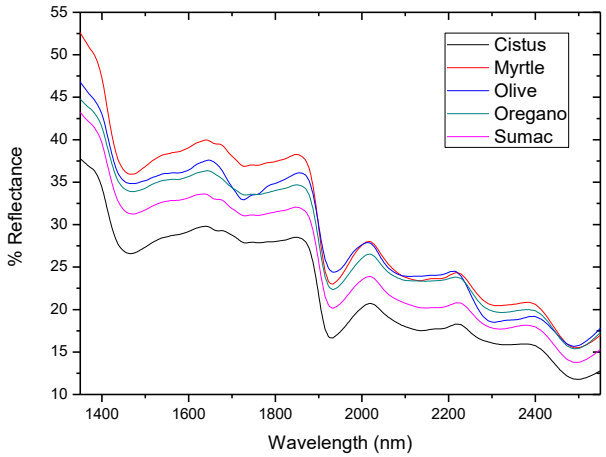

Fig. 2. The average raw NIR spectra (prior to the application of data pre-processing) for cistus, myrtle, olive, oregano and sumac. ***In color for print***

3.2. Model development and validation

Initially, principal component analysis (PCA) was carried out to assess the reliability of the data in generating a robust model focusing on maximum variance. As PCA is an unsupervised technique, the classification does not necessarily relate to the principal component patterns. In this way, it is not prone to overfitting, which is regularly seen with PLS-DA analysis, due to the influence of projecting the data in relation to the best separation of classification groups [37]. **Fig. 3A and 3B** demonstrates the PCA score plot of a multiclass model and a two-class model, respectively. The separation of oregano and adulterated samples is evident in both plots, where clear groupings are observed. The similarities between the multiclass and two-class models, in terms of the specific groupings, particularly within the adulterant group, demonstrates the robustness of the data and good specificity, and therefore warranted further investigation into the predictability of a supervised model.

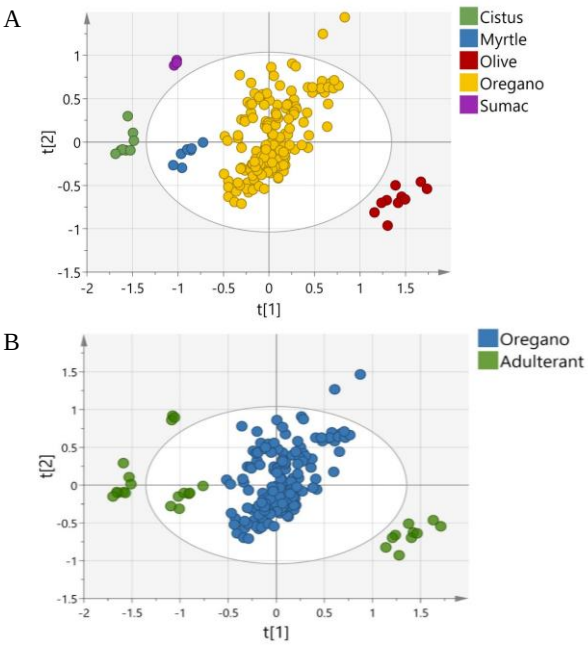

Fig. 3. Unsupervised PCA score plots of the NIRS data; (A) multiclass model and (B) two-class model. All data was pre-processed with SNV and 1DER. ***In color for print***

PLS-DA and OPLS-DA supervised models for the identification of oregano and adulterants were developed using spectral data obtained from analysis using the primary device. Various pre-processing techniques were applied to assess the predictability of the models (**Table 1**). To determine the sensitivity and specificity of each model it was necessary to set a 'cut off' value or limit, over which the sample was deemed authentic oregano. To do this, the predicted values obtained from SIMCA 15 software during the classification, with respect to all the classes within each model, was used. Practically, the larger the predicted value, the greater the probability that the sample belonged to that class. To determine a cut-off value for the whole test set, the highest prediction value, which encompassed the maximum correct predictions (authentic oregano and adulterant samples), was calculated. The cut-off value was calculated for each model separately. With the cut-off value set, the majority of the models give over 90% correct predictions for oregano and the adulterated samples. In relation to the multiclass models, all models demonstrated good predictability, apart from the OPLS-DA model built using raw data (M2). In this case, the prediction classification was 72.8% and 60.5% for the adulterants and oregano, respectively. This correlates with the R2 and Q2 values obtained for the model (0.375 and 0.301, respectively). These values indicate that the model can explain only a small amount of variation and that the predictability is low. This could be due to the effects of noisy or irrelevant data that had not underwent any smoothing or other pre-processing.

The best multiclass model was an OPLS-DA (M8), developed using data transformed with SNV, 1DER and SG. This gave correct predictions of 95.1% and 91.2% for adulterated samples and authentic oregano, respectively. This model also demonstrated a high R2 and Q2 value of 0.90 and 0.86, respectively, indicating a high level of explained variation and good predictability in cross validation. The PLS-DA multiclass model developed with raw data also give high correct predictions of 93.8% for the adulterants and 93.9% for authentic oregano (M1). Interestingly, good predictions were also obtained for the two-class models, including M9 and M10 that were built using raw data. In particular, the OPLS-DA model developed with data transformed with SNV, 1DER and SG give correct predictions of 97.5% and 93.0% for adulterant samples and authentic oregano, respectively (M16). This model provided slightly improved R2 and Q2 values over its multiclass counterpart of 0.97 and 0.99, respectively. As this is a screening approach, the aim is to detect as many adulterated samples as possible. Therefore, when a sample is correctly or incorrectly predicted as being adulterated, the samples will be further assessed via confirmatory analysis [13]. As M16 correctly detected the highest percentage of adulterated samples, it was determined to be the optimal model. The best multiclass and two-class models are shown in **Fig. 4**. As a further indication of model performance and quality, Receiver Operating Characteristic (ROC) curves were created and the Area Under the Curve (AUC) was calculated for the 2-class models. An AUC value of 0.994 was determined for the best 2-class model, further demonstrating its applicability for the determination of oregano authenticity.

Table 1. Predictability of the multiclass and 2-class models determined using an external validation set (n = 195).

Model	IID	Type	Pre-processing	Predictability	
				Adulterant	Oregano
Multiclass	M1	PLS-DA	None	93.8%	93.9%
	M2	OPLS-DA	None	72.8%	60.5%
	M3	PLS-DA	SNV	87.5%	93.0%
	M4	OPLS-DA	SNV	93.8%	93.0%
	M5	PLS-DA	SNV + 1DER	86.4%	97.4%
	M6	OPLS-DA	SNV + 1DER	92.6%	95.6%
	M7	PLS-DA	SNV + 1DER +SG	93.8%	93.0%
	M8	OPLS-DA	SNV + 1DER +SG	95.1%	91.2%
2-Class	M9	PLS-DA	None	95.1%	93.9%
	M10	OPLS-DA	None	95.1%	93.9%
	M11	PLS-DA	SNV	93.8%	93.0%
	M12	OPLS-DA	SNV	93.8%	92.1%
	M13	PLS-DA	SNV + 1DER	92.6%	90.4%
	M14	OPLS-DA	SNV + 1DER	91.4%	94.7%
	M15	PLS-DA	SNV + 1DER +SG	92.6%	95.6%
	M16	OPLS-DA	SNV + 1DER +SG	97.5%	93.0%

SNV; standard normal variate, 1DER; first order derivatives and SG; Savitzky–Golay

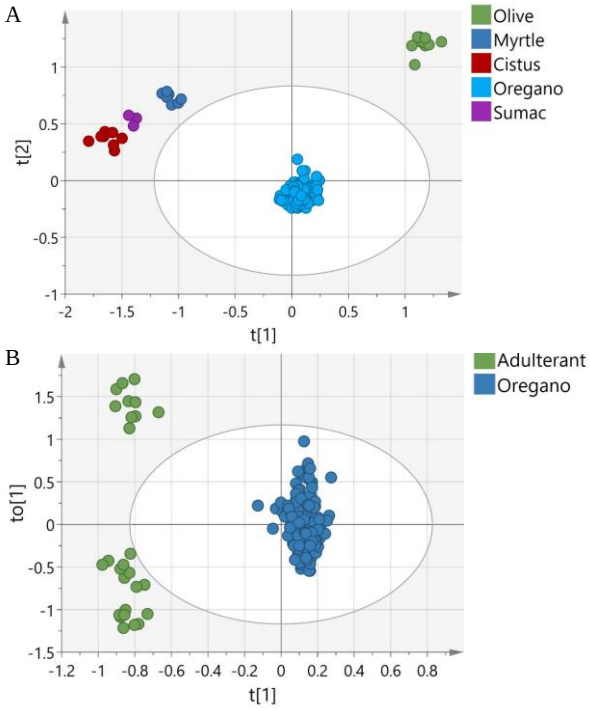

Fig. 4. Supervised OPLS-DA score plots showing SNV, 1DER and SG transformed NIRS data, which resulted in the highest correct predictions. (A) Multiclass model (R2 = 0.9; Q2 = 0.86) and (B) 2-class model (R2 = 0.97; Q2 = 0.99). ***In color for print***

3.3. Model transferability

The development and validation of a robust, handheld analytical approach for use as a screening tool for food authenticity has limited application unless that method can be transferred to many different devices, allowing for their widespread application throughout the supply chains. To demonstrate the potential of model transferability on the NeoSpectra Micro Development kit, in combination with chemometric modelling, a series of experiments were carried out to assess different standardisation approaches. Standardisation of spectral data derived from the secondary device allows for the use of the primary reference set, without the need for these samples to be analysed and a model created on the second device. To calculated the appropriate correction factor for standardisation, two oregano and two adulterated transfer samples were chosen using a random generator on excel. To remove noise, twenty spectra were measured for each sample, using the primary and secondary devices. The mean spectra, prior to any pre-processing (raw), are shown in **Fig. 5**. There are noticeable differences in the intensity, and slight differences in wavelength shift between the spectra from the two devices. However, the characteristic peaks at 1650 nm, 1850 nm, 2000 nm, 2200 nm and 2400 nm, discussed in section 3.1, remain generally consistent. This would suggest that no standardisation or a simplified approach might prove optimal for the model transfer to the secondary device.

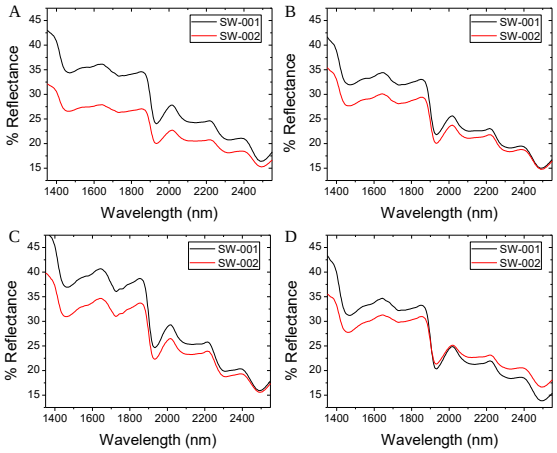

Fig. 5. The mean NIR spectrum of oregano and adulterated transfer samples obtained using the primary (SW-001, black line) and secondary (SW-002, red line) NeoSpectra devices prior to the application of any standardisation methods. (A), (B), (C) and (D) denotes the average spectra for each of the four transfer samples. ***In color for print***

Using the spectral information of the four transfer samples obtained from the two devices, different standardisation approaches were assessed (**Fig. 6**). **Fig. 6A** demonstrates the average spectra of a validation sample obtained on both devices, without standardisation applied. **Fig. 6B-F** displays the raw spectra measured on the primary device (black line), compared to the standardised spectra measured on the secondary device (red line). Again, it is noted that although the spectra may differ in intensity, the characteristic peaks remain identical over all types of standardisation, apart from PDS with a window of 91 (**Fig. 6C**). Visible shifts in the wavelength is evident in **Fig. 6C**, this is most likely because of the large number of spectral points, which make up the transformation matrix. The inclusion of too many wavelengths for standardisation results in high noise, which dominates the spectra [38]. From the analysis of the data, it is difficult to determine the appropriate method, thus, to demonstrate the best model transfer approach, a wider selection of the reference set was used to assess the prediction capabilities of the models developed on the primary device.

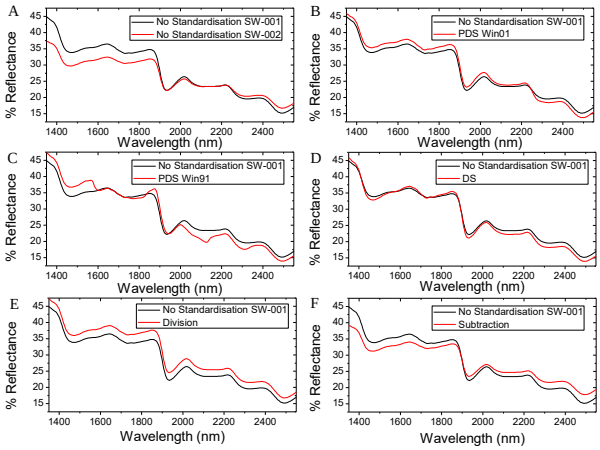

Fig. 6. Effect of standardisation on the average spectra of twenty oregano samples from the validation set assessed on SW-002 (secondary) device. For comparative purposes, the average spectra with no standardisation applied has been shown for SW-001 (primary, black spectra). Standardisation procedures that were applied based on the correction factors calculated using the transfer samples included; (A) no standardisation, (B) Piecewise direct standardisation (PDS) with a window of 1, (C) PDS with a window of 91, (D) direct standardisation (DS), (E) division and (F) subtraction standardisation. ***In color for print***

The model transfer results for predicting 40 validation set samples chosen randomly are summarized in **Table 2**. The results demonstrate varying degrees of predictability, over the model type, pre-processing applied and standardisation approach. Interestingly, the highest correct predictions were obtained using the spectral data measured on SW-002 with no standardisation applied (M5, M11 and M12). M11, a PLS-DA 2-class model, had the highest correct predictions of 90% and 100% for authentic oregano and the adulterated samples, respectively, when measured on the secondary device. This approach has the potential to be successfully used in the model transfer, however it remains to be seen if this result would be the same for a third device. The best multiclass model correctly predicted 70% of the authentic oregano and 100% of the adulterants (M5). The predicted results are poor for the authentic oregano samples when no standardisation is applied and the spectra from SW-002 is tested against M1-4, M6-10 and M13-16 developed on the primary device. The influence of spectral interferences and over-working of the data has most likely had a negative impact on the model predictability.

In general, DS and PDS transformation of the data derived from the secondary device was found to favour the prediction of either oregano or adulterated samples when applied to the reference models. The best predictions for the DS method was determined using the PLS-DA multiclass model (M5). Correct predictions of 60% and 45% was achieved for oregano and adulterated samples, respectively. In contrast to the unstandardised results, DS resulted in the correct prediction of 100% of the authentic oregano samples in the majority of models, but failed to correctly predict the adulterated samples. This poor predictability can be attributed to the utilisation of the whole spectrum of the primary data, in order to fit each spectral point of the secondary data when DS is applied. It is known that spectral variations are limited to smaller regions of the spectrum, thus, each spectral point derived from the secondary data is most likely associated with neighbouring wavelengths as oppose to the full spectrum [34,38]. This may explain why the spectral corrections resulted in poor predictability using the direct standardisation approach for both multiclass and 2-class models.

Various window sizes for PDS were assessed to determine the applicability of the standardisation approach to the model transfer (**Supplementary material SI.1**). Overall, none of the PDS approaches achieved a high enough predictability to be useful in the model transfer. The best predictions were 85% and 40% for oregano and the adulterants, respectively, and were obtained from utilising a window size of 91 (multiclass OPLS-DA models; **Table 2;** M6 and M8). Again, it was noted that PDS favoured the prediction of either authentic oregano or adulterants. Interestingly, this differed between the window sizes, with a window of one demonstrating high correct predictions for the adulterants and a window size of ninety-one correctly predicting the majority of authentic oregano samples. It was anticipated that PDS would provide a greater percentage of correct predictions over DS, as several measurements are collected within a small window and these are incorporated into the transformation matrix [34]. The number of principal components, window size and the number of transfer samples for standardisation is known to have a significant effect on the performance of PDS [39]. Therefore, the experimental design described may have been limited by the use of the PDS approach.

When applying the subtraction standardisation approach, an 85% and 90% correct prediction for oregano and adulterants, respectively, was determined using the 2-class PLS-DA and OPLS-DA models (M11 and M12). All of the multiclass models demonstrated low correct prediction for the authentic oregano samples, however, they were able to correctly predict >95% of the adulterant samples. The division standardisation approach also resulted in an adequate predictability of 75% for oregano and 95% for adulterant samples, using the multiclass OPLS-DA model (M2). It was noted that the majority of the models were able to correctly predict a high percentage of the adulterant samples from the SW-002 derived data that had underwent division standardisation, however, the percentage of correct predictions for authentic oregano was inadequate. It is likely that both subtraction and division standardisation approaches prove successful as only the additive spectral differences are corrected, resulting in the consistency of the remaining spectra [40]. This is, however, dependent on the number and type of transfer samples used.

It is interesting to note that the optimal model, with the highest correct predictions, differed between the validation results obtained on the primary and secondary devices. The predictability was greatest for M8 when assessing the validation set obtained from the primary device (**Table 1**). M8 correctly predicted 97.5% of the adulterants and 93.0% of authentic oregano. In contrast, M11 provided the highest correct predictions of 100% of the adulterants and 90% of oregano samples determined on the secondary device, with no spectral standardisation (**Table 2**). These differences may be due to pre-processing the spectral data with first derivatives during model development. When the validation set is applied from the secondary device, there are likely to be wavelength shifts, and as derivitised models are less robust to instrumental changes, it is more difficult to correct for these [41].

Table 2. Predictability of the multiclass and 2-class reference models developed on the primary device using the validation set data measured on the secondary device with different standardisation approaches applied (n = 40)

				None		DS		PDS Win01		PDS Win91		Subtraction		Division	
Model	ID	Type	Pre-processing	Oreg. %	Adult. %	Oreg. %	Adult. %	Oreg. %	Adult. %	Oreg. %	Adult. %	Oreg. %	Adult. %	Oreg. %	Adult. %
Multiclass	M1	PLS-DA	None	25	100	100	0	40	80	100	0	25	100	65	95
	M2	OPLS-DA	None	20	100	100	0	35	85	100	0	30	100	75	95
	M3	PLS-DA	SNV	65	100	10	100	20	90	10	90	60	95	55	95
	M4	OPLS-DA	SNV	60	100	5	100	20	95	10	90	55	95	45	95
	M5	PLS-DA	SNV + 1DER	70	100	60	45	35	90	75	45	65	95	100	25
	M6	OPLS-DA	SNV + 1DER	15	100	100	10	30	95	85	40	5	100	25	95
	M7	PLS-DA	SNV + 1DER +SG	35	100	100	5	25	95	75	45	30	95	30	95
	M8	OPLS-DA	SNV + 1DER +SG	15	100	100	10	25	95	85	40	10	95	45	95
2-Class	M9	PLS-DA	None	25	100	100	10	50	55	95	20	25	100	50	95
	M10	OPLS-DA	None	100	0	100	0	15	85	55	15	100	0	100	15
	M11	PLS-DA	SNV	90	100	100	0	5	95	100	0	85	95	50	95
	M12	OPLS-DA	SNV	85	100	100	0	5	95	100	0	85	95	60	95
	M13	PLS-DA	SNV + 1DER	40	100	100	0	25	95	100	0	30	100	30	100
	M14	OPLS-DA	SNV + 1DER	55	100	100	0	25	95	100	0	35	95	35	95
	M15	PLS-DA	SNV + 1DER +SG	35	100	100	0	25	95	100	0	45	95	55	95
	M16	OPLS-DA	SNV + 1DER +SG	10	100	100	0	20	95	100	0	15	95	50	95

Oreg. refers to oregano and Adult. refers to the adulterated samples

4. Conclusion

The results obtained in this study have demonstrated the excellent performance of the NeoSpectra Micro NIRS for the determination of oregano authenticity. The development and validation of models using SIMCA 15 software confirmed the performance of the device, with high predictability achieved. The assessment of model transfer methods proved successful, with several approaches found to be applicable. This will be of significance in the future when more devices are assessed, which may require data standardisation due to larger spectral differences. Overall, the work has demonstrated the applicability of the NeoSpectra device to be utilised at various stages of the food supply chain, for the rapid and cost effective screening of oregano authenticity. Further investigations to validate the model transfer method include, assessing more devices and in-field analysis to determine the impact of differences in handling and testing environment on the transferability.

Author statement

C. McVey performed experiments, analysed data, and drafted manuscript. T. McGrath provided scientific oversight, analysed data (model transferability) and reviewed the manuscript. S.A. Haughey and C.T. Elliott reviewed the manuscript.

Acknowledgment

The research reported in this publication was supported by Invest Northern Ireland (NI) Proof of Concept grant "*The future of food authenticity: Rapid handheld in-field detection*" (PoC615).

Conflicts of interest

The authors declare no conflicts of interest.

Supplementary data

Supplementary data to this article can be found online.

References

[1] Stadler, R.H., Tran, L., Cavin, C., Zbinden, P., Konings, E.J.M. AOAC International, 99 (2016) 1135-1144

[2] Barrett, C.B. Actions now can curb food systems fallout from COVID-19. Nat. Food (2020) published online at https://www.nature.com/articles/s43016-020-0085-y.pdf (accessed 15/04/2020)

[3] EU Rapid Alert System for Food and Feed (RASFF). Online food safety register accessed at https://webgate.ec.europa.eu/rasff-window/portal/?event=SearchByKeyword&NewSearch=1&Keywords=horse%20meat (accessed 20/05/2020)

[4] Taylor, G. Anal. Bioanal. Chem. 411 (2019) 7053–7054

[5] Galvin-King, P., Haughey, S.A. and Elliott, C.T. Food Control, 88 (2018) 85-97

[6] Global Market Insights. Seasonings market size by type. Accessed online at https://www.gminsights.com/request-sample/detail/3242?utm_source=globenewswire.com&utm_medium=referral&utm_campaign=Paid_globenewswire (accessed 12/03/2020)

[7] BRC-FDF-FSSA. Guidance on authenticity of herbs and spices- industry best practice on assessing and protecting culinary dried herbs and spices. British Retail Consortium, Food and Drink Federation, Seasoning and Spice Association. Accessed online at https://www.fdf.org.uk/corporate_pubs/guidance-herbsandspices.pdf (assessed 24/09/2019)

[8] Silvis, I.C.J., van Ruth, S.M., van der Fels-Klerx, H.J. and Luning, P.A. Food Control, 81 (2017) 80-87

[9] Riviere, J. E., & Buckley, G. J. Ensuring safe foods and medical products through stronger regulatory systems abroad. Washington, DC, USA: National Academies Press, 2012. Retrieved from http://doi.org/10.17226/13296

[10] PwC and SSAFE. Food fraud vulnerability assessment. Accessed online at https://www.pwc.com/gx/en/services/food-supply-integrity-services/assets/pwc-food-fraud-vulnerability-assessment-and-mitigation-november.pdf (assessed 02/07/2019)

[11] SSAFE Food Fraud tool - AgriFood - PwC (EN). Nederland. Retrieved from http://www.pwc.nl/en/agrifood/ssafe-food-fraud-tool.html (assessed 24/04/2020)

[12] Nestle. Food Fraud Prevention. Economically-motivated adulteration. Accessed online at https://www.nestle.com/sites/default/files/asset-library/documents/library/documents/suppliers/food-fraud-prevention.pdf (assessed 24/01/2019)

[13] Black, C., Haughey, S., Chevallier, O.P., Galvin-King, P., Elliott, C.T. Food Chemistry, 210 (2016)551-557

[14] Wielogorska, E., Chevallier, O., Black, C., Galvin-King, P., Delêtre, M., Kelleher, C.T., Simon A.Haughey, S., Elliott, C.T. Food Chemistry, 239 (2018) 32-39

[15] Silletti, S., Morello, L., Gavazzi, F., Gianì, S., Braglia, L. and Breviario, D. Food Chemistry, 271 (2019) 410-418

[16] Manning, L., and Soon, J.M. Food Policy, 49 (2014) 23-32

[17] Oliveira, M.M., Cruz-Tirado, J.P., Roque, J.V., Teófilo, R.F. and Barbin, D.F. J. of Food Compost. Anal. 87 (2020) 103403

[18] Grassi, S., Casiraghi, E. and Alamprese, C. Food Chemistry, 243 (2018) 382-388

[19] Guelpa, A., Marini, F., du Plessis, A., Slabbert, R. and Manley, M. Food Control (Part B), 73 (2017) 1388-1396

[20] Schmutzler, M., Beganovic, A., Bohler, G. and Huck, C.W. Food Control, 57 (2015) 258-267

[21] Haughey, S.A., Galvin-King, P., Malechaux, A, and Elliott, C.T. Anal. Methods, 7 (2015) 181

[22] Bergman, E.L., Brage, H., Josefson, M., Svensson, O. and Sparén, A. J. Pharm. and Biomed. Anal. 11 (2006) 89-98

[23] Fearn, T. J. Near Infrared Spec. 9 (2001) 229–244

[24] Feudale, R.N., Woody, N.A., Tan, H., Myles, A.J., Brown, S.D. and Ferré, J. Chemomet. Intell. Lab. Syst. 64 (2002) 181–192

[25] Zhang, F., Zhang, R., Ge, J., Chen, W., Yanga, W. and Du, Y. Anal. Methods, 210 (2018) 2169-2179

[26] Greensill, C.V. and Walsh, K.B. J. Near Infrared Spect. 10, 1 (2002) 27-35

[27] Pu, Y.Y., Sun, D.W., Riccioli, C., Buccheri, M., Grassi, M., Cattaneo, T.M.P. and Gowen, A. Food Anal. Methods, 11 (2017) 1021–1033

[28] Wold, S., Antti, H., Lindgren, F. and Öhman, J. Chemomet. Intell. Lab. Syst. 44 (1998) 175-185

[29] Tan, H.W. and Brown, S.D. J. Chemomet. 15 (2001) 647-663

[30] Wang, A., Yanga, P., Chen, J., Wua, Z., Jia, Y., Ma, C and Zhan, X. Infrared Phy. Techn. 103 (2019) 103046

[31] McGrath, T.F., Haughey, S.A., Patterson, J., Fauhl-Hassek, C., Donarski, J., Alewijn, M., van Ruth, S. and Elliott, C.T. Trends Food Sci. Tech. 76 (2018) 38-55

[32) van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A. Smilde, A.K. and van der Werf, M.J. BMC Genomics, 7 (2006) 142

[33] de Noord, O.E. Chemomet. Intell. Lab. Syst. 25, 2 (1994) 85-97

[34] Wang, Y., Veltkamp, D.J. and Kowalski, B.R. Ana. Chem. 63 (1991) 2750-2756

[35] Liu, Y., Cai, W. and Shao, X. Anal. Chim. Acta, 836 (2014) 18-23

[36] Workman, J. The Handbook of Organic Compounds: Methods and Interpretations. NIR Band Assignments for Organic Compounds, Polymers and Rubbers, 1, 197-208 (2000) London: Academic Press

[37] Guo, G., Rösch, P., Popp, J. and Bocklitz, T. J. Chemomet. Special Issue: Conferentia Chemometrica, 34 (2019) e3202

[38] Griffiths, M.L., Svozil, D., Worsfold, P. and Evans E.H. J. Anal. Atomic Spec. 21 (2006) 1045–1052

[39] Zhang, F., Wanchao, C., Zhang, R., Ding, B., Yao, H., Ge, J., Lei Ju, L., Yang, W. and Du, W. Chemomet. Intell.Lab. Syst. 171 (2017) 234-240

[40] Cogdill, R.P., Anderson, C.A. and Drennen, J.K. AAPS PharmSciTech. 39 (2004) E284-297

[41] Xu, L., Zhou, Y.P., Tang, L.J., Wu, H.L., Jiang, J.H., Shen, G.L. and Yu, R.Q. Anal.Chim. Acta. 616 (2008) 138-143