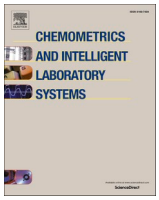
Chemometrics and Intelligent Laboratory Systems 277 (2026) 105809

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



Calibration transfer between benchtop and handheld near-infrared instruments for predictive models of commercial milk powder samples using cubic spline interpolation, piecewise direct standardization and artificial neural networks

Áine M. Ní Fhuaráin [a,*], Gonçalo M. Guedes [b], Ming Zhao [a], António C.S. Ferreira [b], Colm P. O'Donnell [a], Aoife A. Gowen [a,**]

[a] UCD School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin 4, D04 C1P1, Ireland
[b] Centro de Biotecnologia e Química Fina (CBQF), Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Rua Diogo Botelho 1327, Porto, 4169-005, Portugal

A B S T R A C T

Milk powder manufacture extends the shelf-life of milk which facilitates an increase in export opportunities, reduces storage costs and provides an ingredient for food applications. Using Near-Infrared (NIR) spectroscopy predictive models to predict compositional properties of milk powders is important for process and quality control applications. There is a need for calibration transfer, as new instruments are commissioned, including inexpensive, portable handheld instruments. Calibration transfer from benchtop to inexpensive, compact handheld instruments would increase quality measurements without the costly and time-consuming process of validating new calibration models using wet chemistry reference data. In this paper, NIR spectra of 70 samples of Full Fat Milk Powder (FMP) and 59 samples of Skim Milk Powder (SMP) were acquired on a benchtop and two handheld instruments of the same model and specification. Partial Least Squares (PLS) regression and Partial Least Squares-Discriminant Analysis (PLS-DA) were used to develop regression and classification models. Cubic spline interpolation, Piecewise Direct Standardization (PDS) and Artificial Neural Networks (ANNs) were used to perform calibration transfer on spectra from the handheld instrument to the benchtop instrument with the range cropped to match the handheld instrument. The second handheld instrument with the same specifications was then used to take spectra of a subset of the powders to investigate if the calibration transfer models were robust when used with independent instruments of the same type. ANNs yielded a lower Root Mean Square Error Predictions ($RMSE_P$) than PDS for the independent handheld instrument, when the transferred spectra are inputted into the master models.

1. Introduction

The global production of Whole Milk Powder (WMP) and SMP is predicted to increase by 11.48% and 14.34%, respectively, between 2019 and 2029 [1]. Milk powder is used as an ingredient in many food applications including beverages, baking ingredients, biscuits and ice cream [2] and for standardization applications in dairy processing [3]. Reference wet chemistry methods for determining compositional properties for milk powder quality control, such as the Kjeldahl method for protein composition determination, are laborious, time-consuming and waste-generating [4]. NIR spectroscopy measures the broad overtone and combination bands of the higher frequency modes of fundamental vibrations in the range of 800−2500 nm [5,6]. NIR spectroscopy is rapid, non-destructive, inexpensive and is environmentally friendly as it generates little to no chemical waste [7]. Currently, in the dairy sector, NIR instruments are used for screening to determine and control macro-compositional properties of dairy products [8]. The International Organisation for Standardisation (ISO) published an international standard to provide guidance on using NIR spectrometry for measuring milk and milk product quality [9]. The adoption of NIR in food manufacture facilitates the achievement of improved digitalization and sustainability metrics [10,11]. Czaja et al. (2025) [10] discussed the pivotal role of NIR in a circular economy, where instead of using and then discarding products and materials, they are regenerated or circulated on the basis of eliminating waste and pollution. Recycling food waste streams can be enhanced through continuous monitoring or fingerprinting. NIR will complement digital platforms, decision-making tools and digital twins with robust data [10].

Calibration transfer between NIR instruments is vital as most instruments, even when they are very similar, generate unique spectral

* Corresponding author.
** Corresponding author.
E-mail addresses: aine.nifhuarain@ucdconnect.ie (Á.M. Ní Fhuaráin), aoife.gowen@ucd.ie (A.A. Gowen).

https://doi.org/10.1016/j.chemolab.2026.105809
Received 18 January 2026; Received in revised form 6 July 2026; Accepted 9 July 2026
Available online 9 July 2026

responses, making calibration model transfer between them challenging [12]. Improved calibration transfer between instruments could save the dairy industry a lot of time and effort in quality control applications by eliminating the need to develop an entire calibration when a new instrument is commissioned. It would also allow spectra taken from significantly lower cost, portable handheld instruments to be transferred to a calibration model developed from a benchtop instrument with higher specification optics [13] and a higher signal-to-noise ratio [14]. Calibration transfer methods for spectroscopy were reviewed by Fearn et al. (2001) [15] and Workman et al. (2018) [16]. There are three main approaches to calibration transfer including (i) using preprocessing to reduce the variability between instruments and also using combined data from numerous instruments when calibrating, (ii) improving calibrations using bias and slope changes and (iii) by changing the spectra with methods such as PDS [15,16]. There are various methods of performing calibration transfer using mathematical correction or standardization methods. Cubic spline interpolation has previously been used in calibration transfer studies [17]. PDS creates a structured transformation matrix with the spectra of samples measured with the two instruments, which can be used to transform spectra between instruments [16]. PDS is one of the most widely-used calibration transfer methods due to its local character and multivariate nature enabling intensity differences, wavelength shifts and peak broadening to be corrected simultaneously. PDS can sometimes even have better transfer results than full set recalibration for calibrations from a lower quality to a higher quality instrument [18]. PDS does have some limitations such as when choosing the ranks of local PLS models, where the number of eigenvalues for each local model is predetermined based on a tolerance value [18]. This can lead to artifacts in the transferred spectra if the local ranks are too high which can lead to overfitting. For this reason, preprocessing is recommended before PDS [18]. Other methods of calibration transfer include orthogonal signal correction, orthogonal projection, Procrustes analysis, finite impulse response and canonical correlation analysis among many others [16]. There have also been some attempts to use ANNs for calibration transfer [18–20], but some of these are quite dated. Although the literature regarding milk powder is focused more on classification and adulteration detection [8,21–24] than calibration transfer, previous studies have investigated calibration transfer of regression models between a benchtop and a handheld NIR instrument for applications other than milk powders [25,26]. To the best of the authors' knowledge, this is the first study to compare calibration transfer strategies, including ANNs, between a benchtop and two handheld instruments of the same specifications to predict compositional properties and for classification between different types of milk powder. The objective of this study is to investigate 1) if spectra can be transferred from a handheld instrument to a benchtop instrument for a milk powder application, 2) if these transferred spectra can be useful in protein and fat regression models and classification models, 3) if calibration transfer can be sensor-independent and 4) if PDS or ANNs are superior for these objectives.

2. Materials and methods

2.1. Milk powder samples

129 milk powder samples (70 FMP and 59 SMP) manufactured over a 12 month period were obtained from a dairy processor in Ireland. For the classification models, there were 126 samples for each instrument (n = 126) after spectral outlier removal and removal of one mislabelled sample. Some protein and fat values were not available, leaving fewer samples for the regression models than the classification models. The protein models were developed with 116 samples and fat with 117 samples.

2.1.1. Protein and fat measurements

An at-line FT-NIR benchtop instrument (Q-Interline A/S Stengårdsvej 7, DK-4340 Tølløse) was used to determine the protein and fat content of milk powder samples collected from the production line. An average value of each sample was calculated from two measurements of each sample. The three replicates were averaged before any preprocessing, outlier detection and splitting was carried out. The results are the predicted values which were derived from an embedded linear predictive model that was employed by the company. Outlier removal was also performed on the compositional properties, leaving a smaller number of samples. The number of samples in the calibration and validation set, the mean, and range for each property regression model are shown in Table 1.

2.1.2. Spectral measurements

For calibration transfer, a total of three instruments were used, namely a master laboratory benchtop MPA II (Bruker, Germany) and two slave handheld NIR-G-S1 instruments (Innospectra, Taiwan). In this paper the master benchtop instrument is referred to as BT, the first handheld instrument as HH1 and the second handheld instrument as HH2. The number of wavelengths, spectral ranges, acquisition time and approximate purchasing costs are outlined in Table 2. Soft tissues with Isopropyl Alcohol (IPA) (≥99.5%, reagent grade, Thermo Scientific) were used to clean the optical lenses between sample measurements for the HH1 and HH2 measurements. The BT samples were measured in separate vials and therefore no cleaning was required. A background scan was taken on each instrument before spectra acquisition. A Spectralon (SphereOptics, Herrsching, Germany) white tile was used as a reference for the HH1 instrument. For the HH2 spectra, a factory reference was used, and these spectra were taken approximately six months after the spectra were taken on the HH1 instrument. Three experimental replicates were measured for each sample. The mean of these replicates were calculated for model development. A subset of 41 (21 FMP and 20 SMP) samples from the full dataset were measured using the HH2 instrument.

2.1.3. Data analysis and model development

2.1.3.1. PCA and outlier removal. Data analysis was carried out using MATLAB R2024b (Mathworks, Natick, MA, USA) in conjunction with PLS Toolbox 9.5 (Eigenvector Research, Inc., Manson, Washington, USA). For the Kennard-Stone method and the cubic spline interpolation and extrapolation functions, the inbuilt functions on MATLAB were used, and for PDS, PLS-DA and PLS, PLS Toolbox was used. Mahalanobis distance calculations were used to identify outliers in the spectral data. PCA was performed on the spectra with SNV and second order Savitzky-Golay (3rd order polynomial, window size of 11) preprocessing to correct scattering effects and improve resolution of overlapping bands. The data were also mean centred before applying PCA. Then, the Mahalanobis distance calculation was performed on the first 3 PC scores (98.90% of the variance for the BT, 98.39% of the variance for HH1, and 98.49% of the variance for HH2). A threshold of the 0.99 quantile was used for outlier determination as according to the method used by Frizzarin et al. (2021) [27]. One spectral outlier each was identified for the BT instrument and the HH1 instrument (Fig. 1). Both outliers were removed from the BT and HH1 datasets to ensure the calibration transfer model had the same samples in each row for the purpose of calibration transfer. Outlier detection was performed independently for each instrument because instrument-specific spectral characteristics define distinct outlier spaces, such that a sample identified as anomalous on one instrument may not necessarily be anomalous on the other. While differences in outlier occurrence between instruments may themselves provide insight into transfer behaviour, the objective here was to remove instrument-specific spectral artifacts while preserving the paired structure required for transfer modelling. There were no outliers identified in the HH2 dataset (Fig. 1). Values greater than three SDs from the mean were removed from the protein and fat measurements, as in

Table 1
Number of samples, mean, Standard Deviation (SD) and range of milk powder protein and fat measurements for the regression models.

Compositional Parameter	Number of Samples	Number of Samples in the Calibration Set	Number of Samples in the Test Set	Mean (%)	SD (%)	Range (%)
Protein	116	93	23	29.09	6.18	16.50-36.40
Fat	117	94	23	15.95	14.68	0.40-31.80

Table 2
Number of wavelengths, spectral range, acquisition time and for BT, HH1 and HH2 instruments.

Instrument	Number of Wavelengths	Spectral Range (nm)	Acquisition Time (s)	Cost (€)
Master Benchtop (BT)	949	865-2531	40	Approximately 50k
Slaves Handheld (HH1 and HH2)	228	900-1700	7.56	<2k

Visentin et al. (2015) [28]. This was done separately for FMP and SMP. Three outliers were removed for protein and two outliers were removed for fat. After outlier removal, PCA was performed on the BT spectra and then the HH1 and HH2 spectra combined.

2.1.3.2. Master model development. The Kennard-Stone method [29] was used to split the benchtop data into 80% calibration set and 20% test set, with the same samples then used for the HH1 calibration and test set in order to match the rows of the BT dataset for calibration transfer. Before preprocessing, for cubic spline interpolation, PDS and ANN model development, the range of the BT was cropped to match the range of the HH1 and HH2 instruments. Seven preprocessing combinations were applied separately: none; SNV; first order Savitzky-Golay; second order Savitzky-Golay; SNV and first order Savitzky-Golay (3rd order polynomial, window size of 15); SNV and second order Savitzky-Golay (3rd order polynomial, window size of 15); and MSC, to test which one gave the highest accuracy/lowest $RMSE_{CV}$) in the models for classification and regression, respectively. The data were also mean centred before applying PLS or PLS-DA. The number of LV were selected to optimize each model and to prevent overfitting. This was done in conjunction with venetian-blind cross validation (with 10 splits and a blind thickness of 1). For each model, the number of latent variables was chosen using a Number of Latent Variables vs Root Mean Square Error of Calibration plot. The number of latent variables was chosen as the number of latent variables corresponding to the lowest point of the plot before it rose again. This was done to avoid overfitting. For the regression models, separate models were created for each property, protein and fat. For the classification model, there were two classes, one for each product, FMP and SMP.

2.1.3.3. Cubic spline interpolation. Then, the inbuilt MATLAB function for cubic spline interpolation was used on the HH1 calibration and test set and the HH2 dataset. Then, the HH1 calibration set, HH1 test set and HH2 dataset were inputted into the BT master model for protein, fat or classification.

2.1.3.4. PDS. Then, a PDS model was created using the BT calibration set and HH1 calibration set, while transforming the HH1 spectra at the same time. Then, the HH1 test set and HH2 dataset were inputted into the PDS model that was created with the BT calibration set and HH1 calibration set. After this, the HH1 calibration set, HH1 test set and HH2 spectra were inputted into the BT master model for protein, fat or classification. Fig. 2 shows the calibration transfer process for the calibration and test set of the HH1 spectra.

In PDS the response **r** of the standardization sample taken at wavelength j on the master instrument is related to the wavelengths present in a small window around j taken on the slave instrument as described by Feudale et al., 2002 [18]. As mentioned in the introduction, preprocessing is recommended before PDS. Therefore, SNV was used. The data were not mean centred before applying PDS. The difference between the directly measured data and the data subjected to calibration transfer was assessed using a normalized difference metric (as included in Table 3 and Tables S5 and S6). The normalized difference was calculated as the Euclidean norm of their element-wise difference divided by the Euclidean norm of the master dataset (Equation (1)).

$$\textit{normalized difference} = \frac{||x_1 - x_2||}{||x_1||}$$

Equation (1) Normalized difference between two datasets, where x_1 is the master dataset, x_2 is the transferred dataset being compared against it and $||\cdot||$ denotes the Euclidean norm.

Scaling by the norm of the master dataset expresses the difference yields a dimensionless value where smaller values indicate closer agreement between the two datasets. Various window sizes were trialled for PDS in order to achieve the lowest difference.

2.1.3.5. ANNs. For the ANN calibration transfer, a shallow feed forward fully-connected neural network, with an input layer, hidden layer and an output layer was created using the Deep Network Designer application on MATLAB. Hyperparameter tuning was performed to optimize the model with the following parameters: max epochs (range: 1500-2000); mini batch size (4-16); learning rate (0.001-0.0001); validation frequency; dropout layer (range: 0.1-0.5); number of layers (range: 2-5); linear or non-linear ANN; and no preprocessing or SNV preprocessing. The optimized network architecture consisted of 228 input features, a fully connected hidden layer with 50 neurons, and an output layer of 654 neurons, followed by a regression layer. The network included two dropout layers to prevent overfitting. The first dropout layer had a rate of 0.1 and the second dropout layer had a rate of 0.4. A ReLU activation function was inserted after the hidden layer, making the model non-linear. The model was trained using the Adam optimizer with 1500 epochs, a mini-batch size of 6, an initial learning rate of 0.001 and a validation frequency of 30 iterations. The loss function used was Half Mean Squared Error (HMSE). For data augmentation, zero-mean Gaussian noise with a SD of 0.002 was added to the input spectra and small random perturbations ($\pm$0.002) added. These hyperparameters were chosen after some optimization trials. Training data is reshuffled at every epoch. The HH1 calibration set was used as the input spectra, and the BT calibration set as the output spectra. SNV was applied to both input and output spectra before ANN training. The data were not mean centred. The dataset was split into 70% for training and 30% for validation. Due to the size of the training set being small, to evaluate the sensitivity of model performance to random data portioning and network initialization, multiple tests were performed to determine if the architecture is robust. Fifteen models were trained with random weight initialization and random splits. For each run, the training and validation loss together with the RMSE were recorded. Mean values and standard deviations across the fifteen independent runs were used to quantify the reproducibility and robustness of model performance rather than relying on a single train/validation split. After the ANN model was built, the HH1 calibration and test set and the HH2 dataset were inputted into it. Then, the output from the ANN model for the HH1 calibration and test set and the HH2 dataset were inputted into the BT master models.

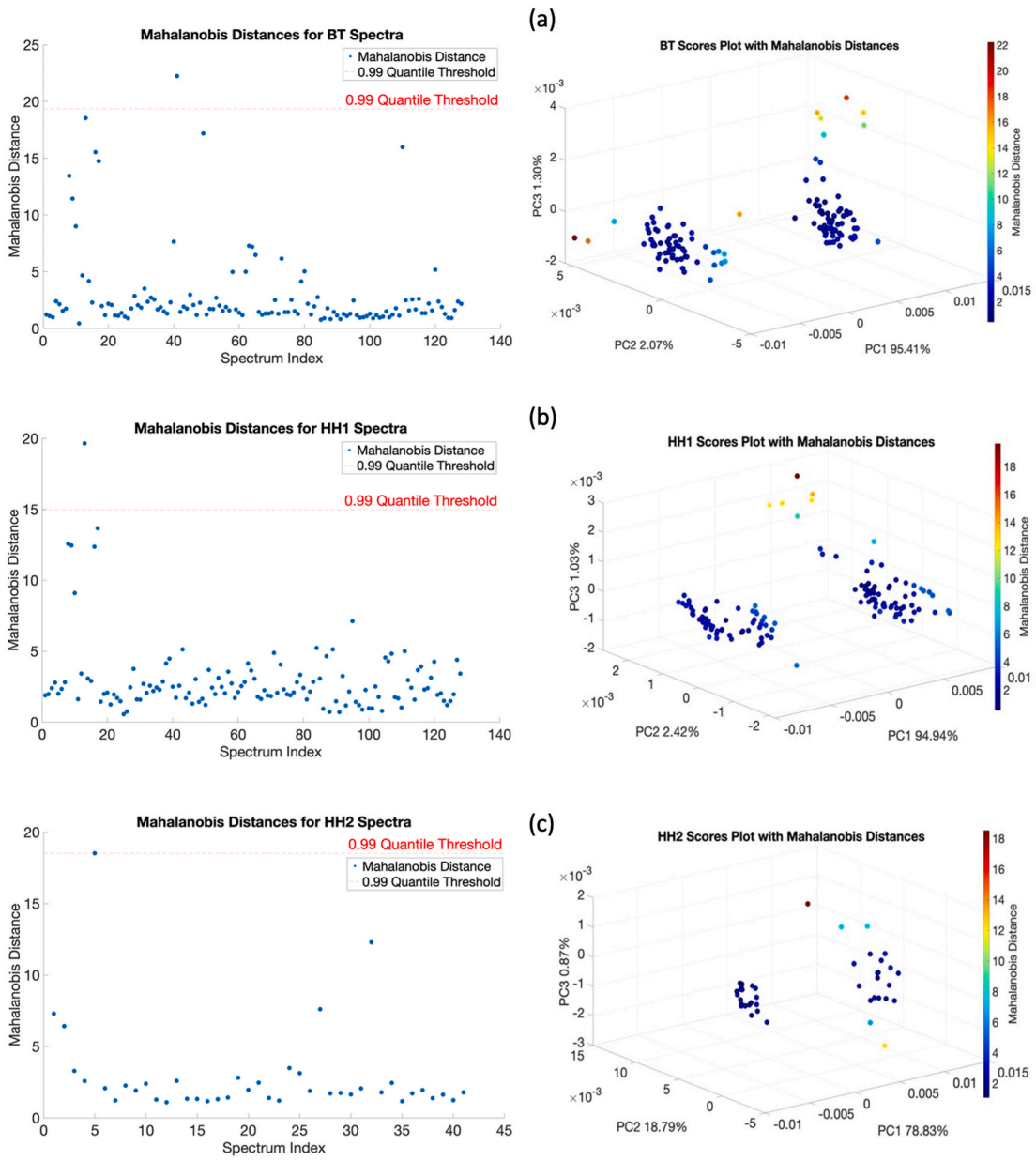


Fig. 1. Quantile threshold plots and Mahalanobis distance plots for the (a) BT spectra, (B) HH1 spectra and (c) HH2 spectra.

2.1.3.6. Correlation coefficients. To provide quantitative spectral-level evaluation of the calibration transfer methods, mean spectral correlation coefficients were computed between the mean reconstructed spectrum and the mean benchtop reference spectrum for both HH1 and HH2, and for both ANN and PDS transfer methods.

2.1.3.7. Wavelength-resolved transfer errors. To provide spectral-level evaluation of the calibration transfer methods, wavelength-resolved mean error profiles were computed for both HH1 and HH2 by comparing the mean reconstructed spectrum to the mean benchtop spectrum at each wavelength, for both ANN and PDS transfer methods.

3. Results and discussion

3.1. Spectra and PCA

For the exploratory PCA analysis, only SNV preprocessing was used, so that the loadings would be interpretable. The spectra are shown in Fig. 3. From these plots it would be assumed that the HH1 and HH2 datasets could not classify between the powders or predict protein and fat as well as the BT spectra, due to fewer peaks in the handheld spectra range, especially as (further described below), the BT spectra have more peaks corresponding to fat content. However, in this study it was found that only the protein regression model was affected by the spectral range.

Fig. S1 in the Supplementary Material shows HH1 and HH2 spectra on the same plot for visual comparison. The spectra are very similar for most of the spectral range until approximately 1600 nm, where the HH2 spectra become somewhat distorted in comparison to the HH1 spectra. As is observed in Fig. 4, a score plot of PCs 1 (95.68%) and 2 (2.31%) of the BT data could separate FMP and SMP. This is an indicator that the BT instrument can be used to create well-performing classification models for these samples.

Then PCA was performed on the spectra from HH1 and HH2 (Fig. 4). First, the two spectral arrays were combined, then SNV preprocessing was applied. The HH1 and HH2 spectra do not form one cluster, therefore we can already suspect a difference in the spectra. This is

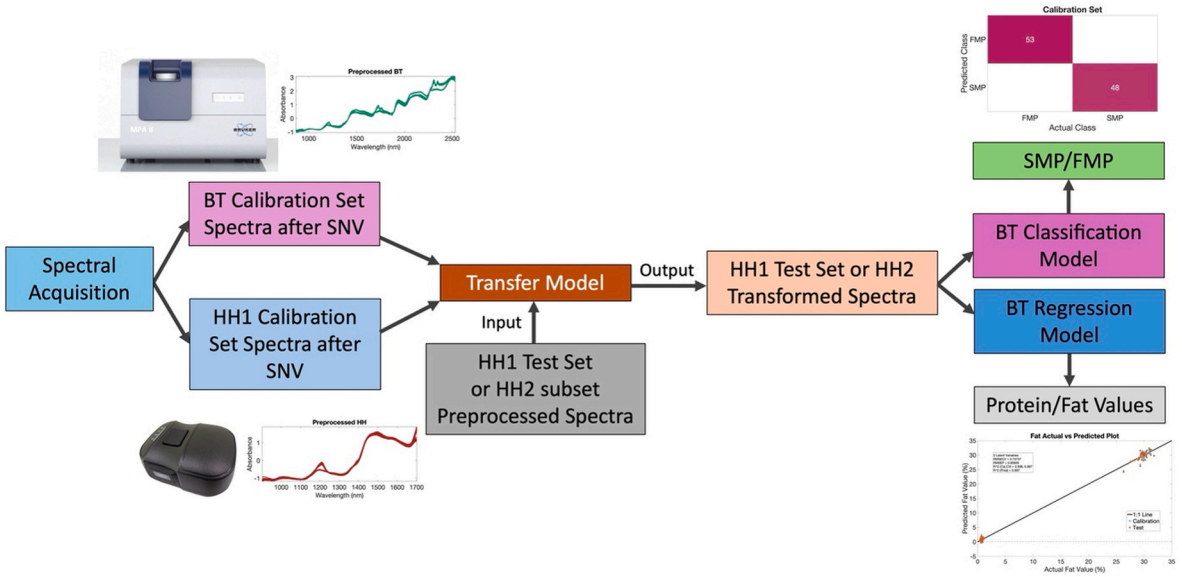


Fig. 2. High-level schematic of the calibration transfer process showing the HH1 test or HH2 subset set being transformed and inserted into the BT model.

Table 3
Spectral differences after PDS without preprocessing, after SNV and with various window sizes for the BT and HH1 spectra for the protein dataset.

Difference between BT and HH1 with no preprocessing	Difference between BT and HH1 with SNV preprocessing	Window Size
0.0816	0.0158	11
0.0723	0.013	21
0.0635	0.0112	31

common between two instruments of the same specifications, although they should technically be identical. Although a suspected outlier appears in the SMP HH2 data, the Mahalanobis distance threshold method did not flag this as an outlier as in Fig. 1.

The loading plots for the first and second loadings are shown in Fig. 5 for BT, HH1 and HH2. Although the peak assignment is challenging due to the SNV preprocessing, the loadings for both instruments show peaks corresponding to protein and fat, which may explain the variance [30, 31]. The peaks at approximately 1725, 1765, 2310 and 2347 nm for fat explain the ν C–H first overtone, ν C–H + δ C–H and the ν_s CH_2 + δ = CH_2 and 1215 nm explain the ν C–H second overtone for protein. The first

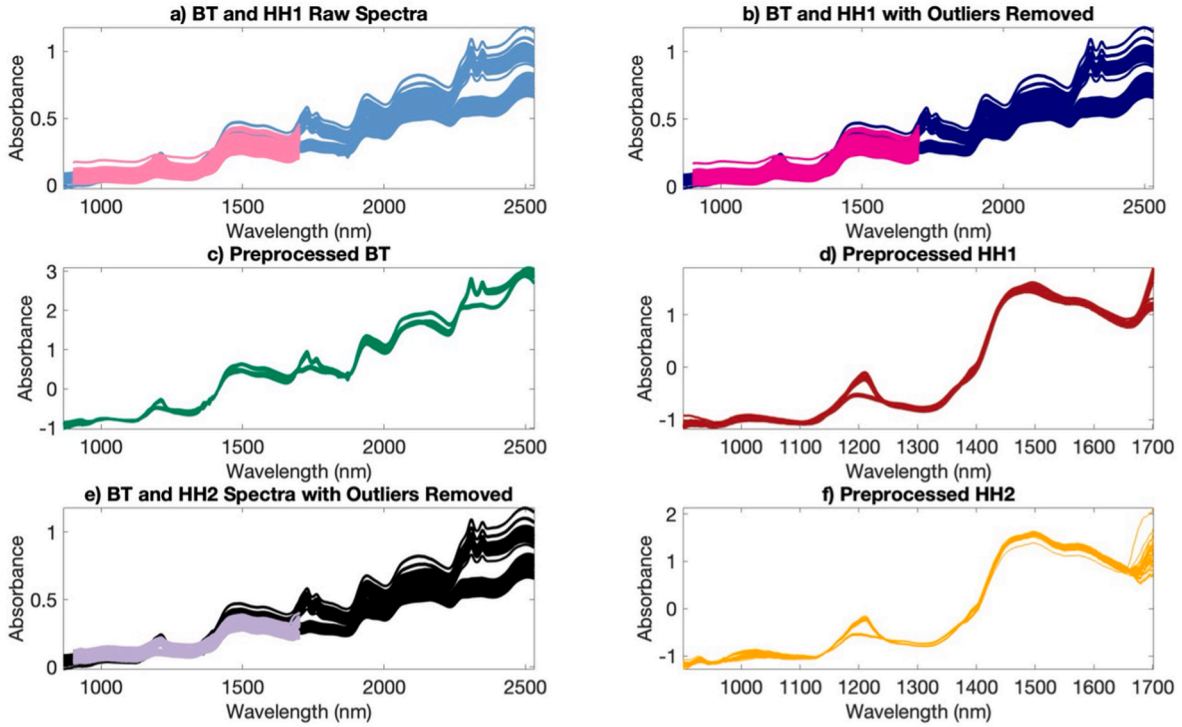


Fig. 3. (a) BT (blue) and HH1 (pink) raw spectra, (b) BT (blue) and HH1 (pink) spectra with outliers removed, (c) BT spectra with SNV preprocessing, (d) HH1 spectra with SNV preprocessing, (e) BT (black) and HH2 (purple) spectra with outliers removed and (f) HH2 spectra with SNV preprocessing. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

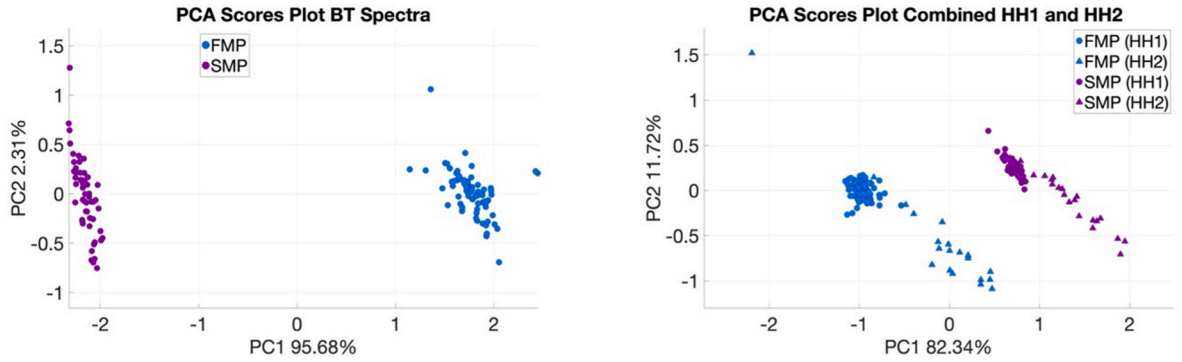


Fig. 4. PCA of the with SNV preprocessing of the BT spectra after SNV and the HH1 and HH 2 spectral matrices combined.

and second HH2 loadings show a peak at 1001 nm which corresponds to N–H for protein [32]. The peak at approximately 929 nm in the first loading for the BT, HH1 and HH2 could correspond to the C–H for lipids [32]. It should be noted that cropping the BT spectra to match the HH range for calibration transfer results in the exclusion of several analytically relevant absorption bands. Specifically, major water combination bands near 1940 nm, fat combination bands near 2310 and 2347 nm and protein combination bands in the 2050-2180 nm region are removed after cropping. Their exclusion may limit model accuracy, however, for the case of examining the performance of the calibration transfer techniques, as is the goal of this paper, this is not overly concerning.

Mean spectral coefficients between the reconstructed and mean benchtop spectra are presented in Table S1 in the Supplementary Information. For HH1, both ANN and PDS achieved $r = 1.0$ with the mean benchtop spectrum, indicating that both methods reconstruct spectra that are essentially identical in shape to the true benchtop spectra for the instrument on which they were trained. For HH2, both methods maintained very high correlation with the mean benchtop spectrum, with the ANN achieved $r = 0.998$ and PDS achieving $r = 0.9978$. While both values indicate strong spectral similarity, the marginally higher correlation of the ANN for HH2 is consistent with the master model prediction results, where the ANN outperforms PDS for protein and fat on this instrument. The small but meaningful difference in correlation between ANN and PDS for HH2, combined with the greater instability of the PDS wavelength-resolved error profile in the 1200-1400 nm region suggests that the ANN produces a more faithful spectral reconstruction when generalising to an unseen instrument.

3.2. Regression

3.2.1. Master models

A PLS regression model was created for each property, protein and fat, using the BT calibration set and BT test set. Although the models have two different clusters for FMP and SMP, 'global' models like these have been reported as useful previously [33]. For protein, the best model was achieved using SNV and 4 LV, resulting in an $RMSE_{CV}$ of 0.73%, an $RMSE_P$ of 0.46%, an R^2 (Cal, CV) of 0.988, 0.986 and an R^2 (pred) of 0.994. For fat, the best BT model was trained with SNV and 1st derivative preprocessing (3rd order polynomial, window size of 15) on the spectral matrix and 3 LV resulted in an $RMSE_{CV}$ of 0.76%, a $RMSE_P$ of 0.76%, an R^2(Cal, CV) of 0.998, 0.997 and an R^2 (pred) of 0.997. It should be noted that these models were created using protein and fat values from previous spectral calibrations, and not wet chemistry reference measurements, due to availability constraints. Therefore, the work in this study does not include classical calibration transfer. The accuracy of the models is not the focus of this study, rather comparing the effectiveness of the various transfer methods for sensor-independent calibration transfer.

Product specific models were also created, however they are not reported in the main text as the protein and fat ranges were very narrow. As well as this, the PDS and ANN calibration transfer did not work as well as for the combined models. The product-specific models and their associated calibration transfer are included in Table S7 in the Supplementary Material.

3.2.2. Cubic spline interpolation

The calibration transfer using cubic spline interpolation results were compared against the baseline models. These results are presented in Table S2 of the Supplementary Material for the HH1 calibration and test sets and HH2 dataset. The cubic spline interpolation is the least computationally-intensive method for calibration transfer, requiring only a short script on MATLAB. However, the results are not able to match the $RMSE_{CV}$ (0.73%) or $RMSE_P$ (0.46%) of the master models, especially for the HH2 model. For the HH1 instrument, the model for fat showed acceptable performance after using cubic spline interpolation. The $RMSE_P$ values for the transformed HH1 calibration and test sets

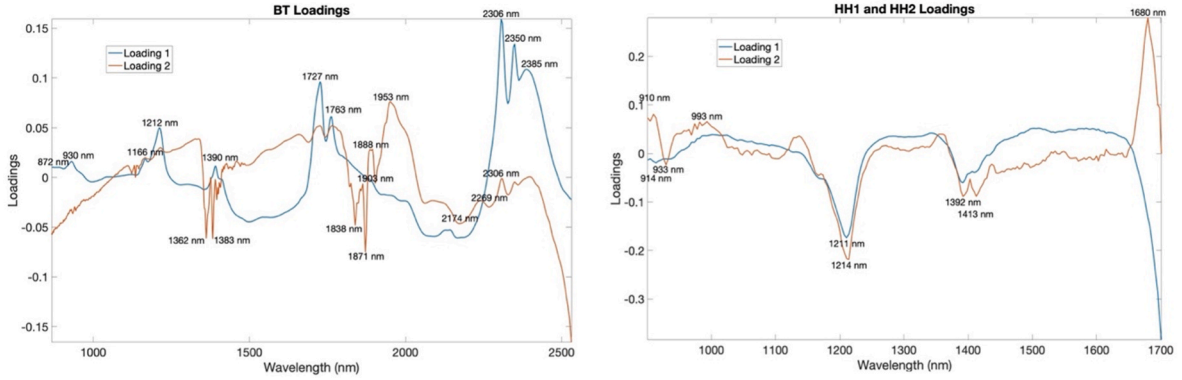


Fig. 5. PCA loading 1 and 2 plot for BT spectra and combined HH1 and HH2 spectral datasets.

(1.78% and 1.61%, respectively) remained reasonably low compared to the $RMSE_{CV}$ and $RMSE_P$ of the BT calibration set (0.76% and 0.46%, respectively), indicating reasonable transferability. However, the protein model had a higher $RMSE_P$ for calibration and test set (3.71% and 3.61%, respectively). These poor results are also reflected in the R^2 value (0.706), showing little predictive ability. For the HH2 dataset, the $RMSE_P$ values for fat (9.76%) and protein (6.80%) were much higher than for the HH1 instrument, showing that cubic spline interpolation has its limitations.

3.2.3. Protein

3.2.3.1. PDS. As detailed previously, cubic spline interpolation was not successful for transforming the HH1 and HH2 spectra for protein and fat into BT spectra to be used in the master BT PLS regression models. To try to improve on these results, PDS is investigated as a technique for more accurate calibration transfer of the spectra. For protein, the HH1 calibration set spectra were transformed to BT spectra calibration set using PDS, creating a calibration transfer PDS model in the process. The differences from the PDS are shown in Table 3 with and without SNV preprocessing and with three different window sizes. The smallest difference came from using SNV preprocessing with a window size of 31 for PDS. The lowering of difference from using preprocessing for PDS is consistent for protein and fat, and for classification as well (Tables S5 and S6 of the Supplementary Material).

The $RMSE_P$ (transformed HH1 calibration set) is only 0.22 higher than the $RMSE_{CV}$ (BT) (0.73%) showing that the calibration transfer works well. The results for the PDS transformations can be found in Table 4. The $RMSE_P$ (transformed HH1 test set) is only 0.56 higher than the $RMSE_p$ (BT) (0.46%), again showing that the PDS transfer for HH1 works well, even with the test set. Then, the HH2 spectra were inputted into the PDS model created by the calibration sets of the BT spectra and the HH1 spectra (Fig. 6). The $RMSE_P$ here was very high at 7.42% compared to the BT $RMSE_P$ of 0.46%, showing that PDS does not have the capabilities to transfer spectra from multiple sensors of the same specifications. The predictions are shifted away from the original model (Fig. 6).

3.2.3.2. ANNs. For protein, although PDS demonstrated a good result in transferring HH1 spectra into BT spectra, its performance decreased when inputting HH2 spectra into the same PDS model. To address this, ANNs were used. The epoch vs loss plots across fifteen repeated runs are shown in Fig. S2 in the Supplementary Material. The consistent convergence behaviour and overlapping loss trajectories across runs indicate that model performance is not strongly dependent on the particular random initialization or data split. The values of loss and RMSE per run are shown in Tables S3 and S4 in the Supplementary Material. The training loss is very consistent across runs (SD of 0.018 on a mean of 0.25, so about 7% relative variation), while validation loss is noticeably more variable (SD of 0.031 on a mean of 0.15, about 21% relative variation). The same pattern shows in RMSE where the training RMSE SD is about 2.5% relative while validation RMSE SD is about 12.6% relative. The training behaviour is stable across initializations and splits, however the validation performance is more sensitive to which 30% of samples are in the held-out set. The validation loss/RMSE is consistently lower than training loss/RMSE, across all 15 runs without exception. This is a known and explainable artifact of dropout-regularized networks, not evidence of overfitting.

For HH1, after using ANNs for calibration transfer, the $RMSE_P$ was higher for both the HH1 calibration (1.70%) and test set (1.52%) than from using PDS. Fig. 7 shows the HH2 transfer to the BT spectra using an ANN. The transformed HH2 spectra match the BT spectra very well, however it seems that it cannot replicate the noise in the BT spectrum at approximately 1400 nm.

ANNs achieved significantly lower $RMSE_P$ values than PDS for transforming the HH2 spectra. ANNs achieved an $RMSE_P$ of 1.03% compared to 7.42% using PDS (Fig. 8). Using ANNs also reduced the shift for HH2 (Fig. 8). This is a very promising result for the protein regression models, as the low $RMSE_P$ (1.03%) value that is comparable to the BT $RMSE_{CV}$ of 0.73% shows that this model could potentially be used in industry. This shows that a subset of the samples taken on a different handheld instrument can still be transferred successfully, which would cut costs and improve sustainability for industry. This is even possible when a factory reference is used with the second instrument and not the white tile reference as is used with the first handheld

Table 4
Summary table of results for calibration transfer of regression models.

Calibration Transfer Method	Description of model	Property	$RMSE_{CV}$ (%)	$RMSE_P$ (%)	R [2] (Cal, CV)	R [2] (Pred)
-	BT model	Protein	0.73	0.46	0.988, 0.986	0.994
-	HH1 model	Protein	0.81	0.60	0.987, 0.983	0.990
Cubic spline interpolation	HH1 calibration set transformed to BT	Protein	0.73	3.71	0.988, 0.986	0.717
Cubic spline interpolation	HH1 test set transformed to BT	Protein	0.73	3.61	0.988, 0.986	0.706
Cubic spline interpolation	HH2 transformed to BT	Protein	0.73	9.76	0.988, 0.986	0.324
PDS	HH1 calibration set transformed to BT	Protein	0.73	0.95	0.988, 0.986	0.977
PDS	HH1 test set transformed to BT	Protein	0.73	1.02	0.988, 0.986	0.971
PDS	HH2 transformed to BT	Protein	0.73	7.42	0.988, 0.986	0.930
Linear ANN	HH1 calibration set transformed to BT	Protein	0.73	1.60	0.988, 0.986	0.934
Linear ANN	HH1 test set transformed to BT	Protein	0.73	1.36	0.988, 0.986	0.952
Linear ANN	HH2 transformed to BT	Protein	0.73	1.95	0.988, 0.986	0.965
Non-linear ANN	HH1 calibration set transformed to BT	Protein	0.73	1.70	0.988, 0.986	0.943
Non-linear ANN	HH1 test set transformed to BT	Protein	0.73	1.52	0.988, 0.986	0.966
Non-linear ANN	HH2 transformed to BT	Protein	0.73	1.03	0.988, 0.986	0.972
-	BT model	Fat	0.76	0.76	0.998, 0.997	0.997
-	HH1 model	Fat	0.42	0.33	1.000, 0.999	0.999
Cubic spline interpolation	HH1 calibration set transformed to BT	Fat	0.76	1.78	0.998, 0.997	0.995
Cubic spline interpolation	HH1 test set transformed to BT	Fat	0.76	1.61	0.998, 0.997	0.997
Cubic spline interpolation	HH2 transformed to BT	Fat	0.76	6.80	0.998, 0.986	0.992
PDS	HH1 calibration set transformed to BT	Fat	0.76	0.81	0.998, 0.997	0.997
PDS	HH1 test set transformed to BT	Fat	0.76	0.66	0.998, 0.997	0.998
PDS	HH2 transformed to BT	Fat	0.76	7.03	0.998, 0.997	0.785
Linear ANN	HH1 calibration set transformed to BT	Fat	0.76	2.35	0.998, 0.997	0.989
Linear ANN	HH1 test set transformed to BT	Fat	0.76	2.67	0.998, 0.997	0.973
Linear ANN	HH2 transformed to BT	Fat	0.76	3.47	0.998, 0.997	0.986
Non-linear ANN	HH1 calibration set transformed to BT	Fat	0.76	1.83	0.998, 0.997	0.995
Non-linear ANN	HH1 test set transformed to BT	Fat	0.76	2.44	0.998, 0.997	0.985
Non-linear ANN	HH2 transformed to BT	Fat	0.76	2.18	0.998, 0.997	0.990

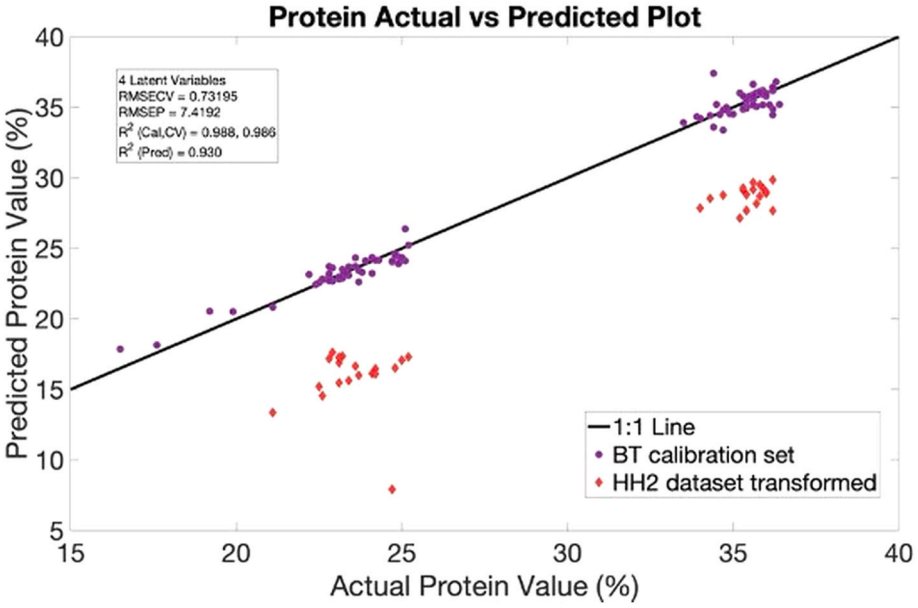


Fig. 6. Actual vs predicted plot for the HH2 spectra transformed using HH1 PDS model plotted with the BT calibration set for protein.

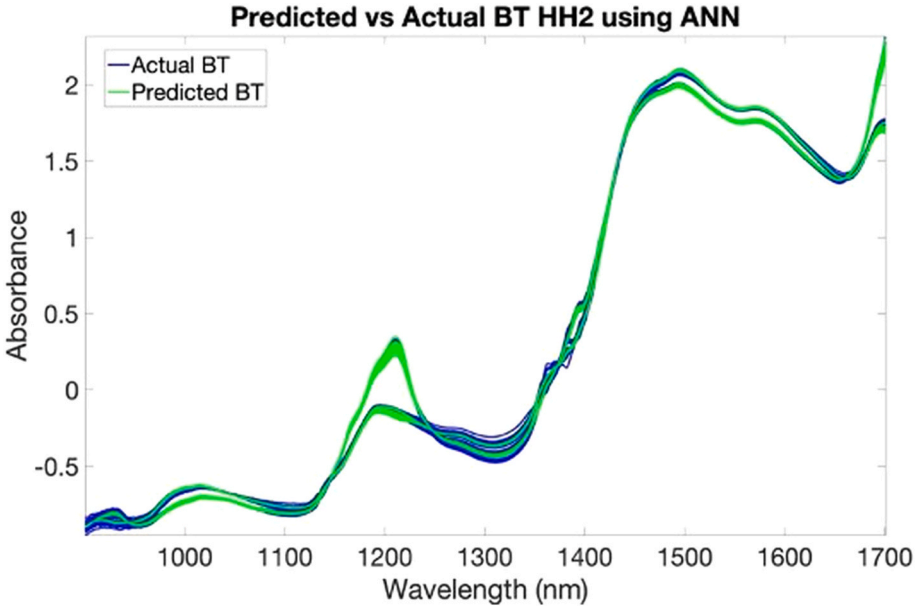


Fig. 7. Actual vs predicted BT spectra using an ANN for transfer of HH2 spectra.

instrument, which could partly contribute to the observed differences between HH1 and HH2. There is also a six month time interval between measurements on the HH1 and HH2 instruments, adding to the instrument-to-instrument variation. The ability of the ANN to maintain accurate calibration transfer under these conditions therefore demonstrates its robustness for quality control industrial applications. According to these results, ANNs are suitable for sensor-independent calibration transfer, while PDS is not.

Overall, the protein results are very promising with the transformed HH1 and HH2 spectra inputted into the BT calibration set model. ANNs are more accurate for the HH2 spectra, however for the original HH1 calibration and test set, PDS is more accurate. Overall, ANNs give the best result for calibration transfer of the HH2 dataset. The successful transfer of the HH2 spectra suggests that calibration transfer may be sensor independent. However, further work with testing across more instruments, larger sample sets, varying measurement times and operating conditions is needed to verify the sensor independent potential of this calibration transfer. As well as this, using reference measurements from wet-chemistry and not an NIR instrument is necessary for further work to verify model performance.

3.2.4. *Fat*

3.2.4.1. PDS. For fat, as with the protein model, a PDS calibration transfer model was created using the BT and HH1 calibration sets. The differences in spectra with various preprocessing applied are shown in Table S5 of the Supplementary Material. As with the classification model and the protein regression model, the smallest difference was for SNV with a window size of 31. The results are promising, with the HH1 calibration set showing a $RMSE_P$ of 0.81% compared to the BT master $RMSE_{CV}$ of 0.76% and $RMSE_P$ of 0.76%. For the HH2 spectra, the $RMSE_P$

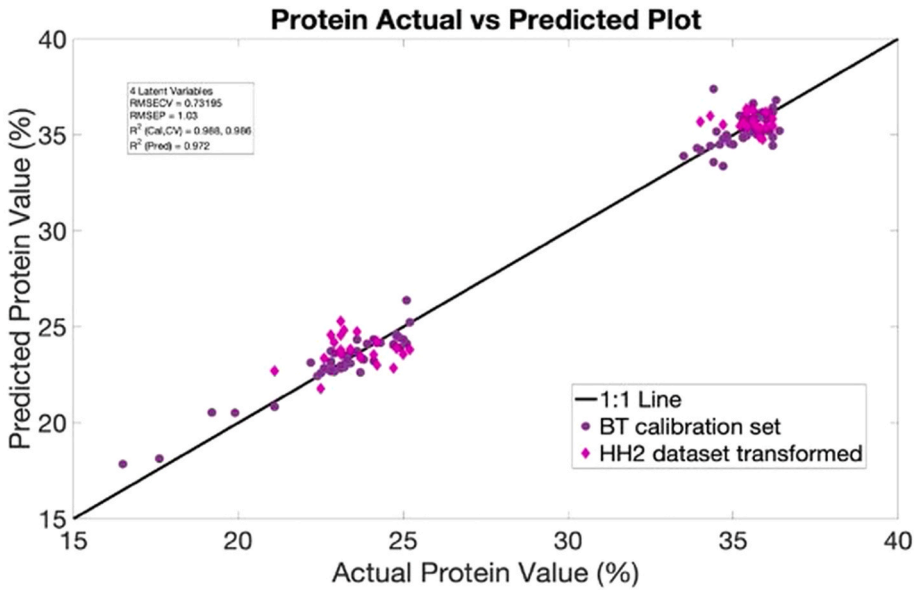


Fig. 8. Actual vs predicted plot for the HH2 spectra protein dataset transformed with the ANN inputted into the protein BT calibration set plotted with the BT calibration set.

of the model at 7.03% was much higher than the $RMSE_P$ of the BT model at 0.76% and the transformed HH1 calibration and test set at 0.81% and 0.66% respectively. This shows that PDS does not work well for transforming the HH2 spectra to the BT spectra for the master fat model, which is consistent with the result for protein.

3.2.4.2. ANNs. For fat, for the HH1 calibration set, the $RMSEP_P$ value (1.83%) is higher than for the PDS (0.81%). For the HH1 test set, the result also worsened, with an $RMSE_P$ of 2.44%. Therefore, for fat, as with protein, PDS is superior for transfer of the HH1 spectra. For the HH2 spectra, the $RMSE_P$ value after transfer is 2.18%. This is a reduction in error of 4.85. Therefore, ANNs are superior for transfer of the HH2 spectra for fat. This is consistent with the results for protein, where PDS was superior for HH1 transfer, and ANNs were superior for HH2 transfer. The transfer of HH2 using ANN is slightly worse than the protein dataset, where the reduction was 6.38 and the actual vs predicted plot (Fig. 8) showed less of a shift than in Fig. 9.

The results of each model are listed in Table 4.

3.3. Classification

The BT model achieved 100% accuracy with 1 LV and SNV preprocessing. The calibration and test set confusion matrix had no false predictions, as can be seen in Fig. S3 of the Supplementary Material.

Cubic spline interpolation was first trialled to attempt calibration transfer between the BT and HH1 spectra. The HH2 spectra were also transformed using cubic spline interpolation. When the BT classification model was applied to the HH1 calibration set and HH2 spectra after cubic spline interpolation, 100% accuracy was achieved for both.

The differences in spectra with various preprocessing and PDS applied are shown in Table S6 of the Supplementary Material. For PDS, when the BT model was applied to the transformed calibration set of the

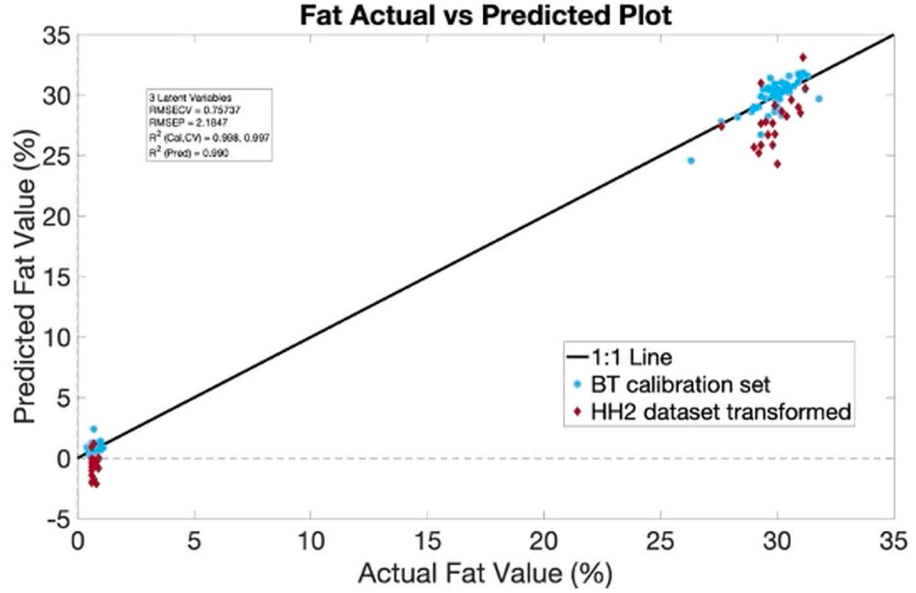


Fig. 9. Actual vs predicted plot for the HH2 spectra transformed with the ANN inputted into the BT calibration set plotted with the BT calibration set for fat.

HH1 spectra, the accuracy was again 100%. This proves that a PDS model can successfully be used to transform the HH1 into BT spectra and give the same results as the original BT spectra in the classification model. To confirm this further, the PDS model created with the calibration sets of the BT and HH1 instruments was applied to the test set of the HH1 spectra. The accuracy was once again 100%. The HH2 spectra were inputted into the PDS model. These transformed spectra were then inputted into the BT classification model. The specificity and sensitivity are 1 and 0 respectively. This model is inaccurate as it predicted every sample as FMP. This may be due to either the white reference being different, small differences between instruments or the samples having decreased protein and increased moisture values when measured with the HH2 instrument. This is a disadvantage for industry as a new calibration transfer PDS model could have to be built when new instruments are bought if conditions are not kept exactly the same or if the instruments are slightly different despite being the exact same type. It is surprising that the cubic spline interpolation performs better than PDS for classification as for regression, PDS was superior.

Using an ANN for calibration transfer of HH1 calibration and test set and HH2 spectra gave 100% accuracy. As with the regression, ANNs are therefore a superior method over PDS for classification for the HH2 calibration transfer. For classification, cubic spline interpolation or ANNs are recommended for sensor-independent calibration transfer.

3.4. Wavelength-resolved errors

Wavelength-resolved mean error profiles for ANN and PDS are presented in Fig. S4 of the Supplementary Information. For HH1, the ANN and PDS mean error profiles overlap almost completely across the entire spectral range of 900-1700 nm, indicating that both methods perform equivalently in spectral space for the instrument on which they were trained. A consistent pattern is observed: positive error in the shorter wavelength region (900-1200 nm), crossing zero at approximately 1200 nm and again at approximately 1400 nm, with negative error dominating beyond 1450 nm.

For HH2, while both ANN and PDS follow the same general error pattern observed for HH1, the PDS mean error profile shows considerable instability and noise, particularly in the 1200-1400 nm region, in contrast to the ANN, which produces a more stable correction across the full spectral range. The ANN appears better able to generalise its spectral correction to HH2, producing a more stable transfer across the full wavelength range. This spectral-level observation provides a mechanistic explanation for the superior prediction performance of the ANN for HH2, and highlights a key limitation of local linear transfer methods when applied to instruments outside the training domain.

Of particular note is the 1600-1700 nm region, previously identified as a source of spectral distortion between HH1 and HH2 (Fig. S1). In this region, both methods show comparable performance after transfer, suggesting that both ANN and PDS are able to partially accommodate the instrument-specific divergence in this region. Nevertheless, the greater instability of PDS across the 1200-1400 nm range, which encompasses spectral features relevant to protein and determination in milk powder, provides a spectral-level explanation for the inferior master model prediction performance of PDS for HH2. The ANN produces a more stable and consistent spectral correction across the full wavelength range, which appears to translate into better generalisation to an unseen instrument in terms of compositional prediction accuracy.

3.5. Sustainability

The sustainability of an analytical method can be improved in a number of ways, for example through using green solvents or none at all, miniaturization or reducing energy consumption [34]. The work reported in this paper employs an eco-friendly analytical method that reduces the use of solvents from the wet chemistry methods for protein and fat. The effectiveness of the proposed calibration transfer method would lead to a reduced need for additional reference methods when purchasing new spectroscopic instruments. The only solvent used is approximately <10 mL of IPA for cleaning the window of the handheld instruments (IPA is classed as a green solvent) [35]. Along with using green solvents and reducing solvent use, this paper demonstrates the utility of handheld instruments for analytical measurements that would traditionally be performed using a benchtop instrument. As well as reducing cost, the fact that these handheld instruments can be used in-situ instead of transporting samples from the processing line to the quality control laboratory, means that transport time and costs are reduced and sustainability is increased [34]. The analytical method in this paper facilitates an increase in sustainability and fulfils the three fundamentals of sustainability [34], as milk powder production helps meet the societal need for nutrition, the small amount of 'green' solvent used replaces wet chemistry methods that use large amounts of solvents and the economy is benefited by increasing export opportunities with milk powder.

4. Conclusions

The need for rapid, low cost, portable and sustainable methods for milk powder quality control is increasing. Calibration transfer is possible for regression and classification models for milk powder samples between a master benchtop instrument and two of the same slave handheld instruments. Various methods of calibration transfer were trialled: cubic spline interpolation; PDS; and ANNs. High accuracy and low $RMSE_P$ values were achieved for the calibration transfer of the HH1 and HH2 spectra for classification and protein and fat regression models. ANNs gave results that show that the HH2 instrument can be used in the BT protein regression model, showing that ANNs for calibration transfer are reasonably robust against time, differences between instruments of the same specification and also varying white reference measurement (i.e. white tile vs factory) method. Therefore, ANNs can facilitate sensor-independent calibration transfer. This could save industry significant time and costs by eliminating the need to develop new calibrations when commissioning a new handheld spectroscopic instrument for quality control purposes. However, further work before industry adoption should include reference data instead of predictions from another NIR instrument to validate these results. Calibration transfer between benchtop and handheld instruments for milk powder samples also fulfils the three fundamentals of sustainability. Adding additional data across more seasons would improve the robustness of the models more. As well as this, investigating the standard error of laboratory is needed in the future.

CRediT authorship contribution statement

Áine M. Ní Fhuaráin: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft. **Gonçalo M. Guedes:** Conceptualization, Methodology, Visualization, Writing – review & editing. **Ming Zhao:** Methodology, Resources, Writing – review & editing. **António C.S. Ferreira:** Funding acquisition, Resources, Writing – review & editing. **Colm P. O'Donnell:** Funding acquisition, Resources, Writing – review & editing. **Aoife A. Gowen:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Writing – review & editing.

Funding

This work has been supported by the Dairy Processing Technology Centre funded through Enterprise Ireland. Grant Agreement Number: TC-2020-0028. Additional travel support was provided by SensorFint Cost Action [CA19145] through a Short-Term Scientific Mission (STSM).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2026.105809.

Data availability

Data will be made available on request.

References

[1] A. Bista, N. McCarthy, C.P. O'Donnell, N. O'Shea, Key parameters and strategies to control milk concentrate viscosity in milk powder manufacture, Int. Dairy J. 121 (2021) 105120, https://doi.org/10.1016/j.idairyj.2021.105120.

[2] Y. Li, X. Jia, Z. Wang, Z. He, M. Zeng, J. Chen, Characterizing changes in maillard reaction indicators in whole milk powder and reconstituted low-temperature pasteurized milk under different preheating conditions, J. Food Sci. 87 (1) (2022) 193–205, https://doi.org/10.1111/1750-3841.15989.

[3] S. Grassi, L. Strani, C. Alamprese, N. Pricca, E. Casiraghi, G. Cabassi, A FT-NIR process analytical technology approach for milk renneting control, Foods 11 (1) (2022) 33, https://doi.org/10.3390/foods11010033.

[4] F.A. Iñón, S. Garrigues, M. De La Guardia, Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques, Anal. Chim. Acta 513 (2) (2004) 401–412, https://doi.org/10.1016/j.aca.2004.03.014.

[5] P. Larkin, Infrared and Raman Spectroscopy: Principles and Spectral Interpretation, second ed., Elsevier, Amsterdam, 2018 https://doi.org/10.1016/C2010-0-68479-3.

[6] Y.Y. Pu, C. O'Donnell, J.T. Tobin, N. O'Shea, Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing, Int. Dairy J. 103 (April) (2020) 104623, https://doi.org/10.1016/J.IDAIRYJ.2019.104623.

[7] Á.M. Ní Fhuaráin, C.P. O'Donnell, J. Luo, A.A. Gowen, A review on MIR, NIR, fluorescence and Raman spectroscopy combined with chemometric modeling to predict the functional properties of raw bovine milk, ACS Food Sci. Technol. 4 (10) (2024) 2258–2271, https://doi.org/10.1021/acsfoodscitech.4c00130.

[8] T.M.P. Cattaneo, S.E. Holroyd, The use of near infrared spectroscopy for determination of adulteration and contamination in milk and milk powder: updating knowledge, J. Near Infrared Spectrosc. 21 (5) (2013) 341–349, https://doi.org/10.1255/jnirs.1077.

[9] ISO/IDF, Milk and Milk Products - Guidelines for the Application of near Infrared Spectrometry. ISO 21543:2020. IDF 201:2020, International Dairy Federation, Belgium, 2020.

[10] T.P. Czaja, S.B. Engelsen, Why nothing beats NIRS technology: the green analytical choice for the future sustainable food production, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. (April 2024) 325, https://doi.org/10.1016/j.saa.2024.125028, 2025.

[11] D. Cozzolino, The contribution of digital and sensing technologies and big data towards sustainable food supply and value chains, Sustain. Food Technol. (2024) 181–187, https://doi.org/10.1039/d4fb00317a.

[12] A. Folch-Fortuny, R. Vitale, O.E. de Noord, A. Ferrer, Calibration transfer between NIR spectrometers: new proposals and a comparative study, J. Chemom. 31 (3) (2017) 1–11, https://doi.org/10.1002/cem.2874.

[13] I.R. Rukundo, M.G.C. Danao, R.B. Mitchell, C.L. Weller, Evaluation of predictive performance of PLS regression models after being transferred from benchtop to handheld NIR spectrometers, Biosyst. Eng. 218 (2022) 245–255, https://doi.org/10.1016/j.biosystemseng.2022.04.014.

[14] V.H. Da Silva, J.J. Da Silva, C.F. Pereira, Portable near-infrared instruments: application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer, J. Pharm. Biomed. Anal. 134 (2017) 287–294, https://doi.org/10.1016/j.jpba.2016.11.036.

[15] T. Fearn, Standardisation and calibration transfer for near infrared instruments: a review, J. Near Infrared Spectrosc. 9 (4) (2001) 229–244, https://doi.org/10.1255/jnirs.309.

[16] J.J. Workman, A review of calibration transfer practices and instrument differences in spectroscopy, Appl. Spectrosc. 72 (3) (2018) 340–365, https://doi.org/10.1177/0003702817736064.

[17] J. Eliaerts, N. Meert, P. Dardenne, F. Van Durme, V. Baeten, N. Samyn, K. De Wael, Evaluation of a calibration transfer between a bench top and portable Mid-InfraRed spectrometer for cocaine classification and quantification, Talanta 209 (August 2019) (2020) 120481, https://doi.org/10.1016/j.talanta.2019.120481.

[18] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemom. Intell. Lab. Syst. 64 (2) (2002) 181–192, https://doi.org/10.1016/S0169-7439(02)00085-0.

[19] X.Y. Yap, K.S. Chia, N.A.S. Suarin, Adaptive artificial neural network in near infrared spectroscopy for standard-free calibration transfer, Chemom. Intell. Lab. Syst. 230 (June) (2022) 104674, https://doi.org/10.1016/j.chemolab.2022.104674.

[20] F. Despagne, B. Walczak, D.L. Massart, Transfer of calibrations of near-infrared spectra using neural networks, Appl. Spectrosc. 52 (5) (1998) 732–745, https://doi.org/10.1366/0003702981944157.

[21] H. Chen, C. Tan, Z. Lin, T. Wu, Classification and quantitation of milk powder by near-infrared spectroscopy and mutual information-based variable selection and partial least squares, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 189 (2018) 183–189, https://doi.org/10.1016/j.saa.2017.08.034.

[22] S.R. Karunathilaka, B.J. Yakes, K. He, J.K. Chung, M. Mossoba, Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: a study using eleven potential adulterants, Heliyon 4 (9) (2018) e00806, https://doi.org/10.1016/j.heliyon.2018.e00806.

[23] M.R.C. Inácio, M.D.F.V. De Moura, K.M.G. De Lima, Classification and determination of total protein in milk powder using near infrared reflectance spectrometry and the successive projections algorithm for variable selection, Vib. Spectrosc. 57 (2) (2011) 342–345, https://doi.org/10.1016/j.vibspec.2011.07.002.

[24] L.J. Mauer, A.A. Chernyshova, A. Hiatt, A. Deering, R. Davis, Melamine detection in infant formula powder using Near- and mid-infrared spectroscopy, J. Agric. Food Chem. 57 (10) (2009) 3974–3980, https://doi.org/10.1021/jf900587m.

[25] L. Salguero-Chaparro, B. Palagos, F. Peña-Rodríguez, J.M. Roger, Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer, Comput. Electron. Agric. 96 (2013) 202–208, https://doi.org/10.1016/j.compag.2013.05.007.

[26] A.L. Bizerra Brito, A.V. Pereira Santos, K.D. Tavares Melo Milanez, M. José Coelho Pontes, L.F. Bezerra Lira Pontes, Calibration transfer of flour NIR spectra between benchtop and portable instruments, Anal. Methods 9 (21) (2017) 3184–3190, https://doi.org/10.1039/c7ay00391a.

[27] M. Frizzarin, T.F. O'Callaghan, T.B. Murphy, D. Hennessy, A. Casa, Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets, J. Dairy Sci. 104 (12) (2021) 12394–12402, https://doi.org/10.3168/jds.2021-20812.

[28] G. Visentin, A. McDermott, S. McParland, D.P. Berry, O.A. Kenny, A. Brodkorb, M. A. Fenelon, M. De Marchi, Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows, J. Dairy Sci. 98 (9) (2015) 6620–6629, https://doi.org/10.3168/jds.2015-9323.

[29] A.R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1) (1969) 137–148.

[30] X. Wang, C. Esquerre, G. Downey, L. Henihan, D. O'Callaghan, C. O'Donnell, Development of chemometric models using Vis-NIR and raman spectral data fusion for assessment of infant formula storage temperature and time, Innov. Food Sci. Emerg. Technol. 67 (January) (2021) 102551, https://doi.org/10.1016/j.ifset.2020.102551.

[31] B.G. Osborne, T. Fearn, P.T. Hindle, Practical NIR Spectroscopy: with Applications in Food and Beverage Analysis, second ed., Longman Scientific & Technical: Essex, England; New York, 1993.

[32] S.E. Holroyd, The use of near infrared spectroscopy on milk and milk products, J. Near Infrared Spectrosc. 21 (5) (2013) 311–322, https://doi.org/10.1255/jnirs.1055.

[33] R. Tange, M.A. Rasmussen, E. Tairac, R. Bro, Application of support vector regression for simultaneous modelling of near infrared spectra from multiple process steps, J. Near Infrared Spectrosc. 23 (2) (2015) 75–84, https://doi.org/10.1255/jnirs.1149.

[34] J. Płotka-Wasylka, H.M. Mohamed, A. Kurowska-Susdorf, R. Dewani, M.Y. Fares, V. Andruch, Green analytical chemistry as an integral part of sustainable education development, Curr. Opin. Green Sustain. Chem. 31 (2021) 100508, https://doi.org/10.1016/j.cogsc.2021.100508.

[35] M. Tobiszewski, S. Tsakovski, V. Simeonov, J. Namieśnik, F. Pena-Pereira, A solvent selection guide based on chemometrics and multicriteria decision analysis, Green Chem. 17 (10) (2015) 4773–4785, https://doi.org/10.1039/c5gc01615k.