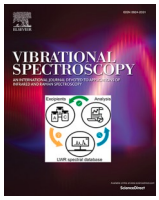
Vibrational Spectroscopy 132 (2024) 103667

Contents lists available at ScienceDirect

Vibrational Spectroscopy

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



Calibration transfer between different spectrometers by wavelength correspondence

Erik Tengstrand *, Lars Erik Solberg, Katinka Dankel, Tiril Aurora Lintvedt, Nils Kristian Afseth, Jens Petter Wold

Nofima AS – Norwegian Institute of Food, Fisheries and Aquaculture Research, PB 210, NO-1431, Ås, Norway

ARTICLE INFO

Keywords:
Calibration transfer
Spectroscopy
NIR
Raman
Chemometrics

ABSTRACT

In this paper we present a method for transferring calibrations between different spectrometers based on assigning wavelength correspondence. It has been tested for near-infrared (NIR) and Raman spectroscopic instruments, and three examples are included in the paper. The calibration transfer is done in three steps: first wavelength correspondence is established. Second, PLS models are built and tuned for the new spectrometer. Third, the PLS models are slope and bias corrected. The advantages with this approach are that it does not require transfer samples and that there is only one parameter to tune: the number of PLS components. While a few samples with reference values are required for the tuning, it is fewer than methods with multiple parameters that need to be tuned.

1. Introduction

Vibrational spectroscopy for analytical measurements has become very common in industry because of two advantages: it is fast, and it is cheap. Most other versatile chemical analysis methods are more expensive and far slower. Near-infrared spectroscopy (NIRS) can provide information about sample traits and composition in milliseconds and is being used for in-line sorting of for instance food, pharmaceuticals, and plastics. Raman spectroscopy is usually slower, and requires seconds to do the same, which is still orders of magnitude faster than other chemical analyses such as chromatography. As added advantages, the vibrational spectroscopic methods require little to no sample preparation and are non-destructive. The trade-off is that they require extensive calibration work. This calibration can be time-consuming and expensive. Since many vibrational spectroscopic methods are used in industry, there are often multiple spectrometers calibrated for the same analysis. Either because of location: the analysis needs to be done in several plants or factories; or because of capacity: if large quantities of small items need to be sorted, even millisecond speeds are insufficient if only a single spectrometer is used. Even spectrometers of the same brand may be sufficiently different, so that a calibration developed on one spectrometer has lower accuracy when used on another spectrometer. When calibrations are used on spectrometers of different brands, models usually lose more accuracy or fail completely. The expensive and time-consuming calibration combined with multiple spectrometers calibrated for the same analysis makes calibration transfer very valuable, since both time and money can be saved by reusing the calibration. Datasets are also increasingly available for download on the internet. If those can be utilized, new applications can also be developed at a reduced cost.

Calibration transfer is the process of using the same data to calibrate multiple spectrometers. For the calibration transfer the different spectrometers are usually labeled: Source, master, or parent for the spectrometer which the data used for modelling is collected on. Target, slave, or child for the spectrometer which the transferred model is deployed on. Throughout this article, the terms source and target spectrometer are used.

There are several strategies for transferring calibrations [1,2], and there is a plethora of methods for those strategies [3]. Perhaps the most noteworthy method is piecewise direct standardization (PDS) [4]. PDS uses a set of transfer samples, i.e. samples that have been measured by both spectrometers, to model each wavelength (e.g. 1000 nm) in the source spectrometer using a few nearby wavelengths (e.g. 990–1010 nm) in the target spectrometer. A PLS model is made for each wavelength in the source spectrometer. Spectra acquired on the target spectrometer are then used with the wavelength models to predict what the spectra from the same sample would look like on the source spectrometer, and the new spectra are then used with the original model. The reverse, called reverse PDS, is when the source spectra are instead

* Corresponding author.
E-mail addresses: tengstrand@gmail.com, erik.tengstrand@nofima.no (E. Tengstrand).

https://doi.org/10.1016/j.vibspec.2024.103667
Received 8 December 2023; Received in revised form 14 February 2024; Accepted 20 February 2024
Available online 27 February 2024

converted to the target spectrometer, and a new model is built. PDS has been used in numerous publications [5–11]. PDS should be considered the state of the art. There are many newer methods published, but PDS still sees very widespread use. Partially because of its relative simplicity.

Another common method is slope and bias correction (SBC). It relies on the spectrometers being standardized [1]. The standardization procedure ensures that the spectrometers produce almost identical spectra, for instance with respect to wavelength correspondence and intensity. Sometimes small differences between the spectrometers cause small residual errors, normally slope and bias errors. In response to these errors, slope and bias correction has been used successfully. It is a simple method, but it is becoming obsolete: some modern spectrometers are standardized well enough that slope and bias correction is unnecessary. Spectrometers of different brands are not standardized with regards to each other, so slope and bias correction fails.

In cases where the spectrometers are of the same brand and made by the same manufacturer, calibration transfer is usually an easy challenge: many manufacturers take care to standardize their spectrometers to simplify calibration transfer. When the spectrometers are from different manufacturers, or even of different types, e.g. grating, scanning, or Fourier-transform spectrometers, calibration transfer becomes more challenging. Differences in optical setups and sample handling can further increase the challenge.

In response to these challenges, many calibration transfer methods have been developed, some of which are quite complex. PDS mentioned above requires selecting the number of components for the PLS models, perhaps at each target spectral variable, and this may lead to discontinuities as stated in the literature [1]. Other methods that directly handle instruments with differing spectral support include trimmed-scores regression (TSR [12]) and principal components canonical correlation analysis (PC-CCA [13]). TSR formulates the problem as one of missing data and attempts to build the "missing" calibration set from existing observations. While not requiring parameters, the pseudo-inverses involved in the algorithm likely make the method sensitive to the size and selection of data. PC-CCA on the other hand builds a transform between "common" subspaces as identified by the CCA process. There are hence several steps with hyperparameter optimization involved. Again, it is expected that the method will require a number of samples both for the identification of subspaces and for estimating the regressions. If one assumes a prior alignment of variables using interpolation, then other methods are open for comparison. For instance, simply re-building the calibration by combining calibration data from the source and target datasets if reference values are present. This method is sometimes called "update" or "global" [1]. A close neighbour would be to explicitly suppress sensitivity to differences between domains such as with the calibration transfer by generalized least-squares [14] or null-augmented regression [15]. With both examples there are considerations to address such as how to weigh the samples in the target domain in the former, and how much of the difference to use in the latter two.

In this article we provide a comparatively simple calibration transfer method and show how it can solve the calibration transfer problem for both NIR and Raman spectroscopy, even when the source and target spectroscopic instruments have large differences. The method could be qualified as "simple" both because there are no tuning parameters, because it is expected to work with comparatively few samples (univariate, linear regression) and because it does not require standardization samples.

There can be many differences between different spectrometers: wavelength correspondence, number of spectral variables, spectral resolution, signal intensity, wavelength range, and environmental sensitivity, to name the most important. Correcting all the differences independently can be very challenging. The method presented in this article is based on the assumptions that the wavelength correspondence is the dominant model error, and that most other differences end up creating slope and bias errors, or overfitting errors. By correcting for the type of error, rather than for the difference between the spectrometers, the calibration transfer is simplified. Making fewer corrections lead to more robust results with less data required, which is ideal in most industrial settings. Wavelength correspondence can be assigned using linear interpolation based on wavelengths provided by the spectrometer vendors. It can also be done with wavelength references such as xenon, argon, or mercury lamps, or specialized wavelength reference tiles.

More complex approaches exist to handle more difficult problems. Some, like PDS, use standardization samples which potentially suffer from the disadvantage of having to measure identical samples on both devices (distance between devices, if one device is broken). Others, as argued above, may require more samples and possibly expert supervision. The proposed method is a natural combination of existing procedures. The novelty is more in advocating the idea that a simple approach with a clear description of the errors it is compensating for is valuable information for the application of calibration transfer, particularly in a field where new methods are presented on a regular basis with a trend of increasing complexity.

2. Materials and methods

2.1. Data

In this work we consider three cases. The details of the spectrometers used in the three cases are summarized in Table 1.

The example data used in Case 1 *Salmon* was collected from previously published work on estimation of fatty acids in salmon. Fat rich fish, such as salmon, is a major source of marine fatty acids in the human diet. Docosahexaenoic acid (DHA) is one of the most important omega-3 fatty acids which has been connected with health effects for human consumption [16–19] as well as fish welfare and fillet quality [20,21]. It is therefore an important quality parameter for the farming industry.

In this case, the source data was acquired on 618 samples of ground salmon using a Kaiser RamanRXN2 Multi-channel Raman analyzer (Kaiser Optical Systems, Ann Arbor, MI, USA), with a 500–1800 cm^{-1} Raman shift range and a 4 cm^{-1} resolution [22]. The target data was collected on 51 intact salmon fillets using a MarqMetrix All-in-One (AIO) Raman system with a 100–3250 cm^{-1} Raman shift range and a 6.5 cm^{-1} resolution [23]. The analyte chosen, the omega-3 fatty acid DHA, was quantified using GC/MS [22,23].

In *Case 2 Bean flour* the samples were prepared by dry fractionation of flours from different varieties of fava beans [24,25]. With a steadily growing population and new consumer trends, bean flour protein has been identified as a nutritious and inexpensive vegetable source of protein. Using pin mill separation, the protein content of the flour can be increased to the desired concentration, and NIR spectroscopy can be used to monitor the product composition.

In this case, the source and target spectra were acquired in reflectance mode on the same samples, and two target data sets were compared. The source data was collected using a FOSS NIR Systems XDS Rapid Content Analyzer (FOSS Analytical, Hillerød, Denmark) in the wavelength region 400–2500 nm with a 0.5 nm resolution. The *Target 1* data was collected using a Prediktor Spektron 1700 (Prediktor, Fredrikstad, Norway) inline NIR spectrometer with a wavelength range of x nm and a resolution of x nm. The *Target 2* data was acquired using a VISUM Palm (Iris, Catalonia, Spain) handheld NIR instrument with a wavelength range of 980–1600 nm and a resolution of 1.5 nm. The protein contents of the samples were measured as 6.25 times the nitrogen content determined using Dumas method.

In Case 3 *Dried salted cod (Clipfish)* Clipfish is a traditional product where the fish is first salted and then dried. The main quality feature is the amount of water in the fish. It has previously been shown that it is possible to estimate the water content by the use of hyperspectral interactance imaging in the NIR region 760–1040 nm [26]. This was done with the Qvision500 industrial scanner (TOMRA sorting solutions, Leuven, Belgium), The imaging approach was useful for proper sampling

Table 1
the spectrometers used for the different cases, and their respective characteristics.

Case	Data set	Instrument	Wavelength region	Optical resolution	Digital resolution	Sample	Analyte	Analyte range (%)
Case 1	Source	Kaiser RamanRXN2	500–1800 cm^{-1}	4 cm^{-1}	0.3 cm^{-1}	Ground salmon fillet	DHA fatty acid	5–9%
Case 1	Target	MarqMetrix AIO Raman system	100–1800 cm^{-1}	6.5 cm^{-1}	2 cm^{-1}	Salmon fillet	DHA fatty acid	2–7%
Case 2	Source	FOSS NIR Systems XDS	400–2500 nm	2 nm	0.5 nm	Bean flour	Protein	28–78%
Case 2	Target 1	Prediktor Spektron NIR sensor	1020–1700 nm	8 nm	3 nm	Bean flour	Protein	28–78%
Case 2	Target 2	VISUM Palm handhelt NIR	980–1600 nm	3 nm	1.5 nm	Bean flour	Protein	28–78%
Case 3	Source	QVision Scanner	760–1040 nm	20 nm	20 nm	Dried salted fish	Water	42–57%
Case 3	Target	SmartSensor	760–1280 nm	17 nm	17 nm	Dried salted fish	Water	42–57%

since the water is unevenly distributed in the fish. The interactance measurement mode is needed to measure in depth in the fish, also to obtain representative spectra of the average water content. In the clip-fish industry there is a need for a portable/handheld instrument for water content assessment. We are then limited to point interactance measurements and that these measurements are done at six different sites on the fish to obtain a satisfactory measurement of the average water content. Such measurements has been done with a prototype NIR system (SmartSensor) previously described by Wold et al. [27]. In this study it is of interest to see if it is possible to successfully transfer a calibration from the imaging system to the point based system.

3. Calibration transfer method

1. The first step of this calibration transfer method is assigning the wavelength correspondence between the two spectrometers. In this paper it was done using the wavelength correspondences supplied by the spectrometer vendors, and linear interpolation between the pixels. The linear interpolation was done using MATLAB and the interp1 function, with the default settings. The interp1 function takes the spectra, and the wavelengths of the two spectrometers and does linear interpolation between the datapoints in the spectra to estimate the spectral intensities at the wavelengths in the target spectrometer. If the wavelength correspondences supplied by the vendors are inaccurate, a wavelength reference can be used instead.
2. When the wavelength ranges for the target and source spectrometers are different, only the overlapping range is selected.
3. The adjusted spectra from the source spectrometer are then used to make a new model with the number of PLS components selected using data from the target spectrometer. The same preprocessing methods that were used for the original models on the source spectrometers are used on the target models.
4. The predictions are then slope and bias corrected with linear regression: a linear regression is made using a few samples (8−15) on the target spectrometer. The predictions on the target spectrometer are then multiplied by the slope, and the bias of the linear regression model is added to the predictions. This is both to correct for differences in intensities from the different spectrometers, and to correct for differences in the optics and geometries in the spectral acquisition setup.

4. Validation of model transfer

In all examples in this article, the target dataset acts as the test set, with a few samples used for selecting the number of PLS components and performing the slope-and-bias correction. In case 2 and Case 3, all samples have been measured on both instruments. The models based on those datasets were validated by holding out the same samples from both the target and the source spectrometer while the models were cross-validated. When the impact of the number of samples was tested, a random set of samples was used for selecting the number of PLS components and performing the slope-and-bias correction. The remaining samples were used for validation. This was repeated multiple times to assess the variation due to the selected samples. Fig. 1 illustrates the sample split between calibration, tuning, and validation sets.

The transferred models were also compared with native models using both the source and target datasets, i.e. models where both the calibration and test set are from the same spectrometer. In all cases the data was split randomly into a calibration set, a tuning set, and a test set. The different models were created with the same split. For case 2 and Case 3, which have the same samples in the source and target datasets, the split is the same for the source and the target. For the case 2 and Case 3, 8 samples were used for tuning. For Case 1, 15 samples were used for tuning, due to the larger differences between the datasets. PLS models with up to 10 components were created using the calibration set. The tuning set was used to select the number of components and slope and bias correct the model. The performance was then assessed using the test set. This was repeated 100 times, and the average performance is provided in Table 2. For all transferred models, the PLS regression used the entire wavelength-ranges that were overlapping for the target and source spectrometers.

5. Results and discussion

Fig. 2 shows how Raman spectra from Case 1 compare before and after the wavelength adjustment. Spectra from Source and Target have different intensities, different number of variables, and different Raman shift range. Without transforming the spectra, models based on the source instrument will fail on the target instrument. The spectra shown are from different samples; the two datasets have no samples in common. The wavelength correspondence is in this case assigned with linear interpolation using the wavelengths provided by the spectrometer vendors. In the examples, i.e. Fig. 2, the spectra after wavelength correspondence adjustment are also normalized. That is not strictly necessary as it does not impact the model performance, but it makes the manual interpretation easier.

Fig. 3 shows the performance of estimating the concentration of the fatty acid DHA for different number of components for the two spectrometers used in Case 1. The source and target spectrometers have different numbers of optimal PLS components. This is often the case when spectrometers are very different. In this case, another contributing reason for the different number of components used is the differences in composition between the samples. In the source dataset, the salmon have been given the same feed. In the target dataset, acting as an external test set, the salmon have been given different feeds, changing the fatty acid composition. For datasets where the sample composition is the same, there can still be overfitting due to changes between spectrometers, if they are different enough. For spectrometers that are sufficiently similar, and where the sample sets have the same composition, the tuning of the PLS model again for the calibration transfer is not necessary. Fig. 3 also shows a higher performance for the transferred model compared to the source model. The reason for this is that the

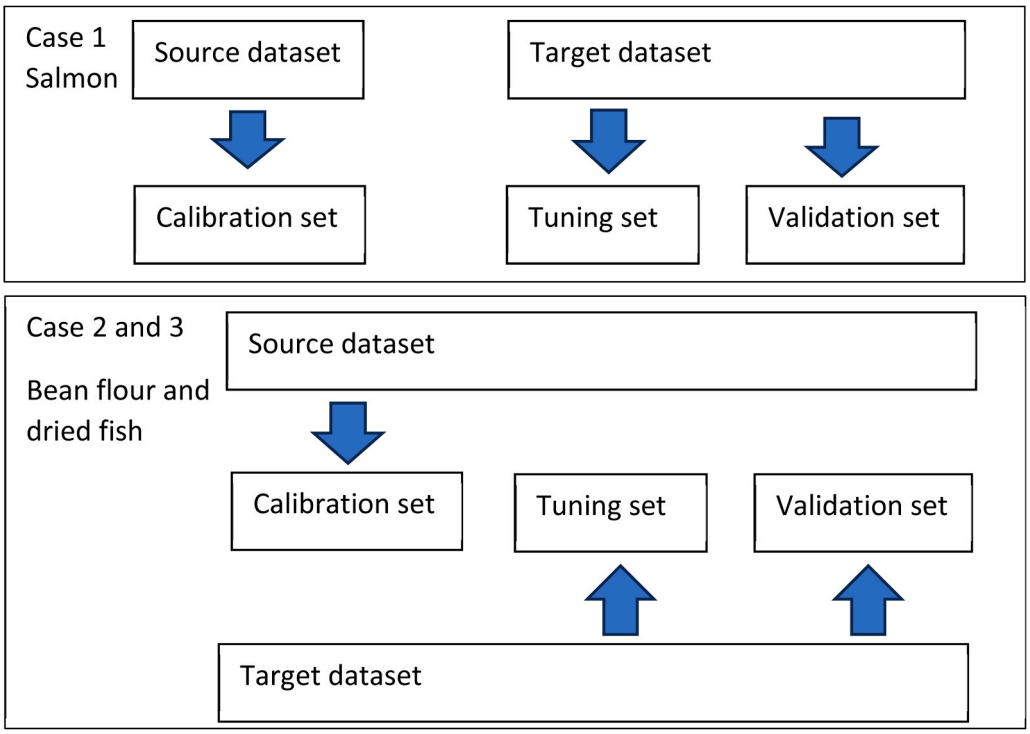


Fig. 1. An illustration of how the datasets are split into sets for calibration, tuning, and validation.

Table 2
The number of PLS components used (LVs) performances (R^2 (coefficient of determination), RMSEP (root mean square error of prediction)) of the transferred models compared to the native models on both the source and targets instruments, the source model on the parts of the source spectra that was available of the target instruments, as well as the performance when PDS was used for the calibration transfer.

Case	model original Source data		model on source spectra overlapping with target spectra				Test transfer model on test data		model of test data obtained by target instrument			Test model transferred by PDS on test data		
	LVs	R^2	RMSEP	LVs	R^2	RMSEP	LVs	R^2	RMSEP	LVs	R^2	RMSEP	LVs	R^2	RMSEP
Case 1	7	0.66	0.33	7	0.66	0.33	4	0.91	0.39	3	0.92	0.36			
Case 2 Target1	6	0.97	2.6	6	0.97	2.7	5	0.96	3.9	7	0.96	3.3	3	0.82	8.4
Case 2 Target2	6	0.97	2.6	6	0.97	2.7	2	0.84	7.2	5	0.91	5.5	5	0.66	10.7
Case 3	5	0.86	1.4	5	0.86	1.4	5	0.75	1.9	5	0.83	1.7	5	0.70	2.8

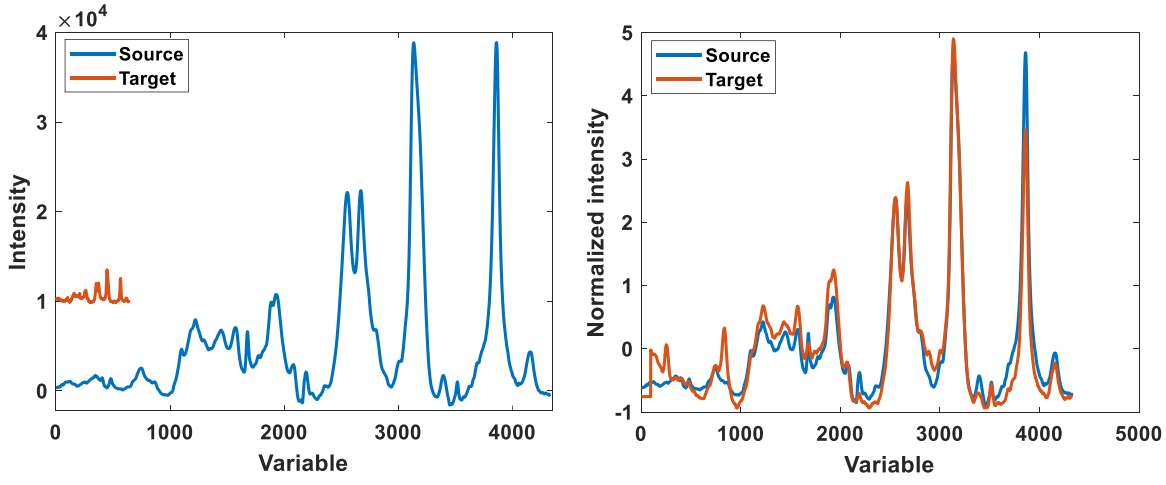


Fig. 2. Raman spectra from Case 1: from the source spectrometer where the calibration was originally developed on, and the target spectrometer that we want to deploy the model on. The figure shows these spectra plotted by variable, which is how a model is applied. The target spectrum has been offset for clarity. The right figure shows the same spectra, where the target spectrum has been interpolated to match the source spectrometer. To simplify the comparison between spectra, they have been normalized by SNV. Source and target spectra are from different samples.

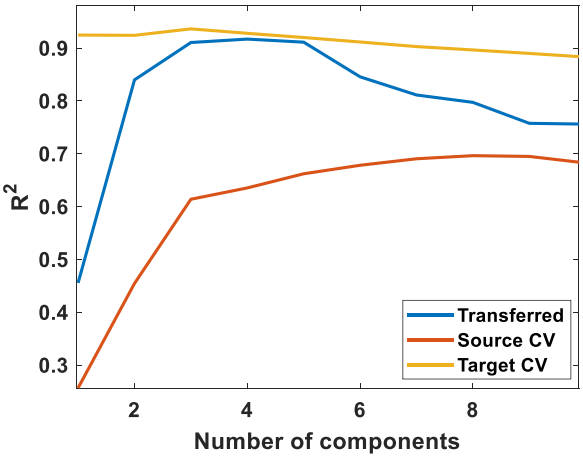


Fig. 3. This figure shows the performance (R^2 (coefficient of determination)) of the DHA (Docosahexaenoic acid) predictions in Case 1 depending on the different number of components for the transferred model, and the cross-validated (CV) performance for the source and target datasets.

target data shows much wider range in DHA, which means that the errors are relatively lower. The performance of the transferred model is not quite as high as the performance of the native target model. Part of this is due to differences in the samples, but there is likely a calibration transfer error contributing as well. In this case, there are clearly numbers of components that yield models with lower performance. In large part, this is due to the differences between the samples rather than the spectrometers: the source dataset has concentrations of DHA ranging from 5% to 9%, while the target dataset has concentrations ranging from 2% to 7%. Since the model needs to be retuned, there is a risk that the wrong number of components is used. This was tested by drawing samples at random from the target dataset and used to tune the number of components. If the standard deviation of the reference values for the drawn samples was higher than 1% DHA, this was considered too low and new samples were drawn. This was repeated 10000 times to assess how reliably good PLS parameters can be found given different number of samples. With 10 samples, the optimal number of components (3, 4, or 5) we found 80% of the time, and a passable number of components (2−6) were found 91% of the time. With 15 samples, the percentages were instead 90% and 96%, and with 20 samples they were 95% and 98%.

In these tests, the samples have been selected randomly. It is often possible to select samples more systematically. Kennard-Stone [28] has been successfully used in many applications, particularly with PDS. When selecting samples methodically rather than randomly, the performance of calibration transfer by wavelength correspondence also increases for the more complicated datasets. Case 1 also reaches stable performance with 8 samples. For case 2 and Case 3, the performance is also improved, but the change is very small since those datasets are less challenging. With PDS, sample selection is still important since PDS solves the differences between the instruments by modelling the difference in the spectra. Calibration transfer by wavelength correspondence solves the differences by direct measurements of the two instruments, either supplied by the vendors or by measuring a wavelength reference such as a xenon lamp. One important consideration is how much sample variation is required for the calibration transfer. All examples here are from the food sector. Many cases in the food sector have seasonal variation. If the seasonal variation needs to be covered for the first calibration transfer, it can become very cumbersome. Calibration transfer by wavelength correspondence does not require the seasonal variation to be covered for the initial transfer, but it can be prudent to update the transfer if the seasonal variation is significant.

In cases where the optical setups and sample handling are the same, the slope and bias errors will be much smaller, possibly to the point that the slope correction is not necessary. Using the method presented in this article, a bias correction is usually necessary, even with identical optical setups due to small errors in the assigned wavelength correspondence. Fig. 4 shows the predictions and references for Case 1. The performance is good for the transferred model, despite the differences between the spectrometers and the different range of DHA.

Case 1 have different samples, and the spectra were obtained at different times, at different locations. This is a common occurrence, and therefore a suitable test for a calibration transfer method for industrial use.

While this is not always possible to do in industry, some calibration transfer methods do require spectra from the same samples obtained from all spectrometers. Case 2 and Case 3 fulfill that criterion, which makes these datasets useful for comparing different transfer methods. They are also very useful in assessing the transfer error itself: since the samples are the same, differences in performance come from the calibration transfer. Fig. 5 shows how the performance changes with number of components in case 2. The source and the target 1 spectrometers are both larger scale instruments, with similar design but from different vendors. The target 2 spectrometer is a handheld spectrometer, and the performance behaves very differently compared to the other two. Fig. 6 shows the spectra, and despite any similarities in design, the spectra from the three spectrometers look very different. The two similar spectrometers have very similar performance until very high number of components are used. The handheld spectrometer has similar performance for the first two components, just slightly lower than the two larger spectrometers. For models using more than three components, the performance collapses completely for the handheld spectrometer, the R^2 dropping from 0.85 to 0–0.4 depending on the number of components. This shows that the model reoptimization step can be important, unless the spectrometers are sufficiently similar. Fig. 7 shows the predicted versus reference for the three spectrometers. Despite the differences between the three spectrometers, the performance for the native model on the source instrument and the transferred models on the target instruments ended up similar.

The target 1 spectrometer had similar performance regardless of how many components were used, so the number of samples used for the model tuning is essentially irrelevant: regardless of the selected number of components, the performance is fine. The target 2 spectrometer behaves more like Case 1 in that it has an acceptable range of 1–3 PLS components. With 5 samples for tuning, selected randomly, the optimal range was reached 88% of the time. With 10 samples for tuning, the optimal range was reached 99% of the time. A large reason that fewer

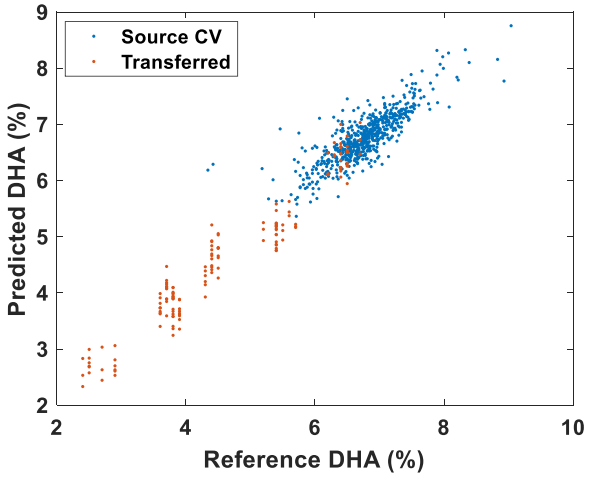


Fig. 4. A plot of the DHA (Docosahexaenoic acid) predictions and references for the source and target datasets for Case 1. The source spectrometer has cross validated predictions, while the target spectrometer shows the predictions of the test set.

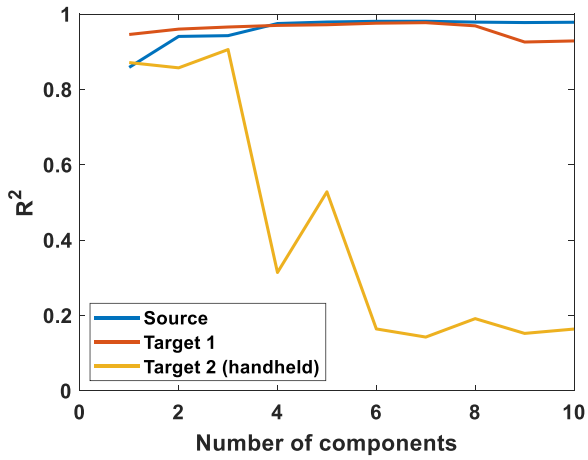


Fig. 5. The cross-validated transfer performance (R^2 (coefficient of determination)) for case 2. The same samples have been measured on the source and two target spectrometers. The data from the source spectrometer has been used for modelling with seven-fold cross-validation. It is the same model, from the source spectrometer, that has been used on all three spectrometers. The same samples have been held out in the two target spectrometers.

samples were required for this dataset, compared to Case 1 is that this dataset uses the same samples, so there is no overfitting error that complicates the situation. The more different the new samples are compared to the old samples; the more samples will be required for the tuning. For the target 1 spectrometer, the optimization curve was generally flat if at least 8 samples were used. When fewer samples were used, the optimization curve would occasionally have strong differences depending on the number of PLS components.

Fig. 8 shows how the performance changes with the number of PLS components used when PDS was used for transfer. This shows the same issue with a reduced performance for the transferred models when too many components are used, but even more than in Fig. 5. A part of this is due to the selection of transfer samples: here they were selected at random. This was repeated several times, with differences in performance every time. The target 2 spectrometer was more sensitive to the sample selection. This highlights the importance of sample selection and parameter tuning in calibration transfer. Parameter tuning and sample selection can be costly processes in terms of number of samples and level of validation required. For situations where multiple transfers are necessary, calibration transfer methods that have fewer parameters to tune can be optimal.

Case 3 uses interactance NIR. Here the interactance depth can interfere, and it has a shorter wavelength range, and lower wavelength resolution than is commonly used in NIR. Fig. 9 shows the performance

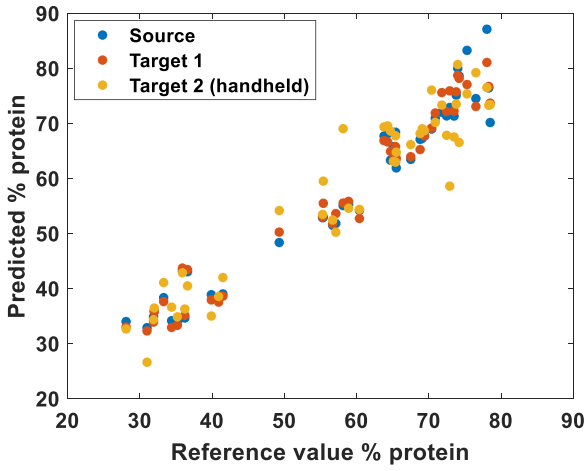


Fig. 7. The predicted versus reference for the three spectrometers in case 2. For the source data, it is the native model. For the target1 and 2, it is the transferred models.

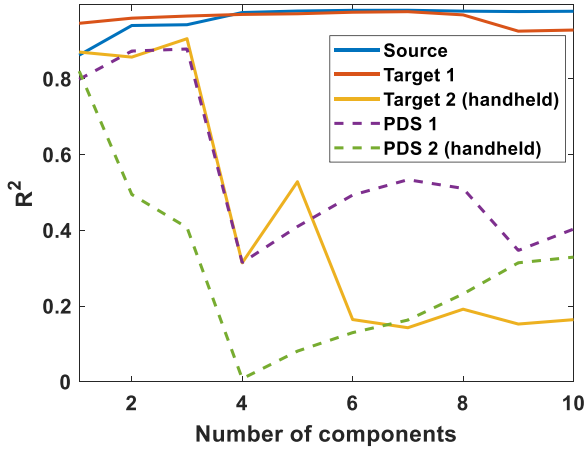


Fig. 8. The performance (R^2 (coefficient of determination)) depending on the number of components for case 2 when PDS was used for the transfer.

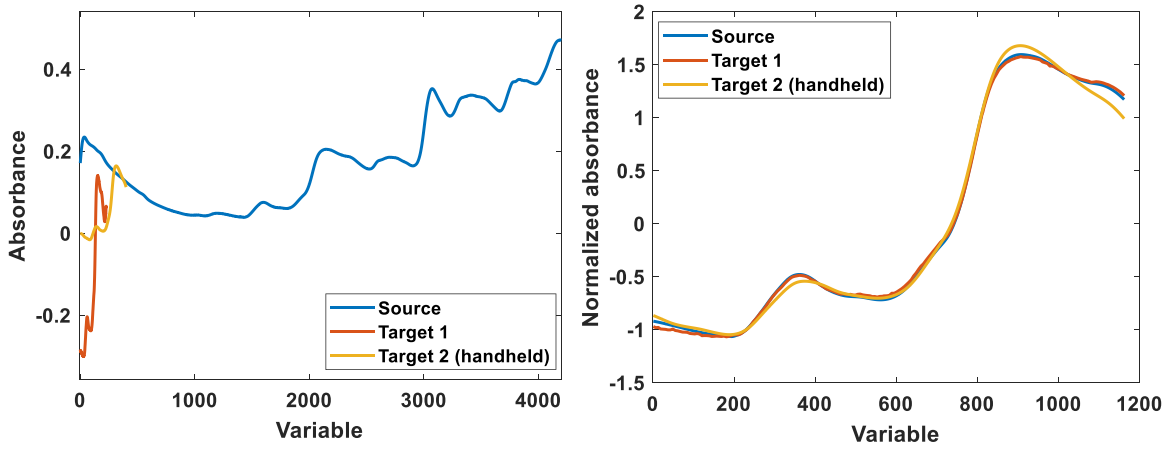


Fig. 6. Spectra from the same sample from the three different spectrometers in case 2, before (left) and after (right) wavelength correction and normalization with SNV. The figure shows these spectra plotted by variable, which is how a model is applied. The wavelength range in the source spectrometer is reduced to the overlapping range after the correction.

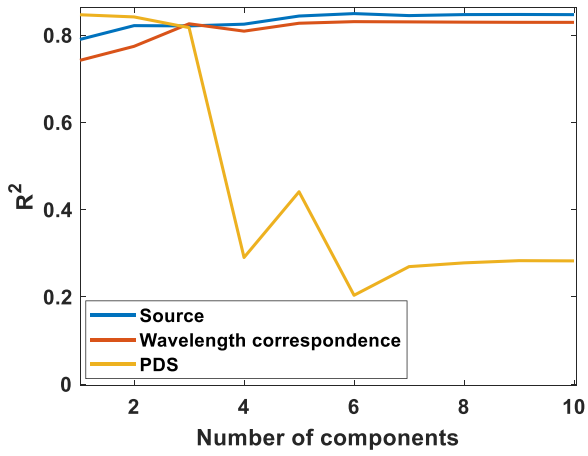


Fig. 9. The performance (R^2 (coefficient of determination)) depending on the number of components for Case 3, including the performance for PDS.

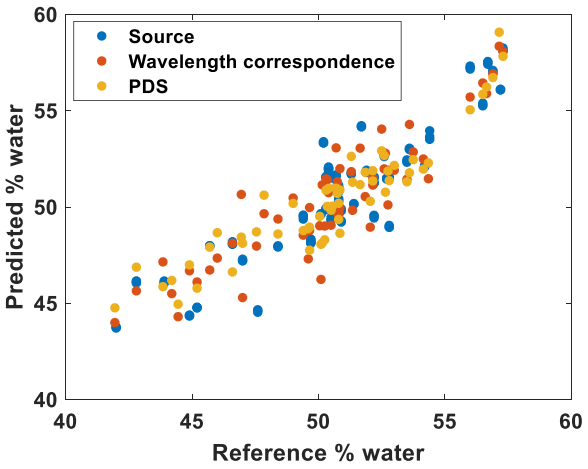


Fig. 10. The predicted versus reference for Case 3.

depending on the number of components used both for calibration transfer by wavelength correspondence and by PDS. It shows similar behavior as case 2: PDS was more sensitive to overfitting. The target spectrometer measures several depths simultaneously. PDS was more sensitive to the interactance depth as well. For PDS, this becomes a question of robustness, and validation: is there enough time and samples to select good settings to achieve the best performance?

In this figure the highest performance is achieved by PDS. Depending on which samples, and how many samples were selected for model tuning, the performance of PDS varied strongly. For case 2, calibration transfer by wavelength correspondence was consistently higher. Calibration transfer by wavelength correspondence has lower performance than expected for Case 3 because are some errors in the wavelengths supplied by least one of the instrument manufacturers. PDS only needs the approximate wavelengths of each spectrometer, since it compares one pixel in one spectrometer to several pixels in the other spectrometer. Fig. 10 shows the predicted versus reference for the source dataset, and the different transfer methods. The performances for both methods are good. Ultimately, the choice of calibration transfer methods should be decided by practical considerations: are the instruments located at the same place? How much time can be spared for sampling, spectral acquisition, referencing and validation? How important is the model performance? How important is model robustness?

Table 2 compares the performances of the transferred models to the native models on both the source and target spectrometers, as well as the performance of the source data after the wavelengths that are missing in the target data have been removed. The most notable observation is that the performances of the transferred models are lower than new models trained on the target spectrometers. It is plausible that larger differences between the source and target spectrometers will lead to a larger decrease in performance. All cases except the Case 3 require fewer PLS components for the transferred models compared to the models trained on the source dataset. Case 2 target 2 also require fewer PLS components in the transferred models compared to the models trained on the target dataset. This illustrates how the models sometimes overfit to the spectrometer they are trained on. Case 1 shows higher R^2 for the transferred models compared to the models applied to the source dataset. This is due to the target dataset covering a much wider analyte range, while having comparable absolute errors, making the relative errors lower. While the performances were lower for the transferred models, the transfer worked for all cases in the sense that RMSE/R2 where not much worse than models built on target and better than those obtained by PDS, despite the differences between the target and source spectrometers, and without using spectra from the same samples from both spectrometers. This simplifies calibration transfer in increasingly common scenarios where it is difficult or even impossible to acquire transfer samples, such as when spectrometers are located far away, or the use of models from a damaged or broken spectrometer, or creating a new application using a dataset downloaded from the internet.

6. Conclusions

Even very simple calibration transfer methods can yield good results, even between spectrometers that have large differences. Here calibration transfer was done by assigning wavelength correspondence, reoptimizing the model to compensate for overfitting to the spectrometers and differences in covariances in the two datasets, and then slope and bias correction for differences in optical setups and sample handling. Despite its simplicity, this calibration transfer method had good performance for both NIR and Raman, even for datasets that had differences in both sample handling and sample composition. Since this method does not require access to the source spectrometer, or samples measured on both the target and the source spectrometer, it can be used in cases when the source spectrometer is far away, broken, or otherwise unavailable.

CRediT authorship contribution statement

Jens Petter Wold: Writing – review & editing, Resources, Funding acquisition, Data curation. **Lars Erik Solberg:** Writing – review & editing. **Katinka Dankel:** Writing – review & editing, Resources, Data curation. **Tiril Lintvedt:** Writing – review & editing, Resources, Data curation. **Nils Kristian Afseth:** Writing – review & editing, Resources, Funding acquisition, Data curation. **Erik Tengstrand:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

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.

Data Availability

Data will be made available on request.

Acknowledgements

Financial support from the Norwegian Fund for Research Fees on Agricultural Products through the project “Precision Food Production”

(no. 314111) and from the Norwegian Research Council through the project "SFI Digital Food Quality" (no. 309259) is greatly acknowledged.

References

[1] T. Fearn, Standardisation and calibration transfer for near infrared instruments: a review, J. Infrared Spec. *9* (4) (2001) 229–244. DOI 10.1255/jnirs.309.

[2] R.N. Feundale, N.A. Woody, H.W. Tan, A.J. Myles, S.D. Brown, J. Ferre, Transfer of multivariate calibration models: a review, Chemom. Intell. Lab *64* (2) (2002) 181–192.

[3] P. Mishra, R. Nikzad-Langerodi, F. Marini, J.M. Roger, A. Biancolillo, D. N. Rutledge, S. Lohumi, Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spectrometers? The answer is not always, Trac-Trend Anal. Chem. *143* (2021). DOI: ARTN 116331. 10.1016/j.trac.2021.116331.

[4] Y.D. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, Anal. Chem. *63* (23) (1991) 2750–2756. DOI10.1021/ac00023a016.

[5] E.L. Bergman, H. Brage, M. Josefson, O. Svensson, A. Sparen, Transfer of NIR calibrations for pharmaceutical formulations between different instruments, J. Pharm. Biomed. *41* (1) (2006) 89–98, 10.1016/j.jpba.2005.10.042.

[6] Y. Sulub, R. LoBrutto, R. Vivilecchia, B.W. Wabuyele, Content uniformity determination of pharmaceutical tablets using five near-infrared reflectance spectrometers: a process analytical technology (PAT) approach using robust multivariate calibration transfer algorithms, Anal. Chim. Acta *611* (2) (2008) 143–150, https://doi.org/10.1016/j.aca.2008.02.016.

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

[8] L.J. Li, W.Q. Huang, Z.L. Wang, S.Q. Liu, X. He, S.X. Fan, Calibration transfer between developed portable Vis/NIR devices for detection of soluble solids contents in apple, Postharvest. Biol. Tec. *183* (2022). DOI: ARTN 111720. 10.1016/j.postharvbio.2021.111720.

[9] E. Bouveresse, D.L. Massart, Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration, Chemom. Intell. Lab *32* (2) (1996) 201–213. Doi 10.1016/0169-7439(95)00074-7.

[10] L. Petillot, F. Pewny, M. Wolf, C. Sanchez, F. Thomas, J. Sarrazin, K. Fauland, H. Katinger, C. Javalet, C. Bonneville, Calibration transfer for bioprocess Raman monitoring using Kennard Stone piecewise direct standardization and multivariate algorithms, Eng. Rep. *2* (11) (2020). DOI: ARTN e12230. 10.1002/eng2.12230.

[11] M.C. Alamar, E. Bobelyn, J. Lammertyn, B.M. Nicolai, E. Molto, Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple, Postharvest Biol. Tec. *45* (1) (2007) 38–45, https://doi.org/10.1016/j.postharvbio.2007.01.008.

[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). DOI: ARTN e2874. 10.1002/cem.2874.

[13] X.Q. Fan, H.M. Lu, Z.M. ZhangDirect calibration transfer to principal components via canonical correlation analysis 181 Chemometr. Intell. Lab.2018, , 21–28 10.1016/j.chemolab.2018.08.006..

[14] M.H. Wise BM, M. Høy, R. Bro, P.B. Brockhoff, Calibration transfer by generalized least squares, Seventh Scandinavian Symposium on chemometrics SSC7 (2001) 19–23.

[15] E. Andries, J.H. Kalivas, A. Gurung, Sample and feature augmentation strategies for calibration updating, J. Chemom. *33* (1) (2019). DOI: ARTN e3080. 10.1002/cem.3080.

[16] P.C. Calder, Marine omega-3 fatty acids and inflammatory processes: Effects, mechanisms and clinical relevance, Mol. Cell Biol. L *1851* (4) (2015) 469–484, https://doi.org/10.1016/j.bbalip.2014.08.010.

[17] A. Eilander, D.C. Hundscheid, S.J. Osendarp, C. Transler, P.L. Zock, Effects of n-3 long chain polyunsaturated fatty acid supplementation on visual and cognitive development throughout childhood: a review of human studies, Prostag Leukotr Ess. *76* (4) (2007) 189–203, https://doi.org/10.1016/j.plefa.2007.01.003.

[18] M.M. Rogero, P.C. Calder, Obesity, inflammation, toll-like receptor 4 and fatty acids, Nutrients *10* (4) (2018). DOI: ARTN 432. 10.3390/nu10040432.

[19] R.N. Thota, J.J.A. Ferguson, K.A. Abbott, C.B. Dias, M.L. Garg, Science behind the cardio-metabolic benefits of omega-3 polyunsaturated fatty acids: biochemical effects vs. clinical outcomes, Food Funct. *9* (7) (2018) 3576–3596. DOI: 10.1039/c8fo00348c.

[20] E. Lutfi, G.M. Berge, G. Bæverfjord, T. Sigholt, M. Bou, T. Larsson, T. Morkore, O. Evensen, N.H. Sissener, G. Rosenlund, et al., Increasing dietary levels of the n-3 long-chain PUFA, EPA and DHA, improves the growth, welfare, robustness and fillet quality of Atlantic salmon in sea cages, Brit J. Nutr. *129* (1) (2023) 10–28.

[21] M. Bou, G.M. Berge, G. Baeverfjord, T. Sigholt, T.K. Ostbye, B. Ruyter, Low levels of very-long-chain n-3 PUFA in Atlantic salmon (Salmo salar) diet reduce fish robustness under challenging conditions in sea cages, J. Nutr. Sci. *6* (2017) e32. DOI: 10.1017/jns.2017.28 From NLM PubMed-not-MEDLINE.

[22] S.S. Horn, B. Ruyter, T.H.E. Meuwissen, B. Hillestad, A.K. Sonesson, Genetic effects of fatty acid composition in muscle of Atlantic salmon, Genet Sel. Evol. (2018) *50*. DOI: ARTN 23. 10.1186/s12711-018-0394-x.

[23] T.A. Lintvedt, P.V. Andersen, N.K. Afseth, K. Heia, S.K. Lindberg, J.P. Wold, Raman spectroscopy and NIR hyperspectral imaging for in-line estimation of fatty acid features in salmon fillets, Talanta (2023) *254*. DOI: ARTN 124113. 10.1016/j.talanta.2022.124113.

[24] Ertesvåg, I.L. Undersøkelse av tørrfraksjonering av åkerbønner dyrket i Norge. Norwegian University of Life Sciences, 2020. ⟨https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/2721569⟩.

[25] C.S. do Carmo, P. Silventoinen, C.T. Nordgard, C. Poudroux, T. Dessev, H. Zobel, A. K. Holtekjolen, K.I. Draget, U. Holopainen-Mantila, S.H. Knutsen, S. Sahlstrom, Is dehulling of peas and faba beans necessary prior to dry fractionation for the production of protein- and starch-rich fractions? Impact on physical properties, chemical composition and techno-functional properties, J. Food Eng. (2020) *278*. DOI: ARTN 109937. 10.1016/j.jfoodeng.2020.109937.

[26] J.P. Wold, I.R. Johansen, K.H. Haugholt, J. Tschudi, J. Thielemann, V.H. Segtnan, B. Narum, E. Wold, Non-contact transflectance near infrared imaging for representative on-line sampling of dried salted coalfish (bacalao), J. Infrared Spec. *14* (1) (2006) 59–66. DOI: DOI 10.1255/jnirs.587.

[27] J.P. Wold, M. O'Farrell, P.V. Andersen, J. Tschudi, Optimization of instrument design for in-line monitoring of dry matter content in single potatoes by NIR interaction spectroscopy, Foods *10* (4) (2021).

[28] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1) (1969), 137-&.