Review article

Transfer of multivariate calibration models: a review

Robert N. Feudale [a], Nathaniel A. Woody [a], Huwei Tan [a], Anthony J. Myles [a],
Steven D. Brown [a,*], Joan Ferré [b]

[a]Chemometrics Research Group, Department of Chemistry and Biochemistry, University of Delaware, Newark, DE 19716, USA
[b]Department de Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Pl. Imperial Tarraco, 1, 43005 Tarragona, Spain

Abstract

Multivariate calibration models are of critical importance to many analytical measurements, particularly for spectroscopic data. Generally, considerable effort is placed into constructing a robust model since it is meant to be used for extended periods of time. A problem arises, though, when the samples to be predicted are measured on a different instrument or under differing environmental factors from those used to build the model. The changes in spectral variations between the two conditions may make the model invalid for prediction in the new system. Various standardization and preprocessing methods have been developed to enable a calibration model to be effectively transferred between two systems, thus eliminating the need for a full recalibration. This paper presents an overview of the different methods used for calibration transfer and a critical assessment of their validity and applicability. The focus is on methods for transfer of near-infrared (NIR) spectra.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Review, Calibration transfer; Multivariate standardization; PDS; DS; FIR; OSC; SBC; Near-infrared spectroscopy; PLS; Signal processing

1. Introduction

Multivariate calibration is a very useful tool for extracting chemical information from spectroscopic signals. It has been applied to various analytical techniques, but its importance has been manifested in near-infrared (NIR) spectroscopy. The most commonly used multivariate methods for chemical analysis are partial least squares (PLS) regression and principal component regression (PCR), where factors that relate to variation in the response measurements are regressed against the properties of interest. Ideally, each factor added to the model would describe variation relevant for predicting property values. In NIR spectra, however, the first few factors that describe majority of the spectral variation are usually accounting for baseline shifts and various instrumental effects.

A practical limitation to multivariate calibration occurs when an existing model is applied to spectra that were measured under new environmental conditions or on a separate instrument. Even if identical samples were measured, the spectral variation of the two response matrices that is captured by the model

* Corresponding author. Tel.: +1-302-831-6861; fax: +1-302-831-6335.
E-mail address: sbd@udel.edu (S.D. Brown).

will differ. For this reason, a model developed on one instrument can generally not be used on spectra of a second instrument to estimate property values. A solution to this calibration transfer problem is to re-measure every sample and construct a new model for newly acquired spectra. This is not a practical solution since there is substantial cost and time associated with building a robust calibration. An alternative is to apply chemometrics techniques to correct for instrumental and environmental differences, thereby making the model transferable and avoiding full recalibration.

Various methods for calibration transfer exist in the literature, and most have been discussed in earlier [1–5] and recent [6,7] reviews. The objective of this paper is to update the modern transfer methods and give a critical assessment of their validity and applicability. Most of the literature reviewed here deals with transfer of near-infrared (NIR) data, although standardization methods have also been applied to UV–visible [8–12] spectrophotometry, fluorescence [13] spectroscopy, Raman [14–17] spectroscopy, and electrochemistry [18–20]. It will be assumed throughout this paper that the signal is a spectrum measured at defined wavelengths and that the property of interest is the analyte concentration, although properties other than concentration and other types of signals also fit the discussion.

2. Situations where a model can become invalid

There are several circumstances that can introduce variations in the new spectra that had not been included in the calibration step. The presence of these nonmodeled variations in the new data can lead to strongly biased predictions from multivariate calibration models. Essentially, three situations exist that can render a model invalid. The first includes changes in the physical and/or chemical constitution of the samples. These changes can result from differences in viscosity, particle size, surface texture, etc. that may occur between batches. Batch differences may arise from some change in the starting materials, sample preparation, or experimental design.

The second situation arises from changes in the instrumental response function. These changes can occur when the data to be predicted are collected on

an instrument different from that used to build the model. Although the spectra may have the same basic shape or profile, the measured intensity values from two instruments will generally be different. Also, the replacement of some parts of the instrument or their deterioration over time will produce changes in the instrument response function. The aging of sources, probes, and detectors may lead to instability of the signal over time (drift) or nonlinearities [21,22] in the spectra. NIR data may be additionally problematic due to differences in pathlengths of optical probes and wavelength registration shifts [21].

The third situation that may render a model invalid occurs when the instrument's environment changes over time. Temperature and humidity variations can have a strong influence on measurement values [23] by causing shifts in absorption bands [24] and non-linear changes in absorption intensities [25] on the spectra. Temperature fluctuations may also induce instrumental changes due to thermal expansion, thus affecting the alignment of optical components and shifting the peaks along the wavelength axis [26].

In some cases, the response of the new samples is not greatly affected by the new measurement conditions and the existing model can be used without applying any corrections. When this is not the case, some strategy must be used to avoid biased predictions that result from transferring a calibration model. If the future sources of variation can be identified, they can be incorporated into the model during calibration. The model can also be updated with new samples to account for the nonmodeled variance. Standardization methods can be applied once the model is already in use to account for the new variations by transforming the spectral responses, model coefficients, or predicted values. These topics will be discussed in the next sections. Although calibration transfer is a generic term that refers to the transfer of a model between different conditions, it will be assumed here that the transfer is between different instruments.

3. Strategies that can be used before the model is implemented

This group of calibration procedures relies on instrumental and experimental design techniques to

avoid the need for data transformation. Noncalibrated variations can be avoided by selecting the appropriate instrumental parameters, environmental conditions, and response channels, or by designing the experiment such that transfer conditions are carefully controlled. The strategy chosen to avoid standardization depends on the ability of the analyst to foresee future sources of variation and the feasibility of controlling the experimental parameters.

3.1. Instrument matching

The calibration transfer problem does not exist if two instrument responses are truly alike. Adhihetty et al. [27] achieved the transference of PLS calibration models between two FT-IR spectrometers by carefully controlling experimental parameters, by selecting appropriate wavelengths, and by subtracting the spectra of purge gases. In order to achieve identical responses from two instruments, the analyst must first decide which parameters are most critical for transferability of the model [2], and then decide if time constraints permit such adjustments. Instrument matching is appropriate only when the expert analyst has simultaneous access to both instruments, when the instruments are of the same type, and when the instruments can be continually modified to maintain transferability as the instruments age differently. Such restrictions make this method impractical for most applications.

3.2. Global models

Noncalibrated variations can be minimized or even avoided by intentionally varying instrumental parameters, sample preparation, and environmental factors in the calibration step to include expected sources of variance in the model. Global models incorporate expected variation for a wide range of experimental conditions and, therefore, fewer sources of variation must be strictly controlled in the future data. For example, the prediction results of spectra that were shifted along the wavelength axis improved with the addition of wavelength-shifted spectra to the calibration model [26]. Global models that modeled temperature as an unknown interferent were shown to perform equally as well as models that were calibrated for a specific temperature [28]. Accurate creation of a global model requires the analyst to foresee new sources of variance and the extent of their variability in the data, which is often difficult to determine. Also, global models tend to be very complex (high dimensionality) and predictions from a global model can be less accurate than those from a local model if small nonmodeled variations are present in the signal [29]. If the analyst has knowledge of future variability, it would be more beneficial to minimize this new variance since the model will then be more robust to the specific conditions for which it was developed [30].

3.3. Model updating

Another way of accounting for new variations in the spectra is to rebuild the model with the addition of a few test samples to the old calibration set. The samples added to construct the updated model must accurately describe the variability in the new instrument such that the model spans both the old and new experimental conditions and is valid for both situations [2]. The addition of new samples makes the model more robust to the new measurement conditions and can often lead to better prediction results [17,31]. When the number of calibration samples is very large, the new samples can be weighted to increase their contribution to the model [32]. Model updating is attractive for long-term instrument drift as old calibration samples are slowly eliminated in favor of new samples, thus eliminating the transfer problem. An unattractive feature of this method is that a separate model would be required for multiple instruments. Model updating is very simple to implement and can generally be applied when instrumental differences are not too complicated. For complex situations, a substantial number of samples would be required to capture the variance in the new system, such that model updating approaches full recalibration.

3.4. Sensor selection

A robust model can be built by using variables that are either insensitive or are less sensitive to the variations in the experimental conditions [22]. An example of this approach is the method developed by Mark and Workman [33] for selection of wavelengths for MLR models that are robust to wavelength

shifts. In this algorithm, spectral variables that exhibit little change in regression coefficients when wavelength is varied are selected for model transfer between instruments. Wavelengths selected by simulated annealing were used to transfer a PLS model between two NIR reflectance instruments [34] and between different temperatures [35]. Variables that were sensitive to instrumental differences were removed in order to transfer a model from an off-line to an on-line application for prediction of PET yarn shrinkage [17]. Genetic algorithms have been employed to find a combination of variables that allow a single calibration model to be used with two separate NIR instruments [36]. Wavelength selection can be applied when the spectral differences involve only shifts in the wavelength axis and cannot be used to correct for intensity differences or peak broadening in spectra. Successful application of sensor selection is dependent on finding variables that are insensitive to the offending variation yet sensitive to quantifying the analyte.

4. Standardization methods

Several strategies attempt to deal with noncalibrated variations that are detected after the model is already in use. In these standardization methods, the response function of a secondary instrument is modified to match the response function of the primary instrument. To accomplish this task, one can standardize the regression coefficients, the spectral responses, or the predicted values by mathematical manipulation. This section describes the different methods for instrument standardization.

4.1. Standardization of the model coefficients

A transfer method proposed by Wang et al. [37] involves transforming the original model to a model suitable for the new instrument. This was defined for both classical and inverse calibration models, but only the latter is described here. The inverse model is defined for the vector of property values $\mathbf{y}$ and the response matrix measured on the primary instrument $\mathbf{X}_1$ as

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b}_1 \tag{1}$$

where $\mathbf{b}_1$ is the regression vector computed for the primary instrument. The above equation can also be written for the secondary instrument

$$\mathbf{y} = \mathbf{X}_2 \mathbf{b}_2 = \mathbf{X}_2(\mathbf{b}_1 + \Delta\mathbf{b}) \tag{2}$$

where $\Delta\mathbf{b}$ is the difference vector of $\mathbf{b}_2$ and $\mathbf{b}_1$ which can be expressed as follows:

$$\Delta\mathbf{b} = \mathbf{b}_2 - \mathbf{b}_1 = \mathbf{X}_2^+\mathbf{y} - \mathbf{X}_1^+\mathbf{y} = (\mathbf{X}_2^+ - \mathbf{X}_1^+)\mathbf{y} \tag{3}$$

where the superscript ($^+$) denotes matrix inversion. Analogous equations can be written for a subset of the data by replacing $\mathbf{X}$ with $\mathbf{S}$. Rearranging the above expression yields an estimate of the new regression vector:

$$\mathbf{b}_2 = \mathbf{b}_1 + \Delta\mathbf{b} = \mathbf{X}_1^+\mathbf{y} + (\mathbf{S}_2^+ - \mathbf{S}_1^+)\mathbf{y} \tag{4}$$

The subset of samples chosen to compute the new model does not have to be a subset of the original calibration set, but must provide an estimate of the same regression vector. Since the property vector $\mathbf{y}$ appears in the above expression, generic standards cannot be used to determine the standardization parameters. With the standardized regression vector, the property values of the new test samples can be predicted directly from the responses measured on the secondary instrument. However, a close examination of Eq. (4) reveals that there is no real transfer of information from the primary to the secondary instrument through this relationship, as shown below

$$\mathbf{b}_2 = \mathbf{X}_1^+\mathbf{y} + \mathbf{S}_2^+\mathbf{y} - \mathbf{S}_1^+\mathbf{y} = \mathbf{b}_1 + \bar{\mathbf{b}}_2 - \bar{\mathbf{b}}_1 \tag{5}$$

where $\bar{\mathbf{b}}_2$ and $\bar{\mathbf{b}}_1$ are the estimated regression vectors from the subset of samples. The difference between $\mathbf{b}_1$ and $\bar{\mathbf{b}}_1$ is negligible because $\bar{\mathbf{b}}_1$ is merely an estimate of the correct regression vector, $\mathbf{b}_1$ [5]. Therefore, the standardized regression model $\mathbf{b}_2$ is simply estimated from the subset samples measured on the secondary instrument.

Forina et al. [38] proposed transferring a regression equation between NIR instruments by means of two PLS steps. Essentially, this method involves a direct standardization (Section 4.2.2) of the model coefficients by means of a transformation matrix $\mathbf{F}$:

$$\mathbf{b}_2 = \mathbf{F}\mathbf{b}_1 \tag{6}$$

In this method, both the transformation matrix and regression model are determined by PLS, hence the two-step PLS procedure. The authors determined that the errors of the standardized model were comparable to that of the model obtained directly from the secondary instrument. Comparable errors were obtained because the same response matrices were used to compute both the transformation matrix and regression models. Substituting Eq. (1) into Eq. (6) and given $\mathbf{X}_2^+ = \mathbf{F}\mathbf{X}_1^+$ (direct standardization), the above equation can be re-expressed as

$$\mathbf{b}_2 = \mathbf{F}\mathbf{X}_1^+\mathbf{y} = \mathbf{X}_2^+\mathbf{y} \qquad (7)$$

The proposed standardization method is equivalent to computing a calibration model directly from the spectra of the secondary instrument. Even if $\mathbf{F}$ were estimated from a subset of samples, the above relationship would still hold.

4.2. Standardization of the spectral responses

The general idea of standardization is to model the instrumental differences directly, a task that cannot be accomplished by using concentration information. In order to model the instrumental differences, the spectral responses of a subset of samples measured on the primary instrument are regressed against the same subset measured on the secondary instrument. Thus, changes in the response values between the two instruments can be corrected, and the original model can be used for prediction on the secondary instrument without having to compute new regression coefficients. Both univariate and multivariate methods have been proposed to standardize the spectral responses from a secondary instrument.

4.2.1. Univariate standardization

A simple approach to standardization developed by Shenk and Westerhaus [39–41] involves a single correction factor at every wavelength channel to adjust for intensity differences between spectra. A similar method, called single wavelength standardization (SWS) [13], also corrects for intensity differences that vary across the wavelength region by regressing the responses of both instruments at each channel j:

$$\mathbf{r}_{j,1} = \mathbf{r}_{j,2}b_j \qquad (8)$$

where $\mathbf{r}_{j,1}$ and $\mathbf{r}_{j,2}$ are the jth columns of $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively, and b_j is the regression coefficient for the jth wavelength. These methods, however, can be used to standardize only those response vectors whose wavelength axes are aligned. In order to account for wavelength shifts between spectra, another local correction procedure was developed by Shenk and Westerhaus [41,42]. This patented method corrects the full spectral response function of an instrument by performing a wavelength index conversion followed by a spectral intensity transformation at each wavelength. The wavelength scale is first corrected by correlating the measurements at wavelength i on the primary instrument with those located in a small window around i on the secondary instrument. A second-order polynomial is fitted to the correlation coefficients, yielding a continuous function across the channels in the processing window. The wavelength corresponding to the maximum of the quadratic function is then selected as the corresponding wavelength of the primary instrument. After performing wavelength conversion, the spectral intensity at index j on the secondary instrument is standardized to the intensity at index i of the primary instrument by linear regression:

$$\mathbf{r}_{i,1} = a_i + b_i\mathbf{r}_{j,2} \qquad (9)$$

where a_i and b_i are the offset and slope, respectively, estimated from the transfer samples. Further developments [43] have been made to the spectral intensity correction to improve the results when standardization samples are different than those to be analyzed. A modification of the wavelength conversion step [44] was also attempted, in which a polynomial was fitted to Euclidean distances rather than correlation coefficients since the latter is susceptible to noise. The patented method is suitable for correcting wavelength shifts or linear intensity differences in spectra, but it cannot correct for peak broadening since it assumes that no relationship exists between neighboring correction models. In order to correct for broadening and more complicated effects, multivariate standardization procedures must be used.

4.2.2. Direct standardization

An important development in multivariate transfer is direct standardization (DS) [37], which consists of

directly relating the response of a sample measured with one instrument to its response obtained on another instrument. This linear relationship is described by the transformation matrix $\mathbf{F}$ according to:

$$\mathbf{S}_1 = \mathbf{S}_2 \mathbf{F} \tag{10}$$

where once again $\mathbf{S}_1$ and $\mathbf{S}_2$ are the response matrices of the standardization samples obtained from the primary and secondary instruments, respectively. The transformation matrix is estimated as

$$\mathbf{F} = \mathbf{S}_2^+ \mathbf{S}_1 \tag{11}$$

where $\mathbf{S}_2^+$ is the generalized or pseudo-inverse of $\mathbf{S}_2$. Once $\mathbf{F}$ is calculated, the response vector of a new sample $\mathbf{x}$ is projected to the original measurement space so that its property values can be predicted with the old model:

$$\hat{\mathbf{x}}^{\mathrm{T}} = \mathbf{x}^{\mathrm{T}} \mathbf{F}. \tag{12}$$

Although it is assumed that the relationship between both responses is linear, some nonlinearities can be tolerated in the multivariate regression. A problem with this approach is that a significant amount of chemical information can be modeled in the transformation matrix. When computing $\mathbf{F}$, it is assumed that any change in response values is due to a change in the instrument function. However, any variation in the chemical composition of the samples will also be incorporated into the model, and thus $\mathbf{F}$ will not truly represent instrumental differences. Another problem is that the number of standardization samples chosen to compute the transform is generally much smaller than the number of channels, which can lead to over-fitting when estimating $\mathbf{F}$ since the system is underdetermined. Therefore, $\mathbf{F}$ is typically estimated by means of PCR or PLS regression in order to obtain a least squares solution. Another way to overcome this over-fitting problem is to reduce the number of channels involved in the regression, which forms the basis of piecewise direct standardization.

4.2.3. Piecewise direct standardization

In direct standardization, each wavelength of the primary spectra is related to all wavelengths of the secondary spectra simultaneously. For many transfer applications, particularly those in which the response

of one instrument is shifted along the wavelength axis with respect to another instrument, the spectral correlations are usually limited to smaller regions. For this reason, piecewise direct standardization (PDS) [37] was proposed. In PDS, the response $\mathbf{r}$ of the standardization samples measured at wavelength j on the primary instrument is related to the wavelengths located in a small window around j measured on the secondary instrument:

$$\mathbf{r}_j = \mathbf{R}_j \mathbf{b}_j \tag{13}$$

where $\mathbf{R}_j$ is the localized response matrix of the transfer samples and $\mathbf{b}_j$ is the vector of transformation coefficients for the jth wavelength. The regression vectors calculated for each window in the data are then assembled to form a banded diagonal matrix $\mathbf{F}$ according to

$$\mathbf{F} = \mathrm{diag}(\mathbf{b}_1^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \ldots, \mathbf{b}_j^{\mathrm{T}}, \ldots, \mathbf{b}_k^{\mathrm{T}}) \tag{14}$$

where k is the number of wavelengths. The response of any unknown sample can then be standardized as if it were measured on the primary instrument using Eq. (12). Since PDS operates with a moving window, edge effects can occur where there is insufficient data to form a complete window. For these cases, the ends of the spectra are either removed or estimated by extrapolation [22,45]. The spectral window used for filtering can be symmetrical or asymmetrical around each variable [21,22,24,45]. It was found that symmetrical windows were desirable when the wavelength shifts are negligible [21], whereas asymmetrical windows can improve the transfer when a spectral shift exists [24].

PDS is one of the most widely used transfer methods, and is typically employed as a reference for other novel techniques. Its superiority over other standardization methods can be attributed to its local character and multivariate nature, enabling simultaneous correction of intensity differences, wavelength shifts, and peak broadening. Studies have shown that transfer results with PDS can be better than those of full set recalibration when the signal on one instrument is standardized to that of a higher quality instrument [46]. The PDS algorithm was improved to correct for significant additive differences between spectra by column mean-centering the data prior to standardization [47]. This approach, referred to as

PDSB for baseline correction, is equivalent to including an offset term in the PDS equation [9]. Another modification of the PDS algorithm, called continuous PDS (CPDS) [25,48], was proposed to correct for continuous temperature variation in the spectra. In this method, a separate PDS model is computed to transfer spectra between a calibration temperature and various discrete temperatures. A polynomial is then fitted to the discrete transformation matrices to correct for all temperature differences that lie in the standardization range.

One of the main problems associated with PDS resides in estimating the ranks of local PCR/PLS models. Usually, the number of eigenvalues to retain for local models is predetermined based on a tolerance value [47,49]. However, artifacts in the transferred spectra can be observed when the local ranks are too high, resulting in over-fitting of the data. Bouveresse and Massart [49] developed a procedure to detect and reduce PDS transfer artifacts by examining the spectrum of an independent sample before and after standardization. High residuals between the transferred and original spectra can reveal the presence of erroneous local rank values. For those wavelengths that exhibit high residuals, local PDS models are re-calculated using fewer components. Poor estimation of local rank due to eigenvalue swapping [50] between adjacent windows can also result in discontinuities in the PDS transferred spectra. It was observed that a small change in the input data can cause eigenvalues that are similar in magnitude to swap places, producing a model different from that in a neighboring window. The abrupt change in the shape of the model produces discontinuities in the spectra.

Since much of the difficulty in determining the local rank of PDS models can be attributed to high-frequency noise in the data, modifying the spectra before transfer can prove beneficial. The data can be processed with derivatives [46] and Fourier transforms [10] to reduce noise in the data prior to PDS transfer. The spectra can also be decomposed via PCA, and then reconstructed with only relevant factors to filter noise before transfer with PDS [50]. A new method called wavelet hybrid direct standardization (WHDS) [51] couples wavelet decomposition with DS and PDS to improve transfer results when noise is prevalent in the data. After decomposing the spectra, the approximation (low frequency) and detail

(high frequency) coefficients are reconstructed, transferred separately with PDS and DS, respectively, and then recombined before prediction. This technique was shown to be a more reliable means of standardization than PDS since it is more robust to high-frequency noise and eigenvalue swapping effects.

4.3. Standardization of the predicted values

One of the most widely used methods for correcting predicted values is the simple univariate slope and bias correction (SBC) [52]. In this method, one assumes a linear relationship between the predictions for spectra measured on the secondary instrument and the predictions that would have been obtained if the samples had been measured on the primary instrument. The property values of the standardization samples measured on both instruments are first predicted with the original calibration model developed on the primary instrument.

$$\mathbf{y}_{s,1} = \mathbf{S}_1 \mathbf{b} \tag{15}$$

$$\mathbf{y}_{s,2} = \mathbf{S}_2 \mathbf{b} \tag{16}$$

Then, a linear equation is fitted between the two predictions, using either ordinary least squares or orthogonal least squares [52], and the predicted values for the new samples are corrected for the bias (intercept) and the slope of the regression line:

$$\mathbf{y}_{2,corr} = bias + slope \cdot \mathbf{y}_2 \tag{17}$$

Alternatively, the predictions of a subset of the secondary samples can be regressed against their corresponding reference values in order to compute the transfer model. In general, SBC will work well when the instrumental differences are simple and systematic in all data measured in the new instrument. Since this method is a univariate approach, the results are not affected by artifacts such as those coming from bad local rank determination. However, when the instrumental differences are more complex and differ from sample to sample, procedures such as PDS or other approaches may be needed to obtain acceptable results. Bouveresse et al. [52] implemented a diagnostic tool to determine whether SBC can be applied to a given transfer situation. The procedure involves comparing the residual variances obtained on both

instruments by means of an F-test, and if the residual variances are similar, SBC can be employed. This diagnostic tool was extended to cases in which the residual variances of the two instruments are not independent estimates [53].

Slope and bias correction was used to transfer univariate and simple MLR models for the prediction of protein and moisture in wheat flour between nine diffuse reflectance NIR instruments [54]. Jones et al. [55] used a similar approach to transfer MLR equations developed for the moisture determination in freeze-dried injection products. SBC was shown to perform as well as PDS in correcting instrumental variations over time for the prediction of ethylene concentrations in copolymers [56] and chromium content in tanning sewage [12].

4.4. Transformation before standardization

Standardization methods often model relationships between spectra in the original measurement space. These methods can also be employed to model differences between spectra that have been transformed to another domain. Walczak et al. [57] proposed that standardization be done in the wavelet domain (SWD). This method relates wavelet transform coefficients of a subset of standard samples obtained on two instruments with univariate linear models. The wavelet coefficients of the new spectra are transferred using these models and then converted back to the wavelength domain. Standardization can also be performed in the PC factor space [58]. In this method, the PC scores of the samples are transferred from the secondary to the primary instrument, and then converted back to the original variable space. In both of these algorithms, the spectral variation is compressed which can lead to greater stability in the transfer. However, due to this compression along the wavelength axis, window-based standardization methods such as PDS cannot be employed since the local information is not retained in neighboring variables. Therefore, transfer of information in these domains is limited to univariate models or to direct standardization.

4.5. Other standardization methods

A few attempts have been made to use artificial neural networks (ANNs) for standardization [23,59–

61]. ANNs generally suffer from over-fitting problems because a reasonably sized ANN has far more parameters to be estimated than there are transfer samples available. Duponchel et al. [59] addressed this issue by using a window-based method that creates several local ANN models in a manner analogous to PDS transfer. Although the number of nodes for each local network is greatly reduced, one would have to build as many ANNs as there are spectral windows with this method. Despagne et al. [60] used this idea in a more novel way by treating the response values within the spectral window as individual objects, thus creating an ANN with an artificially high number of training samples that appears to generalize better. The authors included a positional index as an additional input for each window to help globalize the local model. Another problem with ANNs is that the model is usually optimized by the ability of the network to reproduce a spectrum rather than its ability to minimize prediction errors. Optimizing based on spectral residuals places identical value on all variables, even though they may not be equally important for prediction. Goodacre et al. [61] used a neural network in a more direct way to transfer mass spectral data between two spectrometers. A neural network was created with 150 input masses and 150 output masses and a variable number of hidden nodes connecting the two input layers. Not surprisingly, they had to reduce the number of hidden nodes, and thus the number of parameters, in the network in order to produce acceptable results.

Other approaches to calibration transfer are based on maximum likelihood principal component analysis (MLPCA) [62] and positive matrix factorization (PMF) [63]. In both of these methods, the spectra to be transferred are treated as missing values that need to be estimated from an augmented matrix. The transferred spectra are first represented by zeros, and then reconstructed through an iterative filling process that is guided by the standardization subset. Calibration transfer was also accomplished using the Kalman [64] and Wiener [65] filters. The Kalman filter is a recursive, digital filter that operates by predicting a new state vector at any given time based on information obtained from previous measurements. It is difficult to make a critical assessment of any of these methods since they have been used very sparsely for calibration transfer problems. The trans-

fer methods based on MLPCA and PMF are likely to suffer from convergence problems since they deal with vast amounts of missing data. The use of the Kalman filter in correcting for drift in analytical systems [66] seems to have potential for updating calibration models over time.

5. Selection of the standardization samples

The approaches to instrument standardization require that the spectra of a few, well-chosen samples be measured on both instruments before transfer parameters can be computed. The samples used to establish the transfer parameters must be representative enough to describe the differences between the instruments, and should ideally span the entire experimental domain.

The original algorithm used for sample selection is based on choosing samples with high leverages [37]. The sample with highest leverage is first removed from the calibration matrix and is then orthogonalized against every other sample. This selection and removal sequence is repeated until the desired number of samples has been extracted. However, this technique is very sensitive to heterogeneities in the distribution of the data [49] since it selects samples based on leverage. The Kennard–Stone (KS) algorithm [67], on the other hand, allows for the selection of samples from all parts of the data space for nonhomogeneous distributions. This method begins by finding the two samples with the highest Euclidean distance between them, calculated from either their spectra or their PC scores [19]. The sample with the smallest distance from the previously selected samples is retained, and the procedure is repeated. If needed, the KS algorithm can be initialized with the spectrum that is closest to the mean of the data matrix, which is the most representative sample [59]. Another selection method is based on finding a combination of samples that maximizes the smallest inter-point distance [49], but generally requires much computation time. The tools developed by Jouan-Rimbaud, et al. [68] can be used to compare numerically the representativity of standardization and calibration data [69].

Unfortunately, none of these methods has a criterion for deciding what number of transfer samples is sufficient to perform the standardization. The optimal number of transfer standards is usually determined by examining validation results for a series of standards, and then choosing the number that minimizes prediction errors [24]. In some cases, the number is selected by practical considerations, such as the maximum number of samples available for transfer. Commonly, transfer samples are selected from those measured under the initial conditions and then re-measured under the new set of conditions. The transfer parameters can also be computed with a subset selected from the test samples and then re-measured on the first instrument [52,70], which is advantageous since the samples used for standardization are very representative of those used for prediction. Another approach is to use generic standards, such as polystyrene [45] or blends of pure material [58], to determine the transfer parameters. This approach should overcome the problems of sample degradation and should be applicable to various analytes. The main problem with this approach is that the spectra of the generic standards must be similar to those of the validation samples such that their spectral changes will reflect the changes encountered by real samples. If the spectral changes were produced by temperature fluctuations, only subset samples should be used and not generic standards since temperature effects are analyte dependent [71].

6. Nonstandardization methods

Section 4 discussed ways of standardizing the response of one instrument to that of another instrument to avoid reconstructing a new calibration model. Although these methods work well for solving many transfer problems, they are useless for applications in which the same transfer samples cannot be measured on both instruments. Measuring identical samples on both instruments is not problematic when the instruments are located in the same laboratory or within some local distance. However, it may be difficult or even impossible to re-measure the standards when one instrument is at a remote location or when the samples are chemically unstable. When transfer standards are not available, preprocessing methods must be employed. The methods discussed here correct for intensity changes only in the spectra and not for wavelength shifts.

The most commonly used preprocessing techniques are derivatives. First and second derivatives are used to remove baseline offsets and linearly sloped baselines (scattering), respectively. Since derivatives are calculated by taking the difference between adjacent spectral points, noise in the data is usually magnified. For this reason, the Gap [72] or Savitzky–Golay [73] algorithms that smooth the data prior to differentiation are typically employed. Another commonly used preprocessing method is multiplicative signal correction (MSC) [74–76]. In this technique, each spectrum is regressed against a reference spectrum, usually the mean of the calibration set, and the spectrum is then corrected by subtracting the intercept and dividing by the slope of the regression line. Generally, MSC works well with spectra that are fairly linear in concentration and with the spectra of samples that are chemically similar. However, these preprocessing methods offer minimal improvements in transfer results since they assume that the baseline and slope differences are constant throughout the entire wavelength region. For most spectroscopic applications, this assumption does not hold.

A variation of the MSC algorithm, known as finite impulse response (FIR) filtering [77], was proposed to transfer a calibration model when the additive and slope differences vary with wavelength. In FIR filtering, each spectrum is projected to a reference space as in MSC, but the regression is performed within a moving window equivalent to the piecewise MSC [78] approach. The main drawback of the FIR algorithm is the appearance of artifacts in the transferred spectra [77,79], which limits its robustness. The source of these spectral artifacts was found to reside in the calculation of local regression models where correlations between the sample and reference spectra approach linear independence. A modification of the FIR algorithm was recently reported [80] to solve this artifact problem. With this modified algorithm, the transfer between NIR spectrometers [77,81] was significantly improved, and the filtered spectra were free of artifacts [80].

Even with the improvement to the FIR algorithm, transfer applications involving FIR are limited. When applying MSC, there is the potential for loss of chemical information since the sample spectrum is made to resemble the reference spectrum. This loss might be higher in FIR since the sample spectrum

approaches that of the reference spectrum more closely. Although FIR and MSC are very convenient to implement because they do not require transfer standards, they should be treated as standardization methods to avoid information loss. Such is the case with full spectrum standardization (FSS) [13] in which a univariate regression is performed between the spectra of a common sample measured on two instruments. Thus, any differences in the response values can be attributed to instrumental variations. Obviously, FSS is valid for constant changes in additive and multiplicative effects, but it can be used in a piecewise manner like FIR to account for wavelength-dependent differences.

In order to avoid information loss when processing spectra, orthogonal signal correction (OSC) [82] and related versions [83–88] were proposed. The general goal of OSC is to remove the largest systematic variation in $\mathbf{X}$ that is unrelated or orthogonal to $\mathbf{Y}$, thereby simplifying the calibration model. Although OSC was originally developed as a preprocessing method to reduce the number of factors in multivariate regressions, it is now also used to transfer calibrations between instruments [31,81,83]. For calibration transfer, OSC is designed to remove vectors that are orthogonal to $\mathbf{Y}$ and common to both instruments in order to make the model more transferable. In recent literature, though, there has been some attention regarding the validity of OSC as a signal processing method [89,90]. In most cases, filtering with OSC merely reduces the number of factors in the model by the number of OSC components removed without any improvement in model simplicity or prediction errors. Multivariate regression methods such as PLS implicitly model nonlinear effects and other unwanted variation during the calibration step. Therefore, preprocessing with OSC, a PLS-based filtering method, will generally not improve the calibration since it removes variance that has already been modeled. Based on this reasoning, it can be inferred that OSC transfer results should approach those obtained through model updating.

Perhaps one of the most effective processing methods is wavelets. In wavelet analysis, the original signal is decomposed into wavelet coefficients of varying frequency, and the signal to noise ratio can be improved by removing certain frequency contributions that correspond to extraneous information. Wavelets have proven useful for removing baseline

structure that varies between the calibration and test sets [91], which is especially attractive for calibration transfer problems. For most transfer applications involving NIR spectra, the low-frequency coefficients can be removed since they usually relate to instrumental differences [81]. Wavelets are generally valid for many calibration transfer problems because the analytical signal is usually of a different frequency than that of the offending variation, and thus the analytical signal can be isolated from various background noise. However, when different detector responses produce different frequency components, wavelet analysis breaks down. Nonetheless, wavelets show great promise as alternative transfer methods since they are very versatile and do not require transfer samples.

7. Conclusions

Many of the methods discussed in this review have been successfully applied to various calibration transfer problems. However, none of these methods can truly provide quality transfer results for very complex systems. PDS is probably the best solution for complex systems, but even it is not without its limitations. Neural networks offer some comfort in that they can be used to model nearly any nonlinear effect, but unless the number of parameters to be estimated is minimized, the network will ultimately over-fit. More research is required in the area of nonstandardization methods because it is usually not possible to remeasure the same samples on multiple instruments. Wavelets are very encouraging as transfer methods since they are widely applicable and do not require transfer samples. The choice of transfer methodology will depend upon the application at hand since no single method is optimal for all situations. Although there is no real guideline for selecting the best technique, the comments suggested in this review can be used as a starting point.

Acknowledgements

References

[1] T. Dean, T. Isaksson, NIR news 4 (1993) 8–9.
[2] T. Dean, T. Isaksson, NIR news 4 (1993) 14–15.
[3] O.E. De Noord, Chemom. Intell. Lab. Syst. 25 (1994) 85–97.
[4] E. Bouveresse, D.L. Massart, Vibr. Spectrosc. 11 (1996) 3–15.
[5] T. Dean, B.R. Kowalski, in: S.D. Brown (Ed.), Computer Assisted Analytical Spectroscopy, Wiley, Chichester, UK, 1996, pp. 175–187.
[6] X.L. Chu, H.F. Yuan, W.Z. Lu, Spectrosc. Spec. Anal. 21 (2001) 881–885.
[7] T. Fearn, J. NIRS 9 (2001) 229–244.
[8] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, E. Rovira, Appl. Spectrosc. 49 (1995) 593–597.
[9] F. Sales, M.P. Callao, F.X. Rius, Chemom. Intell. Lab. Syst. 38 (1997) 63–73.
[10] C.S. Chen, C.W. Brown, S.C. Lo, Appl. Spectrosc. 51 (1997) 744–748.
[11] D. Ozdemir, R. Williams, Appl. Spectrosc. 53 (1999) 210–217.
[12] F. Sales, A. Rius, M.P. Callao, F.X. Rius, Talanta 52 (2000) 329–336.
[13] L. Norgaard, Chemom. Intell. Lab. Syst 29 (1995) 283–293.
[14] C.K. Mann, T.J. Vickers, Appl. Spectrosc. 53 (1999) 856–861.
[15] H.N. Quang, M. Jouan, N.Q. Dao, Anal. Chim. Acta 379 (1999) 159–167.
[16] H. Swierenga, A.P. de Weijer, R.J. van Wijk, L.M.C. Buydens, Chemom. Intell. Lab. Syst. 49 (1999) 1–17.
[17] H. Swierenga, A.P. de Weijer, L.M.C. Buydens, J. Chemom. 13 (1999) 237–249.
[18] A. Herrero, M.C. Ortiz, Anal. Chim. Acta 348 (1997) 51–59.
[19] F. Sales, M.P. Callao, F.X. Rius, Analyst 124 (1999) 1045–1051.
[20] F. Sales, M.P. Callao, F.X. Rius, Analyst 125 (2000) 883–888.
[21] J. Lin, S.C. Lo, C.W. Brown, Anal. Chim. Acta 349 (1997) 263–269.
[22] H. Swierenga, W.G. Haanstra, A.P. de Weijer, L.M.C. Buydens, Appl. Spectrosc. 52 (1998) 7–16.
[23] F.W. Koehler, G.W. Small, R.J. Combs, B.B. Knapp, R.T. Kroutil, Anal. Chem. 7 (2000) 1690–1698.
[24] J. Lin, Appl. Spectrosc. 52 (1998) 1591–1596.
[25] F. Wulfert, W.T. Kok, O.E. de Noord, A.K. Smilde, Anal. Chem. 72 (2000) 1639–1644.
[26] D. Ozdemir, M. Mosley, R. Williams, Appl. Spectrosc. 52 (1998) 1203–1209.
[27] I.S. Adhihetty, J.A. McGuire, B. Wangmaneerat, T.M. Niemczyk, D.M. Haaland, Anal. Chem. 63 (1991) 2329–2338.
[28] F. Wulfert, W.T. Kok, A.K. Smilde, Anal. Chem. 70 (1998) 1761–1767.
[29] F. Despagne, D.L. Massart, Anal. Chem. 72 (2000) 1657–1665.
[30] O.E. De Noord, Chemom. Intell. Lab. Syst. 23 (1994) 65–70.
[31] C.V. Greensill, P.J. Wolfs, C.H. Spiegelman, K.B. Walsh, Appl. Spectrosc. 55 (2001) 647–653.

[32] C.L. Stork, B.R. Kowalski, Chemom. Intell. Lab. Syst. 48 (1999) 151–166.

[33] H. Mark, J. Workman, Spectroscopy 3 (1988) 28–36.

[34] H. Swierenga, P.J. de Groot, A.P. de Weijer, M.W.J Derksen, L.M.C. Buydens, Chemom. Intell. Lab. Syst. 41 (1998) 237–248.

[35] H. Swierenga, F. Wulfert, O.E. deNoord, A.P. deWeijer, A.K. Smilde, L.M.C. Buydens, Anal. Chim. Acta 411 (2000) 121–135.

[36] D. Ozdemir, M. Mosley, R. Williams, Appl. Spectrosc. 52 (1998) 599–603.

[37] Y.D. Wang, D.J. Veltkamp, B.R. Kowalski, Anal. Chem. 63 (1991) 2750–2756.

[38] M. Forina, G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Corti, P. Conti, R. Giangiacomo, C. Galliena, R. Bigoni, I. Quartari, C. Serra, D. Ferri, O. Leoni, L. Lazzeri, Chemom. Intell. Lab. Syst. 27 (1995) 189–203.

[39] J.S. Shenk, M.O. Westerhaus, W.C. Templeton, Crop Sci. 25 (1985) 159–161.

[40] J.S. Shenk, M.O. Westerhaus, Crop Sci. 31 (1991) 1694–1696.

[41] J.S. Shenk, M.O. Westerhaus, NIR News 4 (1993) 13–15.

[42] J.S. Shenk, M.O. Westerhaus, U.S. Patent 4866644. Sept. 12 (1989).

[43] E. Bouveresse, D.L. Massart, P. Dardenne, Anal. Chem. 67 (1995) 1381–1389.

[44] M. Forina, C. Casolino, Quim. Anal. 18 (1999) 61–71.

[45] Y.D. Wang, B.R. Kowalski, Appl. Spectrosc. 46 (1992) 764–771.

[46] Y.D. Wang, M.J. Lysaght, B.R. Kowalski, Anal. Chem. 64 (1992) 562–564.

[47] Z.Y. Wang, T. Dean, B.R. Kowalski, Anal. Chem. 67 (1995) 2379–2385.

[48] H.K. Barring, H.F.M. Boelens, O.E. deNoord, A.K. Smilde, Appl. Spectrosc. 55 (2001) 458–466.

[49] E. Bouveresse, D.L. Massart, Chemom. Intell. Lab. Syst. 32 (1996) 201–213.

[50] P.J. Gemperline, J.H. Cho, P.K. Aldridge, S.S. Sekulic, Anal. Chem. 68 (1996) 2913–2915.

[51] H.W. Tan, S.D. Brown, J. Chemom. 15 (2001) 647–663.

[52] E. Bouveresse, C. Hartmann, D.L. Massart, I.R. Last, K.A. Prebble, Anal. Chem. 68 (1996) 982–990.

[53] M. Forina, C. Casolino, Quim. Anal. 18 (1999) 49–59.

[54] B.G. Osborne, T. Fearn, J. Food Technol. 18 (1983) 453–460.

[55] J.A. Jones, I.R. Last, B.F. MacDonald, K.A. Prebble, J. Pharma. Biomed. Anal. 11 (1993) 1227–1231.

[56] S. Macho, A. Rius, M.P. Callao, M.S. Larrechi, Anal. Chim. Acta 445 (2001) 213–220.

[57] B. Walczak, E. Bouveresse, D.L. Massart, Chemom. Intell. Lab. Syst. 36 (1997) 41–51.

[58] K.S. Park, Y.H. Ko, H. Lee, C.H. Jun, H. Chung, M.S. Ku, Chemom. Intell. Lab. Syst. 55 (2001) 53–65.

[59] L. Duponchel, C. Ruckebusch, J.P. Huvenne, P. Legrand, J. NIRS 7 (1999) 155–166.

[60] F. Despagne, B. Walczak, D.L. Massart, Appl. Spectrosc. 52 (1998) 732–745.

[61] R. Goodacre, E.M. Timmins, A. Jones, D.B. Kell, J. Maddock, M.L. Heginbothom, J.T. Magee, Anal. Chim. Acta 348 (1997) 511–532.

[62] D.T. Andrews, P.D. Wentzell, Anal. Chim. Acta 350 (1997) 341–352.

[63] Y.L. Xie, P.K. Hopke, Anal. Chim. Acta 384 (1999) 193–205.

[64] P. Teppola, S.P. Mujunen, P. Minkkinen, Chemom. Intell. Lab. Syst. 45 (1999) 371–384.

[65] J.J. Siska, C.R. Hurburgh, J. NIRS 9 (2001) 107–116.

[66] S.C. Rutan, E. Bouveresse, K.N. Andrew, P.J. Worsfold, D.L. Massart, Chemom. Intell. Lab. Syst. 35 (1996) 199–211.

[67] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137–148.

[68] D. Jouan-Rimbaud, D.L. Massart, C.A. Saby, C. Puel, Anal. Chim. Acta 350 (1997) 149–161.

[69] F. Despagne, D.L. Massart, M. Jansen, H. van Daalen, Anal. Chim. Acta 406 (2000) 233–245.

[70] E. Bouveresse, D.L. Massart, P. Dardenne, Anal. Chim. Acta 297 (1994) 405–416.

[71] Y.D. Wang, B.R. Kowalski, Anal. Chem. 65 (1993) 1301–1303.

[72] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, Appl. Spectrosc. 51 (1997) 240–246.

[73] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.

[74] P. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 39 (1985) 491–500.

[75] A.J. Iversen, T. Palm, Appl. Spectrosc. 39 (1985) 641–646.

[76] T. Isaksson, T. Naes, Appl. Spectrosc. 42 (1998) 1273–1284.

[77] T.B. Blank, S.T. Sum, S.D. Brown, S.L. Monfre, Anal. Chem. 68 (1996) 2987–2995.

[78] T. Isaksson, B. Kowalski, Appl. Spectrosc. 47 (1993) 702–709.

[79] S.T. Sum, S.D. Brown, Appl. Spectrosc. 52 (1998) 869–877.

[80] H.W. Tan, S.T. Sum, S.D. Brown, Appl. Spectrosc. (in press).

[81] P. Geladi, H. Barring, E. Dabakk, J. Trygg, H. Antti, S. Wold, B. Karlberg, J. NIRS 7 (1999) 251–264.

[82] S. Wold, H. Antti, F. Lindgren, J. Öhman, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[83] J. Sjoblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold, Chemom. Intell. Lab. Syst. 44 (1998) 229–244.

[84] C.A. Andersson, Chemom. Intell. Lab. Syst. 47 (1999) 51–63.

[85] J.A. Westerhuis, S. de Jong, A.K. Smilde, Chemom. Intell. Lab. Syst. 56 (2001) 13–25.

[86] T. Fearn, Chemom. Intell. Lab. Syst. 50 (2000) 47–52.

[87] S. Wold, J. Trygg, A. Berglund, H. Antii, Chemom. Intell. Lab. Syst. 58 (2001) 131–150.

[88] J. Trygg, S. Wold, J. Chemom. 16 (2002) 119–128.

[89] J.A. Fernández Pierna, D.L. Massart, O.E. de Noord, Ph. Ricoux, Chemom. Intell. Lab. Syst. 55 (2001) 101–108.

[90] O. Svensson, T. Kourti, J.F. MacGregor, J. Chemom. 16 (2002) 176–188.

[91] H.W. Tan, S.D. Brown, J. Chemom. 16 (2002) 228–240.