

# A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy

Jerome J. Workman Jr.<sup>1,2</sup>

Applied Spectroscopy  
2018, Vol. 72(3) 340–365  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0003702817736064  
journals.sagepub.com/home/asp



## Abstract

Calibration transfer for use with spectroscopic instruments, particularly for near-infrared, infrared, and Raman analysis, has been the subject of multiple articles, research papers, book chapters, and technical reviews. There has been a myriad of approaches published and claims made for resolving the problems associated with transferring calibrations; however, the capability of attaining identical results over time from two or more instruments using an identical calibration still eludes technologists. Calibration transfer, in a precise definition, refers to a series of analytical approaches or chemometric techniques used to attempt to apply a single spectral database, and the calibration model developed using that database, for two or more instruments, with statistically retained accuracy and precision. Ideally, one would develop a single calibration for any particular application, and move it indiscriminately across instruments and achieve identical analysis or prediction results. There are many technical aspects involved in such precision calibration transfer, related to the measuring instrument reproducibility and repeatability, the reference chemical values used for the calibration, the multivariate mathematics used for calibration, and sample presentation repeatability and reproducibility. Ideally, a multivariate model developed on a single instrument would provide a statistically identical analysis when used on other instruments following transfer. This paper reviews common calibration transfer techniques, mostly related to instrument differences, and the mathematics of the uncertainty between instruments when making spectroscopic measurements of identical samples. It does not specifically address calibration maintenance or reference laboratory differences.

## Keywords

Calibration, transfer, alignment, multivariate, method comparison, bias and slope, uncertainty

Date received: 2 June 2017; accepted: 2 September 2017

## Introduction

One may discuss two main scenarios regarding the subject of calibration transfer to include: (1) the concept of using the same calibration model over time on multiple instruments using similar samples and measurement conditions; or (2) for using the same calibration model to adapt to new measurement and sample conditions, such as changes in natural product phenotype, chemical composition, temperature, sample presentation geometry, variation in instrumentation methods, and the like. The first scenario may be referred to simply as “calibration transfer between instruments” and the second, “calibration maintenance.” This review specifically addresses calibration transfer between instruments, and within that scenario more specifically the subject of calibration transfer involving instrument differences and methods to identify and

compensate for those differences. The topic of calibration maintenance is a significant topic of its own and is not addressed here in the allotted space for this article. In addition, in the scope of this review, limited equations are given for the most well-used transfer algorithms, with more extensive equations shown for instrument evaluation methods.

It is noted that multiple papers addressing calibration maintenance and the issues involved have been published.

<sup>1</sup>Unity Scientific, Milford, MA, USA

<sup>2</sup>National University, Department of Health and Human Services, La Jolla, CA, USA

### Corresponding author:

Jerome J. Workman Jr, Unity Scientific, Milford, MA and National University, Department of Health and Human Services, La Jolla, CA, USA.  
Email: jworkman2002@gmail.com

One paper discusses the concepts of calibration space and its relationship to a specific calibration model. The convolution of the initial sample matrix variables and instrument environment variables determine the conditions for the primary model, and as these conditions change a dynamic modeling maintenance approach is required.<sup>1</sup>

Calibration transfer in its purest form allows a calibration to be moved from one instrument to another with statistically retained accuracy and precision.<sup>2,3</sup> Ideally, a model developed on one instrument would provide the statistically identical analysis on the same sample set when used on other instruments, without bias or slope adjustments, without additional product samples for standardization, and without product recalibration.<sup>4</sup>

A variety of textbooks and papers exist describing the general methods and history of calibration transfer. Among these are references describing the details of the background chemometrics, as well as the instrumentation and data types involved.<sup>5,6</sup> One comprehensive calibration transfer review paper presents an overview of the different mathematical methods used for calibration transfer and a critical assessment of their validity and applicability.<sup>7</sup> The focus of that review paper is on a description of methods used for transfer of calibrations for near-infrared (NIR) spectra. A second early, yet prominent, review of calibration transfer thoroughly describes chemometric methods used for most aspects of calibration transfer.<sup>8</sup> That review classifies transfer methods as being represented by three main approaches, to include: (1) developing more robust calibrations, derived by combining data from multiple instruments and by using spectral pre-treatments of that data; (2) by modifying calibrations using bias and slope adjustments (the traditional method); and (3) by adjusting spectra, using such methods as piecewise direct standardization (PDS) combined with a patented method that requires spectra from transfer sample sets measured on each instrument involved in the transfer experiment.<sup>9</sup>

A comprehensive book chapter presents a review of calibration transfer methods and a critical assessment of the more common methods in terms of applicability and capabilities.<sup>10</sup> This previously published review presents an overview of calibration transfer methods and a critical assessment of the more common methods in terms of applicability and capabilities. There is a focus on specific methods for calibration transfer using NIR spectra, and it is noted that the methodology described is widely applicable to other analytical measurement techniques. Comprehensive reviews of chemometric methods, including calibration transfer, have been published since 1980 as part of the Fundamental Reviews in Analytical Chemistry.<sup>11</sup> One of the most cited of such fundamental chemometrics reviews, which delves into multiple topics related to calibration and calibration transfer methods, is found in Lavine et al.<sup>12</sup>

## Instrument Strategies for Calibration Transfer

### *Comparison of Transfer Methods*

A patented technique for transferring calibration models from a reference analytical instrument to a target analytical instrument is demonstrated. In this technique, transfer samples are measured on a reference instrument, producing a reference instrument response for each sample. These measurements are repeated for the target instrument, to produce a target instrument response for each transfer sample. One then generates transfer coefficients capable of performing a multivariate estimation of the reference instrument responses for the transfer samples from the target instrument responses for those samples. The transfer coefficients may then be used to convert a target instrument response for an unknown sample into the equivalent response for the reference instrument. The patent describes piecewise, classical, and inverse transfer techniques. The target instrument responses for the transfer samples are combined with reference instrument responses for the full set of calibration samples to derive a multivariate prediction model for the target instrument.<sup>13</sup>

A research paper describes the transfer of calibration models between NIR spectrometric instruments using three different standardization sets representing different sample chemistry types to be used as transfer standards. The first set contains samples that are very similar to the agricultural samples from three different sets to be analyzed, the second set contains generic biological and chemical standards, and the third transfer sample set contained pure organic and inorganic chemicals.<sup>14</sup> Another research paper describes several standardization strategies for calibration transfer methods, as well as an analysis of problems associated with the choice of the standardization samples.<sup>15</sup>

### *Instrument Alignment and Correction*

There are a number of research publications describing strategies and techniques for instrument alignment and correction. In these approaches the instrument wavelength or wavenumber (X) and photometric (Y) axes are carefully maintained as compared to physical standards such that instrument drift and the differences between instrument wavelength and photometric axes scales are minimized.<sup>4,16–19</sup> Full spectrum instruments and research associated with basic instrument measurement technology, instrumentation alignment, mathematical calculations, measuring agreement between instruments, and applying wavelength and photometric standards are described in a series of papers. The current state of commercial NIR spectrophotometers and their comparative accuracies are presented in this series. The performance of different instrument types, such as monochromator, Fourier transform (FT), and diode array-based instruments are

compared in detail with data comparisons shown in tabular and graphical forms.<sup>3,20–25</sup> The technical issues associated with multivariate calibration transfer (or calibration transfer) for spectroscopic instruments using absorption spectroscopy are addressed in a detailed chemometrics monograph book chapter. The chapter describes the methods for evaluating transfer effectiveness.<sup>2</sup>

### Filter Instrument Calibration Transfer

There have been many approaches to correcting instruments for calibration transfer. In one instance, filter instruments were used and a method demonstrated for computing calibrations that are robust against filter differences and different bandwidths of filters. The technique allows the development of general wavelength positions and regression coefficients that are robust against small variations in center wavelengths for filters and for wavelength variations that also exist between full scanning instruments.<sup>26</sup>

More recent work describing a similar approach uses a univariate procedure to correct for spectral measurements on a secondary instrument. The method refers to using isolated variables (i.e., specific wavelengths) rather than full spectra or spectral windowing for calibration and prediction. Given a set of transfer samples, a robust regression modeling procedure is used to obtain a model with low sensitivity with respect to changes in the univariate wavelength channels selected. The proposed method was reported to compare favorably with PDS as a method of transfer for octane and naphthenes in gasoline and for oil in corn.<sup>27</sup>

### The Master Instrument Concept

The utilization of NIR spectroscopy by the forage and feed industry was initially limited by the requirement to individually calibrate each instrument. Calibration transfer thus became a necessity for practical application and implementation of NIR over time. In the initial application of the terminology for calibration transfer involved references to a “master instrument” concept, where 60 sealed samples were used for measurements on primary (“master”) and target (“slave”) instruments for the transfer process. These inappropriate terms were initially used to refer to primary and secondary or parent and child instruments, respectively. One monochromator was chosen to be the master instrument. Calibration equations developed on it were transferred to six secondary/child instruments. The standard errors of difference (SED), including bias, among instruments for crude protein, acid detergent fiber, neutral detergent fiber, lignin, and in vitro dry matter disappearance were less than the comparative SED among laboratories for the reference chemical analysis. It was concluded that satisfactory calibration transfer could be accomplished with

ten samples. This was the most successful and practical method for calibration transfer in the mid-1980s.<sup>28</sup> The term “master instrument” is still widely used.

## Mathematical Correction or Standardization Methods

### Piecewise Direct Standardization (PDS)

A calibration transfer method termed PDS was applied to a set of gasoline samples measured on two different NIR spectrometers.<sup>29</sup> A calibration transfer method, PDS, was applied to a set of two-component samples measured on the same ultraviolet–visible (UV–Vis) spectrometer with the use of a cuvette cell having a 10 mm pathlength and a fiber optic probe with a 2 mm pathlength. Piecewise direct standardization proceeds by determining a structured transformation matrix using the spectra of a few samples measured with both devices. This transformation matrix can then be used to transform any spectrum measured on one device to that obtained on another device, thereby making the calibration model transferable between devices.<sup>30</sup>

More detailed chemometric improvements to the basic technique of PDS for NIR spectrometric instruments is further described for multivariate calibration transfer in a 1996 paper.<sup>31</sup>

One of the more common practical methods applied for field calibration transfers is PDS. The algorithm used for PDS determines the structured transformation matrix using the spectra of a set of transfer samples measured on both parent and child instruments. The resulting transformation matrix is used to transform any spectrum measured on any spectrophotometer, thereby allowing calibration model transfer.

For the PDS computation, the following Eqs. 1–3 are applied to convert the child instrument response matrix to the parent instrument response matrix.

$$A_{Pj} = A_{Cj}b_j \quad (1)$$

where  $A_{Pj}$  is the response of the standardization samples measured on the parent instrument at wavelength  $j$ ;  $A_{Cj}$  is the response for the standardization samples measured on the child instrument at wavelength  $j$ ; and  $b_j$  is the transformation or correction vector of coefficients for the  $j$ th wavelength.

The transformation matrix for PDS is computed using Eq. 2.

$$\hat{T}_{PDS} = \text{diag}(b_1^T, b_2^T, \dots, b_j^T, \dots, b_k^T) \quad (2)$$

where  $k$  is the total number of wavelengths in the standardization samples measured on both parent and child instruments.

Finally, the response prediction equation for any unknown sample measured on the child instrument is estimated using Eq. 3.

$$\hat{\mathbf{a}}^T = \mathbf{a}^T \hat{\mathbf{T}}_{PDS} \quad (3)$$

where the superscript  $T$  is the matrix transpose and  $\hat{\mathbf{a}}$  represents the estimated or predicted value for the measured sample  $\mathbf{a}$ .

### Orthogonal Signal Correction (OSC)

Orthogonal signal correction was applied to NIR spectra that were used in a calibration for the water content in a pharmaceutical product. Partial least squares (PLS) calibrations were then compared to other calibration models with uncorrected spectra, models with spectra subjected to multiplicative signal correction, and several other transfer methods. The performance of OSC was on the same level as for PDS and spectral offset correction for each individual instrument and PLS models with both instruments included.<sup>32</sup>

The main goal of OSC is to determine and remove from the spectral data,  $\mathbf{A}$ , the part of information which is not correlated with the concentration,  $\mathbf{Y}$ , within a limited number of orthogonal scores ( $\mathbf{t}$ ) and loadings ( $\mathbf{p}'$ ). Orthogonal signal correction is designed to remove variation from the spectral data,  $\mathbf{A}$ , that is orthogonal to the reference data  $\mathbf{Y}$ . The orthogonal variation is modeled by additional components for  $\mathbf{A}$  and results in the decomposition (Eq. 4).

$$\mathbf{A} = \mathbf{t} \times \mathbf{p}' + \mathbf{t}_o \mathbf{p}'_o + \mathbf{e} \quad (4)$$

where  $\mathbf{t}_o$  and  $\mathbf{p}'_o$  represent the scores and loadings for the orthogonal component and  $\mathbf{e}$  the residual. By removing the  $\mathbf{Y}$  orthogonal variation from the data via  $\mathbf{A} - \mathbf{t}_o \mathbf{p}'_o$ , OSC maximizes correlation and covariance between the  $\mathbf{A}$  and  $\mathbf{Y}$  scores to achieve both good prediction and interpretation.

### Orthogonal Projection

Dual-domain regression models using wavelet prism decomposition paired by orthogonal projection were used for transfer of NIR spectra.

The resulting model is robust against typical and routine instrument variations. The method, which does not require property values from the secondary data set, is tested on three NIR data sets, and is compared with orthogonal projection in the wavelength domain, OSC, and with model updating (MUP) approaches. The method was reported to be a competitive transfer method to each of these approaches.<sup>33</sup>

### Procrustes Analysis (PA)

In analytical chemistry, it is necessary to form instrument-dependent calibration models. Problems such as instrument drift, repair, or use of a new instrument create a need for recalibration. Since recalibration can require considerable costs and cause time delays, methods for calibration transfer have been developed. One paper shows that many of these approaches are based on the statistical procedure known as Procrustes' analysis (PA). Successful transfer by PA methods is shown to involve translation (mean-centering), rotation, and stretching (warping) of instrument responses to make them match.<sup>34</sup>

### Finite Impulse Response (FIR)

In one paper, four calibration transfer techniques were compared. Three established techniques, (1) finite impulse response (FIR) filtering, (2) generalized least squares weighting (GLSW), and (3) PDS, were evaluated. It was found that a fourth technique, baseline subtraction, was the most effective for calibration transfer. Using as few as 15 transfer samples, predictive capability of the analytical method was maintained across multiple instruments and major instrument maintenance.<sup>35</sup> Previously, a standard-free method using the FIR filter was successfully employed to transfer the NIR spectra of caustic brines, analgesics, and terpolymer resins. This paper carries the FIR transfer method one step further, leading to an improved algorithm that makes the transfer more robust and general by avoiding transfer artifacts in the filtered spectra.<sup>36</sup>

### Maximum Likelihood Principal Component Analysis (MLPCA)

A calibration transfer method, called MLPCA, is analogous to conventional principal component analysis (PCA), but incorporates measurement error variance information in the decomposition of the multivariate data.<sup>37</sup>

### Variable Successive Projections Algorithm (VSPA)

The VSPA used to select variables for building robust transferable multiple linear regression (MLR) models is investigated. Robust MLR models were compared to a primary instrument PLS model employing PDS to correct the transfer instrument spectra. In both data sets, the mean prediction errors at the transfer or child instrument for the robust VSPA-MLR and the PLS-PDS models were comparable, and slightly better for VSPA-MLR. The proposed approach appears to be a valid alternative to the commonly used PLS-PDS technique.<sup>38</sup>

### *Compressed Wavelet Domain Standardization (WTDS)*

Direct standardization in the compressed wavelet domain is referred to as WTDS. Calibration transfer in the compressed wavelet domain permits a reduction in processing time for the analysis of large spectral data sets with reported success in calibration transfer.<sup>39</sup>

### *Canonical Correlation Analysis (CCA)*

A method based on CCA for calibration model transfer was developed for calibration transfer. In this work, two real NIR data sets were tested in a comparative study between the CCA and PDS. A comparison of methods was positive for the CCA technique over PDS.<sup>40</sup>

Another paper describing calibration transfer using CCA coupled with wavelet transform (WTCCA) is applied to NIR spectra. It is noted that both changes in instrument and sample characteristics cause variation from the initial calibration conditions; this causes increased prediction errors with the application of the initial or primary calibration model. Both instrument and sample changes must be considered in the complete calibration transfer scenario. Canonical correlation analysis coupled with wavelet transform was reported to reduce root mean square error of prediction (RMSEP) by mitigating some of the variation caused by both the changes in instrument and sample states. Furthermore, WTCCA was also reported to be applicable to calibration transfer for on-line and in-line applications.<sup>41</sup>

### *Positive Matrix Factorization (PMF)*

Another approach to calibration transfer has been developed based on a factorization technique, PMF. Positive matrix factorization was developed and initially applied to environment data analysis. It has important differences from PCA. Since it is a least squares approach to solving the factor analysis problem, it can use subjective weights for individual data points and thereby make it possible to include uncertain data points like missing or below detection limit values in the analysis.<sup>42</sup>

### *Spectral Regression (SR)*

A calibration transfer method for NIR spectra based on SR is proposed. A comparative study of a spectral method and PDS for standardization on two benchmark NIR data sets is presented with results similar to PDS with background correction.<sup>43</sup>

A procedure for the transfer of the regression equation in NIR spectroscopy from a first instrument to a second instrument is presented. The procedure uses PLS

regression twice: in the first step to compute the relationship between the spectra of transfer samples of the two instruments; and in the second step to compute the regression equation (relationship between chemical variables and spectral variables) of the first instrument. Spectra were recorded with four different instruments, in four different laboratories with transfer results, as standard error of prediction (SEP), reported to be comparable to the original calibration.<sup>44</sup>

### *Wavelet Packet Transform Standardization (WPTS)*

A standardization algorithm, WPTS, based on a wavelet packet transform (WPT) and entropy criteria, is proposed to transfer spectra between two NIR spectrometers. By comparison, the proposed WPTS can reach a transfer performance comparable to WTS1, which is the best existing method, and better than WTS2 and classical PDS by a simpler technique of variable selection with a lower computational cost of calibration transfer.<sup>45</sup>

### *Stacked Partial Least Squares (SPLS) Regression*

An SPLS regression and stacked dual-domain (SDD) regression analysis was applied with commonly used techniques for calibration transfer to improve predictive performance from transferred multivariate calibration models. The paper reports that the predictive performance for gasoline samples for three conventional calibration transfer methods, PDS, OSC, and MUP, requiring standards measured on both instruments, was significantly improved.<sup>46</sup>

### *Alternating Trilinear Decomposition (ATLD)*

Alternating trilinear decomposition is used to model the changes in composition (or reference data), spectral response, and instrument differences. Because the relative intensity of each instrument is defined as a spectral difference between instruments, the spectra measured on different instruments can be standardized using a relative intensity correction scheme. Near-infrared spectra for corn and tobacco leaf samples were measured using three instruments with results showing that spectra measured on one instrument can be correctly predicted using PLS models built with the spectra measured on other instruments.<sup>47</sup>

### *Support Vector Regression (SVR)*

Support vector regression (SVR) is a nonlinear regression method used for multivariate calibration. Support vector regression may be used to transfer models between instruments subject to different environmental conditions by applying a transfer learning method. In this method, the



difference of the SVR regression vectors between old and new conditions is penalized by a  $L_2$  norm penalty, which can prevent the new SVR model from changing too much. This method uses the commonality of SVR models for different measurement conditions, and reduces the need for training samples when building a new SVR model. This approach may be used to update an SVR model quite easily when new instrument or sample conditions occur. This method was reported to improve prediction errors when compared to basic MUP and PDS. The approach is optimized when the standardization samples are selected from the primary model support vectors.<sup>48</sup>

### Dual Model based on Predicted Results Ratios

A dual-model method builds a primary calibration model and computes a correction model based on the ratios of the predicted results from the different instruments. The secondary instrument predictions are corrected to be equivalent to those of the primary model on the primary instrument. The results show that a precise correction model is obtained and the model can be used to correct the predictions of the primary model. The prediction error after the correction is reported to be equivalent to that of the primary model.<sup>49</sup>

### Proximal Methods

Proximal regularizers, a regularization framework from the field of convex optimization, are applied to standardize instrument functions over different wavelength ranges. In this approach, penalty terms are assigned to a direct standardization function to constrain the transfer matrix and the resulting transferred spectra, including sparsity and smoothness. This approach was reported as effective for application of a narrow wavelength range to a transfer or target instrument with a significantly broader wavelength range.<sup>50</sup>

## Miscellaneous Transfer Methods

### Multiplicative Signal Correction (MSC)

Two MSC algorithms are compared for the standardization of data from two NIR spectrometers. Absorbance spectra were measured in the range of 1000–2200 nm for a set of 45 jet fuel samples. Data from one instrument were standardized to match data from a second instrument using windowed MSC (W-MSC) and moving window MSC (MW-MSC). For W-MSC user-defined windows were selected and for MW-MSC the window size was optimized based on a two-step procedure: (1) assigning a cutoff window to avoid over-processing; and (2) selection of a specific window size based on sample leverage.<sup>51</sup>

For the basic MSC algorithm, each spectrum in a set of spectra represents the dependent variable (as a vector) in a

least squares regression case, where the mean spectrum is the independent variable (also a vector). From this computation, both the slope and intercept MSC variables are computed. Thus, the general least squares formula in standard notation is as Eq. 5.

$$\mathbf{a}_i = o_i + b_i \bar{\mathbf{a}}_j + \mathbf{e}_i \quad (5)$$

For this least squares linear regression equation,  $\mathbf{a}_i$  is an individual spectrum  $i$  from the data set,  $\bar{\mathbf{a}}_j$  is the mean spectrum of the set of spectra,  $o_i$  is the bias or offset, and  $\mathbf{e}_i$  is the residual spectrum.

The estimated or corrected MSC spectrum is computed using the computed slope ( $b_i$ ) and intercept ( $o_i$ ) constants as Eq. 6.

$$\hat{\mathbf{a}}_i(\text{MSC}) = \frac{(\mathbf{a}_i - o_i)}{b_i} \quad (6)$$

Multiplicative signal correction approximates a regression estimate of the mean spectrum with the remaining unexplained residual spectrum representing the spectral differences between the spectra within a set. This residual difference is assumed to contain much of the chemical information differences between samples in the spectrum set.

### Classification Methods

Spectrometer instruments are characterized by classifying their spectra into previously defined clusters. The spectra are mapped to the clusters and a classification is made based on similarity of extracted spectral features to one of the previously defined clusters. Calibration models for each cluster are provided to compensate for instrumental variation. A simplified method of calibration transfer maps clusters to each other, so that a calibration transferred between clusters models only the difference between the two clusters, substantially reducing the complexity of the model for transfer success.<sup>52</sup>

### Synthetic Calibration Spectra

A spectrum simulation method is described for use in the development and transfer of multivariate calibration models from NIR spectra. By use of previously measured molar absorptivities and solvent displacement factors, synthetic calibration spectra are computed using only background spectra collected with the spectrometer for which a calibration model is desired. The resulting synthetic calibration set is used with PLS regression to form the calibration model. This methodology is demonstrated for use in the analysis of physiological levels of glucose (1–30 mM) in an aqueous matrix containing variable levels of alanine, ascorbate, lactate, urea, and triacetin.<sup>53</sup>

### *Dispersive to FT Transfer*

Near-infrared calibration models were developed for the determination of content uniformity of pharmaceutical tablets. Data used in this study were acquired from five NIR instruments manufactured by two different vendors. One of these spectrometers is a dispersive-based NIR system while the other four were FT-based. The transferability of the optimized PLS calibration models developed on the primary instrument (A) located in a research facility was evaluated using spectral data acquired from secondary instruments B, C, D, and E. Instruments B and E were located in the same research facility as spectrometer A, while instruments C and D were located in a production facility 35 miles away. The same set of tablet samples were used to acquire spectral data from all instruments. To minimize the effect of instrument variability, calibration transfer techniques such as PDS and wavelet hybrid direct standardization (WHDS) were used.<sup>54</sup>

A principal component regression (PCR) calibration model developed on a scanning NIR spectrophotometer (using a standard cuvette) was transferred to a FT-NIR spectrophotometer (using a fiber optic probe) for the determination of NaCl in aqueous solutions by using the PDS method. It was reported that only two transfer samples were needed in obtaining a good calibration transfer.<sup>55</sup>

For calibration transfer using FT instruments, many approaches rely on the measurement of a subset of calibration transfer samples on each instrument, an approach that is often not practical. Specific to FT-designed instruments, an approach is proposed that models the optical variations between spectrometers caused by self-apodization, shear, and detector misalignment, such as off-axis field of view (FOV). These typical FT instrument variations are examined and equations describing their effects are given. Shear and off-axis detector FOV are the cause of nonlinear distortions for both the photometric amplitude and wavenumber axes. This type of distortion, demonstrated in FT instruments, cannot be accommodated for by the application of standard wavenumber or photometric alignment or correction methods. One novel approach has broad applicability to a variety of optical spectrometers for improvement of both multivariate calibration transfer and instrument standardization.<sup>56</sup>

### *Dispersive to Handheld Transfer*

The objective of this study was to assess the potential of a calibration transfer from the Foss NIRSystem 6500 to the Polychromix Phazir, a MEMS-based handheld spectrometer. The results show that good calibration models were obtained for various feed properties (i.e., fat, fiber, protein, and starch) developed on a Foss NIRSystem 6500, based on a spectral database of 9164 samples transferred to a Polychromix Phazir handheld spectrometer.<sup>57</sup>

### *Temperature Compensation*

Calibration transfer between spectra measured at different temperatures was investigated for NIR spectroscopic determination of NaCl in aqueous solutions. A PCR calibration model developed at 23.0 °C was transferred to 28.5 °C by using a PDS method.<sup>58</sup>

### *Two-Dimensional Method Transfer*

A method is described for calibrating two-dimensional responses measured on multiple instruments or on a single instrument under different operating conditions. From computer simulation and experimental data, it was found that the design of the standard sample is crucial for the parameter estimations and response standardization.<sup>59</sup>

## **Instrument Performance and Calibration Transfer**

Calibration transfer is a multistep process. The basic spectra are initially measured on at least one instrument (i.e., the parent, primary, or master instrument) and combined with the corresponding reference chemical information (i.e., reference or actual primary chemistry values) for the development of all calibration models. These models are verified and maintained on the initial calibration instrument and are transferred to other instruments (i.e., child, secondary, or transfer instruments) to enable analysis using the child instruments with minimal intervention and recalibration. Note that a perfect calibration transfer would require no intervention on the child or transfer instrument, or stated another way, the entire issue of calibration transfer between instruments disappears if the instruments or transformed spectra are precisely alike.

If instruments are the "same," then one sample placed on any of the instruments will predict or report precisely the "same" result at any time. Since instruments are not alike, and in fact change over time, calibration transfer techniques are applied to produce the best attempt at calibration model or data transfer. All spectroscopic techniques measure the volume fractions of the various components of a mixture, not necessarily the weight fractions (unless densities and mixing properties are identical). While the ideal approach would be to use volume fraction units to express the reference laboratory values, this is unlikely due to industrial familiarity with specific tests and units of measure. Current practice across many industries includes test procedures that specify how analysis is to be done and how the results are to be reported; this will take time to change. Therefore, the reference values may be some prescribed analysis method representing the weight fraction of materials, the volume percent of composition, or some arbitrarily defined unit. Thus, the nonlinearity caused by differences in the spectroscopy and the reported reference values must be compensated for to some extent, using calibration practices. This often involves additional wavelength information

or additional factors when applying multivariate calibration methods.

In practice, the current procedures for multivariate calibration transfer, or simply calibration transfer, is the use of a set of software algorithms, and physical materials (or standards) measured on multiple instruments, used to move calibrations from one instrument to another. If the basic method for calibration transfer does not produce satisfactory results, the user simply begins to measure more samples on the transfer (i.e., child) instrument until the model is updated to the new instrument based on the specific child instrument characteristics.

### Types of Spectrophotometers

Spectrometers or spectrophotometers, for analysis methods from UV through IR, and Raman spectroscopy are produced in many design types. There are instruments based on the grating monochromator with mechanical drive, grating monochromator with encoder drive, the Michelson interferometer in various forms, dispersive gratings with array detectors, interference and linear variable filter (LVF) instruments with single detectors, linear variable filters with array detection, MEMS based Fabry–Perot interferometers, digital transform actuator technologies (DMD-DLP), acousto-optic tunable filters (AOTF), Hadamard transform spectrometers, laser diodes, tunable lasers, various interference filter types, multivariate optical computing, and others.

Transfer of sophisticated multifactor PLS models across instruments of different design types is not common using any approach. The requirement of precise spectral shapes across instruments requires more sophisticated transforms than basic  $x$ - (wavelength or frequency) and  $y$ -axes (photometric value) corrections. Multivariate calibrations are developed for many data dimensions and require delicate instrument adjustments across this multidimensional space to fit measured data precisely. The center wavelength, data spacing, photometric response, photometric linearity, resolution, instrument line shape and symmetry, and other parameters must be nearly identical for multivariate calibrations to yield equivalent prediction results across different spectrometers and different design types of spectrometers.

### Instrument Comparison and Evaluation Methods

One of the essential aspects for determining the efficacy and quality of calibration transfer is to make certain the spectrometer instrumentation is essentially the same or as similar as possible before the calibration transfer experiment. Many standard tests are used to determine likeness between spectrophotometers. The specific tests outlined here are applicable to spectrometers using different wavelength or frequency ranges. Eight basic instrument tests and a summary evaluation are described as a standard method

to determine instrument measurement performance, including: wavelength/wavenumber accuracy; absolute wavelength/wavenumber accuracy; wavelength/wavenumber repeatability; absorbance/response accuracy; absorbance/response repeatability; photometric linearity; photometric noise; signal averaging (noise) tests; and an instrument line shape (ILS) test. If carefully conducted, these experiments provide specific information for diagnosing mechanical, optical, and electronic variations associated with design or manufacturing limitations. They provide immediate and objective data for correcting and refining instrumentation quality and likeness using either engineering-based or mathematics-based corrections.

### Instrument Optical Quality Performance Tests

The following series of tests is used to qualify instrument performance and to determine which issues are problematic due to deficiencies in instrument design features. These tests are related to likeness in measurement performance between instruments and to overall accuracy, and precision (as repeatability and reproducibility).

#### Wavenumber or Wavelength Accuracy

The wavenumber accuracy of a spectrophotometer is determined by using a suitable reference standard. For this test, repeat measurements of the same wavenumber standard are performed for a specific period and number of replicates. The mean spectrum wavenumber position ( $\bar{v}_i$ ) is calculated for each wavenumber measurement for the first-derivative zero crossings for scan-to-scan replicates, or an alternative peak selection algorithm. The wavenumber accuracy is reported as the average wavenumber difference value for the scan-to-scan set as  $\bar{v}_i$ .

$$\bar{v}_i = \frac{\sum_{i=1}^n (v_i - v_{ref})}{n} \quad (7)$$

where  $\bar{v}_i$  is the average wavenumber difference value for the scan-to-scan set;  $v_i$  are individual wavenumber shifts of the first-derivative zero-crossover position for replicate number of scans  $i$ ;  $v_{ref}$  is the reference wavenumber position for the standard, and  $n$  is the number of replicate measurements.

#### Absolute Wavenumber or Wavelength Accuracy

The absolute average wavenumber accuracy is reported as the absolute average wavenumber difference value for the scan-to-scan set as  $|\bar{v}_i|$ .

$$|\bar{v}_i| = \frac{\sum_{i=1}^n |v_i - v_{ref}|}{n} \quad (8)$$



where  $|\bar{v}_i|$  is the average absolute wavenumber difference value for the scan-to-scan set;  $v_i$  are individual wavenumber shifts of the first-derivative zero-crossover position for replicate number of scans  $i$ ;  $v_{ref}$  is the reference wavenumber position for the standard, and  $n$  is the number of replicate measurements.

### Wavenumber or Wavelength Repeatability (Precision) Test

The wavenumber repeatability or precision of the spectrophotometer is verified using a suitable reference standard. For this test, repeat measurements of the same reference standard sample are made. The wavenumber repeatability, as  $s_i$ , is computed as the standard deviation (SD) of the replicate measurements.

$$s_i = \sqrt{\frac{\sum_{i=1}^n (v_{ij} - \bar{v}_i)^2}{n-1}} \quad (9)$$

where  $s_i$  is the SD for the scan-to-scan wavelength precision or repeatability for scan-to-scan measurements, and where  $v_i$  are individual wavenumber positions of the first-derivative zero-crossover for replicate number of scans  $i$ ;  $\bar{v}_i$  is the average wavenumber value for the scan-to-scan set; and  $n$  is the number of replicates.

### Absorbance/Response Accuracy Test

The absorbance response accuracy of a spectrophotometer is measured using an acceptable transmittance or reflectance standard. For this test, repeat measurements of the same photometric standard are completed.

The absorbance accuracy is reported as the average absorbance difference value for the scan-to-scan set as  $\bar{A}_i$ .

$$\bar{A}_i = \frac{\sum_{i=1}^n (A_i - A_{ref})}{n} \quad (10)$$

where  $\bar{A}_i$  is the average absorbance difference value for the scan-to-scan set;  $A_i$  are individual absorbance measurements for replicate number of scans  $i$ ;  $A_{ref}$  is the reference absorbance values for the standard at the specific measurement wavenumbers, and  $n$  is the number of replicate measurements.

### Absorbance/Response Repeatability (Precision) Test

The absorbance repeatability or precision of the spectrophotometer is verified using a suitable reference standard. For this test, repeat measurements of the same reference

standard sample are made. The absorbance repeatability, as  $As_i$ , is computed as the SD of the replicate measurements.

$$As_i = \sqrt{\frac{\sum_{i=1}^n (A_i - \bar{A}_i)^2}{n-1}} \quad (11)$$

where  $As_i$  is the SD for the scan-to-scan wavelength precision or repeatability for scan-to-scan measurements, and where  $A_i$  are individual absorbance values for specific wavenumbers for replicate number of scans  $i$ ;  $\bar{A}_i$  is the average absorbance value for the scan-to-scan set; and  $n$  is the number of replicates.

### Photometric Linearity Test

The photometric linearity of the spectrophotometer by using a set of reference neutral density filters or suitable reflectance standards. Plot the observed response against the expected reference response. The slope ( $b$ ) of the line for reference ( $x$ ) versus measured ( $y$ ) data should be  $1.00 \pm 0.05$  and the intercept ( $a$ ) should be  $0.00 \pm 0.05$ . Calculate the slope and intercept using the reference material measurements of 5, 10, 20, 40, 60, and 80% reflectance/transmittance; or 1.3, 1.0, 0.70, 0.40, 0.22, 0.10 A, respectively.

The results are reported as a graph of the measured linear response against the expected response at the measured wavenumbers.

$$b = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2} \text{ and } a = \bar{y} - b\bar{x} \quad (12)$$

Note the slope ( $b$ ) and intercept ( $a$ ) for the data set of  $x_i$  (actual) and  $y_i$  (measured) pairs of measurements for each wavenumber position.

### Photometric Noise Test

The photometric noise is completed by using a reference neutral density filter or reflectance standard at either 0.0 or 1.0 A or approximately 100% reflection/transmission or 10% reflection/transmission, respectively.

The photometric noise is computed as the SD of the spectral response. This statistic is calculated for a set of spectra averaged over a standard measurement period (e.g., 30 s) using Eq. 11. It is reported for the standard value measured, e.g., 0.0 or 1.0 A.

### Instrument Line Shape (ILS) Test

The bandpass or line shape is important in transferring calibrations and in maintaining sufficient resolution in the

spectrometer to provide reliable spectral data and peak positions. The fine structure of product samples is maintained when the line shape is constant. The spectrometers must maintain constant line shape over the spectral range for transfer of calibrations. The integrity and structure of the measured spectra depend on an instrument with reliable, sufficient, and repeatable line shape. This may be measured using specialized equipment and methods as described in Workman.<sup>6</sup>

### Summary Specifications for Instrument “A likeness” Testing

General specifications for comparing instruments should meet basic minimum pre-established criteria for calibration transfer. Specific calibrations and prediction performance criteria will vary and so then will the comparative instrument performance metrics (Table 1).

### Mathematical Tests for Bias and Slope Significance Following Calibration Transfer

In the process of transferring calibrations from a parent to a child instrument, one may take four different fundamental strategies for matching the predicted values across instruments. Each of these strategies varies in complexity and efficacy: (1) adjusting the calibration model (i.e., the regression or  $b$ -vector); (2) adjusting the instrument as it measures spectra (i.e., the  $x$ - and  $y$ -axes) and other parameters; (3) adjusting the spectra (using various spectral transformations, such as matching  $x$ - and  $y$ -axes and adjust apparent lines shapes via smoothing; or (4) adjusting the final predicted results (via bias or slope adjustments).

### Comparisons When Transferring Calibrations

For spectroscopic-based measurements using multivariate calibrations, one may compare the standard or reference concentrations for a set of samples to the spectroscopic-based predicted values. One may also compare the response of the parent instrument to that of the child instrument. In making these comparisons, one may perform statistical tests for bias, correlation, and slope. A statistically significant difference in bias should result in a change of the bias. A statistically significant result in correlation or slope should result in a basic multivariate recalibration. One should investigate the cause of slope changes between instruments and make corrections to the instrument or spectra.

### Bias or Slope Adjustments of Predicted Results Across Parent and Child Instruments

A significant bias between parent and child predicted values results mainly from instrumental differences. Other sources

of significant bias changes between reference values and spectroscopy-based predicted values are due to chemical or spectral interferences. These differences cause significant bias in the measured analyte concentration due to the effect(s) of an additional component, changes in a physical property of the sample, or from an analytical measurement anomaly.<sup>60</sup>

### Bias (Means) One-Sample $t$ -Test Between Parent and Child Instruments

The concept of mean differences (or bias) requires a test of significant to determine whether a bias adjustment in predicted results should be implemented. An appropriate statistical test will determine if the variation in the mean values (bias) between sample sets of predicted values is within expected random fluctuation for a normally distributed population of measurements. The larger the sample set used to test calibration transfer bias, the more accurate the estimate of the bias value. Thus, using 20 test samples would give a more accurate estimate of the bias than ten samples. The smaller the standard error of the mean (SEM), the greater the confidence will be of the true bias value (Eqs. 14 and 15). If one uses 20 rather than ten samples for bias checking, a more powerful estimate of the true bias occurs by a factor of  $\sqrt{20}/\sqrt{10} = 1.41$ . This is true only if there is no slope difference. Note: it is acceptable to make the statistically significant bias correction even if there is a slope difference.

The statistical test used to determine bias significance is a simple parametric one-sample  $t$ -test. For this test, one need only have the average predicted mean value for a set of samples on the parent instrument and the set of predicted values for the same sample set measured on the child instrument. The mean value for the set of reference samples measured on the parent (i.e., calibration) instrument is designated as  $\bar{X}_{Parent}$ . Note for this test only the mean value is known. The mean value for the same set of reference samples measured on the child (i.e., transfer) instrument is designated as  $\bar{X}_{Child}$ . For this test, we are setting our test hypotheses as follows:

$$\begin{aligned} H_o : \bar{X}_{Parent} &= \bar{X}_{Child} \\ H_A : \bar{X}_{Parent} &\neq \bar{X}_{Child} \end{aligned}$$

Then, compute the SD for the child instrument ( $s_{Child}$ ) for the sample predicted data on the measured set of samples as in Eq. 13:

$$s_{Child} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - 1}} \quad (13)$$

where  $\bar{y}_i$  is the mean predicted value for the sample set for the child instrument;  $y_i$  is the individual predicted value for the set of test samples for the child instrument; and  $n$  is the number of samples used for the calculation.

Then, compute the SEM for the child instrument, which is found by using Eq. 14:

$$SEM = \frac{s_{Child}}{\sqrt{n}} \quad (14)$$

Here, we are testing whether the predicted value mean is statistically the same for both the parent and child instruments. There are many variations of testing mean differences, but this is a basic test for comparing means when the sample size of two groups is identical. This test determines whether the average predicted values are statistically the same for the test set used.

Furthermore, we note that the reference values are not assumed to be known for the test set. For this  $t$ -test statistic, we are able to compute our  $t$ -test for mean bias significance using Eq. 15:

$$t = \frac{\bar{X}_{Child} - \bar{X}_{Parent}}{SEM} \quad (15)$$

If this resulting  $t$ -test is greater than the  $t$  critical value for  $n-1$  degrees of freedom, we then know the bias is significant (accept  $H_A$ ) and it should be changed. If the  $t$ -value computed is less than the critical value of  $t$ , it is not significant (we accept  $H_0$ ) and should not be changed. If the bias is significant, then test the difference between the slopes of the two lines or compare their correlation coefficients. Note that  $n$  in this case is the number of samples in the test set, for example,  $n = 20$ .

### Bias (Means) Two-Sample $t$ -Test Between Parent and Child Instruments

If the predicted values are known for the set of test samples from both the parent and child instruments, one uses a parametric two-sample  $t$ -test. For this example, there is an identical number of samples for each bias test, so the sample sizes for each test of the parent and child instruments are identical. No assumption is made that the reference values are known for the set of test samples. For this test, the mean predicted value for the set of reference samples measured on the parent (i.e., calibration) instrument is designated as  $\bar{X}_{Parent}$ . The mean predicted value for the same set of reference samples measured on the child (i.e., transfer) instrument is designated as  $\bar{X}_{Child}$ . The hypotheses tests are the same as in the one-sample  $t$ -test:

$$H_0 : \bar{X}_{Parent} = \bar{X}_{Child}$$

$$H_A : \bar{X}_{Parent} \neq \bar{X}_{Child}$$

With this information, the SDs for both the parent ( $s_{Parent}$ ) and child ( $s_{Child}$ ) predicted data on the measured set of samples are computed using Eqs. 16 and 17, respectively.

$$s_{Parent} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{n-1}} \quad (16)$$

$$s_{Child} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n-1}} \quad (17)$$

where  $\bar{x}_i$  and  $\bar{y}_i$  are the mean predicted values for the sample set for each instrument; and  $x_i$  and  $y_i$  are the individual predicted values for the set of test samples for each the parent and child instruments, respectively.

Then, the  $t$ -test is computed for mean bias significance using Eq. 18:

$$t = \frac{\bar{X}_{Child} - \bar{X}_{Parent}}{\sqrt{\frac{s_{Child}^2}{n_c} + \frac{s_{Parent}^2}{n_p}}} \quad (18)$$

If this  $t$ -test value is greater than the  $t$  critical value for  $n_c + n_p - 2$  degrees of freedom, the bias is significant (accept  $H_A$ ), and thus the bias should be changed. If the  $t$ -value computed is less than the critical value of  $t$ , it should not be changed and is not significant (we accept  $H_0$ ). If the bias is significant, then test the difference between the slope of the two lines or compare their correlation coefficients for significance.

### Comparing the Correlation Coefficients Between Parent and Child Instruments Using the ( $r$ -to- $z$ Transform) Significance Test

The slope should not be changed to adjust the predicted values following calibration transfer. However, the slope significance should be computed as an indication of a need to recalibrate the instrument using a new multivariate model. For this test, the correlation coefficients of the sample set for the parent instrument are compared to the correlation coefficient of the same sample set on the child instrument; this correlation is termed Pearson's  $r$  statistic.

To compare the correlation coefficients between the parent and child instrument predicted values to test whether there is a significant difference between them, regress the parent instrument predicted values (as the  $x$ , reference, or independent values) against the child instrument predicted values (as the  $y$ , or dependent values). For the resulting correlation, expected to be near 1.0, we actually use a value of  $r = 0.995$  for testing. Our hypotheses

tests for a suitable correlation between the parent and child instrument predictions is given as:

$$H_o: r > 0.995; r=1.0$$

$$H_A: r < 0.995; r \neq 1.0$$

With this set of hypotheses tests we use the following Eq. 19 to compute the Z test statistic:

$$Z_{Test} = \frac{|Z_{Child} - Z_r|}{\sqrt{\frac{1}{n_c-3} + \frac{1}{n_p-3}}} \quad (19)$$

This  $Z_{Test}$  is computed and compared for  $\alpha=0.05$ , and yields a Z critical value of 1.96. So, if the computed  $Z_{Test}$  statistic is greater than 1.96, the alternate hypothesis ( $H_A: r \neq 1.0$ ) is accepted. This indicates that the instrument predictions are not alike and the actual correlation is different between the child and parent instruments. If the  $Z_{Test} < 1.96$  the correlations are alike and assumed to be near 1.0, so the null hypothesis is accepted ( $H_o: r = 1.0$ ). If the null is accepted, we state that the correlation is essentially the same (or 1.0) for the child versus parent predicted values.

### Slope Significance Limit Test Between Parent and Child Instruments

If the child or transfer instrument is expected to perform identically to the parent instrument for predictions, one may test this assumption by computing the confidence limits for comparison between the parent and child predicted values using criteria computed only from the parent instrument. For a slope significance test, one observes the slope change between the parent instrument predicted values for a set of test samples over time. The slope changes are retained for analysis.

For this calculation, compute the predicted values for the set of 20 test samples on the parent instrument one week apart. Thus, we would have two sets of predicted values designated by different measurement times as  $x_i$  for time 0 (the reference values) and for the same set of samples measured one week later designated as  $y_i$ . (Note these test samples must be chemically stable over the one-week period.)

For the computation of the acceptable slope confidence interval (CI) for the child instrument to be considered alike to the parent instrument, compute three basic sets of numbers. These are: (1) the set of parent predicted values at time 0 ( $x_i$ ); (2) the set of parent predicted values at week 1 ( $y_i$ ); and (3) the set of predicted values from the simple linear regression between these  $x_i$  and  $y_i$  values, designated as  $\hat{y}_i$ . Note that  $n$  is the number of  $x_i, y_i$  pairs (i.e., 20 for this example). From these three sets of values, we can

compute the SD of the residuals for the predicted values, from the regression ( $\hat{y}_i = b + mx_i$ ) as:

$$s_{y/x} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} \quad (20)$$

Then, compute the SD for the desired slope as:

$$s_m = \frac{s_{y/x}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2}} \quad (21)$$

Lastly, compute the confidence limit for the slope, given as:

$$CI_m = m \pm t \times s_m \quad (22)$$

where  $t$  is the critical value of the t distribution, two-tailed test, at  $\alpha=0.05$  and with degrees of freedom as  $n-2$ . This value can be found in any table of the critical values of the t distribution.

Thus, the test criteria for testing a child instrument slope following calibration transfer would be within the confidence limits of the test measured on the same parent instrument over approximately this time interval. To complete this parent and child comparison, one would designate the parent instrument results as  $x_i$  and the child measured results as  $y_i$ . Then the slope would be computed for the regression and must fall within the computed confidence limits for the time 0 and week 1 parent tests. This is a test of equivalence in slope between parent and child predicted values.<sup>61</sup>

### Approaches for Developing Global Models

Various methods have been proposed to produce the universal model or a calibration that is mostly robust against standard instrument differences or changes with time as are common to commercial instruments of today. These have been referred to as robust models, or global models. For computing a robust model, various experimental designs have been constructed to better represent the product, reference values, and instrument calibration space and to include typical changes and interferences that should be included within the model for it to be broadly applicable. Using this approach, one might design a factorial experiment for the composition of the calibration set to include multiple variations typically encountered during routine analysis. A list of some of these variations may consist of differences in sample pathlength, sample cup, sample temperature, sample moisture content, flow rate, particle size, interferent content, instrument type, constituent ratios,

sampling parameters, and others.<sup>62</sup> These approaches will work for a period of time until the instrument drifts or the product or constituent chemistry changes. These types of changes are expected and thus routine recalibration (i.e., MUP) would be required as a standard procedure if the chemistry or instrument changes are considered significant. A method for selection of specific robust wavelengths in MLR models that are more forgiving toward wavelength differences in interference filter based instruments has been demonstrated to be effective.<sup>26</sup>

A robust method to reduce the effect of interference on NIR measurements has been demonstrated. This pre-processing method is termed OSC; as it removes variation from the spectral data,  $x$ , that is orthogonal to  $y$ , it maximizes correlation and covariance between the  $x$  and  $y$  scores to achieve more accurate predictions in NIR spectroscopy.<sup>63,64</sup>

### *Augmenting Models Over Time*

If instrument differences are significant, the predicted values between parent and child instruments will be unacceptably large. In these cases, one may begin to collect more samples on the child instrument and then re-compute the multivariate model using the majority of high leverage sample data from the new instrument. This in effect uses the calibration transfer as a temporary solution or bridge to building an accurate model on the child instrument. Such a practice is often used when the instrument differences are too significant for direct and accurate calibration transfer. Augmentation using noise simulation with characterized noise sources may be added during the modeling step to improve prediction results and stability.<sup>65,66</sup>

### *Sample Selection to Improve Spectral Data*

Sample selection methods have been used and perfected since the beginning of chemometric methods and spectroscopy. There are many methods and a variety of nomenclatures for these sample selection techniques. The purpose of such methods is to remove the redundancy in spectral data such that the most repetitive samples do not have excessive influence on the regression model. This provides a basis such that the regression line more appropriately fitted to the extreme samples, including those with high and low analyte concentrations. Such methods of sample selection include: random subset selection, manual subset selection, spectral subtraction methods for “uniqueness” tests, stratified sample selection, discriminant based selection techniques using spectral distances, correlation matching techniques, and others. These methods are described in more detail in Workman et al.<sup>67</sup> One of the first successful approaches for sample selection was a process that used spectral subtraction to remove the unusual spectra from all of the other spectra described in Honigs et al.<sup>68</sup> Even for

early research, these methods significantly improved the SEP for multiple constituents in forage analysis.<sup>69</sup>

### *Spectral Data Transformation*

This process consists in altering spectral data from the child instrument to be more like that measured on the parent instrument. Piecewise direct standardization has been used most often for this procedure.<sup>70</sup>

### *Special Standardization Mathematical Approaches*

This approach has been discussed and referenced in the following paper.<sup>21</sup>

### *Locally Weighted Regression (LWR) Methods*

As a global or universal calibration approach, LWR or local regression methods use spectral data and corresponding reference data to build a “local” calibration using only those samples near the unknown or test sample spectra. For example, the unknown spectrum is measured and the sample spectra most like the unknown are selected from a resident database. The multivariate calibration model is then computed using only the local samples. The samples can be down-weighted for use in the regression model based on distance from the unknown sample. This allows quite accurate prediction analysis when a variety of samples and instrument type data is incorporated into a spectral database. The first description of the use of this method for spectroscopy is referenced in Naes et al.,<sup>71</sup> based on original work from the statistics community.<sup>72</sup>

### *Use of Indicator Variables*

A method has been used previously that simultaneously optimizes the calibration for multiple instruments and provides  $t$ -tests for the differences between them. The method creates the calibration by running the samples on several instruments (the more the better). All the data are added into the calibration and indicator variables are used between the instruments. With this approach one obtains the best calibration that is optimized for all instruments, the coefficients of the indicator variables are the biases between the instruments, and the  $t$ -tests computed from this method are valid for the corresponding bias values.<sup>73</sup>

### **Measuring the Uncertainty Between Instruments Following Calibration Transfer**

In any specific calibration transfer situation, if the prescribed method for calibration transfer does not produce satisfactory results, the user simply begins to measure more samples directly using the child (transfer) instrument



**Table 1.** General specifications for comparing instruments.

Instrument parameter	Expected deviation
Wavenumber accuracy	Accuracy must be $\pm 0.1 \text{ cm}^{-1}$ in agreement with calibration (reference) as absolute maximum deviation
Absolute wavenumber accuracy	Absolute deviation must be less than $\pm 0.8 \text{ cm}^{-1}$ versus NIST reference standard for multiple wavenumbers
Wavenumber repeatability	$< 0.01 \text{ cm}^{-1}$ ( $1\sigma$ )
Absorbance/response accuracy	Absolute deviation must be less than $\pm 0.02\%T$ versus NIST traceable measurement of $0.09\%T$ (3.046 AU) to $0.10\%T$ (3.000 AU) for specific Reference Standards at $\sim 1333 \text{ nm}$ ( $7500 \text{ cm}^{-1}$ ) and $\sim 2222 \text{ nm}$ ( $4500 \text{ cm}^{-1}$ ), respectively. Accuracy must be $\pm 0.01\%T$ ( $\pm 0.02 \text{ AU}$ ) in agreement with calibration (reference) parent instrument (absolute maximum deviation)
Absorbance/response repeatability	$< 0.001 \text{ AU}$ ( $1\sigma$ )
Photometric linearity	Slope: $1.00 \pm 0.02$ ; Intercept: $\pm 0.02$ (Absolute) for method
Photometric noise	$\text{RMS} < 0.001 \text{ AU}$ for $0.1\%T$ ND filter
Instrument line shape (ILS) test	
Center wavelength measured	Accuracy must be $\pm 0.05 \text{ cm}^{-1}$ in agreement with calibration (reference) line using interpolation or center of mass peak picking method
FWHM measured (% deviation)	$< 5\%$ deviation from calibration (reference) parent instrument (absolute maximum deviation)
Asymmetry measured (% deviation)	$< 5\%$ deviation as per prescribed method from calibration (reference) parent instrument (absolute maximum deviation)

until the model is updated based on the child or transfer instrument characteristics. Such transfer of calibration is also accomplished by recalibration on the child instrument or by blending samples measured on multiple instruments into a single calibration. Although the blending approach improves method robustness (or ruggedness) for predicted results across instruments using the same calibration, it is not relevant for all applications, for analytes having small net analyte signals, or for achieving the optimum accuracy.

### *How to Tell if Two Instrument Predictions are Statistically Alike*

The main question when comparing parent to child instrument predictions, or a reference laboratory method to an instrument prediction, or results from two completely different reference methods, is how to know if the differences are meaningful or significant and when they are not. There is always some difference expected, since an imperfect world allows for a certain amount of “natural” variation. However, when are those differences considered statistically significant differences or when are the differences too great to be acceptable? There are a number of reference papers and guides to tell us how to compute differences, diagnose their significance, and describe the types of errors involved between methods, instruments, and analytical techniques of many

types. Whether the analytical method is based on spectroscopy and multivariate calibration methods, other instrumental methods, or even gravimetric methods.<sup>74</sup> Youden and Steiner<sup>74</sup> describe collaborative laboratory tests, the ranking of laboratories for accuracy, outlier determination, ruggedness tests for methods, and diagnosing the various types of errors in analytical results.

Table 2 illustrates simulation data that are representative of spectroscopy data following multivariate calibration. These data could refer to results from different analytical methods or instruments, designated as A and B; for example, if A and B are designated from two different instruments for 20 different samples. The original data from a calibrated instrument is designated as A. The results of data transferred to another instrument are represented by B. There are duplicate measurements for A as A1 and A2, and for B as B1 and B2, respectively. From these data we will perform an analysis and look for levels of uncertainty and acceptability for the analytical results.

For comparison of the data in Table 2, A1 and B1 data are selected, noting that one might compare multiple analyses for multiple instruments as a more powerful test of the estimated SD of a measurement result for any method or instrument combination. However, the approach shown here is adequate for estimates of uncertainty for typical analysis and calibration transfer applications. Here, the National Institute of Standards and Technology (NIST)

**Table 2.** Data used for illustration for instruments A and B.

Sample no. (n)	A1	A2	B1	B2
1	12.4	12.2	12.1	12.0
2	12.9	12.5	12.5	12.6
3	14.0	14.4	13.9	13.7
4	16.0	16.3	15.9	15.7
5	13.2	13.0	12.9	12.8
6	12.8	12.4	12.7	12.9
7	14.5	14.8	14.9	14.5
8	13.0	13.3	13.4	13.7
9	13.6	13.3	13.5	13.4
10	12.7	12.4	12.6	12.9
11	14.2	14.5	14.6	14.1
12	16.3	16.6	16.4	16.2
13	17.8	17.9	17.9	17.4
14	18.0	18.3	18.5	18.5
15	14.5	14.2	13.9	13.5
16	17.2	17.0	17.5	17.7
17	14.4	14.1	14.6	14.8
18	15.2	15.4	15.7	15.5
19	16.6	16.7	16.5	16.3
20	13.5	13.2	13.1	13.4

nomenclature is used.<sup>75,76</sup> We note the following set of equations:

$$y = f(X_1, X_2, \dots, X_N) \quad (23)$$

where  $y$  is the estimated analytical value for any sample as a function of a series of measurement quantities such as  $X_1, X_2, \dots, X_N$ ; and where each  $X_i$  is an independent observation (or measurement). Note that when using this nomenclature, the A1 and B1 measurements for each sample are denoted as  $X_i$  measurements. Also note the value ( $y_i$ ) for each sample measurement is estimated as the sample mean from  $N$  independent measurements and is denoted as  $X_{i,k}$ , giving the following relationship:

$$y_i = \bar{X}_i = \frac{1}{N} \sum_{k=1}^N X_{i,k} \quad (24)$$

The estimated analytical value ( $y_i$ ) is the mean for a number of measurements of that sample ( $\bar{X}_i$ ) using the analytical method prescribed. It then follows that the standard uncertainty  $u(X_i)$  with reference to the measured values ( $X_i$ ) is equal to the estimated SD of the mean:

$$u(X_i) = s(\bar{X}_i) = \left( \left( \frac{1}{n(n-1)} \sum_{k=1}^N (X_{i,k} - \bar{X}_i)^2 \right) \right)^{\frac{1}{2}} \quad (25)$$

So, to apply this to the data illustration from Table 2, using A1 and B1 as  $X_1$  and  $X_2$ , therefore:

$$u(y_i) = u(X_i) = s(\bar{X}_i) = \left( \frac{1}{n(n-1)} \sum_{k=1}^N (X_{i,k} - \bar{X}_i)^2 \right)^{\frac{1}{2}} \quad (26)$$

where  $u(y_i)$  is the estimated standard uncertainty for a series of measurements on multiple samples where the mean value of the measurements for each sample is used for comparison. The equations above are often used for multiple measurements of a single physical constant. When a single measurement pair as A1 and B1 is used for each sample, variance is computed for each of the 20 samples and results pooled to give an estimate of standard uncertainty  $u(y_i)$  as:

$$u(y_i) = u(X_i) = s(\bar{X}_i) = \left( \frac{1}{n(n-1)} \sum_{k=1}^N (\sigma_{i,k})^2 \right)^{\frac{1}{2}} \quad (27)$$

Compiling these results yields the following for the standard uncertainty reported as  $u(y_i) = 0.226$ .

## Relative Standard Uncertainty

This is denoted as  $u_r(y_i) = u(y_i)/|y_i|$  and so applying the previously computed results we note that  $y_i$  is equal to 14.648 and  $|y_i| = 14.648$  as the mean of all A1 and B1 values. Therefore, the  $u(y_i) = 0.226$ , and so the relative standard uncertainty is reported as:

$$u_r(y_i) = u(y_i)/|y_i| = 0.226/14.648 = 0.0154 \quad (28)$$

**Confidence Levels Reported.** The confidence levels would be expressed as follows:

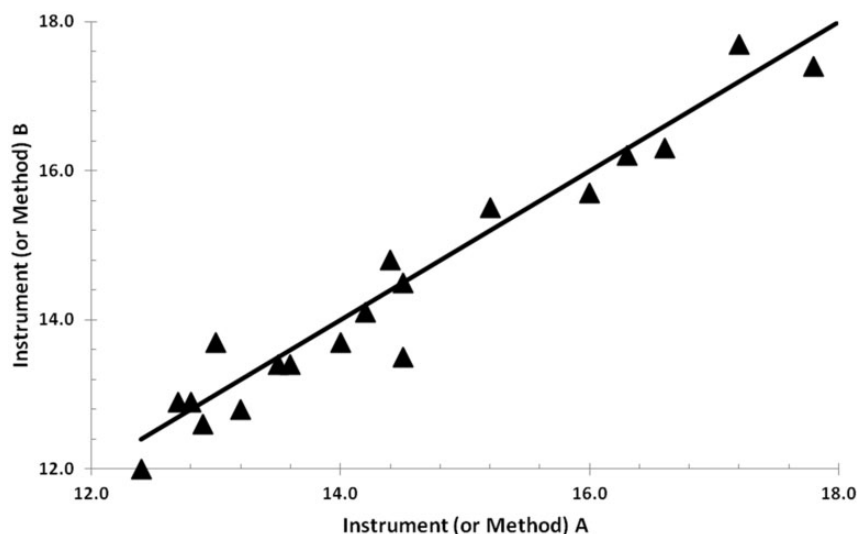
$$\text{For } 68\% : y_i \pm u(y_i) = 14.648 \pm 0.226 = 14.42 - 14.87 \quad (29)$$

$$\text{For } 95\% : y_i \pm 2 \times u(y_i) = 14.648 \pm 0.452 = 14.20 - 15.10 \quad (30)$$

Thus, the expression of certainty for this example of data would be as:

$$y = 14.648 \pm 0.226U \text{ or } y = 14.648(0.226)U \quad (31)$$

Therefore, for the analyst the final step is deciding if this is a satisfactory result across the two instruments.



**Figure 1.** Instrument A1 (x-axis) compared to instrument B1 (y-axis) for a set of samples, and with data points compared to perfect line of equality.

### Bland–Altman Test as a Measure of Limits of Agreement

As a second analysis of the data in Table 2, one of the definitive techniques used for method comparisons within the highly regulated clinical sciences is termed the Bland–Altman method.<sup>77</sup> For a comparison of this method, two single measurements are taken for each sample on each instrument as A1 and A2. The first measurement is used for comparison, the second for a repeatability study. For this method of comparison, a line of equality plot is made to compare two methods or two instruments. The various x,y data points are plotted against a perfect straight line of equality. The authors make the point that correlation ( $r$ ) measures the strength of the relationship between two variables, but it does not measure the agreement between them. Perfect agreement is indicated by the data points lying directly on the line of equality. A perfect correlation is indicated if the points lie along any straight line (Fig. 1). The authors emphasize that: (1) correlation indicates the strength of a relationship between variables—not that the analytical results agree; (2) a change in scale does not affect correlation, but drastically affects agreement; (3) correlation depends upon the range of the true quantity (analyte) in the sample; (4) tests of significance are mostly irrelevant between two similar analytical methods, and (5) data in poor agreement analytically can be highly correlated (Fig. 2).

A Bland–Altman plot (Fig. 3), extremely familiar to clinical analysts, demonstrates a good visual comparison technique to evaluate the agreement between two methods or instruments. The x-axis (abscissa) for each sample is represented by the average value for each sample obtained from

the comparison results (using two methods or two instruments). The y-axis (ordinate) for each sample is represented by the difference between one method and the second method (or instruments A and B in our example) for each sample. Such a plot uses the mean and  $\pm 2$  SDs as the upper and lower comparison thresholds.

To assess if the data are in close enough agreement for analytical purposes between A1 and B1, the bias or mean difference ( $\bar{d}$ ), the SD of the differences ( $s$  or SD), and the expected “limits of agreement” are computed. These are expressed as  $\bar{d} \pm 2 \times s$  for a 95% confidence level.

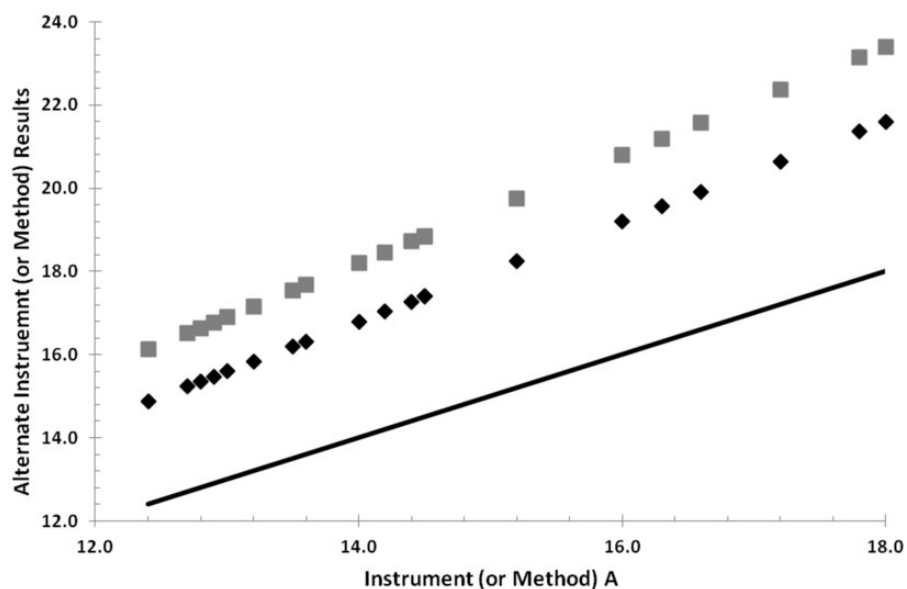
The mean difference is computed as the average of all the differences between the comparative instruments, for each sample as:

$$\bar{d}_i = \frac{\sum_{i=1}^n (A_i - B_i)}{n} \quad (32)$$

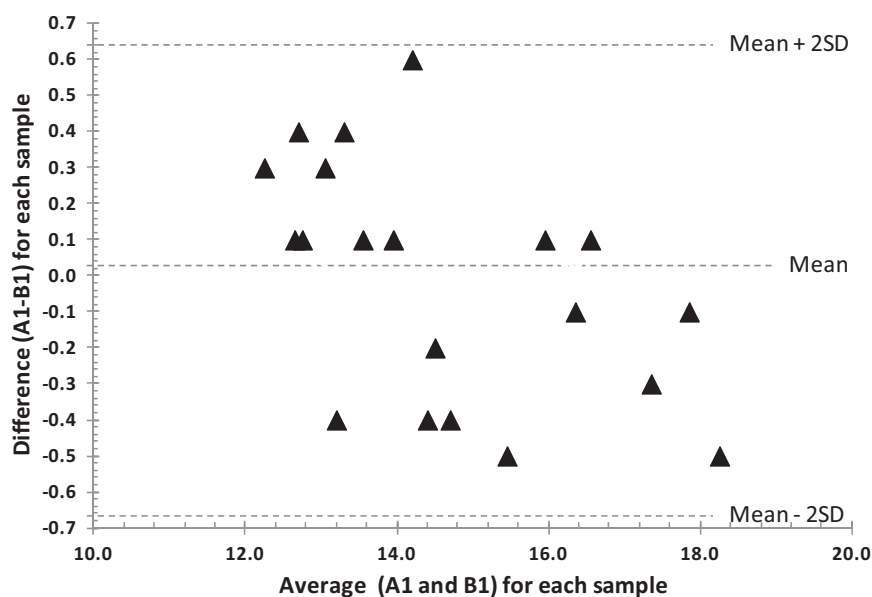
where  $\bar{d}_i$  is the mean difference between each sample for instrument or method  $A_i$  and instrument or method  $B_i$  for each sample  $i$ ;  $A_i$  is the individual result for the A instrument and  $i$ th sample;  $B_i$  is the individual results for the B instrument and  $i$ th sample; and  $n$  is the number of sample pairs used for comparison.

The SD for this comparison for the set of samples is computed as:

$$s_i = \sqrt{\frac{\sum_{i=1}^n (A_i - B_i)^2}{2n}} = \sqrt{\frac{\sum_{i=1}^n D_i^2}{2n}} \quad (33)$$



**Figure 2.** Showing the line of equality (solid line) for a perfect agreement as well as two examples of perfectly correlated data that are not in analytical agreement (black and gray squares). This illustrates that correlation as an imperfect representation of agreement between methods.



**Figure 3.** The Bland-Altman plot indicating the difference for a set of samples plotted against the mean for each sample for instruments or methods A1 and B1.

where  $s_i$  is the SD of differences between the two instruments or methods;  $D_i$  is the individual differences for the  $i$  number of samples for A and B instruments.

Therefore, for Table 2 data for A1 and A2, the following analysis results. Bias is  $\bar{d} = -0.015$ ; SD for all samples is  $s = 0.226$ ; 95% confidence level is  $\bar{d} + 2s = 0.438$ ; and  $\bar{d} - 2s = 0.468$ .

If the number is considered too large for a 95% confidence of the result agreement, then these instrument and calibration combinations are not considered equivalent. On the other hand, if these limits of agreement are acceptable for the application where they are used this is an acceptable method. In a clinical situation, a physician determines the level of accuracy or agreement required for

critical intervention decision-making; this would be analogous to a process control supervisor or analytical scientist assessing the acceptable level of agreement between comparative methods in order to use the alternate method or instrument as a substitute for a primary method.

## Wavelength/Wavenumber Difference Considerations

Wavelength or wavenumber precision and accuracy, formally termed repeatability and reproducibility, within a single instrument and between multiple instruments over time, is essential for optimum transfer of quantitative calibrations, and for the continuous use of databases in qualitative search techniques. To maintain the integrity of data, it is essential that the wavelength axis be stable within instruments and identical across different instruments over changing measurement conditions. This is a difficult task for the basic commercial instrument designers and manufacturers in terms of making spectrometers highly precise and accurate on the wavelength axis. Questions arise as to the engineering tolerances required for high precision spectrometers in terms of mechanical, optical, and electronic components. Also, essential to the precise wavelength alignment are the exact procedures required to properly calibrate the wavelength axis of the spectrometer based on measurements of known reference standards.<sup>23,78</sup>

### *Different Approaches to Alignment of the Wavelength/Wavenumber Axis*

Due to the inherent uncertainties of standard reference materials (SRMs), one may ask if they are precise enough to be useful for aligning the wavelength axis of spectrometers over time.<sup>79</sup> One may inquire if the uncertainties inherent in SRMs are too large for the precision alignment required for today's spectrometers. For example, NIST SRM 1920a reported a published uncertainty of  $\pm 1$  nm.<sup>80,81</sup> With large uncertainties in published wavelength peak positions, referred to as measurands, would it not be most reasonable to ensure that wavelength accuracy is within inherent NIST uncertainties when using SRMs, but strive to be far closer in agreement for instrument-to-instrument matching?

A seminal paper on the use of rare earth oxides (REOs) for use as wavelength standards in NIR is familiar to many spectroscopists.<sup>80</sup> One important wavelength standard made for use in the NIR region has been SRM 1920a, consisting of a mixture of REOs. This SRM has had known stability problems related to variation in wavelength scale, and as such has been replaced with SRM 2036 glass.<sup>82</sup> SRM 1920a is not a particularly stable material and wavelength registration changes based on temperature have been published as ( $\pm 0.15$  nm). In the certification of SRM 1920a,

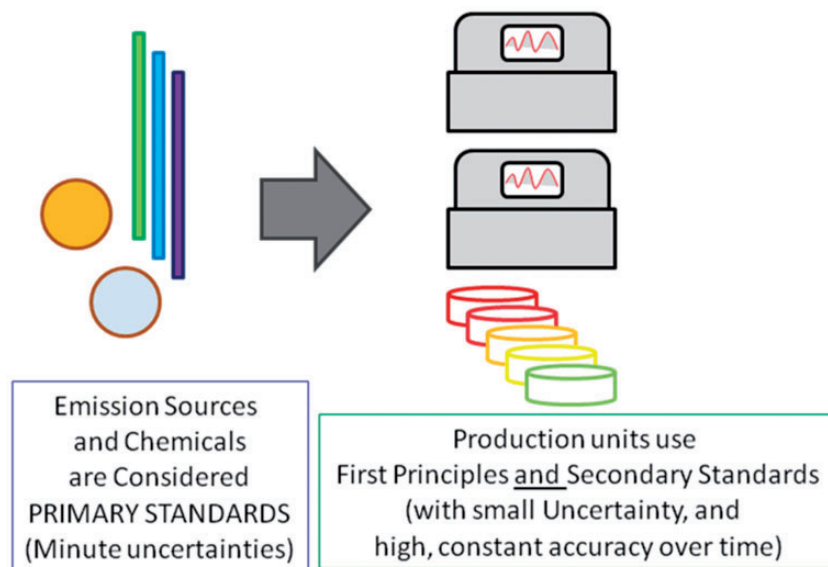
NIST also experienced challenges with their secondary Cary Model spectrometer and their manual method of peak selection resulting in the large reported uncertainty. In calibrating this SRM, NIST reportedly used the precise NBS/NIST spectrometer to measure three wavelengths (viz., 1012.9, 1260.7, and 1535.5 nm). The Cary instrument was reported to exhibit a deviation of  $\pm 1$  nm during the test measurement period. For the NIST spectrometer, emission reference lamps included the use of Ne at 703.24 nm, Xe at 881.94 and 979.97 nm, Hg at 1013.97 and 1529.58 nm, and Kr at 1816.73 nm. The use of such atomic line spectra for wavelength registration produces the lowest uncertainty.<sup>83</sup>

For the longer NIR wavelength region, 1-2-4 trichlorobenzene was used to mark the 2152.60 nm position. Toluene (used for IR at 3290.8, 3422.0, 3484.0, 5147.2, 5381.8, and 5549.0 nm) and carbon disulfide (2224.0 nm) have also been recommended for wavelength calibration in early published work.<sup>84,85</sup> The task of selecting appropriate standards for certified peak wavelength positions is critical for improved x-axis (wavelength) alignment and calibration of spectrometers. Wavelength alignment technology is dependent on selecting a suitable set of reference peak wavelengths and stable materials to align each spectrometer system using best available measurement practices. If wavelength precision must approach  $\pm 0.05$  nm and accuracy across the entire spectral region is better than  $\pm 0.2$  nm for 95% CI, a series of technological improvements must be implemented in the future.

So, what if a master instrument, with its own idiosyncratic scanning features, can be duplicated? One might analyze a paired reference standard with the master instrument and align the newly manufactured instrument to the set of wavelength peak position values found using the master instrument. This approach was used in the early development of NIR instruments in an attempt to make them more similar.<sup>26,86,87</sup> The problem with this approach is the entire world may have a different set of nominal wavelength values (measurands) than the master instrument. Spectrometers change over time and conditions and wavelength alignment drifts. A common spectrometer may demonstrate inherent nonlinearity in its wavelength axis, which does not allow accurate alignment throughout the complete spectral range. So, to consistently align the wavelength axis with other instrument types, and even to align a single instrument over time, requires not only repeatability, but also accuracy indicated by agreement with reference standard peak positions (i.e., measurands).

These reference peak positions are provided by a set of known physical standards (physical in the sense of primary lines that are universally reproducible). Such standards may be primary emission standards (i.e., atomic lines), physical standards of mixed rare earth glasses, or even organic liquids at standardized pathlength and temperature. The





**Figure 4.** Illustration of instrument calibration using first principles methodology. Instrument performance parameters may be corrected in real-time using primary standards to calibrate resident secondary standards. One or both types of standards are included within instruments.

measurands for calibrating wavelength scales would optimally be obtained using first principles whenever possible as this approach is universal and based on sound physical science and metrology, which by definition is fixed to a specific known uncertainty (Fig. 4). The national laboratories around the globe use this approach as much as possible.<sup>76,88</sup> Other materials have been proposed to calibrate instruments that may be used as physical standards, such as etalons. The use of etalons for wavelength alignment was successfully implemented for a process diode array spectrometer.<sup>89–93</sup> The instrument was extremely precise but too expensive for market acceptance, as the spectrograph had originally been designed for the harsh conditions of an earth orbiting satellite.

To determine a unique set of peak wavelength values based on an arbitrary master instrument does not serve our purposes adequately over time. One would ideally expect high precision with universal accuracy so that databases now and in the future may be aligned with any instrument technology and be carried reproducibly into the future. In addition, using a first principles approach allows one to standardize the data based on known wavelength positioning at any time. First principles wavelength alignment may be defined electronically and thus no physical instrument or object is required to transfer the technology necessary for accurate wavelength axis calibration. What is required to make all this work are repeatable and well-characterized peak wavelength standards, either emission sources, pure compounds, or solid REO glasses with known peak wavelength positions, stability, and well-defined uncertainty. Requirements are for glasses, crystalline

polymers, and pure chemical solutions with an uncertainty of better than  $\pm 0.1$  nm for multiple wavelengths. For IR and Raman systems, the uncertainty should be better than  $\pm 0.1\%$  or better than  $0.05 \text{ cm}^{-1}$  (wavenumber).<sup>94</sup> If a spectrometer is repeatable, but not reproducible (accurate), then database and calibration transfer becomes an issue of trying to align the wavelength axis to some unknown or poorly defined system.

#### *National Institute of Standards and Technology Uncertainty Number for Standard Reference Materials (SRMs)*

The NIST uncertainty is a measure of uncertainty unlike the typical manufacturer reports as either precision or accuracy. For example, NIST uncertainty has been calculated from Eq. 34.

$$U = \sqrt{A^2 + B^2 + C^2} \quad (34)$$

where  $A$  equals twice the largest SD of measurement of multiple emission lines or primary standards (measurands) versus the spectrometer measured values over a period of 31 days (or some designated period);  $B$  = twice the SD of the uncertainty in the peak selection method used; and  $C$  is the maximum variation in the standard due to temperature changes over a specified range and conditions. (Note: The number two [twice] is referred to as the  $k$ -value also known as the coverage factor, see SRM 2036.<sup>76,82</sup>)

In testing SRM 1920a, the reported error for multiple wavelengths using emission lines for the NIST

**Table 3.** Commercial instruments using identical SRM 1920a material (wavelength measurement comparison using different instrument models from different manufacturers).

NIST (measurand)	Instrument A	Instrument B	Instrument C	Instrument D	Instrument E	Instrument F	Instrument G
798.0	798.27	798.30	797.95	NA	798.51	NA	NA
886.5	886.73	885.85	886.05	NA	886.22	NA	NA
1132.9	1133.31	1133.39	1132.17	1132.46	1134.02	1134.02	1132.55
1261.8	1261.85	1261.71	1261.07	1261.54	1261.74	1261.74	1261.10
1320.2	1318.39	1320.21	1320.01	1320.71	1317.25	1315.61	1320.56
1534.6	1534.83	1533.97	1534.37	1535.89	1533.86	1533.86	1535.66
1681.4	1682.13	1680.69	1681.23	1682.54	1681.37	1681.37	1682.34
1757.6	1757.83	1756.83	1757.59	1758.02	1756.55	1756.67	1758.02
1847.3	1847.05	1846.36	1846.87	1847.58	1846.26	1846.42	1847.58
1935.5	1935.15	1935.42	1935.69	1935.86	1934.46	1934.42	1935.86
2006.3	2009.13	2006.70	2006.69	2010.14	2009.45	2010.26	2010.30
Mean difference	−0.23	0.24	0.22	−0.79	0.22	0.36	−0.71
SD	1.09	0.52	0.35	1.27	1.53	2.24	1.35

spectrometer was  $\pm 0.265$  nm for all the data points measured over 31 days. The NIST actual reported uncertainty for SRM 1920a is equal to Eq. 35.

$$U = \sqrt{0.57^2 + 0.41^2 + 0.15^2} = 0.75 \text{ nm} \quad (35)$$

This value was rounded up to  $\pm 1$  nm considering the added uncertainty of the Cary spectrometer measurements and rare earth mixture stability over the lifetime of the SRM.

A more detailed description of uncertainty computations and rationale with respect to wavelength SRMs is found in the SRM references as noted.<sup>8,76,82</sup> There are significant complexities to the question of uncertainty in reference materials and so the reader is referred directly to these citations. One may also obtain more detailed information on the history of the international metrical committees responsible for defining the technical details for uncertainty associated with measurement results; the reader is referred to Appendices C and D in the SRM 2036 reference<sup>82</sup> for historical details.

### Commercial Instrument Wavelength Data

A set of measurement data was taken as outlined in Table 3. These are real data measured on a series of commercially calibrated NIR spectrometers using the identical sample and sample holder to measure a variety of performance characteristics. The wavelength measurement data across the A–G designated instruments consists of seven commercial instrument models manufactured by four different companies. The identities remain anonymous for the purpose of this demonstration. The models used for this experiment include dispersive monochromators, a diode-array-based instrument, and FT-NIR systems.

Table 3 demonstrates the instrument models A–G as individual columns, with the different nominal wavelength values (measurands) for the SRM 1920a standard material in the first column. The results of a set of standard spectra are shown in each column demonstrating the peak positions as determined using a five-point cubic spline, peak interpolation method.<sup>95</sup> No data processing or smoothing was applied to the spectra, but after the manufacturer's recommended calibration the reference material was measured using three replicates. For Table 4, a single spectrum from each instrument was selected for identification of peak positions. The reported wavelengths are shown as compared to the NIST nominal wavelengths (measurands) for a nominal bandwidth of 10 nm. The last two rows show the mean difference and SD between the measured values and the measurand values for each peak, respectively, for the complete set of instrument measurements at each wavelength. From Tables 3 and 4 it may be observed that there is significant variation from commercial NIR instruments and published nominal wavelength positions for this reference standard.

The results may be analyzed from a measure of uncertainty viewpoint. For wavelength “precision” (SD of error between the measured and measurand values) the SD of difference ( $s_i$ ) is computed for each measured wavelength as Eq. 36:

$$s_i = \sqrt{\frac{\sum_{j=1}^n (\lambda_i - \lambda_{ref})^2}{n}} \quad (36)$$

where ( $n$ ) is the number of replicate wavelength measurements at each peak, ( $\lambda_i$ ) is the measured wavelength peak position for each replicate measurement, and ( $\lambda_{ref}$ ) is the

**Table 4.** Results of commercial instruments on SRM 1920a for each nominal wavelength (measurand 10 nm bandwidth values as top row).

Instrument	798.0	886.5	1132.9	1261.8	1320.2	1534.6	1681.4	1757.6	1847.3	1935.5	2006.3
A	798.27	886.73	1133.31	1261.85	1318.39	1534.83	1682.13	1757.83	1847.05	1935.15	2009.13
B	798.30	885.85	1133.39	1261.71	1320.21	1533.97	1680.69	1756.83	1846.36	1935.42	2006.70
C	797.95	886.05	1132.17	1261.07	1320.01	1534.37	1681.23	1757.59	1846.87	1935.69	2006.69
D	NA	NA	1132.46	1261.54	1320.71	1535.89	1682.54	1758.02	1847.58	1935.86	2010.14
E	798.51	886.22	1134.02	1261.74	1317.25	1533.86	1681.37	1756.55	1846.26	1934.46	2009.45
F	NA	NA	1134.02	1261.74	1315.61	1533.86	1681.37	1756.67	1846.42	1934.42	2010.26
G	NA	NA	1132.55	1261.10	1320.56	1535.66	1682.34	1758.02	1847.58	1935.86	2010.30
Mean difference	0.257	-0.287	0.231	-0.264	-1.237	0.034	0.267	-0.241	-0.426	-0.234	2.653
SD	0.33	0.43	0.73	0.40	2.19	0.79	0.68	0.65	0.67	0.62	3.04

Pooled SDs of error for all wavelengths measured = 1.303 nm.

measurand or reference wavelength value. Therefore the SD at each wavelength ( $s_i$ ) is calculated as Eq. 36.

The mean difference between the measured and measurand values for wavelength “accuracy” is determined by Eq. 37.

$$\text{Wavelength MEAN DIFF} = \bar{\lambda}_i - \lambda_{ref} \quad (37)$$

where ( $\bar{\lambda}_i$ ) is the replicate average peak wavelength position for the instrument measurements for each reference peak; and ( $\lambda_{ref}$ ) is the reference (measurand) wavelength position for the standard at each peak wavelength.

### Relative Standard Uncertainty Across Commercial Instruments

In referring to Table 4, the pooled SD of error between all measured wavelength values as compared to the NIST measurand values across all commercial instruments for the identical standard sample is calculated using the pooled SDs and is computed to be 1.303 nm.

The confidence levels for a randomly selected commercial spectrometer would be expressed as follows:

For .68 % :  $y_i \pm u(y_i) = \text{Nominal (Measurand)}$

NIST Wavelength for SRM  $\pm 1.303$  nm

For .95 % :  $y_i \pm 2 \cdot u(y_i) = \text{Nominal (Measurand)}$

NIST Wavelength for SRM  $\pm 2.606$  nm

The expression of uncertainty for this example of data would be as:

$y = \text{Nominal (Measurand) NIST Wavelength for SRM} \pm 1.303 \text{ nm}$  or  $y = \text{Nominal (Measurand) NIST Wavelength for SRM} (1.303)U$

Table 3 results indicate that for many multivariate equations where precise wavelength alignment (or agreement) is

critical for accurate calibration transfer, with minimal bias and slope changes, one can envision a large technical difficulty in transferring multivariate calibration models from one current instrument type to another and one instrument manufacturer to another.

Use the above equation for wavelength SD of error between the measured values and the measurand values and the following equation for instrument wavelength repeatability. Using the mean spectrum wavelength position ( $\bar{\lambda}_i$ ) for replicate measurements at each peak ( $\lambda_{ij}$ ), the repeatability is computed as Eq. 38.

$$s_i = \sqrt{\frac{\sum_{j=1}^n (\lambda_{ij} - \bar{\lambda}_i)^2}{n - 1}} \quad (38)$$

where ( $s_i$ ) is the SD for the scan-to-scan wavelength (as precision or repeatability) for the instrument under the specific measurement conditions, and where ( $\lambda_{ij}$ ) are individual wavelength peak positions for replicate measurement  $i$ ; ( $\bar{\lambda}_i$ ) is the average peak position value for the replicate scan set; and  $n$  is the number of replicate measurements.

The mean spectrum wavenumber position ( $\bar{\lambda}_i$ ) is calculated as Eq. 39:

$$\bar{\lambda}_i = \frac{\sum_{i=1}^n \lambda_i}{n} \quad (39)$$

From the Table 3 data we may compute the following tabular results (as Table 4).

### Photometric Difference Considerations

Photometric or amplitude accuracy and precision, as reproducibility and repeatability, respectively are essential for

building consistent large databases over time for use in qualitative searches or quantitative multivariate analysis. If the instrument in use is inconsistent in terms of linearity and photometric or amplitude accuracy, the analytical precision and accuracy will be jeopardized over time. Photometric and linearity drift over time within a single instrument or between instruments creates errors and variation in the accuracy of measurements using databases collected over time with different photometric registrations.

Photometric precision and accuracy, when measured using specific protocols, are termed photometric repeatability and reproducibility. The photometric stability within a single instrument and between multiple instruments over time is important for seamless transfer of multivariate calibrations and for the unbiased application of qualitative search libraries. The stability in the wavelength axis of spectrophotometers is the most critical parameter.<sup>24</sup> Data integrity depends on the accuracy and precision of the  $x$  (i.e., wavelength or wavenumber), and  $y$  (i.e., photometric or amplitude) axes. There are multiple factors affecting the photometric stability of spectrophotometers, these include: alignment and mechanical tolerances of optical elements; detector noise characteristics and amplitude; detector linearity; detector electronics, including gain and offset settings; signal amplifier noise; and sample presentation distance, repeatability, and reproducibility.<sup>23,78</sup>

Accurate and precise photometric axis alignment is a complex issue for commercial instrument manufacturers and design engineers. As always, the quality of manufacturing, design tolerances, quality of components, and maintenance of stringent system calibration protocols are required for photometric stability. As is the theme of this paper, the appropriate use of primary and secondary reference standards and materials is also essential to superior metrical performance using spectroscopy.

### *Different Approaches to Alignment of the Photometric Axis*

Approaches to aligning the photometric axis of any spectrophotometer may involve measuring a standard reference sample and correcting the instrument photometric values to those values obtained by measuring the same sample on a highly accurate and traceable spectrophotometer. This is basically a sound approach, but the question becomes how often should the recalibration be performed, and how much difference between measured and reference standard values is acceptable before one recalibrates the photometric axis. When photometric or amplitude reference materials with known reference values and spectral shapes are used, then instruments may always be corrected in near real-time to a known reference standard, and this will always be true even in the future. For a correction using a stable, physical reference approach, there is only

dependency on natural laws and physics, not on unstable materials and instrumentation.

The master instrument concept was acceptable for its time,<sup>28,86,87</sup> but metrical requirements are best served by first principle calibration. The measurands (or actual nominal values) for calibrating photometric scales would optimally be obtained using first principles whenever possible as this is a universal approach and is based on pure physical science and metrology, which by definition is fixed to a specific known uncertainty. The national laboratories around the globe use this approach as much as possible.<sup>76,88</sup>

A protocol to correct photometric values measured using a spectrophotometer would ideally provide high precision with universal accuracy so that expensive and time-consuming databases created over time are aligned with any instrument technology, which could be carried reproducibly into the future. In addition, using a first principles approach allows one to standardize the data based on known photometric values at any time. First principles photometric calibration may be defined electronically and thus no physical instrument or object is required to transfer the technology required for photometric adjustments or calibration.

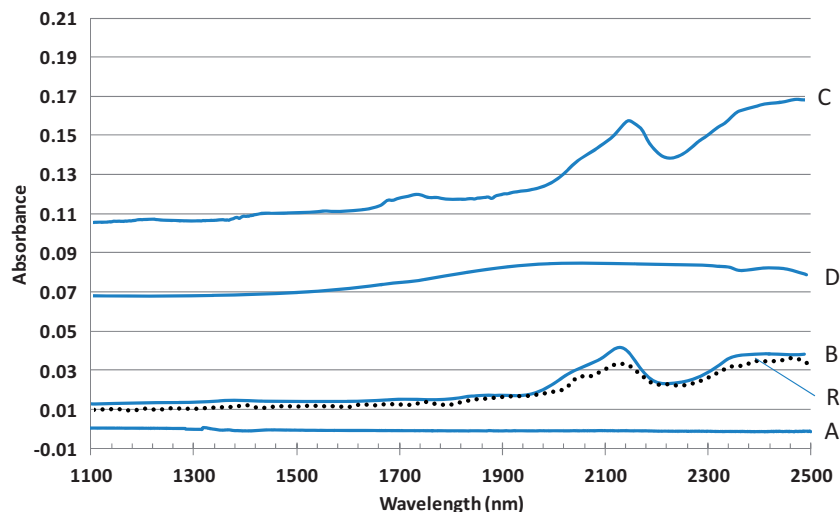
To create a first principles photometric calibration protocol, one requires repeatable and well-characterized photometric material standards, with known stability, and with well-defined uncertainty. To have minimal effects for calibration transfer, absorbance or photometric response accuracy should have an absolute deviation of less than  $\pm 0.02\%T$  (or  $R$ ) versus a NIST traceable (or calibrated) spectrophotometer measurement at  $99.5\%T$  (or  $R$ ). Photometric accuracy must be  $\pm 0.01\%T$  (or  $R$ ) in agreement with calibration (reference) instrument (absolute maximum deviation) across all wavelengths. Absorbance/response repeatability of  $\pm < 0.0001A$  ( $1\sigma$ ) would be optimal.

### *National Institute of Standards and Technology Uncertainty Number for Photometric SRMs*

National Institute of Standards and Technology uncertainty is a measure of uncertainty unlike the typical manufacturer report as either precision or accuracy. For example, NIST uncertainty has been calculated using Eq. 40.<sup>80</sup>

$$U = \sqrt{a^2 + b^2 + c^2} \quad (40)$$

where  $a$  = twice the largest SD of measurement of multiple wavelength photometric values (measurands) versus the spectrometer measured values over a period of one month (or other designated period of several weeks duration);  $b$  = twice the SD of the uncertainty in the photometric data measurement method used (note: the noise on any photometric measurement adds uncertainty to the absolute photometric values); and  $c$  is the maximum variation in the reference standard due to humidity, material instability,



**Figure 5.** An identical Fluorilon R99 sample measured on four different (i.e., A–D) commercial NIR spectrophotometers as compared to a reference (R) spectrometer measurement (dotted line).

photometric bleaching, or temperature changes over a specified range and conditions.

### Measurement of Photometric Accuracy

A comparison of the leading NIR spectrophotometer performance characteristics demonstrates the state of the art for wavelength and photometric accuracy and precision in modern NIR instrumentation. For this paper, an identical sample of Fluorilon R99 was measured across four different instrument models and manufacturers; the same sample in the same sample holder is used for all the measurements. (Avian Technologies LLC). The identical sample was measured on a double monochromator system carefully calibrated using NIST SRM and other calibration protocols to obtain the reference (or measurand) values. The example in Fig. 5 shows the results for Fluorilon R99, approximately 97% reflectance standard. The dotted line spectrum indicates the reference instrument measurements.

It is immediately obvious that only one manufacturer is carefully matching the instrument photometric values to the reference standard values. Even so, only one of the instruments matches the standard with sufficient minimum accuracy. This indicates that data collected on any of the instrument models tested will result in significant photometric differences that must be compensated for if using databases collected across the various instrument models and manufacturers.

The mean difference between the measured and measurand values for Photometric “accuracy” is determined by Eq. 41.

$$\text{Photometric MEAN DIFF} = \bar{A}_i - \bar{A}_{ref} \quad (41)$$

**Table 5.** Average photometric values for instrument models A–D compared to a reference spectrophotometer across a range of 1100–2500 nm, reported as R.

Instrument No.	Ave. measured (R)	Ave. reference (R)	Mean difference (R)
A	0.9950	0.9662	0.0288
B	0.9620	0.9662	–0.0042
C	0.7490	0.9662	–0.2172
D	0.8363	0.9662	–0.1300

where  $(\bar{A}_i)$  is the average measured reference material photometric value for the test instrument in absorbance units (A) over the entire wavelength range; and  $(\bar{A}_{ref})$  is the average measured reference material photometric values for the calibrated reference instrument in absorbance units (A) over the entire measurement wavelength range. The results are reported as: Mean Difference or Photometric “accuracy.”

Table 5 demonstrates the average accuracy for each of the instrument models A–D over the full wavelength range of 1100–2500 nm compared to the reference calibrated spectrophotometer (e.g., Perkin-Elmer Lambda 9/19 UV-Vis-NIR spectrophotometer). Note that instrument B is within a minimum specified deviation of ( $\pm 0.01$  R) across the full spectral range. This result indicates that it is a useful practice to calibrate each instrument based on a known photometric standard.

### Conclusion

General specifications for comparing instruments should meet basic minimum pre-established criteria for calibration



transfer. These criteria are based on the actual use of the spectrophotometer and the accuracy required for measurements for any specified application. General requirements for each spectrometer should depend on the use, application, and overall performance requirements. When extremely high performance and “aliqueness” of instrument-recorded spectra are required then comparison statistics are more important. Specific calibrations and prediction performance criteria will vary and so then will the comparative instrument performance metrics.

There are many conventional and unconventional approaches to calibration transfer. However, the fact remains that significant differences in the instrument response between parent and child instruments causes significant variation in predicted results following calibration transfer. If instrument spectral profiles can be made statistically alike between instruments the transfer issue disappears. The additional challenges of relating specific reference laboratory results to results predicted using spectroscopy is another ongoing area of discovery and represents yet another problem yet to be resolved.

There are well-accepted methods for comparing instrument performance and these have been applied to analytical results in commerce and clinical analysis. When describing instrument differences in terms of uncertainty; one notes that NIST uncertainty is very well-defined and does not relate to a simple measurement set with a single SD computation. Although the within-instrument wavelength repeatability or precision might be considered reasonable for commercial NIR spectrometers, the uncertainty of the wavelength registration as measured using solid reference materials and comparing those results to measurand values shows significant differences.

Although the within instrument wavelength repeatability or precision for photometric measurements might be considered reasonable for commercial NIR spectrometers, the accuracy as measured using a stable Fluorilon R99 reference material and comparing those results to reference values is also significantly different across instrument platforms. The basic physics of making instruments more alike and calibrating them to original measurement conditions based upon first principles will be a step in the right direction for calibration transfer.

## References

1. P. Shahbazikhah, J.H. Kalivas. “A Consensus Modeling Approach to Update a Spectroscopic Calibration”. *Chemom. Intell. Lab. Syst.* 2013. 120: 142–153.
2. J.J. Workman Jr. “The Essential Aspects of Multivariate Calibration Transfer”. 40 Years of Chemometrics—From Bruce Kowalski to the Future. Washington D.C.: American Chemical Society, 2015, pp.257–260.
3. H. Mark, J. Workman. “Calibration Transfer, Part I”. *Spectroscopy*. 2013. 28(2): 24–37.
4. J.J. Workman Jr. “First Principles of Instrument Calibration”. *NIR news*. 2016. 27(3): 12–15.
5. H. Mark, J. Workman. *Chemometrics in Spectroscopy*, 1st ed. Boston, MA: Elsevier/Academic Press, 2010.
6. J.J. Workman. “Calibration Transfer”. In: J. Workman, editor. *The Concise Handbook of Analytical Spectroscopy: Physical Foundations, Techniques, Instrumentation, and Data Analysis*. New Jersey; Singapore: World Scientific Publishing-Imperial College Press, 2016, Volume 3: Pp. 185–188. Volume 4: Pp. 159–162.
7. R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, et al. “Transfer of Multivariate Calibration Models: A Review”. *Chemom. Intell. Lab. Syst.* 2002. 64(2): 181–192.
8. T. Fearn. “Standardization and Calibration Transfer for Near Infrared Instruments: A Review”. *J. Near Infrared Spectrosc.* 2001. 9(4): 229–244.
9. J. Shenk, M. Westerhaus. *Optical Instrument Calibration System*. US Patent 4866644. Filed 1986. Issued 1989.
10. S.D. Brown. “Transfer of Multivariate Calibration Models: A Review”. *Chemom. Intell. Lab. Syst.* 2002. 64(2): 181–192.
11. B.R. Kowalski. “Chemometrics”. *Anal. Chem.* 1980. 52(5): 112R–122R.
12. B. Lavine, J. Workman. “Chemometrics”. *Anal. Chem.* 2008. 80(12): 4519–4531.
13. B.R. Kowalski, D.J. Veltkamp, Y.D. Wang. *Calibration Transfer for Analytical Instruments*. US Patent 5459677. Filed 1991. Issued 1995.
14. E. Bouveresse, D.L. Massart, P. Dardenne. “Calibration Transfer Across Near-Infrared Spectrometric Instruments Using Shenk’s Algorithm: Effects of Different Standardisation Samples”. *Anal. Chim. Acta*. 1994. 297(3): 405–416.
15. O.E. De Noord. “Multivariate Calibration Standardization”. *Chemom. Intell. Lab. Syst.* 1994. 25(2): 85–97.
16. J. Workman, B. Kowalski, P. Mobley. “The Design of an Expert Calibration System for Spectroscopic Based Process Analytical Chemistry”. In: *Proceedings of the 40th Annual ISA Analysis Division Symposium*. Toronto, ON: April 25–27, 1995. Pp. 97.
17. J. Workman. “Diode Array Process Instrumentation in the Near Infrared”. In: C. Burgess, D.G. Jones, editors. *Spectrophotometry, Luminescence and Colour; Science and Compliance*. Amsterdam: Elsevier Science, 1995, pp.369–384.
18. D. Tracy, R. Hoult, A. Ganz. *Standardizing and Calibrating a Spectrometric Instrument*. US Patent 5303165, Filed 1992. Issued 1994.
19. J. Workman Jr, H. Mark. “Optimizing the Regression Model: The Challenge of Intercept–Bias and Slope ‘Correction’”. *Spectroscopy*. 2015. 30(7): 32–37, 48.
20. J. Workman Jr, J.P. Coates. “Multivariate Calibration Transfer: The Importance of Standardizing Instrumentation”. *Spectroscopy*. 1993. 8(9): 36–42.
21. J. Workman Jr, H. Mark. “Calibration Transfer, Part II: The Instrumentation Aspects”. *Spectroscopy*. 2013. 28(5): 12–25.
22. J. Workman Jr, H. Mark. “Chemometrics in Spectroscopy Calibration Transfer, Part III: The Mathematical Aspects”. *Spectroscopy*. 2013. 28(6): 28–35.
23. J. Workman Jr, H. Mark. “Calibration Transfer, Part IV: Measuring the Agreement Between Instruments Following Calibration Transfer”. *Spectroscopy*. 2013. 28(10): 24–33.
24. J. Workman Jr, H. Mark. “Calibration Transfer, Part V: The Mathematics of Wavelength Standards Used for Spectroscopy”. *Spectroscopy*. 2014. 29(6): 18–27.
25. J. Workman Jr, H. Mark. “Calibration Transfer, Part VI: The Mathematics of Photometric Standards Used for Spectroscopy”. *Spectroscopy*. 2014. 29(11): 14–21.
26. H. Mark, J. Workman. “A New Approach to Generating Transferable Calibrations for Quantitative Near-Infrared Spectroscopy”. *Spectroscopy*. 1988. 3(11): 28–36.
27. R.K.H. Galvão, S.F.C. Soares, M.N. Martins, M.F. Pimentel, et al. “Calibration Transfer Employing Univariate Correction and Robust Regression”. *Anal. Chim. Acta*. 2015. 864: 1–8.

28. J.S. Shenk, M.O. Westerhaus, W.C. Templeton. "Calibration Transfer Between Near Infrared Reflectance Spectrophotometers". *Crop Science*. 1985. 25(1): 159–161.
29. Y. Wang, B.R. Kowalski. "Calibration Transfer and Measurement Stability of Near-Infrared Spectrometers". *Appl. Spectrosc.* 1992. 46(5): 764–771.
30. C.S. Chen, C.W. Brown, S.C. Lo. Calibration Transfer from Sample Cell to Fiber-Optic Probe". *Appl. Spectrosc.* 1997. 51(5): 744–748.
31. E. Bouveresse, D.L. Massart. "Improvement of the Piecewise Direct Standardization Procedure for the Transfer of NIR Spectra for Multivariate Calibration". *Chemom. Intell. Lab. Syst.* 1996. 32(2): 201–213.
32. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, et al. "An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra". *Chemom. Intell. Lab. Syst.* 1998. 44(1): 229–244.
33. D. Poerio, S. Brown. EXPRESS: Dual-Domain Calibration Transfer by Orthogonal Projection. *Appl. Spectrosc.* 2017. DOI: 10.1177/0003702817724164.34.
34. C.E. Anderson, J.H. Kalivas. "Fundamentals of Calibration Transfer Through Procrustes Analysis". *Appl. Spectrosc.* 1999. 53(10): 1268–1276.
35. R.P. Cogdill, C.A. Anderson, J.K. Drennen. "Process Analytical Technology Case Study, Part III: Calibration Monitoring and Transfer". *AAPS PharmSciTech.* 2002. 6(2): E284–E297.
36. H. Tan, S.T. Sum, S.D. Brown. "Improvement of a Standard-Free Method for Near-Infrared Calibration Transfer". *Appl. Spectrosc.* 2002. 56(8): 1098–1106.
37. D.T. Andrews, P.D. Wentzell. "Applications of Maximum Likelihood Principal Component Analysis: Incomplete Data Sets and Calibration Transfer". *Anal. Chim. Acta.* 1997. 350(3): 341–352.
38. F.A. Honorato, R.K.H. Galvão, M.F. Pimentel, B. de Barros Neto, et al. "Robust Modeling for Multivariate Calibration Transfer by the Successive Projections Algorithm". *Chemom. Intell. Lab. Syst.* 2005. 76(1): 65–72.
39. J. Yoon, B. Lee, C. Han. "Calibration Transfer of Near-Infrared Spectra Based on Compression of Wavelet Coefficients". *Chemom. Intell. Lab. Syst.* 2002. 64(1): 1–14.
40. W. Fan, Y. Liang, D. Yuan, J. Wang. "Calibration Model Transfer for Near-Infrared Spectra Based on Canonical Correlation Analysis". *Anal. Chim. Acta.* 2008. 623(1): 22–29.
41. J. Bin, X. Li, W. Fan, J.H. Zhou, C.W. Wang. "Calibration Transfer of Near-Infrared Spectroscopy by Canonical Correlation Analysis Coupled with Wavelet Transform". *Analyst.* 2017. 142(12): 2229–2238.
42. Y. Xie, P.K. Hopke. "Calibration Transfer as a Data Reconstruction Problem". *Anal. Chim. Acta.* 1999. 384(2): 193–205.
43. J. Peng, S. Peng, A. Jiang, J. Tan. "Near-infrared Calibration Transfer based on Spectral Regression". *Spectrochimica Acta, Part A.* 2011. 78(4): 1315–1320.
44. M. Forina, G. Drava, C. Armanino, R. Boggia, et al. "Transfer of Calibration Function in Near-Infrared Spectroscopy". *Chemom. Intell. Lab. Syst.* 1995. 27(2): 189–203.
45. T.A.N. Chao, L.I. Menglong. "Calibration Transfer Between Two Near-Infrared Spectrometers Based on a Wavelet Packet Transform". *Analytical Sciences.* 2007. 23(2): 201–206.
46. W. Ni, S.D. Brown, R. Man. "Data Fusion in Multivariate Calibration Transfer". *Anal. Chim. Acta.* 2010. 661(2): 133–142.
47. Y. Liu, W. Cai, X. Shao. "Standardization of Near Infrared Spectra Measured on Multi-Instrument". *Anal. Chim. Acta.* 2014. 836: 18–23.
48. Y. Binfeng, J. Haibo. "Near-Infrared Calibration Transfer via Support Vector Machine and Transfer Learning". *Analytical Methods.* 2015. 7(6): 2714–2725.
49. J. Wang, Z. Li, Y. Wang, Y. Liu, et al. "A Dual Model Strategy to Transfer Multivariate Calibration Models for Near-Infrared Spectral Analysis". *Spectroscopy Letters.* 2016. 49(5): 348–354.
50. T. Boucher, M.D. Dyar, S. Mahadevan. "Proximal Methods for Calibration Transfer". *J. Chemometr.* 2017. 31: e2877.
51. K.E. Kramer, R.E. Morris, S.L. Rose-Pehrsson. "Comparison of Two Multiplicative Signal Correction Strategies for Calibration Transfer Without Standards". *Chemom. Intell. Lab. Syst.* 2008. 92(1): 33–43.
52. K.H. Hazen, T.B. Blank, S. Monfre, T.L. Ruchti. Method of Characterizing Spectrometer Instruments and Providing Calibration Models to Compensate for Instrument Variation. US Patent 6864978. Filed 2000. Issued 2005.
53. Y. Sulub, G.W. Small. "Spectral Simulation Methodology for Calibration Transfer of Near-Infrared Spectra". *Appl. Spectrosc.* 2007. 61(4): 406–413.
54. 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.* 2008. 611(2): 143–150.
55. J. Lin, S.C. Lo, C.W. Brown. "Calibration Transfer from a Scanning Near-IR Spectrophotometer to a FT-Near-IR Spectrophotometer". *Anal. Chim. Acta.* 1997. 349(1–3): 263–269.
56. T.D. Ridder, B.J.V. Steeg, G.L. Price. "Robust Calibration Transfer in Noninvasive Ethanol Measurements, Part I: Mathematical Basis for Spectral Distortions in Fourier Transform Near-Infrared Spectroscopy (FT-NIR)". *Appl. Spectrosc.* 2014. 68(8): 852–864.
57. J.A.F. Pierna, P. Vermeulen, B. Lecler, V. Baeten, et al. "Calibration Transfer from Dispersive Instruments to Handheld Spectrometers". *Appl. Spectrosc.* 2010. 64(6): 644–648.
58. J. Lin. "Near-IR Calibration Transfer Between Different Temperatures". *Appl. Spectrosc.* 1998. 52(12): 1591–1596.
59. B.R. Kowalski, Y. Wang. Calibration Transfer for Second Order Analytical Instruments. US Patent 5559728. Filed 1994. Issued 1996.
60. "Interference Testing in Clinical Chemistry; Approved Guideline". In: CLSI document EP7-A2, Pennsylvania, 2005. 2nd ed.
61. H. Mark. Principles and Practice of Spectroscopic Calibration, 1st ed. New York, NY: John Wiley and Sons, Inc, 1991.
62. D. Abookasis, J. Workman. "Direct Measurements of Blood Glucose Concentration in the Presence of Saccharide Interferences Using Slope and Bias Orthogonal Signal Correction and Fourier Transform Near-Infrared Spectroscopy". *J. Biomed. Opt.* 2011. 16(2): 027001.
63. S. Wold, H. Antti, F. Lindgren, J. Ohman. "Orthogonal Signal Correction of Near-Infrared Spectra". *Chem. Intell. Lab. Syst.* 1998. 44: 175–185.
64. H.C. Goicoechea, A.C. Oliveri. "A Comparison of Orthogonal Signal Correction and Net Analyte Preprocessing Methods. Theoretical and Experimental Study". *Chem. Intell. Lab. Syst.* 2001. 56: 73–81.
65. Y. Sulub, R. LoBrutto, R. Vivilecchia, B. Wabuyele. "Near-Infrared Multivariate Calibration Updating Using Placebo: A Content Uniformity Determination of Pharmaceutical Tablets". *Vibrational Spectroscopy.* 2008. 46(2): 128–134.
66. M.J. Sáiz-Abajo, B.H. Mevik, V.H. Segtnan, T. Næs. "Ensemble Methods and Data Augmentation by Noise Addition Applied to the Analysis of Spectroscopic Data". *Anal. Chim. Acta.* 2005. 533(2): 147–159.
67. J. Workman, P. Mobley, B. Kowalski, R. Bro. "Review of Chemometrics Applied to Spectroscopy: 1985–95, Part I". *Applied Spectroscopy Reviews.* 1996. 31(1–2): 73–124.
68. D.E. Honigs, G.M. Hieftje, H.L. Mark, T.B. Hirschfeld. "Unique-Sample Selection via Near-Infrared Spectral Subtraction". *Anal. Chem.* 1985. 57(12): 2299–2303.
69. J. Workman. "Near Infrared Spectroscopy: Comparison of Random Versus Spectrally Selected Calibration Sets for use in Forage Quality Evaluation". *Proc. 1986 Forage and Grasslands Conference.* Lexington, KY: April 17–18, 1986.
70. B.M. Wise. "Using the PLS\_Toolbox in PAC Applications". *Process Control Qual.* 1993. 5: 73–85.

71. T. Naes, T. Isaksson, B. Kowalski. "Locally Weighted Regression and Scatter Correction for Near-Infrared Reflectance Data". *Anal. Chem.* 1990. 62(7): 664–673.
72. W.S. Cleveland, S.J. Devlin. "Locally Weighted Fitting: An Approach to Fitting Analysis by Local Fitting". *J. Am. Stat. Assoc.* 1988. 83: 596–610.
73. D.A. Burns. "Indicator Variables: How They May Save Time and Money in NIR Analysis". In: D.A. Burns, E.W. Ciurczak, editors. *Handbook of Near-Infrared Analysis*. Boca Raton, FL: CRC Press, 2007. Chapter 14. Pp. 166, 297–306.
74. W.J. Youden, E.H. Steiner. *Statistical Manual of the AOAC*, Arlington, VA, 1984.
75. National Institute of Standards and Technology (NIST). "Standard Uncertainty and Relative Standard Uncertainty". 2017. <http://physics.nist.gov/cgi-bin/cuu/Info/Constants/definitions.html>.
76. B. N. Taylor, C.E. Kuyatt. "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results". NIST Technical Note #1297. 1994. September 1994 edition, Pp. 1–24.
77. J.M. Bland, D.G. Altman. "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement". *Lancet.* 1986. 1: 307–310.
78. Y. Wang, D.J. Veltkamp, B. Kowalski. "Multivariate Instrument Standardization". *Anal. Chem.* 1991. 63(23): 2750–2756.
79. C. Burgess, J. Hammond. "Wavelength Standards for the Near-Infrared Spectral Regions". *Spectroscopy.* 2007. 22(4): 40–48.
80. V.R. Weidner, P.Y. Barnes, K.L. Eckerle. "A Wavelength Standard for the Near Infrared Based on the Reflectance of Rare-Earth Oxides". *J. Research NBS.* 1986. 91(5): 243–253.
81. Standard Reference Material (SRM). 1920a Certificate of Analysis, Near-Infrared Reflectance Wavelength Standard from 740 nm to 2000 nm, National Institute of Standards and Technology, Gaithersburg, MD 20899, Pp. 1–7 (2 February 2006).
82. Standard Reference Material (SRM). 2036 Certificate, Near-Infrared Wavelength/Wavenumber Reflection Standard, National Institute of Standards and Technology, Gaithersburg, MD 20899, Pp. 1–7 (24 January 2013).
83. National Institute of Standards and Technology (NIST). "Basic Atomic Spectroscopic Data". [http://physics.nist.gov/PhysRefData/Handbook/element\\_name.htm](http://physics.nist.gov/PhysRefData/Handbook/element_name.htm).
84. N. Acquista, E.K. Plyler. "Calibrating Wavelengths in the Region from 0.6 to 2.6 Microns". *J. Res. Natl. Bur. Stand.* 1952. 49(1): 13–16.
85. E.K. Plyler, L.R. Blaine, M. Nowak. "Reference Wavelengths for Calibrating Prism Spectrometers". *J. Research NBS.* 1957. 58(4): 195–200.
86. K.G. Nordqvist. Method of Reducing the Susceptibility to Interference of a Measuring Instrument. US Patent No. 4944589. Filed 1988. Issued 1990.
87. J.S. Shenk, M.O. Westerhaus. "New Standardization and Calibration Procedures for NIRS Analytical Systems". *Crop Sci.* 1991. 31: 1694–1696.
88. A. Springsteen. "An Ode to Traceability and Calibration Interval". *Spectroscopy.* 2006. 21(11): 6–9.
89. J. Coates, T. Davidson, L. McDermott. "The Design and Application of Spectrometric Analyzers for the Chemical Process Industry". *Spectroscopy.* 1992. 7(9): 40–49.
90. J. Workman Jr, J. Coates. "The Importance of Standardizing Instrumentation". *Spectroscopy.* 1993. 8(9): 36–42.
91. D. Tracy, R. Hoult, A. Ganz. Standardizing and calibrating a spectrometric instrument. US Patent 5303165. Filed 1992. Issued 1994.
92. J. Workman. "Diode Array Process Instrumentation in the Near Infrared". Oxford Conference on Spectroscopy. *Analytical Spectroscopy Library.* 1995. 6: 369–384.
93. J. Workman, L. McDermott, B. DeSimas, J. Coates. "Requirements for Chemometrics in Process Spectrometers". *J. Process Anal. Chem.* 1996. 2: 444–450.
94. L.M. Hanssen, C. Zhu. "Wavenumber Standards for Mid-Infrared Spectrometry". In: J. Chalmers, P.R. Griffiths, editors. *The Handbook of Vibrational Spectroscopy*. Vol. 1, Chichester: John Wiley and Sons Ltd, 2002, pp.881–890.
95. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. Cambridge: Cambridge University Press, 2007.