Calibration Maintenance and Transfer Using Tikhonov Regularization Approaches

JOHN H. KALIVAS,* GABRIEL G. SIANO, ERIK ANDRIES,†
and HECTOR C. GOICOECHEA

Department of Chemistry, Idaho State University, Pocatello, Idaho 83209 (J.H.K.); InLight Solutions Inc., 800 Bradbury SE, Albuquerque, New Mexico 87106 (E.A.); and Laboratorio de Desarrollo Analítico y Quimiometría, Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe (S3000ZAA), Argentina (G.G.S., H.C.G.).

Maintaining multivariate calibrations is essential and involves keeping models developed on an instrument applicable to predicting new samples over time. Sometimes a primary instrument model is needed to predict samples measured on secondary instruments. This situation is referred to as calibration transfer. This paper reports on using a Tikhonov regularization (TR) based method in both cases. A distinction of the TR design for calibration maintenance and transfer is a defined weighting scheme for a small set of new (transfer or standardization) samples augmented to the full set of calibration samples. Because straight application of basic TR theory is not always possible with calibration maintenance and transfer, this paper develops a generic solution to always enable application of TR. Harmonious (bias/variance tradeoff) and parsimonious (effective rank) considerations for TR are compared with the same TR format applied to partial least squares (PLS), showing that both approaches are viable solutions to the calibration maintenance and transfer problems.

Index Headings: Multivariate calibration; Calibration maintenance; Calibration transfer; Calibration standardization; Tikhonov regularization; Partial least squares; PLS; Augmentation.

INTRODUCTION

Multivariate calibration relates a dependent variable such as a chemical or physical property to independent variables such as spectroscopic measurements by

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ denotes an $m \times 1$ vector of quantitative values of the analyte for m calibration samples, $\mathbf{X}$ symbolizes the $m \times n$ calibration matrix of n predictor variables, and $\mathbf{b}$ represents the $n \times 1$ vector of calibration model coefficients that must be estimated. The $m \times 1$ vector $\mathbf{e}$ indicates normally distributed errors with mean zero and covariance matrix $\sigma^2\mathbf{I}$. Without loss of generality, it is assumed that all data is mean centered. The regression vector is commonly estimated by the methods of partial least squares (PLS), ridge regression (RR), or principal component regression (PCR). For this paper, $\mathbf{y}$ contains analyte concentration information, $\mathbf{X}$ contains spectra measured over n wavelengths or frequencies, and $n \gg m$. The relationship $n \leq m$ is applicable as well, and in this case, multiple linear regression (MLR) could also be used to obtain the regression vector. The goal in calibration is to determine an appropriate estimate of $\mathbf{b}$ ($\hat{\mathbf{b}}$) in order to predict with the best accuracy (minimum bias) and precision (minimum variance) the amount of calibrated analyte present in a future sample $\mathbf{x}$ using $\hat{y} = \mathbf{x}^t\hat{\mathbf{b}}$.[1,2]

Once a model is estimated, the duration of applicability of the model becomes relevant. The current calibration can fail due to uncalibrated spectral features appearing in new samples at a later time; for instance, the calibrated analyte is lower or higher than the calibration concentrations in $\mathbf{y}$ or new spectral responding chemical constituents appear. Depending on the instrument and sample type, other chemical, physical, and environmental influences can cause new spectral features to appear. These include changes in viscosity, particle size, surface texture, pH, temperature, humidity, and pressure. Instrumental effects can also cause a current calibration to fail and these include drift and repairing the instrument with a new source, detector, or other component. Thus, mechanisms are needed to update the current model to include new chemical, physical, environmental, and/or instrumental effects not in the current calibration domain. Connected to this calibration maintenance issue is the calibration transfer problem. Here the concern is using a calibration model developed on a primary instrument(s) to predict the sample composition from a spectrum measured on a secondary instrument.

Calibration maintenance and transfer have been the subjects of numerous studies and are well reviewed.[3–6] Three general approaches are possible. One consists of forming an initial robust model. This approach can be accomplished by using spectral pretreatment methods such as multiplicative scatter correction, finite impulse response filters, derivatives, and/or wavelength selection. An alternative mechanism to forming a robust model is to globally calibrate the model by including all potential chemical, physical, environmental, and/or instrumental effects in the original model through measuring spectra under all possible future conditions. For example, samples at different pH values and temperatures can be used if varying or new pH values or temperatures are expected in the future. However, a difficulty with this approach is the large number of samples needed to span all potential future effects, and for each sample included in $\mathbf{X}$, the corresponding analyte reference value must be determined for $\mathbf{y}$. Obtaining reference values is usually time consuming and costly. Extra columns in $\mathbf{X}$ could be included with values for the different conditions, thereby possibly limiting the total number of samples, e.g., temperature or time for drift.[7,8]

A second approach to calibration maintenance and transfer is adjusting spectra for samples measured on a primary or secondary instrument to fit the original calibration model determined on the primary instrument(s). These methods generally fit under the basic framework of the statistical method known as Procrustes analysis. The general approach of Procrustes analysis is to determine transformation parameters

APPLIED SPECTROSCOPY

that map spectra measured on one instrument to match spectra measured for the same samples on another instrument over the same or different wavelength ranges.[9] One of the more popular methods is piecewise direct standardization (PDS).[10] In this case, a small set of samples (standardization set) must have been measured on a primary instrument at the same time that the calibration samples were measured. Procrustes analysis determines the proper rotation, dilation, and translation to accomplish transformation of the standardization samples. New samples are then transformed and predicted by the primary calibration model. In order to avoid measuring the same samples on the primary and secondary instruments, the mean primary calibration spectrum can be used as the transformation target. In this case, primary calibration samples need to also be transformed to the mean spectrum before forming the primary calibration model and the number of wavelengths measured for samples must be the same for the primary and secondary instruments.

A third general approach, and the focus of this paper, is to update (rebuild) the primary calibration model to properly predict new sample spectra measured on primary or secondary instruments. Having detected an unmodeled source of variance within new samples, one method is to directly add to each spectrum of the primary calibration spectra in X, pure component spectral shapes of the unmodeled effects not present in X (such as drift, chemical substances, or temperature).[11] A regression method is then used to estimate the new updated model. A related design used synthetic spectra for the primary calibration X that were based on a mathematical description of light propagation in skin tissue. To each of these synthetic spectra, a spectrum of a human subject was added and a model was formed to predict glucose for future spectra measured from the same subject.[12] This approach of adding a spectrum of the new condition to each of the calibration primary spectra requires a reference determination for only the one sample in the new condition that is being added. If the new sample does not contain the analyte, e.g., a drift correction, then no reference analysis is needed.

Rather than adding spectra under the new condition to each existing spectrum in the primary calibration set (real or simulated), an alternative to forming an updated model is to augment the original calibration set of X and y with additional calibration samples containing the new variance. In this case, Eq. 1 is written as (ignoring the e term):

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_L \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{L} \end{pmatrix} \mathbf{b} \qquad (2)$$

where L represents an $l \times n$ matrix of spectra measured for l samples in the new conditions or on the secondary instrument and y_L denotes respective concentrations. The updated or new regression vector is obtained by applying a regression method to Eq. 2. A problem with augmenting the original calibration set is that usually many samples are needed for L and this can become time consuming and costly, especially when complex reference methods are needed to form y_L.

Using only a few samples for L to characterize the new conditions has been proposed and studied.[13,14] The selection of samples for the small standardization set is critical to all calibration maintenance and transfer approaches.[15] The goal is to form a standardization set with as few samples as possible and still fully span the new situation, thereby adequately accounting for the new variance(s).

If the standardization set can be measured when the original primary calibration model is formed and also under the new conditions, then spectral differences can be used for L with $y_L = 0$.[13] Using spectral differences eliminates the need to use a reference method to obtain values for y_L. However, long-term stability is required for the standardization set. Another possibility is to use samples without the analyte for L with $y_L = 0$. Such an approach was used with blanks for L measured on the primary instrument during warm-up to adjust for a new instrumental profile and correct for any instrumental drift that may have occurred.[16]

One application of Eq. 2 consisted of using in X spectra measured on laboratory prepared solutions and for L, a small number of samples measured in the new conditions in which the model will subsequently be used.[17] By using a well-designed set of laboratory samples, the model could better characterize the analyte-dependent information and L allowed the regression method to correct for the new conditions, which in this case was a culture medium. Augmenting simulated pure component spectra in X with real spectra for L has also been used for calibration transfer with Eq. 2.[18]

The concept of augmenting original calibration samples with information spanning the new conditions has been alternatively applied to calibration maintenance and transfer by using prediction augmented hybrid methods.[19,20] Characterization of the new conditions occurred by repeatedly measuring the spectrum of a single sample selected from the center of the concentration space. With such an approach, the reference value for y_L must only be determined once. Because the analyte concentration is constant, the eigenvectors from the singular value decomposition (SVD) of the repeat spectra could be used in L.

Other possible spectral artifacts to place in L to desensitize a primary model include mathematical representation of drift, known spectral interferences, a solvent peak, or background such as fluorescence in Raman spectroscopy. However, a problem with Eq. 2 is that if the standardization set is small, then too much emphasis may be applied to the larger original calibration set X. Thus, a weighting scheme was proposed by modifying Eq. 2 to

$$\begin{pmatrix} \mathbf{y} \\ \lambda \mathbf{y}_L \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda \mathbf{L} \end{pmatrix} \mathbf{b} \qquad (3)$$

where λ symbolizes a weight value.[21] Using the regression methods of PLS, PCR, or MLR to estimate b for Eq. 3 requires determination of the respective meta-parameters. Without augmentation (Eq. 1) the meta-parameter for PLS and PCR is the number of basis vectors (latent vectors, factors) and for MLR it is the number and location of wavelengths. With Eq. 3, there is now the additional weight meta-parameter λ. Thus, a problem with Eq. 3 is the lack of an obvious methodology to determine proper weights. To date, selection of a weight value has been based on replication of samples in the standardization set.[15,21] For example, if $\lambda = 1$, then no replication of the standardization set is used, if $\lambda = 2$ then duplicates are augmented, etc. This approach has not always proved satisfactory.

Rather than using multiple L arrays of exact duplicates as a way to set λ (whether L is a single spectrum or a collection of spectra), perturbations of the original primary calibration samples (or a standardization set) with random noise in various combinations were augmented to X with one L or multiple Ls

for different perturbations.[22] Spectra measured under different temperatures were also perturbed and used in **L**. This approach of augmenting **X** with multiple **L** arrays composed of noise perturbations of samples from **X** can be thought of as an ensemble method.[23] With this ensemble method, there is no λ and Eq. 2 is used with a regression method to estimate **b**. Decisions needed for this ensemble method are the number of noisy spectra to augment with and how spectra should be perturbed. The closer the noise structure in **L** mimics the noise structure in **X**, the more desensitized the model should be to the noise. For example, in recent work, **X** was measured at 36 °C and the mean difference spectrum between spectra at 38 and 34 °C were randomly added to the spectra at 36 °C to form multiple **L** arrays.[24]

Equation 3 is actually a representation of Tikhonov regularization (TR) and is fully developed in the next section. Thus, by using a TR approach for calibration maintenance and transfer, the idea of weighting samples in **L** is now put on a firm theoretical foundation and logical criteria to select the weight are developed.

TIKHONOV REGULARIZATION AND VARIATIONS

The most general formulation of TR is expressed as identifying the model coefficients that

$$\min\left[\left\|\mathbf{Xb} - \mathbf{y}\right\|_a^a + \lambda^2 \left\|\mathbf{L}(\mathbf{b} - \mathbf{b}^*)\right\|_b^b\right] \quad (4)$$

where $\|\cdot\|_p$ signifies the regression vector p-norm, e.g., $p = 2$ is the 2-norm or Euclidean norm, a and b represent the same or different norms in the range $1 \leq a, b < \infty$, **L** denotes a regulation operator that enforces the estimate of **b** to belong to the corresponding subspace, $\mathbf{b}^*$ designates the true model coefficients for the analyte, and λ symbolizes the regularization meta-parameter controlling the weight given to the second term.[25–27] For the calibration maintenance and transfer problems, **L** denotes spectra or artifact representations at the new conditions or for the secondary instrument as discussed in the Introduction section. The left term has been labeled a bias (accuracy) indicator and the right term reflects the model size and hence, when $b = 2$, the 2-norm acts as a variance (precision) measure.[28]

Choices for a and b are varied. When $a = b = 2$, the solution to Expression 4 is

$$\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \lambda^2\mathbf{L}^t\mathbf{L})^{-1}(\mathbf{X}^t\mathbf{y} + \lambda^2\mathbf{L}^t\mathbf{Lb}^*) \quad (5)$$

which is also the solution to

$$\begin{pmatrix} \mathbf{y} \\ \lambda\mathbf{Lb}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda\mathbf{L} \end{pmatrix} \mathbf{b} \quad (6)$$

Often $\mathbf{b}^*$ is not known and in the case of spectroscopic analysis where **L** is composed of spectra, then $\mathbf{y}_L \approx \mathbf{Lb}^*$ and Expression 4 reduces to

$$\min\left(\left\|\mathbf{Xb} - \mathbf{y}\right\|_2^2 + \lambda^2\left\|\mathbf{Lb} - \mathbf{y}_L\right\|_2^2\right) \quad (7)$$

In this case, Eq. 6 becomes Eq. 3 with the solution

$$\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \lambda^2\mathbf{L}^t\mathbf{L})^{-1}(\mathbf{X}^t\mathbf{y} + \lambda^2\mathbf{L}^t\mathbf{y}_L) \quad (8)$$

Thus, the weight value empirically set in Eq. 3 in previous work is actually the normal meta-parameter in TR and good mechanisms exists for determining λ.[25,26,28,29] A solution to Eq. 3 can also be obtained by PLS, PCR, etc. This paper also looks at using PLS with Eq. 3 to obtain an estimate of **b**.

Calibration Maintenance and Transfer with a Tikhonov Regularization Approach. For calibration maintenance, the goal is updating an existing model to new conditions such as new spectrally responding species or replacement of the instrument source. In this situation, **L** contains spectra under the new conditions. When spectra in **L** contain the analyte, then reference values of the analyte are needed for $\mathbf{y}_L$. The TR approach estimates **b** to be orthogonal to the new interfering spectral artifacts in **L** that are confounding an accurate prediction of the analyte. Thus, the model is desensitized to the interfering spectral artifacts. Simultaneously, the regression vector usually needs to accurately predict the original calibration samples that do not have the new condition present as well as keeping the size of the regression vector from being too large. Too large a regression vector increases the chance of forming an over-fitted model and prediction variances can escalate.

If a representative subset of samples from the original calibration is available, these samples can be measured under the new condition and difference spectra would be used in **L** and the $\mathbf{y}_L$ values would not be needed. In this case, Expression 7 and Eqs. 3 and 8 become, respectively,

$$\min(\left\|\mathbf{Xb} - \mathbf{y}\right\|_2^2 + \lambda^2\left\|\mathbf{Lb}\right\|_2^2) \quad (9)$$

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda\mathbf{L} \end{pmatrix} \mathbf{b} \quad (10)$$

and

$$\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \lambda^2\mathbf{L}^t\mathbf{L})^{-1}\mathbf{X}^t\mathbf{y} \quad (11)$$

As described in the following section, these two equations and this expression also define RR when $\mathbf{L} = \mathbf{I}$. If samples without the analyte are used for **L**, e.g., matrix-matched blanks, the solvent, background, or pure-component spectra of new artifacts, then Expression 9 and Eqs. 10 and 11 are also applicable. For drift, it should be possible to use generic pseudo-spectra in **L**. For example, drift pseudo-spectra in **L** could be formed as a constant, linear, and/or parabolic as well as cubic or higher order wavelength functions. Using drift pseudo-spectra has been implemented with PCR where the drift pseudo-spectra acted as pseudo-principal components in the augmented eigenvector set.[30,31] Similarly, key eigenvectors from the SVD of **L** with samples containing constant or no analyte, such as spectra from repeatedly measuring the spectrum of a single sample,[19,20] could be used for **L** with $\mathbf{y}_L = \mathbf{0}$.

As with any regression method, the desired regression vector needs to be orthogonal to non-analyte information in **X**. From Expression 9 and Eq. 10 it is easy to observe that the desired regression vector also needs to be orthogonal to the new chemical, physical, environmental, and/or instrumental conditions characterized in **L**. The goal is to try to do this with as few spectra as possible in **L**.

Similar points described about **L** and $\mathbf{y}_L$ for calibration maintenance are applicable to calibration transfer. Additional-

ly, the augmented set in $\mathbf{L}$ can be spectra from multiple instruments to form a model applicable to more than two instruments.

Harmony as the Bias/Variance Tradeoff. When $\mathbf{L} = \mathbf{I}$ and $a = b = 2$, TR is said to be in standard form and is also known as RR with the solution

$$\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \lambda^2\mathbf{I})^{-1}\mathbf{X}^t\mathbf{y} \tag{12}$$

for the equation

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda\mathbf{I} \end{pmatrix} \mathbf{b} \tag{13}$$

or the expression

$$\min(\left\|\mathbf{Xb} - \mathbf{y}\right\|_2^2 + \lambda^2\left\|\mathbf{b}\right\|_2^2) \tag{14}$$

Thus, calibration and transfer in the augmented format is a form of RR and methods used to determine λ for RR should be useful here. It has been shown that the regression vector 2-norm $\|\hat{\mathbf{b}}\|_2$ in Expression 14 is proportional to the prediction variance.[28,32–35] Therefore, in the case of TR in standard form, optimization of Expression 14 is concerned with simultaneous minimization of bias and variance indicators. Such an optimization seeks the most harmonious model: one that is the most Pareto (closest to the origin) in an L-curve plot. The L-curve plot is obtained by plotting for each λ the corresponding regression vector 2-norm against the root mean square error of calibration (RMSEC). That is, in order to select the right λ value in Expression 14, a host of models are formed by varying λ and plotting the respective regression vector 2-norms against RMSEC. In such a plot, an L-shaped curve results and the optimal model (λ) is in the corner of the L-curve near the origin.[25,29] This model represents the best compromise for the bias/variance tradeoff, i.e., the most harmonious model. The L-curve (harmonious, Pareto) plot is also applicable to selecting other model meta-parameters such as the number of factors for PCR or PLS. In general, RR L-curves and regression vectors are commonly similar to those for PCR and PLS.[28]

Modifying Tikhonov Regularization for Calibration Maintenance and Transfer. For calibration maintenance and transfer, $\mathbf{L} \neq \mathbf{I}$ and the structures of $\mathbf{X}$ and $\mathbf{L}$ can have a disruptive impact in Eqs. 8 and 11. Specifically, the inverse operation is not stable and is poorly defined if spectra in $\mathbf{X}$ and $\mathbf{L}$ are collinear and/or $m < n$ ($\mathbf{X}$ and $\mathbf{L}$ are nearly singular defined as the determinate of $\mathbf{X}^t\mathbf{X}$, or $\mathbf{L}^t\mathbf{L}$ is nearly zero or the condition number is large and, hence, poorly conditioned).[36] The method of RR expressed in Eq. 12 provides a mechanism for forcing $\mathbf{X}^t\mathbf{X}$ to be full rank by adding a small number to the diagonal of $\mathbf{X}^t\mathbf{X}$, thereby stabilizing the inverse operation.[37] As long as λ is nonzero, the last n rows of the augmented matrix in Eq. 13 are linearly independent, making the augmented matrix full rank. The greater the value of λ, the greater the degree of orthogonality (or equivalently, the greater the degree of nonsingularity). Because $\mathbf{L} \neq \mathbf{I}$ with calibration maintenance and transfer, the structure of $\mathbf{L}$ has a significant impact on the inverse operation.

In order to stabilize TR for calibration maintenance and transfer, an additional regularization meta-parameter will generally be needed. To accomplish this, Expression 7 and Eqs. 3 and 8 are respectively written as

$$\min\left(\left\|\mathbf{Xb} - \mathbf{y}\right\|_2^2 + \tau\left\|\mathbf{b}\right\|_2^2 + \lambda^2\left\|\mathbf{Lb} - \mathbf{y}_L\right\|_2^2\right) \tag{15}$$

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \\ \lambda\mathbf{y}_L \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \tau\mathbf{I} \\ \lambda\mathbf{L} \end{pmatrix} \mathbf{b} \tag{16}$$

and

$$\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \tau^2\mathbf{I} + \lambda^2\mathbf{L}^t\mathbf{L})^{-1}(\mathbf{X}^t\mathbf{y} + \lambda^2\mathbf{L}^t\mathbf{y}_L) \tag{17}$$

where τ denotes the stabilizing meta-parameter to enhance the degree of nonsingularity for the covariance matrix in the inverse operation. Rather than using TR and requiring the second meta-parameter, PLS or PCR could be directly applied to Eqs. 3 or 10. However, two meta-parameters (factors and λ) are still required.

Inspection of Expressions 7, 9, 14, and 15 provides further understanding of the situation. Because $\mathbf{L}$ is composed of spectra, the physical meaning of the product $\mathbf{Lb}$ results in the last terms of Expressions 7 and 9 actually being prediction errors and, hence, there is no explicit direct minimization of the size of the regression vector as in Expressions 14 and 15. Using the second meta-parameter τ reinforces the direct minimization of the regression vector size.

When $\mathbf{y}_L = \mathbf{0}$, Expression 9 and Eqs. 10 and 11 would be adjusted accordingly to account for the second meta-parameter. However, in this situation, a different approach not requiring the second meta-parameter is possible. The process consists of transforming the general form of TR in Expression 9 to a format such as that in Expression 14.[25] The regression vector is now obtained using a standard RR algorithm (or PLS, PCR, etc.)[38] and then back-transformed to the general form. The back-transformed regression vector desensitized to the artifacts in $\mathbf{L}$ can now be used to predict samples. This paper reports on only using the second meta-parameter incorporated into Expression 9 and Eqs. 10 and 11.

Summarizing, when $\mathbf{L} = \mathbf{I}$ for RR, the inverse operation is stabilized with λ. As soon as spectra are used in $\mathbf{L}$, the second regularization meta-parameter τ will usually become necessary. As $\mathbf{X}$ and $\mathbf{L}$ approach full rank matrices, τ becomes smaller, and in the limit, as $\mathbf{L}^t\mathbf{L}$ approaches $\mathbf{I}$, the value of τ approaches zero. While additional regularization with τ is not needed with the methods of PLS, PCR, or MLR when using Eqs. 3 and 10, other method dependent stabilizing meta-parameters are still needed.

Determining τ, Factors, and λ. The first meta-parameter to optimize for TR is τ. Because the first two terms on the right side of the equality sign in Eq. 17 are similar to RR, the same procedure used with RR will be tested in this paper. The proper τ value is expected to be that in the corner region of the L-curve from a plot of $\|\hat{\mathbf{b}}\|_2$ against RMSEC. This is the τ that provides a good regression vector to predict the calibration data. For PLS, the same plot is used but the number of factors is determined instead of τ. Once τ has been ascertained, the focus now is to establish the λ value providing simultaneous acceptable predictions of the information in $\mathbf{L}$ and $\mathbf{X}$. Noting that $\|\mathbf{Lb} - \mathbf{y}_L\|_2$ is proportional to the root mean square error of $\mathbf{L}$ (RMSEL) and as shown in the Results and Discussion section, the appropriate L-curve capturing this tradeoff is obtained by plotting RMSEL against RMSEC. To better visually characterize the L-curve, log values are plotted.[25]

Further Generalization of Tikhonov Regularization for Calibration Maintenance and Transfer. In recent work, a generalization of TR was proposed for desensitizing models to anticipated spectral artifacts.[39,40] The generalization consists of using a unique weight for each artifact in $\mathbf{L}$ that the model is being updated for (desensitized to). In this case, Eq. 3 is written as

$$\begin{pmatrix} \mathbf{y} \\ \Lambda \mathbf{y}_L \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \Lambda \mathbf{L} \end{pmatrix} \mathbf{b} \qquad (18)$$

where Λ represents an $l \times l$ diagonal matrix of weights λ_i for each artifact in $\mathbf{L}$. This desensitization adaptation of TR is applicable to calibration maintenance and transfer. Not realized was recognition of requiring additional regularization of $\mathbf{X}$ due to the potential instability with solving Eq. 18 depending on the structure of $\mathbf{X}$ and $\mathbf{L}$ (the matrices may not be full rank). When the term $\Lambda \mathbf{y}_L = \mathbf{0}$ and $\mathbf{L} = \mathbf{I}$, Eq. 18 represents generalized RR.[37] Determining individual weights was not attempted in this study. That is, only one weight is used to characterize a global correction for all new chemical, physical, environmental, and/or instrumental effects in $\mathbf{L}$.

If unique correction is needed for multiple individual artifacts, such as drift, new chemical component, or temperature, then $\mathbf{L}$ would have to be composed of spectra uniquely characterizing each specific artifact and Eq. 18 could be used. Two difficulties with Eq. 18 are the necessity of a protocol to determine each λ_i in Λ and the necessity of obtaining spectra uniquely characterizing each artifact. It may be that all artifacts in the new samples can be corrected with one λ value representing a compromise or average weight for the artifacts, much like the selection of PCR or PLS factors with a $\mathbf{Y}$ matrix for multiple analytes.

A multiple weighting scheme could also be used for the ensemble method noted previously in which groups of spectra are uniquely perturbed.[24] Each group of spectra would be weighted separately.

Other Uses of Tikhonov Regularization. The framework of TR expressed by Eq. 3 has found utility in various situations. A diagonal $\mathbf{L}$ matrix equal to the spectral noise relative to each wavelength was used to remove irrelevant wavelengths from the regression vector.[38,41] Because $\mathbf{L}$ was diagonal and hence full rank, the second meta-parameter τ was not needed. The method of TR has been used for smoothing $\mathbf{X}$ as well as the regression vector, and a TR approach was used with self-modeling curve resolution problems.[38,42,43] Replacing the 2-norm on the regression vector with the 1-norm and setting $\mathbf{L} = \mathbf{I}$ has been used to obtain variable (wavelength) selection and has also become known as the least absolute shrinkage and selection operator (LASSO).[44–46] In a limited study, an estimate of $\mathbf{b}^*$ was set to the pure-component spectrum of the analyte with $\mathbf{L} = \mathbf{I}$ in Eqs. 5 and 6.[47] Recently, grey component analysis based on TR was developed for explorative data analysis[48] and TR related approaches have been used for cancer classification with gene expression data.[49]

PARSIMONY BY THE EFFECTIVE RANK

The effective rank (ER) is a measure of the degrees of freedom being used to fit the model, and hence, the smaller the ER, the more parsimonious the model. Several measures for ER have been proposed,[25,50–54] some of which have been evaluated in Refs. 28 and 55. One approach uses a common basis set and the other two are independent of the basis set. The common basis set approach defines ER as a function of filter factors where filter factors are computed for each eigenvector from the SVD of the calibration spectra.[25] Filter factors have been defined for PLS, PCR, and RR but not for the modifications of TR and PLS proposed in this paper.

A method to estimate ER not dependent on a particular basis set is based on leave-one-out cross-validation and has been identified as pseudo-degrees of freedom.[54] Another basis set independent method, and the approach taken in this paper, uses a Monte Carlo method and is named the generalized degrees of freedom.[53] The measure does depend on a meta-parameter that sets the magnitude of perturbations to the calibration concentrations in $\mathbf{y}$. However, the resulting model ER is quite invariant to the actual value of this meta-parameter, i.e., a large range of values produce the same results.[28,52,53] This ER is defined as the sum of sensitivities for the fitted values ($\hat{\mathbf{y}}$) from the model relative to the perturbations in respective values of $\mathbf{y}$. The algorithm used in this paper adds to $\mathbf{y}$ normally distributed noise at 1% to form respective noise-perturbed models, followed by obtaining the corresponding $\hat{\mathbf{y}}$ vectors of fitted values. This process is repeated 300 times. For each calibration sample, the slope from the plot of the 300 fitted values against the noise added is obtained and the ER is the sum of all the slopes.

Effective ranks computed from the three methods have been experimentally found to essentially be equivalent for PLS, PCR, and RR.[28,55] As with harmonious curves, an ER measure provides a mechanism for impartial graphical comparison of different modeling methods. This comparison is accomplished by plotting a bias measure such as the error criterion RMSEC and/or root mean square error of validation (RMSEV) against ER. Such plots have been well described.[56]

EXPERIMENTAL

Software. MatLab 7 (The MathWorks, Natick, MA) programs for TR and PLS were written by the authors.

Data Centering. When the concentrations for $\mathbf{y}_L$ are not zero, $\mathbf{X}$, $\mathbf{y}$, $\mathbf{L}$, and $\mathbf{y}_L$ are mean centered to respective means. This local centering approach has been shown to provide improved modeling performance.[57,58] Validation samples measured under the same conditions as $\mathbf{L}$ are centered to the mean of $\mathbf{L}$ prior to prediction. When the concentrations for $\mathbf{y}_L$ are zero, only $\mathbf{X}$ and $\mathbf{y}$ are mean centered and validation samples are mean centered relative to $\mathbf{X}$ before prediction.

Data Sets. *Temperature.* Twenty-two samples composed of water, ethanol, and 2-propanol were measured from 590 to 1091 nm at 1 nm intervals at 30, 40, 50, 60, and 70 °C.[59] Spectra from 850 to 1049 nm were used. Temperature-specific calibration and validation sets described in Ref. 52 were used to form the same respective 13 and 6 sample sets (pure component samples are excluded). Results are presented for primary calibration at 30 °C for ethanol to predict ethanol at 50 °C. To select the standardization set, the Kennard Stone algorithm was applied to the 13-sample calibration set at 30 °C, but measured at 50 °C. The Kennard Stone algorithm utilized selects the first sample closest to the mean and successive samples are furthest from this sample.

Corn. Eighty (80) samples of corn were measured from 1100 to 2498 nm at 2 nm intervals on three near-infrared (NIR) spectrometers designated m5, mp5, and mp6.[60] Reference values are provided for oil, protein, starch, and moisture

content, and protein is the prediction property studied in this paper. For this study, every other wavelength was used for a total of 350 wavelengths. Thirty samples selected using a Kennard Stone algorithm on the 80 samples measured on instrument m5 serve as the primary calibration set. The validation samples are the remaining samples but measured on instrument mp5. The standardization set consist of four samples selected by a Kennard Stone algorithm from the same 30 calibration samples but measured on instrument mp5. Also available with this data set are three spectra measured for a glass standard on m5 and four spectra measured for the same glass standard on mp5. The mean difference spectrum is used for the standardization set.

τ and λ Values. Sixty τ^2 values were used ranging from approximately 1×10^{-15} to 0.5 as determined from the singular value decomposition of the primary spectral calibration covariance matrix. Fifty λ^2 values were used for TR ranging from 1×10^{-5} to 1×10^9. These same values were used for PLS with $\lambda_{PLS} = \lambda_{TR}^2$.

RESULTS AND DISCUSSION

This paper investigates two spectral situations for the standardization set in **L**. One consists of spectra of samples measured at the new conditions (or on a secondary instrument) and the other is difference spectra between the same samples measured when the original calibration model was formed and under the new conditions. Regardless of the type of spectra, samples can or cannot contain the analyte.

When the standardization set is composed of samples containing the analyte, a better representation of the analyte-matrix-effected spectra for the new condition will be captured compared to using standardization samples without the analyte. Varying degrees of the new matrix effects will be represented depending on the level of matrix matching between the standardization samples and future samples. The more analyte-matrix-effecting aspects missing from the standardization set, the more ineffective the standardization set becomes. These statements are true whether spectral differences or spectra at the new conditions are used. Thus, the general focus in this study is using samples with the analyte present in the standardization set. However, when the Kennard Stone algorithm was used to select the standardization set for the temperature data, one of the four samples has no analyte (ethanol). Another exception is the corn data when the glass standards are used. Studies were performed (not reported) with the temperature data where all samples in the standardization set contained no analyte and inferior results were obtained.

Calibration Maintenance.. The temperature data represents the situation in which a primary model is formed under one condition and has to be updated to handle new conditions. The first situation discussed is when measured spectra are used for the standardization set in **L** and, hence, reference ethanol concentrations are needed for $\mathbf{y}_L$.

The first task is to select an appropriate value for τ. Plotted in Fig. 1 is the L-curve to accomplish this selection. Each gray scale represents a different τ value and the + symbol denotes the range of λ values for each respective τ value. There are only a few possibilities for suitable models in the corner region of the L-curve. As long as the selected τ value is in close proximity to the corner region, the exact value of τ is not critical, as final RMSEV values are not appreciably different. The τ value selected for this data set is not only in the corner
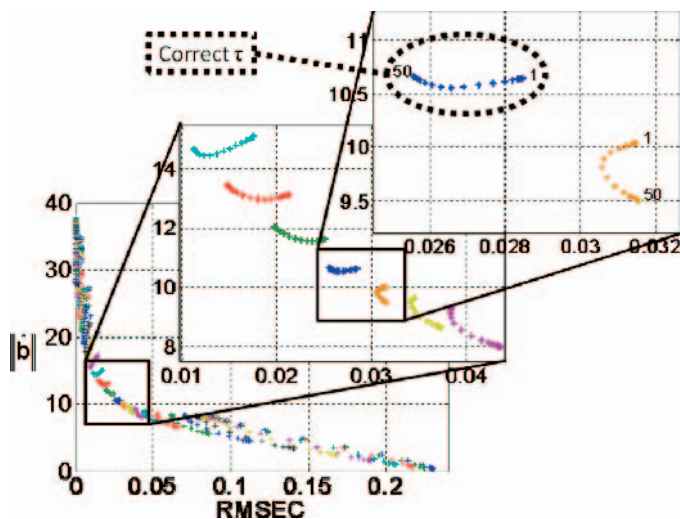

FIG. 1. The temperature data TR L-curve for selecting λ where the standardization data set has spectra and respective ethanol concentrations. Each gray scale represents a different τ value and the + symbol denotes the range of λ values for each respective τ value. Numbers correspond to λ endpoint indices where $\lambda_1 = 3.16 \times 10^4$ and $\lambda_{50} = 0$. See text for further description.

region of Fig. 1 but is also the point at which the RMSEC values decrease as λ_i converges to zero. A good model is when the weight on spectra and concentration values for the new conditions decreases, so does the prediction error of the calibration set. From Fig. 1, it can be seen that in the smaller regression vector 2-norm region and corresponding larger RMSEC portion, the RMSEC values actually increase as less weight is given to the new condition. This trend is exemplified in the second expansion plot and was observed for other situations. The point in the corner portion where the correct RMSEC trend occurs assists in pinpointing the proper corner region of the L-curve. For consistency, this same approach was used to identify the number of factors to use with PLS. Other methods could be used to determine τ values or the number of factors such as a cross-validation process, but these were not studied in this paper. The focus of this work is not determining the best method for meta-parameter selection, but to show the applicability of using a TR approach for calibration maintenance and transfer.

Once a reasonable value for τ or the number of PLS factors has been established, the next meta-parameter value to determine is λ. An L-curve approach is again used by plotting log(RMSEL) against log(RMSEC) and is shown in Fig. 2. Models in the corner region represent those with a reasonable tradeoff between predicting calibration samples and the standardization set with low errors. As λ values decrease in magnitude from the corner, the RMSEC values decrease because the models are evolving to better fit and hence better predict the calibration set. At λ_{39} in Fig. 2, there is essentially convergence to the model best for **X** and the regression vector shape and size are essentially unchanged as λ values further decrease towards λ_{50} equal to zero. Conversely, as λ values increase in magnitude from the corner, the RMSEL values decrease because the model is evolving to better to predict the standardization set. At λ_{27} in Fig. 2, there is essentially convergence to the model best for **L** and the regression vector is again essentially unchanged in shape and size as λ values further increase towards λ_1 equal to 10^9. In the corner regions, these
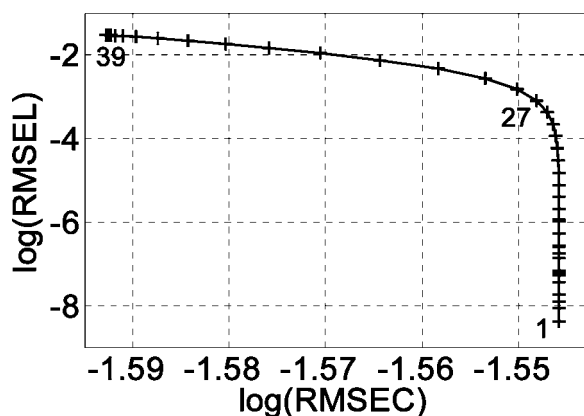
FIG. 2. The temperature data TR L-curve for selecting λ for the situation noted in Fig. 1. Numbers correspond to λ endpoint indices where $\lambda_1 = 3.16 \times 10^4$ and $\lambda_{50} = 0$.

models are reasonably orthogonal to spectral artifacts characterized in the calibration and standardization sets. The approach used in this study was to select the λ value for TR or PLS closest to the corner. However, the approach of just setting λ to some large number for TR or augmented PLS (APLS) could be used. That is, when the emphasis is on the new conditions, setting the value of λ is not critical as long as it is large enough to be in the corner region or beyond with larger values.

Listed in Table I are the RMSEV values describing how well TR and PLS perform. A base value for comparison is using PLS without a standardization set, i.e., no calibration maintenance. To distinguish between the two PLS variations, APLS is used to differentiate from base PLS without any augmentation. As expected, performing no maintenance provides poor predictability at the new temperature. Improved results are obtained with TR and four standardization samples. Prediction error also improves with APLS, but TR appears to provide lower prediction errors at a cost to potential inflated prediction variance due to the larger regression vector 2-norm. Another base value for comparison is how TR and APLS perform compared to the PLS model at 50 °C, i.e., no calibration maintenance is needed as the model is built with a full recalibration using only data at the new condition. In this case, both TR and APLS perform worse than the full recalibration. However, this model can only be used to predict samples at 50 °C and not at the original temperature of 30 °C. Additionally, full recalibration requires more samples.

With fewer samples in the standardization set (two samples), the TR RMSEV value degrades. It is generally accepted that only a few representative samples are needed and four samples

provides acceptable results. The key point in forming the standardization set is to properly span the new condition. In this study (not fully reported) random two, three, and four sample selection was performed several times and the mean RMSEV trend is that three samples are significantly better than two. When four samples are used, there is only a small improvement over three. For a fixed number of samples, some random sample selections showed marked improvements in RMSEV values compared to another random selection. This further confirms the importance of the standardization set properly spanning the new conditions.[15]

When spectral differences are used for **L**, the values for y_L are all zero and Table I shows improved results compared to using spectra with reference values. With spectral difference, TR and APLS perform similarly. Both also perform similarly to full recalibration at the new temperature, although the regression vector norms and ER are greater for TR and APLS implying that greater prediction uncertainties are possible. By augmenting with only a few standardization samples, difference spectra provide an efficient process for updating a primary calibration model to a new condition. Difference spectra allow TR and APLS to better determine a regression vector with improved orthogonality to the new conditions because there is only non-analyte information in **L**. Using spectra with reference values requires TR and APLS to determine what part of **L** is due to the analyte and, simultaneously, what part of **L** is relevant to the non-analyte information.

Because analyte concentration is zero with difference spectra, an SVD was performed on the standardization set and **L** was replaced with a varying number of eigenvectors. Results shown in Table I reveal that using only three eigenvectors improves prediction errors and further reduction in prediction error is gained with two eigenvectors. It is logical to expect some improvement to occur as using a subset of eigenvectors filters some spectral difference noise, allowing TR and PLS to better discern the unique non-analyte information.

With regard to parsimony determined by the ER, the general pattern is that using difference spectra for the standardization set provides models with smaller ERs. The anomaly to this is using difference spectra with APLS, which has the largest ER. This may well be due to the values used for the meta-parameters with APLS. Because there is better representation of the new conditions with difference spectra, the ER is generally smaller.

Calibration Transfer. The corn data represents the situation in which a primary model is formed on one instrument and has to be updated to handle a new instrument. The first situation discussed is using measured spectra for the standardization set in **L** and using protein composition for y_L. From values listed

TABLE I. Prediction results for the temperature data using spectra and spectra differences.

Method	L (No. eig.)	No. samples	RMSEV	‖b̂‖	ER	τ or factors	λ
TR	Spectra	4	0.0408	10.63	5.35	1.20×10^{-2}	5.11
TR	Spectra	2	0.0907	11.90	5.18	9.00×10^{-3}	3.65
APLS	Spectra	4	0.0573	10.23	3.36	4	1.78
TR	Difference	4	0.0272	16.52	3.91	5.07×10^{-3}	3.65
TR	Difference (3)	4	0.0304	11.90	2.97	1.20×10^{-2}	0.35
TR	Difference (2)	4	0.0291	15.74	4.64	5.07×10^{-3}	0.49
APLS	Difference	4	0.0360	10.97	8.51	4	13.33
PLS	No L: Cal at 30 °C	0	0.2548	8.62	3.25	3	0
PLS	No L: Cal at 50 °C	0	0.0379	8.01	3.14	3	0

TABLE II. Prediction results for the corn data using spectra and spectra differences.

Method	L	No. samples	RMSEV	$\lVert\hat{\mathbf{b}}\rVert$	ER	τ or factors	λ
TR	Spectra	4	0.1881	57.46	10.39	1.35×10^{-2}	7.15
APLS	Spectra	4	0.3165	40.71	6.22	4	6.81
TR	Difference	4	0.1744	57.09	7.41	1.35×10^{-1}	14.00
TR	Difference	1	0.1807	54.45	8.98	1.35×10^{-2}	2.61
APLS	Difference	4	0.2825	40.52	10.13	6	26.10
APLS	Difference	1	0.2253	52.94	8.92	6	1.78
LPLS[a]	Spectra	4	0.1776	55.68	9.45	6	0
PLS	No L: Cal on m5	0	0.7570	55.68	9.45	6	0
PLS	No L: Cal on mp5	0	0.1893	64.79	9.01	6	0

[a] LPLS = mp5 validation samples are mean centered to the four mp5 samples in **L** and prediction is by PLS based on the m5 calibration samples (local center for validation).

in Table II, it is observed that the RMSEV TR and APLS values are better than not providing any calibration transfer and predicting spectra measured on instrument mp5 with the primary calibration model on m5. The TR approach appears to accomplish the update better than APLS with respect to the RMSEV values. However, in further analysis of the TR and APLS results, it was found that as λ approaches zero, forming a smaller RMSEC value, the RMSEV is also reduced. Other calibration transfer studies have found that just mean centering the secondary instrument validation set to the mean of the standardization set measured on the secondary instrument is often enough processing to allow the primary calibration model to be used.[61,62] That is, locally centering data to the respective instrument means brings the self-centered mp5 spectra into the m5 self-centered calibration space sufficiently to allow prediction of the mp5 validation samples. As λ decreases, the model is being biased towards the calibration set and when the validation set is appropriately centered by the standardization set, the better the model can then also fit the validation set. This observation implies that the primary difference is instrumental and not chemical or physical. Only local centering was performed in this study and is identified as local PLS (LPLS) in Table II and does indeed provide a slightly better RMSEV value with a smaller regression vector 2-norm than TR and greater differences with APLS. The greater difference with APLS is conjectured to be due to the more limited number of factors to step through compared to the unlimited number of τ values and, hence, a smoother variation in plots.

Changing the Kennard Stone algorithm to select samples furthest from each other and not starting with the sample closest to the mean as the first sample generates even greater TR and APLS RMSEV values of 0.2930 and 0.554, respectively. This observation further implies the importance of the standardization set structure not only to properly span the conditions, but also to capture the mean. If only local centering is used with the temperature data, the four factor LPLS RMSEV and regression vector 2-norm values are 0.0908 and 16.13, respectively. Centering with only the temperature data is not enough. It may be that if a better representative center could be identified for the temperature data, i.e., a centering spectrum that better translates the new condition to the primary model space, then local centering may suffice.

When spectral differences are used for **L**, the values for $\mathbf{y}_L$ are all zero and TR and APLS should now be better able to provide a model orthogonal to the instrumental difference and simultaneously be able to predict the calibration set on instrument m5. Results in Table II show that improved results are indeed obtained and comparable to LPLS as well as just

forming a calibration model on instrument mp5 to only predict validation samples measured on mp5.

Also applied to spectral differences is the situation of using spectral differences for a glass standard measured on both instruments. Tabulated in Table III are the results in which it is observed that equivalent results are obtained with TR and APLS using spectral differences, LPLS, and full recalibration. This is further evidence that the primary difference between spectra measured on instruments m5 and m5p is instrumental. If a chemical, physical, and/or environmental effect were simultaneously occurring, then just using glass standards would not be sufficient to completely characterize the new conditions to provide an appropriate model update. If instrumental changes are occurring in conjunction with chemical, physical, or environmental changes, it is expected that different weights would be needed for each effect as described in Eq. 18. Alternatively, based on observations from the temperature and corn data, it may be possible to just use one λ set to a large number.

Similar ER trends observed with the temperature data are seen with the corn data. Except for APLS, the ER generally reduces with difference spectra. A study was not performed to ascertain the sensitivity of the ER relative to respective meta-parameters.

While not studied, it is possible to use these TR and PLS approaches to provide a model workable on multiple instruments. In this case, the primary calibration data is augmented with spectra measured on different instruments. If the degree of differences between the secondary instruments and the primary instrument is small, one λ value should suffice. If respective differences are unique, it is expected that individual λ values would be required for each secondary instrument to determine a model. Again, it may be possible to just use one λ set to a large number.

CONCLUSION

Results were only compared to APLS, but other methods such as an augmented PCR could be used. This study demonstrates that the TR approach to calibration maintenance and transfer provides the flexibility needed to desensitize a

TABLE III. Prediction results for the corn data using glass standards.

Method	RMSEV	$\lVert\hat{\mathbf{b}}\rVert$	ER	τ or factors	λ
TR	0.2277	55.40	9.07	1.35×10^{-2}	1.87
APLS	0.1838	54.40	9.25	6	1.78

primary model to new conditions once it has been determined that new samples are essentially outliers to the current calibration model. While not part of this paper, methods exist to determine outliers[1,2] and indeed represent a critical aspect of calibration maintenance and transfer.[63]

It has been shown that only a few samples are needed to update the model but these samples should be selected (if possible) based on a selection strategy that maximizes the ability of the standardization set to properly span the new conditions. With a standardization set properly spanning the new situation or secondary instrument, improved results should be obtainable compared to a random selection of samples for the standardization set. The better $\mathbf{L}$ focuses on what the updated regression vectors need to be desensitized (orthogonalized) to, the more accurate the predictions will be.

When instrument differences are the correction focus, simple local centering can often suffice. Using difference spectra provided equivalent results to simple local mean centering. When model updating is for a new chemical, physical, or environmental effect, TR and APLS performed better than just mean centering.

This work reported on a methodology that selects a weight value for λ that compromises prediction of the original full calibration data and the data used for updating. If artifacts in the new conditions are the emphasis, the weight should be set to a large number and no optimization of the weight is needed.

It may be possible to dispense with meta-parameters altogether by using a target optimization scheme as outlined in Refs. 41, 46, and 64. Such an approach would also be useful to determine individual weights in Eq. 18 for respective spectra in $\mathbf{L}$ uniquely characterizing each spectral artifact such as drift, new chemical component, and temperature. The target approach could also be applied to weighting subsets of samples, e.g., a set of samples spanning the drift variance, a set of samples spanning the temperature effect, etc. Determining multiple weights for samples in $\mathbf{L}$ is akin to generalized RR in which a meta-parameter is sought for each variable. The target approach applied to generalized RR should be applicable here and is currently under study. As noted in the Results and Discussion section, it may be that one weight value set to a large number could be used.

In the work presented in this study, a model desensitized to new spectral artifacts over the complete spectral range was sought using TR or APLS. Under current study is calibration transfer and maintenance where key wavelengths are simultaneously sought as part of the model updating process.

ACKNOWLEDGMENTS

1. T. Næs, T. Isaksson, T. Fern, and T. Davies, *A User Friendly Guide to Multivariate Calibration and Classification* (NIR Publications, Chichester, 2002).
2. T. J. Hastie, R. J. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer-Verlag, New York, 2001).
3. O. E. de Noord, Chemom. Intell. Lab. Syst. **25,** 85 (1994).
4. T. Fearn, J. Near Infrared Spectrosc. **9,** 229 (2001).
5. R. N. Feudale, N. A. Woody, H. Tan, A. J. Myles, S. D. Brown, and J. Ferré, Chemom. Intell. Lab. Syst. **64,** 181 (2002).
6. R. P. Cogdill, C. A. Anderson, and J. K. Drennen III, AAPS PharmSciTech **6,** E284 (2005).
7. J. H. Kalivas and B. R. Kowalski, Anal. Chem. **54,** 560 (1982).
8. F. Wülfert, W. T. Kok, O. E. de Noord, and A. K. Smilde, Chemom. Intell. Lab. Syst. **51,** 189 (2000).
9. C. E. Anderson and J. H. Kalivas, Appl. Spectrosc. **53,** 1268 (1999).
10. Z. Y. Wang, T. Dean, and B. R. Kowalski, Anal. Chem. **67,** 2379 (1995).
11. D. M. Haaland, Appl. Spectrosc. **54,** 246 (2000).
12. K. Maruo, M. Oota, M. Tsurugi, T. Nakagawa, H. Arimoto, M. Hayakawa, M. Tamura, Y. Ozaki, and Y. Yamada, Appl. Spectrosc. **60,** 1423 (2006).
13. M. O. Westerhaus, "Improving Repeatability of NIR Calibrations Across Instruments", in *Proceedings of the Third International Near Infrared Spectroscopy Conference,* R. Biston and N. Bartiaux-Thill, Eds. (Agriculture Research Centre Publishing, Gembloux, Belgium, 1991), p. 671.
14. Y. Wang, D. J. Veltkamp, and B. R. Kowalski, Anal. Chem. **63,** 2750 (1991).
15. X. Capron, B. Walczak, O. E. de Noord, and D. L. Massart, Chemom. Intell. Lab. Syst. **76,** 205 (2005).
16. K. E. Kramer and G. W. Small, Appl. Spectrosc. **61,** 497 (2007).
17. M. R. Riley, M. A. Arnold, and D. W. Murhammer, Appl. Spectrosc. **52,** 1339 (1998).
18. Y. Sulub and G. W. Small, Appl. Spectrosc. **61,** 406 (2007).
19. C. M. Wehlburg, D. M. Haaland, D. K. Melgaard, and L. E. Martin, Appl. Spectrosc. **56,** 605 (2002).
20. C. M. Wehlburg, D. M. Haaland, and D. K. Melgaard, Appl. Spectrosc. **56,** 877 (2002).
21. C. L. Stork and B. R. Kowalski, Chemom. Intell. Lab. Syst. **48,** 151 (1999).
22. M. J. Sáiz-Abajo, B. H. Mevik, V. H. Segtnan, and T. Næs, Anal. Chim. Acta **533,** 147 (2005).
23. M. Zhu, Am. Stat. Assoc. **62,** 97 (2008).
24. B. H. Mevik, V. H. Segtnan, and T. Næs, J. Chemom. **18,** 498 (2005).
25. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (SIAM, Philadelphia, 1998).
26. R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier, Amsterdam, 2005).
27. A. Dax, SIAM J. Optimization **2,** 602 (1992).
28. J. B. Forrester and J. H. Kalivas, J. Chemom. **18,** 372 (2004).
29. C. L. Lawson and R. J. Hanson, *Solving Least Square Problems* (SIAM, Philadelphia, PA, 1995).
30. F. Vogt, K. Rebstock, and M. Tacke, Chemom. Intell. Lab. Syst. **50,** 175 (2000).
31. F. Vogt, H. Steiner, K. Booksh, and B. Mizaikoff, Appl. Spectrosc. **58,** 683 (2004).
32. K. Faber and B. R. Kowalski, J. Chemom. **11,** 181 (1997).
33. K. Faber and B. R. Kowalski, Chemom. Intell. Lab. Syst. **34,** 283 (1996).
34. N. M. Faber, X. H. Song, and P. K. Hopke, Trends Anal. Chem. **22,** 330 (2003).
35. J. A. Fernández Pierna, L. Jin, F. Wahl, N. M. Faber, and D. L. Massert, Chemom. Intell. Lab. Syst. **65,** 281 (2003).
36. J. H. Kalivas and P. M. Lang, *Mathematical Analysis of Spectral Orthogonality* (Marcel Dekker, New York, 1994).
37. A. E. Hoerl and R. W. Kennard, Technometrics **12,** 55 (1970).
38. R. DiFoggio, J. Chemom. **19,** 203 (2005).
39. R. DiFoggio, J. Chemom. **21,** 208 (2007).
40. J. H. Kalivas, Anal. Chim. Acta **505,** 9 (2004).
41. F. Stout and J. H. Kalivas, J. Chemom. **20,** 22 (2006).
42. P. H. C. Eilers, Anal. Chem. **75,** 3631 (2003).
43. P. J. Gemperline and E. Cash, Anal. Chem. **75,** 4236 (2003).
44. J. F. Claerbout and F. Muir, Geophysics **38,** 826 (1973).
45. R. Tibshirani, J. R. Statist. Soc. B **58,** 267 (1996).
46. F. Stout, J. H. Kalivas, and K. Héberger, Appl. Spectrosc. **61,** 85 (2007), and references therein.
47. W. Shih, K. L. Bechtel, and M. S. Feld, Anal. Chem. **79,** 234 (2007).
48. J. A. Westerhuis, E. P. P. A. Derks, H. C. J. Hoefsloot, and A. K. Smilde, J. Chemom. **21,** 474 (2007).
49. E. Andries, T. Hagstrom, S. R. Atlas, and C. William, J. Bioinformatics Comp. Biol. **5,** 79 (2007).
50. F. A. O'Sullivan, Stat. Sci. **1,** 502 (1986).
51. R. H. Myers, *Classical and Modern Regression with Applications* (Duxbury, Pacific Grove, 1990), 2nd ed.
52. K. Baumann and N. Stiefl, J. Comp.-Aided Mol. Design **18,** 549 (2004).
53. J. Ye, J. Am. Stat. Assoc. **93,** 120 (1998).
54. H. van der Voet, J. Chemom. **13,** 195 (1999).

55. H. A. Seipel and J. H. Kalivas, J. Chemom. **19,** 64 (2005).
56. J. H. Kalivas and P. J. Gemperline, "Calibration", in *Practical Guide to Chemometrics*, P. J. Gemperline, Ed. (CRC Press Taylor & Francis Group, Boca Raton, FL, 2006), 2nd ed.
57. Y. P. Du, S. Kasemsumran, K. Maruo, T. Nakagawa, and Y. Ozaki, Anal. Sci. **21,** 979 (2005).
58. J. H. Kalivas, J. Chemom. **22,** 227 (2008).
59. F. Wülfert, T. W. Kok, and A. K. Smilde, Anal. Chem. **70,** 1761 (1998).
60. B. M. Wise, N. B. Gallagher, R. Bro, and J. M. Shaver, *PLS_Toolbox 3.0 for use with MATLAB* (Eigenvector Research, Manson, WA, 2003).
61. C. E. Anderson and J. H. Kalivas, Appl. Spectrosc. **53,** 1268 (1999).
62. H. Swieranga, W. G. Haanstra, A. P. de Weijer, and L. M. C. Buydens, Appl. Spectrosc. **52,** 7 (1998).
63. R. D. Guenard, C. M. Wehlburg, R. J. Pell, and D. M. Haaland, Appl. Spectrosc. **61,** 747 (2007).
64. J. H. Kalivas and R. L. Green, Appl. Spectrosc. **55,** 1645 (2001).