Anal. Chem. **2010,** *82,* 3642–3649

Model Updating for Spectral Calibration Maintenance and Transfer Using 1-Norm Variants of Tikhonov Regularization

M. Ross Kunz,[†] John H. Kalivas,*[,†] and Erik Andries[‡]

Department of Chemistry, Idaho State University, Pocatello, Idaho 83209, and Department of Mathematics, Central New Mexico Community College, and Center for Advanced Research Computing, University of New Mexico, Albuquerque, New Mexico 87106

In this study, calibration maintenance confronts the problem of updating a model developed in the primary condition to accurately predict the calibrated analyte in samples measured in new secondary conditions. Calibration transfer refers to updating a model based on a primary instrument to predict samples measured on new secondary instruments. A 2-norm variant of Tikhonov regularization (TR) has been used with spectroscopic data to perform calibration maintenance and transfer where just a few samples measured in the secondary condition/instrument are augmented to the primary calibration data to update the primary model. To achieve improved predictions, in this paper we report on 1-norm penalties to create two novel variants of TR for model updating. To solve the multiple penalty minimization numerical problems involved with the new TR variants, data transformation processes are applied, allowing a least absolute shrinkage and selection operator type algorithm to be implemented. Near-infrared spectra measured under two different temperatures represent the calibration maintenance application, and near-infrared spectra measured on two instruments denote the calibration transfer situation. Compared to TR in the recently developed 2-norm penalty mode, validation sample prediction errors are reduced when the 1-norm TR variant that selects wavelengths is used.

A multivariate spectral calibration model is restricted in prediction capabilities because, over time, sample, environmental, and instrumental conditions change and introduce new variances not accounted (calibrated) for in the original primary model. Calibration maintenance and transfer address the issue of adjusting spectra and/or the primary model such that the model continually predicts with the same quality. Calibration maintenance specifically requires building a model in the original primary condition and then maintaining this model to form accurate predictions of new samples measured in new noncalibrated secondary conditions. New secondary conditions can include temperature, pressure, particle size, pH, humidity, and new spectrally responding species. Calibration transfer involves building a model on a primary instrument and then predicting samples measured on a new instrument (secondary instrument). Additionally, the secondary instrument can be the primary instrument at a later point in time. In addition to basic differences between two different instruments, transfer problems can also include drift in instrument measurements or repairing the instrument with new components. Calibration maintenance and transfer have been subjects of numerous studies and are well reviewed.[1−4]

There are three common ways to perform spectral calibration maintenance and transfer.

The first method is to build an exhaustive model robust enough to accurately predict in future conditions. This type of model building requires many samples to properly span anticipated variances during the lifetime of the model. Alternatively, on an as-needed basis, a large set of samples measured in the secondary conditions can be added to samples measured in the primary conditions. This process essentially amounts to a complete recalibration that can be costly and time-consuming. A related approach to building a robust model is using spectral preprocessing methods such as multiplicative scatter correction,[5,6] finite impulse response filters,[7,8] derivatives, orthogonal correction,[9,10] and wavelength selection.[11−14]

* Corresponding author. E-mail: kalijohn @isu.edu. Phone: (208) 282-2726. Fax: (208) 282-4373.

[†] Idaho State University.

[‡] Central New Mexico Community College and University of New Mexico.

(1) Fearn, T. *J. Near Infrared Spectrosc.* **2001**, *9*, 229–244.
(2) Feudale, R. N.; Woody, N. A.; Tan, H.; Myles, A. J.; Brown, S. D.; Ferré, J. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 181–192.
(3) Cogdill, R. P.; Anderson, C. A.; Drennen, J. K., III. *AAPS PharmSciTech* **2005**, *6*, E284−E297.
(4) Brown, S. D. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S. D., Tauler, R., Walczak, B., Eds.-in-Chief; Kalivas, J. H., Section Ed.; Elsevier: Amsterdam, 2009; pp 345−377.
(5) Martens, H.; Stark, E. *J. Pharm. Biomed. Anal.* **1991**, *9*, 625–635.
(6) Pedersen, D. K.; Martens, H.; Nielsen, J. P.; Engelsen, S. B. *Appl. Spectrosc.* **2002**, *56*, 1206–1214.
(7) Blank, T. B.; Sum, S. T.; Brown, S. D. *Anal. Chem.* **1996**, *68*, 2987–2995.
(8) Tan, H. W.; Sum, S. T.; Brown, S. D. *Appl. Spectrosc.* **2002**, *56*, 1098–1106.
(9) Igne, B.; Roger, J. M.; Roussel, S.; Bellon-Maurel, V.; Hurburgh, C. R. *Chemom. Intell. Lab. Syst.* **2009**, *99*, 57–65.
(10) Zeaiter, M.; Roger, J. M.; Bellon-Maurel, V. *Chemom. Intell. Lab. Syst.* **2006**, *80*, 227–235.
(11) de Groot, P. J.; Swierenga, H.; Postma, G. J.; Melssen, W. J.; Buydens, L. M. C. *Appl. Spectrosc.* **2003**, *57*, 642–648.
(12) Swierenga, H.; Wülfert, F.; de Noord, O. E.; de Weijer, A. P.; Smilde, A. K.; Buydens, L. M. C. *Anal. Chim. Acta* **2000**, *411*, 121–135.
(13) Ozdemir, D.; Williams, R. *Appl. Spectrosc.* **1999**, *53*, 210–217.
(14) Honorato, F. A.; Galvão, R. K. H.; Pimentel, M. F.; de Barros Neto, B.; Araújo, M. C. U. A.; de Carvalho, F. R. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 65–72.

10.1021/ac902881m
Published on Web 03/30/2010

The second approach to calibration maintenance and transfer is to transform spectra measured in the secondary condition to appear as if the samples were measured in the primary condition. Such processes require a small set of samples (standardization set) to be measured in both conditions and allow mapping from one spectral domain to another, e.g., from mid-infrared to near-infrared (NIR) or from a low-resolution instrument to a high-resolution instrument.[15,16] Thus, this second approach is limited to conditions that alter spectra (wavelength shifts, intensity changes, and/or baseline offsets) and is not applicable when new variances arise such as new spectrally responding species or the analyte concentration in a new sample is outside the primary condition concentration range. Further limiting this spectral transformation method is the need to have a stable standardization set for the same samples to be measured under primary and secondary conditions. A recent variation of orthogonal preprocessing in conjunction with spectral transformation is dynamic orthogonal projection where a few samples are measured in the secondary condition and, using kernel functions, these spectra are modified to appear as if they were measured in the primary condition.[10] Spectral differences between the actual secondary and the same samples estimated as primary spectra are then used in the orthogonal preprocessing. With this approach, measurement in the primary condition is not required, eliminating the need for standardization set stability.

A third approach is to infuse into the primary model information describing the secondary condition, thereby providing a mechanism by which the primary model can be updated to predict in the new conditions. To account for new secondary conditions, spectra of samples measured in the secondary conditions are augmented to the primary spectra and an updated model is obtained. As noted previously, many samples are generally needed to span the new condition, whether it be chemical, physical, and/or instrumental differences requiring correction, and this amounts to a full recalibration. Recent work using a Tikhonov regularization (TR) model updating process only requires a few samples measured in the secondary condition.[17] In this paper we report on two new TR variants, one of which includes wavelength selection as part of the model updating. Because it has been shown that wavelength selection alone can accomplish calibration maintenance and transfer,[11–14] it is expected that TR with wavelength selection will significantly improve the capabilities of TR. The new approaches are applied to two NIR spectral data sets. One is concerned with calibration maintenance where the primary model formed at 30 °C needs to be updated to handle samples measured at 50 °C. The other data set deals with calibration transfer. While model updating with TR does not easily lend itself to some unique situations where spectral transformation processes are applicable, e.g., adjusting spectra measured in one spectral domain to appear as if the samples were measured in another spectral domain, model updating with TR can correct for situations when new variances arise such as new spectrally responding species or the analyte concentration in a new sample is outside the primary condition concentration range.

(15) Balke, S. T.; Lew, R. *Appl. Spectrosc.* **1993**, *47*, 1747–1750.
(16) Zhang, L.; Noda, I.; Wu, Y. *Appl. Spectrosc.* **2009**, *63*, 112–119.
(17) Kalivas, J. H.; Siano, G. G.; Andries, E.; Goicoechea, H. C. *Appl. Spectrosc.* **2009**, *63*, 800–809.

VARIANTS OF TIKHONOV REGULARIZATION

The standard relationship for multivariate calibration used to develop the primary model is given by

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

where $\mathbf{y}$ represents an $m \times 1$ vector of quantitative analyte values (such as concentration) for m calibration samples, $\mathbf{X}$ represents the $m \times n$ matrix of spectra of n wavelengths or frequencies for the primary condition, and $\mathbf{b}$ represents the $n \times 1$ vector of primary model coefficients. The $m \times 1$ vector $\mathbf{e}$ indicates normally distributed errors with mean zero and covariance matrix $\sigma^2\mathbf{I}$. When $n \gg m$, the common spectral situation, methods such as ridge regression (RR), partial least squares (PLS), or principal component regression (PCR) can be used to estimate $\mathbf{b}$ by $\hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y}$, where $\mathbf{X}^{+}$ is the respective pseudoinverse of $\mathbf{X}$.[18,19] The model vector $\mathbf{b}$ must also be able to accurately predict analyte values in any new sample spectrum, $\mathbf{x}_{\text{new}}$, by $\hat{\mathbf{y}}_{\text{new}} = \mathbf{x}_{\text{new}}^{t}\hat{\mathbf{b}}$. Wavelengths can be selected, and if a small enough set is selected such that $n \leq m$ and $\mathbf{X}$ is well conditioned, then multiple linear regression (MLR) can be used to estimate the model vector.[18,19]

It is possible to measure only a few new samples in the secondary condition over the same wavelengths as those measured for the primary calibration samples, represented as $\mathbf{M}$ and $\mathbf{y}_{\mathbf{M}}$, and augment this new information to the original primary calibration data if these new samples are properly weighted by λ. Ignoring the $\mathbf{e}$ term, eq 1 can be modified to include the weighted augmentation as

$$\begin{pmatrix}\mathbf{y}\\ \lambda\mathbf{y}_{\mathbf{M}}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{M}\end{pmatrix}\mathbf{b} \tag{2}$$

Without the weight metaparameter (tuning parameter) λ, one runs into issues of the new calibration model being more influenced by the original primary conditions.

Equation 2 can be solved by PLS, PCR, or some other method. In this case, eq 2 is reformatted to $\mathbf{y}_{\text{A}} = \mathbf{X}_{\text{A}}'\mathbf{b}$, where the A subscript denotes the augmented arrays in eq 2 and the slanted prime represents spectra modified relative to PLS, PCR, or some other method. For PLS or PCR, the modified $\mathbf{X}$ is the respective projected matrix.[18,19,21] Essentially, PLS or PCR determines the model vector satisfying the expression $\min \|\mathbf{X}_{\text{A}}'\mathbf{b} - \mathbf{y}_{\text{A}}\|_2^2$, where $\|...\|_p$ denotes the vector p-norm and the subscript 2 symbolizes the vector 2-norm (Euclidean norm). In recent work on calibration maintenance and transfer, satisfactory results were obtained by using PLS with eq 2.[17]

TR 2. It is also possible to solve eq 2 by a TR approach. Specifically, a solution is sought that satisfies

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \lambda^2\|\mathbf{Mb} - \mathbf{y}_{\mathbf{M}}\|_2^2) \tag{3}$$

(18) Næs, T.; Isaksson, T.; Fern, T.; Davies, T. *A User Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, U.K., 2002.
(19) Hastie, T. J.; Tibshirani, R. J.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer-Verlag: New York, 2009.

In applying expression 3 to calibration maintenance and transfer, it was found that an additional tuning parameter, η, was needed, and expression 3 becomes

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \eta^2\|\mathbf{b}\|_2^2 + \lambda^2\|\mathbf{Mb} - \mathbf{y_M}\|_2^2) \qquad (4)$$

where η represents a penalty on the model 2-norm.[17] The first two terms in expression 4 are RR (another form of TR), and hence, expression 4 can be thought of as an RR-based method with an additional 2-norm penalty on predicting a few samples measured in the secondary condition. In applications of TR in the RR format, the first term of expression 4 has been labeled a bias (accuracy) indicator and the middle term reflects the model size and acts as a variance (precision) merit.[20,21] Note that when $\mathbf{M} = \mathbf{I}$, the identity matrix, and $\mathbf{y_M} = \mathbf{0}$ in expression 3, this expression reduces to RR. Using TR in expression 4 will be referred to as TR applying the 2-norm (TR 2) to the new spectral conditions.

The TR variant in expression 4 estimates $\mathbf{b}$ such that it is desensitized to the new noncalibrated spectral artifacts in $\mathbf{M}$ that are confounding an accurate analyte prediction. Simultaneously, the model vector needs to accurately predict the original primary calibration samples lacking the new secondary condition as well as keeping the size of the model vector from being too large. Too large a model vector increases the chance of forming an overfitted model, and prediction variances can escalate.

The quality of the model from TR 2 depends on the values selected for metaparameters η and λ, just like the PLS, PCR, or RR model quality depends on the number of basis vectors (factors, latent vectors) for PLS and PCR or the ridge value for RR. In this work, the L-curve approach is used to select metaparameter values as described in the Model Selection and Results and Discussion sections.

TR 2−1. Changing the regression model 2-norm penalty to a 1-norm in expression 4 forms the following expression:

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \tau\|\mathbf{b}\|_1 + \lambda^2\|\mathbf{Mb} - \mathbf{y_M}\|_2^2) \qquad (5)$$

where τ symbolizes the 1-norm tuning parameter. For this paper, expression 5 shall be referred to as TR applying the 2-norm and the 1-norm (TR 2−1) to the new spectral conditions. To avoid confusion with η being used for weighting the model 2-norm penalty, the τ symbol is used to clarify weighting the 1-norm penalty. Both tuning parameters in expression 5 are determined by an L-curve approach.

The 1-norm acts similarly to the 2-norm in expression 4 in that the 1-norm guards against over- or underfitting. However, unlike the 2-norm, the 1-norm restricts solutions of $\mathbf{b}$ to be sparse and, hence, performs wavelength selection. The least absolute shrinkage and selection operator (LASSO) results from expression 5 when the third term is removed.[19,22−24] Thus, expression 5 can be thought of as LASSO (wavelength selection) with an additional penalty minimizing prediction error for a few samples measured in the secondary conditions. The LASSO approach by itself (the first two terms in expression 5) has been used for multivariate calibration with wavelength selection.[24,25] Expression 5 also contains a method known as the elastic net when $\mathbf{M} = \mathbf{I}$ and $\mathbf{y_M} = \mathbf{0}$,[19,26] and this TR process has also been studied for multivariate calibration with wavelength selection.[25]

Because it has been shown that (1) TR 2 can accurately perform calibration maintenance and transfer[17] and (2) TR 2−1 includes wavelength selection capabilities (coupled with the fact that wavelength selection alone has been shown to accomplish calibration maintenance and transfer[11−14]), TR 2−1 should be an improvement over TR 2. An advantage of using TR 2−1 over a wavelength selection algorithm such as a genetic algorithm is that model formation and wavelength selection occur simultaneously and, hence, less empirically set algorithm parameters are needed. This advantage and others are described in ref 25.

TR 1. Alternatively, the 1-norm penalty can be applied to predict the new samples measured in the secondary conditions in expression 4 instead of the 2-norm. In this case, the following expression results:

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \eta^2\|\mathbf{b}\|_2^2 + \tau\|\mathbf{Mb} - \mathbf{y_M}\|_1) \qquad (6)$$

This variation of TR shall be referred to as TR applying the 1-norm (TR 1) to the new spectral conditions and can be thought of as RR with an additional 1-norm penalty on predicting samples in the secondary condition. Expression 6 also contains the elastic net in the special case with $\mathbf{M} = \mathbf{I}$ and $\mathbf{y_M} = \mathbf{0}$. As with TR 2 and TR 2−1, the L-curve approach is used to determine good values for η and τ.

ALGORITHMS

The model vector for expression 4 is obtained by the least-squares solution $\hat{\mathbf{b}} = (\mathbf{X}^t\mathbf{X} + \eta^2\mathbf{I} + \lambda^2\mathbf{M}^t\mathbf{M})^{-1}(\mathbf{X}^t\mathbf{y} + \lambda^2\mathbf{M}^t\mathbf{y_M})$. Using ranges for the metaparameters η and λ, the resulting models are evaluated using the L-curve approach to assess appropriate metaparameter values. This process is described in the Model Selection section.

The least-angle regression (LAR) algorithm[27] is one of several algorithms that can be used for LASSO and is written as

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \tau\|\mathbf{b}\|_1) \qquad (7)$$

Note that the LAR algorithm does not directly use a τ value. Instead, each iteration corresponds to a nonzero model vector coefficient in $\mathbf{b}$ being added or removed from the model vector in the previous iteration. The value $\tau = 0$ forms the classical least-squares solution (model vector in the last iteration of the LAR algorithm). In this paper, the LAR algorithm is used to solve expressions 5 and 6 with the modification that the columns of $\mathbf{X}$ are not scaled to mean zero and standard deviation 1 (autoscaled) as originally suggested for LAR.[27] Instead, all data are mean centered for all algorithms. In previous work with LAR for multivariate calibration with wavelength selection and autoscaled

(20) Forrester, J. B.; Kalivas, J. H. *J. Chemom.* **2004**, *18*, 372–384.
(21) Kalivas, J. H. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S. D., Tauler, R., Walczak, B., Eds.-in-Chief; Kalivas, J. H., Section Ed.; Elsevier: Amsterdam, 2009; pp 1−32.
(22) Claerbout, J. F.; Muir, F. *Geophysics* **1973**, *38*, 826–844.
(23) Tibshirani, R. *J. R. Stat. Soc. B* **1996**, *58*, 267–288.
(24) Stout, F.; Kalivas, J. H.; Héberger, K. *Appl. Spectrosc.* **2007**, *61*, 85–95.
(25) Ottaway, J.; Kalivas, J. H.; Andries, E. *Anal. Chim. Acta*, submitted for publication.
(26) Zou, H.; Hastie, T. *J. R. Stat. Soc. B* **2005**, *67*, 301–320.
(27) Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. *Ann. Stat.* **2004**, *32*, 407–499.

data, it was found that inappropriate wavelengths were selected; i.e., wavelengths with the best sensitivity were not selected in simulated situations as all wavelengths were seen to have equal sensitivity after scaling.[25]

For TR 2−1 written as expression 5, the data are reformatted to fit expression 7 to allow using LAR. This consists of augmenting $\mathbf{X}$ and $\mathbf{y}$ in expression 5 with $\lambda\mathbf{M}$ and $\lambda\mathbf{y_M}$ to form

$$\mathbf{X}_A = \begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{M}\end{pmatrix}$$

and

$$\mathbf{y}_A = \begin{pmatrix}\mathbf{y}\\ \lambda\mathbf{y_M}\end{pmatrix}$$

to form

$$\min(\|\mathbf{X}_A\mathbf{b} - \mathbf{y}_A\|_2^2 + \tau\|\mathbf{b}\|_1) \quad (8)$$

and the LAR algorithm can now be used.

For TR 1 written as expression 6, the data were also reformatted and, additionally, transformed prior to using the LAR algorithm. Similar to expression 5, $\mathbf{X}$ and $\mathbf{y}$ are first augmented with $\eta\mathbf{I}$ and $\mathbf{0}$ to form

$$\mathbf{X}_A = \begin{pmatrix}\mathbf{X}\\ \eta\mathbf{I}\end{pmatrix}$$

and

$$\mathbf{y}_A = \begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}$$

Using the augmented data, expression 6 now becomes

$$\min(\|\mathbf{X}_A\mathbf{b} - \mathbf{y}_A\|_2^2 + \tau\|\mathbf{M}\mathbf{b} - \mathbf{y_M}\|_1) \quad (9)$$

Next, using processes described in refs 28−30, the augmented data are transformed, and expression 9 is now written as

$$\min(\|\bar{\mathbf{X}}_A\bar{\mathbf{b}} - \bar{\mathbf{y}}_A\|_2^2 + \tau\|\bar{\mathbf{b}}\|_1) \quad (10)$$

which can now be solved by the LAR algorithm. The bar indicates transformed data, and the LAR estimated model vector $\bar{\mathbf{b}}$ must be transformed back to the estimated model vector $\hat{\mathbf{b}}$ that is used to obtain prediction concentrations. While this transformation process is new to using TR in 1-norm, transformations have been used with TR written as expression 3 and with PLS and PCR to apply derivative operators and obtain smooth model vectors as well as with expression 3 to apply penalty weights based on spectroscopic noise.[40]

(28) Elden, L. *BIT* **1982**, *27*, 487–502.
(29) Lawson, C. L.; Hanson, R. J. *Solving Least Square Problems*; SIAM: Philadelphia, 1995.
(30) Hansen, P. C. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*; SIAM: Philadelphia, 1998.

MODEL SELECTION

The L-curve approach to determine suitable tuning parameters and, hence, the final model is well described.[21,29–31] The basis of the approach for RR, PLS, PCR, or TR 2 is plotting the model vector 2-norm (or any measure related to prediction variance or model complexity) against the root-mean-square error of calibration (RMSEC) (or any measure characterizing prediction accuracy or bias). In such a plot, an L-shaped curve is obtained with each point on the curve representing a model based on a different metaparameter and the better model is in the corner region nearest the origin. This model represents the best compromise for the bias/variance trade-off, i.e., the most harmonious model. For TR 2 represented by expression 4, such L-curves were used to sequentially determine tuning parameters η and λ[17] and the process was recently improved.[32] Noting that TR 2−1 and TR 1 in expressions 5 and 6, respectively, are reformatted to expressions 8 and 10, respectively, the focus is on the 1-norm penalties, and hence, the model vector 1-norm is used instead of the 2-norm in forming the L-curves. Each expression still requires determination of two metaparameters, and described in the Results and Discussion section is the process used to sequentially ascertain respective metaparameters.

EXPERIMENTAL SECTION

Software. MATLAB 7.8 (The MathWorks, Natick, MA) programs for TR were written by the authors.

Data Centering. Data are mean centered using the local mean centering approach. Local mean centering is the process whereby $\mathbf{X}$, $\mathbf{y}$, $\mathbf{M}$, and $\mathbf{y_M}$ are each mean centered to the respective means. Validation spectra measured in the same secondary conditions as $\mathbf{M}$ are mean centered to the mean of $\mathbf{M}$. Locally mean centered is best used when specific conditions of the new samples are known and has been shown to provide improved modeling performance.[33,34] Under certain conditions, the samples in $\mathbf{M}$ may have been measured under both the primary and secondary conditions. In this case, difference spectra can be used for $\mathbf{M}$, $\mathbf{y_M}$ is set to the zero vector, and only $\mathbf{X}$ and $\mathbf{y}$ are mean centered. Validation samples in the secondary condition are now centered to the mean of $\mathbf{X}$.

Data Sets. *Temperature.* A total of 22 samples composed of water, ethanol, and 2-propanol were measured from 590 to 1091 at 1 nm intervals at 30, 40, 50, 60, and 70 °C.[35] Spectra from 850 to 1049 nm are used in this paper. Temperature-specific calibration and validation sets described in ref 35 were used to form the same respective 13- and 6- sample calibration and validation sets (pure component samples are excluded). Results are presented in this paper for primary calibration of ethanol at 30 °C to predict ethanol at 50 °C. To select the samples for $\mathbf{M}$, the Kennard Stone algorithm[36] 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

(31) Aster, R. C.; Borchers, B.; Thurber, C. H. *Parameter Estimation and Inverse Problems*; Elsevier: Amsterdam, 2005.
(32) Kunz, M. R.; Ottaway, J.; Kalivas, J. H.; Andries, E. *J. Chemom.*, in press.
(33) Du, Y. P.; Kasemsumran, S.; Maruo, K.; Nakagawa, T.; Ozaki, Y. *Anal. Sci.* **2005**, *21*, 979–984.
(34) Kalivas, J. H. *J. Chemom.* **2008**, *22*, 227–234.
(35) Wülfert, F.; Kok, T. W.; Smilde, A. K. *Anal. Chem.* **1998**, *70*, 1761–1767.
(36) Bouveresse, E.; Hartmann, C.; Massart, D. L.; Last, I. R.; Prebble, K. A. *Anal. Chem.* **1996**, *68*, 982–990.

samples are further from this sample. These sample sets are the same as those used in ref 17.

Corn. A total of 80 samples of corn were measured from 1100 to 2498 at 2 nm intervals on three NIR spectrometers designated m5, mp5, and mp6.[37] For this study, every other wavelength is used for a total of 350 wavelengths. Reference values are provided for oil, protein, starch, and moisture content, and protein is the prediction property studied in this paper. A total of 30 samples selected using a Kennard Stone algorithm on the 80 samples measured on instrument m5 serve as the primary calibration set. Validation samples are the remaining samples, but measured on instrument mp5. A total of 4 samples for **M** were selected by a Kennard Stone algorithm from the same 30 calibration samples, but measured on instrument mp5. These sample sets are the same as those used in ref 17.

Metaparameter λ, η, and τ Values. For TR 2−1 with the temperature data, 290 λ values are used ranging from 0 to 100 in equal linear increments. For the corn data, λ varied from 0 to 20 in equal linear increments over 100 values. For TR 1 with the temperature data, 100 η values are used ranging from 0 to 0.7 in exponential increments. For the corn data, values ranged from 0 to 5 in exponential increments over 100 values. Since LAR is being used, numerical values of τ are not involved, and instead, an LAR iteration (model) must be selected. Successive LAR models approach the least-squares solution and correlate to an increasing τ value. The η and λ values for TR 2 were those determined best in ref 17.

RESULTS AND DISCUSSION

Results for TR 2 have recently been published[17] and are compared to those developed for TR 2−1 and TR 1. Each TR approach is investigated using four samples in **M** and difference spectra of the four samples in **M**. It is possible to evaluate differences as the samples selected for **M** have actually been measured in both the primary and secondary conditions for both data sets. As previously noted, values for $\mathbf{y_M}$ are zero with difference spectra. It was observed that the best results for TR 2 were obtained with difference spectra as the unique differences between the primary and secondary conditions could be better characterized and weighted to form a model vector with greater orthogonality to the secondary condition.[17]

Calibration Maintenance. The temperature data set represents calibration maintenance as a primary model can be formed at one temperature (condition) and then updated to predict at a new temperature. The process to select TR 2−1 tuning parameters is first described followed by a description of the method used for TR 1. After selection of respective metaparameters, results are discussed for final selected models in conjunction with results from TR 2 and RR with no calibration maintenance.

TR 2−1. In determining acceptable metaparameters for TR 2−1 in expression 5, it is desirable to minimize both the RMSEC and RMSEM. The first metaparameter to determine is τ (LAR model). Shown in Figure 1a,b are L-curves for RMSEC (samples at 30 °C) and RMSEM (samples at 50 °C) with the model 1-norms where each L-curve represents a different λ value and each point on a curve is an LAR model. As λ increases in value (greater weight on predicting the samples in **M**) the L-curves move further

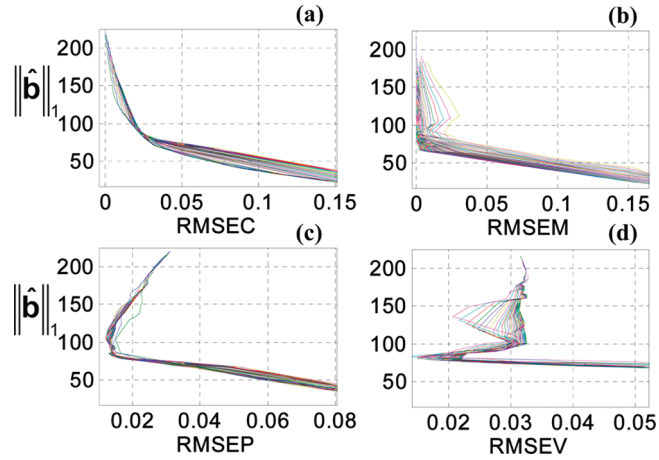

Figure 1. TR 2−1 temperature 1-norm L-curves where each curve represents a different λ value (increasing from right to left) and points on the curves correspond to different LAR models with model number (τ) increasing with the 1-norm: (a) RMSEC for the primary calibration samples at 30 °C, (b) RMSEM for the four **M** samples measured in the secondary conditions at 70 °C, (c) RMSEP for the validation samples measured in the primary condition at 30 °C, (d) RMSEV for the validation samples measured in the secondary conditions at 70 °C. Note that (d) would not be available for true new samples and is only shown to display how the RMSEV varies with the metaparameters.

to the left in Figure 1b, and hence, with each increase in λ, the RMSEM is reduced. From Figure 1a, it is observed that the models essentially converge at a certain 1-norm value. It turns out that this is also true with the RMSEM plot in that, as λ increases, the L-curves converge within a tolerance and the LAR models converge as well. Using the 1-norm of convergences from Figure 1a where the RMSEC and 1-norm are minimized as a guide (the corner region of the RMSEC L-curve with 1-norm at approximately 89), the LAR model number is selected from the RMSEM L-curve in Figure 1b as that furthest to the left. To achieve the final model (final λ) with the least amount of overfitting for a fixed τ, each model vector is analyzed over the full λ range at the selected LAR model number (τ). Plotted in Figure 2 are the model vectors, and the final λ is the smallest value where the model vectors first converge to the same shape. Beyond this point, the RMSEM slightly decreases while the 1-norm slightly increases.

Plotted in Figure 1d are the L-curves for the validation samples measured at 50 °C showing the models acceptable for predicting the primary calibration samples, and simultaneously the secondary samples in **M** are also applicable to predicting the new samples in the secondary condition. Similarly, using the six validation samples but measured in the primary condition to form the prediction errors (RMSEP) plotted in Figure 1d, these samples are also best predicted in the selected 1-norm region. This observation suggests that, if possible, samples measured at the primary condition can be left out of the TR model development and used to further assist model selection.

TR 1. Shown in Figure 3 are TR 1 L-curves with RMSEC and model 1-norms. Plotted are the last 5 LAR models formed (out of 11 total LAR models) as these were the leftmost (Pareto) curves. Opposite TR 2−1, each TR 1 L-curve corresponds to an LAR model (τ) and each point making up a curve is an η value. As the value of η increases, the 1-norm increases in value along a given LAR

(37) Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M. *PLS_Toolbox 3.0 for Use with MATLAB*; Eigenvector Research: Manson, WA, 2003.

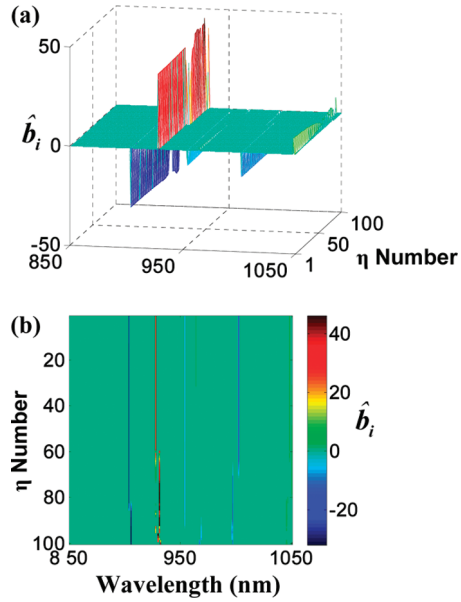

Figure 2. Temperature TR 2−1 model vectors with η values increasing in magnitude as the η number decreases in the plots. At approximately η number 60, model vectors have converged in shape, and beyond this point with decreasing η number (η value increasing in magnitude), only the 1-norm slightly changes.

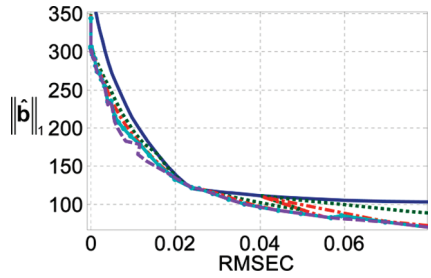

Figure 3. Temperature 1-norm L-curves for TR. Curves are the last five models formed by the LAR algorithm. Points on each are η values increasing in magnitude with an increase in 1-norm.

model L-curve. The sequence is to first determine which LAR model is appropriate and next which η value. The approach used here is to use the LAR L-curve that is most Pareto (closest to the origin) and use the η value in the corner of the L-curve. From Figure 3 it is seen that the L-curves intersect over the narrow 1-norm region, and this fact assists in determining the 1-norm and, hence, η value for the final model. The better LAR model is then the one that maintains the more consistent Pareto structure over the range of η values. While the RMSEM L-curve is useful for TR 2−1, these L-curves provided no new information or assistance in selecting the tuning parameters.

Model Evaluations. Listed in Table 1 for the final selected TR 2−1 and TR 1 models are model evaluation criteria composed of RMSEV, R^2, the 2- and 1-norms of the model vectors, and respective λ and η values. Also listed are the evaluation criteria for TR 2 and RR (without **M**) previously obtained.[17]

Since the objective of calibration maintenance is to predict in the new conditions with as few secondary samples spanning the secondary conditions as possible, four samples are used for **M**. This number was found to be reasonable in the previous work with TR 2.[17,32] With four samples, TR 2−1 provides the smallest

Table 1. Temperature Results

method	**M**	RMSEV	R^2	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_2$	η	λ
TR 2	spectra	0.0408	0.983	10.6	123.0	0.012	5.11
TR 2−1	spectra	0.0253	0.998	47.9	104.0		7.20
TR 1	spectra	0.0571	0.964	10.8	112.6	0.012	
TR 2	difference	0.0272	0.992	16.5	117.3	0.005	3.65
TR 2−1	difference	0.0222	0.994	54.3	107.4		48.0
TR 1	difference	0.0993	0.946	11.9	133.6	0.010	
RR with no **M**, cal at 30 °C		0.2158	0.934	8.79	100.2	0.020	
RR with no **M**, cal at 50 °C		0.0207	0.978	8.15	92.6	0.024	

RMSEV value compared to TR 2 and TR 1. Shown in Figure 4 are the respective model vectors. Notice that, even though a 1-norm is used with TR 1, a sparse model is not obtained. The improved result with TR 2−1 suggests that a sparse solution with selected wavelengths provides a better model-updating process. This agrees with literature that has shown wavelength selection alone can provide good quality calibration maintenance and transfer.[11−14] With TR 2−1, the advantages of TR 2 and wavelength selection are combined. All three TR variants improve on the prediction obtained by RR without calibration maintenance. Only TR 2−1 does as well as using an RR calibration model specifically formed for the secondary condition.

Using difference spectra in **M** with $\mathbf{y}_\mathbf{M}$ set to the zero vector, reductions of the RMSEV values are observed for TR 2 and TR 2−1, but the TR 1 RMSEV value increases. It was previously noted that difference spectra allow TR 2 to better determine a model vector with improved orthogonality to the new conditions because there is only nonanalyte information in **M**.[17] Apparently, this does not hold with TR 1.

Calibration Transfer. The corn data set represents a calibration transfer situation since the primary model is built on one instrument and then has to be updated to use on a new instrument, i.e., instrument m5 to instrument mp5. Table 2 list the results obtained by using the TR 2−1 and TR 1 protocols outlined for the temperature data.

When using spectra for **M**, TR 2−1 provides the best results followed by TR 1, with TR 2 performing the worst of the three. Nevertheless, TR 2 still performs substantially better than no calibration transfer at all and is comparable to using the RR model specifically formed for the new instrument with a full calibration. Interestingly, the RMSEV values for TR 2−1 and TR 1 improve beyond the RR RMSEV value for calibration on mp5 to predict

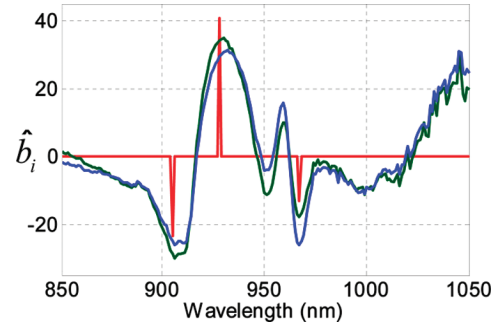

Figure 4. Temperature TR 2 (green), TR 1 (blue), and TR 2−1 (red) model vectors. TR 2 and TR 1 model vectors have been multiplied by 20 for plotting.

Table 2. Corn Results

method	**M**	RMSEV	R^2	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_2$	η	λ
TR 2	spectra	0.1881	0.836	57.5	834.0	0.013	7.15
TR 2−1	spectra	0.1261	0.945	391.2	1032		17.6
TR 1	spectra	0.1413	0.912	76.0	1069	0.006	
TR 2	difference	0.1744	0.866	57.1	846.0	0.014	14.0
TR 2−1	difference	0.1293	0.924	346.4	1054		19.0
TR 1	difference	0.2101	0.896	76.5	1061	0.006	
RR with no **M**, cal on m5		0.9551	0.882	58.5	858.5	0.012	
RR with no **M**, cal on mp5		0.1732	0.864	63.8	923.5	0.008	

validation samples measured on mp5. Plotted in Figure 5 are the respective model vectors with trends similar to those of the temperature data.

When difference spectra are used in **M**, TR 2−1 continues to provide a lower RMSEV value and has approximately the same RMSEV value as with using spectra. Similar to the temperature data, the RMSEV value for TR 1 increases, confirming that difference spectra are probably not useful with TR 1. As also observed with the temperature data, using difference spectra assists TR 2 to form a better predicting model, but for TR 2−1, no improvement is really gained. This is important because it implies that, for model maintenance or transfer, only a few spectra at the new conditions are needed and samples measured in the primary condition are not required.

CONCLUSION

This study demonstrates that TR 2−1 and TR 1 are viable approaches to accomplish calibration maintenance and transfer. It appears that because TR 2−1 also carry outs wavelength selection, it performed the best. While difference spectra assist TR 2 and degrade TR 1 results, the TR 2−1 approach generally did not improve nor degrade, providing further enhanced capabilities to complete calibration maintenance and transfer without requiring stable standardization samples that were measured under the original calibration primary conditions. It was also observed that if enough samples are available in the primary condition, then additional data splitting would allow generation of L-curves for prediction samples in the primary condition to assist in selecting metaparameters.

Developed in this paper is an empirical graphical protocol to determine model metaparameters whereby satisfactory predictions are subsequently obtained. The mechanics of metaparameter selection outlined are consistent for the two data sets studied. To make TR more adaptable, an algorithm is needed to automatically select appropriate metaparameters. While TR 1 provides results comparable comparable to those of TR 2−1, the TR 2−1 method is preferred due to the simpler and, hence, more specific L-curve procedure.

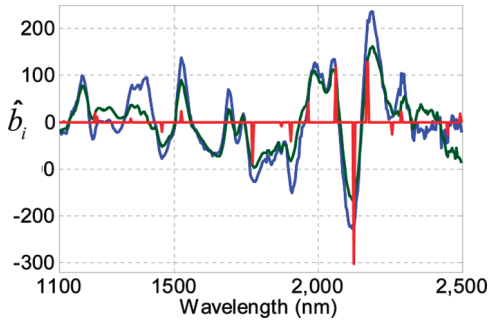

Figure 5. Corn TR 2 (green), TR 1 (blue), and TR 2−1 (red) model vectors. TR 2 and TR 1 model vectors have been multiplied by 20 for plotting.

In this study, samples for **M** were selected from a large set using the Kennard Stone algorithm to obtain a representative subset of samples spanning the secondary condition. Such a larger set of samples measured in the secondary condition is often not available, or else one might as well perform a full recalibration. Recent work using randomly selected samples for **M** has shown that TR 2−1, TR 1, and TR 2 are fairly robust to the actual composition of **M**.[32] In that work, it was also possible to use a consensus modeling approach with TR 2 and TR 2−1.

It should be noted that methods of wavelength selection are usually one of two modes. One is to select individual wavelengths, and the other is determining wavelength intervals (bands).[38] Using TR 2−1 falls into the first category, and evaluated in this paper is the notion of determining a model with a selected subset of wavelengths. There is probably no unique set of wavelengths that is best, and the positive results obtained using consensus modeling approaches for calibration maintenance and transfer with wavelength selection[32] warrant further investigation. The method of TR 2−1 is a simple process that appears to determine a suitable subset of wavelengths. As previously noted, the advantage of using TR 2−1 over a wavelength selection algorithm that requires iterative selection of a wavelength subset, model determination, and model and then evaluation is that the TR 2−1 model is determined and evaluated in conjunction with wavelength selection. Because of this more streamlined approach, less empirically determined algorithm parameters need to be set.

It is possible to form a more encompassing expression that contains all of the above expressions as well as RR, LASSO, and the elastic net, all under specific respective conditions, by writing the following expression:

$$\min(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \lambda^2\|\mathbf{Mb} - \mathbf{y_M}\|_2^2 + \tau\|\mathbf{Mb} - \mathbf{y_M}\|_1) \quad (11)$$

Expression 11 also contains a method known as adaptive LASSO when the second term is removed and in the third term $\mathbf{y_M} = \mathbf{0}$ and **M** is a diagonal matrix.[19,39] Such an approach has recently been studied for multivariate calibration with wavelength selection.[25] Application of expression 11 to calibration maintenance and transfer is not studied in this paper as a third tuning parameter on the model vector 2-norm would probably be required and, hence, complicate tuning parameter determinations. Additional penalty terms could be added as well, such as a model 1-norm penalty to force wavelength selection. With each new penalty in the minimization expression comes the addition of a new tuning parameter and determination of a proper value taking into account the trade-off with the other tuning parameters in the minimization expression.

(38) Brenchley, J. M.; Hörchner, U.; Kalivas, J. H. *Appl. Spectrosc.* **1997**, *51*, 689–699.
(39) Zou, H. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429.
(40) Stout, F.; Kalivas, J. H. *J. Chemom.* **2006**, *20*, 22–33.

Other variants of these expression are possible such as recent work using expression 3 with $\mathbf{M}$ = derivative operators and $\mathbf{y_M} = \mathbf{0}$ to form a smooth model vector $\mathbf{b}$.[40] Additional variations of TR have been reviewed.[17,21]

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CHE-0715179 (cofunded by the MPS Chemistry and DMS Statistics Divisions and by the MSPA Program), and we gratefully acknowledge this support. InLight Solutions, Albuquerque, NM, is also acknowledged and appreciated for partial funding.

Received for review December 17, 2009. Accepted March 11, 2010.

AC902881M