Special Issue Article

Journal of CHEMOMETRICS

Received: 20 November 2009, Revised: 19 January 2010, Accepted: 25 January 2010, Published online in Wiley InterScience: 2010

(www.interscience.wiley.com) DOI: 10.1002/cem.1302

Impact of standardization sample design on Tikhonov regularization variants for spectroscopic calibration maintenance and transfer

Matthew R. Kunz[a], Joshua Ottaway[a], John H. Kalivas[a]* and Erik Andries[b,c]

Multivariate spectroscopic calibration models are only valid to predict samples within the span of the calibration sample space measured relative to the current instrument environment (the primary conditions). Predicting samples in secondary conditions with new variances, same sample variances as the calibration space but a new instrument environment, or both, requires some form of continual model maintenance and/or transfer. Previous work has shown that a Tikhonov regularization (TR) approach is capable of accomplishing both tasks by updating the primary model based on only a few samples (transfer or standardization set) measured under the secondary conditions. A distinction of the TR design for calibration maintenance and transfer is a defined weighting scheme for the small set of standardization samples augmented to the full set of primary calibration samples. Critical to successful calibration maintenance or transfer is the standardization sample set composition, i.e. standardization samples should properly represent that are less secondary conditions. This paper reports on using TR-based methods to investigate this issue and a consensus modeling approach is briefly evaluated. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: multivariate calibration; calibration maintenance; calibration transfer; calibration standardization; Tikhonov regularization; augmentation; wavelength selection; consensus modeling

1. INTRODUCTION

Once a multivariate calibration model has been determined for a set of primary conditions, the applicability of the model over time becomes relevant. Primary conditions are the span of calibration samples (calibration space) and the span of instrumental conditions (wavelength accuracy, photometric response, bandwidths, etc.) and environmental conditions under which calibration samples are measured. Approaches are needed to maintain the primary model to predict under new chemical, physical, environmental, and/or instrumental effects (secondary conditions) not spanned in the primary calibration domain. Associated with calibration maintenance is the calibration transfer problem. In this case, the interest is using a calibration model developed on a primary instrument to predict sample compositions from spectra measured on a secondary instrument(s). The secondary instrument can be the primary instrument at a new point in time.

Calibration maintenance and transfer have been subjects of numerous studies and are well reviewed [1–5]. Three general approaches are possible. One consists of forming a robust primary model usually accomplished by using spectral pretreatment methods such as multiplicative scatter correction, finite impulse response filter, derivatives and/or wavelength selection. Another method to form a robust model is to globally calibrate by including all potential chemical, physical, environmental and/or instrumental variances in the original model through measuring spectra under all possible future conditions. However, a difficulty with this approach is the large number of samples needed to span all potential future variances and the corresponding analyte reference value must be determined for each sample. Obtaining reference values is usually time consuming and costly.

A second approach to calibration maintenance and transfer is adjusting sample spectra measured in the secondary condition or on a secondary instrument to fit the primary calibration model. Usually a small set of samples (standardization set) must be measured in the primary condition at the same time calibration samples are measured. The standardization samples must also be available for measuring in future secondary conditions. Correction terms are determined that transform standardization sample spectra measured in the secondary condition to appear as if the spectra were measured in the primary condition. New samples are then transformed with the correction terms and predicted by the primary calibration model. This approach to calibration maintenance and transfer is limited to conditions that alter spectra (wavelength shifts, intensity changes and/or baseline

* *Correspondence to: J. H. Kalivas, Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA.*
E-mail: kalijohn@isu.edu

a *M. R. Kunz, J. Ottaway, J. H. Kalivas*
Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA

b *E. Andries*
Department of Mathematics, Central New Mexico Community College, Albuquerque, NM 87106, USA

c *E. Andries*
Center for Advanced Research Computing, University of New Mexico, Albuquerque, NM 87106, USA

offsets) and are 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. As noted, limiting this spectral transformation method is the need to have a stable standardization set in order 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 [6]. Spectral differences between the actual secondary samples 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 general approach, and the focus of this paper, is to update (rebuild) the primary calibration model to properly predict sample spectra measured in a secondary condition or on a new instrument. Model updating is not only useful for situations requiring spectral adjustments, but it is also useful when the primary calibration fails due to uncalibrated spectral features appearing in new samples such as the analyte concentration being lower or higher than primary calibration concentrations or when 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 and pH.

Approaches to updating the primary model are varied [7–23]. The methods studied in this paper are three variants of Tikhonov regularization (TR) [24,25] used for calibration maintenance and transfer [26,27]. One of the TR variations is restricted to vector 2-norm minimizations [26] and the other two new adaptations of TR were recently developed to include 1-norm vector minimizations [27]. Results can improve with the 1-norm compared to those reported for TR in 2-norm. Basically, a small set of samples measured in the new secondary condition or instrument is augmented to the larger primary calibration sample set and optimally weighted to form a model desensitized to the new condition or instrument. Critical to forming a good model is the composition of the standardization samples measured in the secondary condition to augment the primary calibration samples. This paper reports on a study investigating the impact of the standardization sample set composition for the three forms of TR.

2. STANDARDIZATION SAMPLES

A problem with the method of transforming spectra measured in a secondary condition to the primary condition is that the small sample subset (standardization set) must be measured in both primary and secondary conditions and, most importantly, measured in the primary condition when the primary calibration samples were measured. Selecting the standardization set from a large number of primary calibration samples and then measuring these same samples under the secondary condition avoids the problem of going back and measuring any new standardization samples under the original primary condition which may no longer exist. However, these standardization samples need to be stable over time. Alternatives to selecting a standardization set include generic reference standards, sealed reference standards and glass standards [28–35]. Generally, it has been found that the more the standardization samples resemble new prediction samples (more prediction sample-like), the better the quality of the transformation process [29–33]. This observation is dependent on the sources of difference between primary and secondary conditions.

When a model is updated, as with TR, the standardization sample set only needs to be measured in the new secondary condition. If these standardization samples are actual samples and not some kind of reference sample such as a sealed ampoule, then by default, the standardization samples are prediction sample-like. Analyte reference values are still needed for these updating samples.

If a large number of samples are available in the secondary condition, the standardization set can be selected from these samples using the Kennard Stone algorithm (a commonly used algorithm where the goal is to select samples spanning the respective space) [36]. In this case, the standardization set should be representative of the new condition. However, the actual method used to select a standardization sample subset can affect the calibration transfer quality and, hence, prediction in the new conditions [11,37]. For example, the transformation quality of the piecewise direct standardization (PDS) algorithm was found to be sensitive to the method of subset selection, while prediction augmented classical least squares/partial least squares (PACLS/PLS) was not [37]. It may be that a consensus (bagging, ensemble, fusion, stacking) model approach [38–43] would provide a better updated model compared to updating a model with one subset, i.e., multiple updated models are formed with different standardization sets and some form of composite prediction is reported from the models deemed acceptable.

The luxury of having a large sample set measured in the secondary conditions to select from is usually not available and hence, the need for model updating procedures robust to the standardization set composition. Two examples of when a large set of samples in the secondary condition is not available are when only the first few samples of a new grade or batch are available or an existing primary model is needed to predict samples in a new growing region of an agriculture product that is geographically specific.

Evaluated in this paper are two promising variations of TR that include 1-norm minimization to ascertain if standardization sample set structure requirements can be relaxed and hence, more applicable to situations when a large set of samples measured in secondary conditions is not available. Because the evaluation is based on randomly selected standardization samples, also assessed is a consensus modeling approach with the multiple random standardization samples.

3. TIKHONOV REGULARIZATION MODIFICATIONS

Multivariate calibration for a primary condition involves relating the 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 are mean centered. For this paper, $\mathbf{y}$ contains analyte concentration information, $\mathbf{X}$ contains spectra measured over n wavelengths or frequencies and $n >> m$. The goal in the primary 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 $\hat{y} = \mathbf{x}^t\hat{\mathbf{b}}$ [44,45]. Methods such as ridge regression and PLS are commonly used and when $n \leq m$, multiple linear regression can also be used.

At some point in time, some form of calibration maintenance and/or transfer is needed. In the case of model updating, the objective is to augment the primary calibration $\mathbf{X}$ and $\mathbf{y}$ with only a few standardization samples containing the new variance and weighted by λ. In this case, Equation (1) is written as (ignoring the $\mathbf{e}$ term)

$$\begin{pmatrix} \mathbf{y} \\ \lambda\mathbf{y_L} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda\mathbf{L} \end{pmatrix}\mathbf{b} \tag{2}$$

where $\mathbf{L}$ represents an $l \times n$ matrix of spectra measured for l samples in the new condition or on the secondary instrument and $\mathbf{y_L}$ denotes respective concentrations.

Using the regression methods of PLS or principal component regression (PCR) to estimate $\mathbf{b}$ in Equation (2) requires determination of respective meta-parameters for the number of basis vectors (latent vectors, factors) and the weight value. Selection of a weight value in past work has been based on replication of samples in the standardization set [11,17]. 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. Recent work applied a method developed for TR based on the *L*-curve to determine acceptable PLS meta-parameters [26]. Presented in this paper is a different approach incorporating the *L*-curve to select meta-parameters. The Results and Discussion section describes the *L*-curves.

3.1. TR in 2-norm

An alternative to using PLS or another method to solve Equation (2) is to apply an algorithm to Equation (2) that satisfies

$$\min\left(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \lambda^2\|\mathbf{Lb} - \mathbf{y_L}\|_2^2\right) \tag{3}$$

where $\|\|_p$ signifies the vector p-norm, e.g. $p = 2$ is the 2-norm or Euclidean norm in Expression (3). This expression represents a TR variation [26]. The desired model vector in Equation (2) and Expression (3) needs to be as orthogonal as possible to non-analyte information in $\mathbf{X}$ as well as orthogonal to new chemical, physical, environmental and/or instrumental conditions characterized in $\mathbf{L}$. In this way, the model is desensitized to the interfering spectral artifacts. The goal is to try and do this with as few spectra as possible in $\mathbf{L}$.

It is not necessary to have standardization samples with the analyte present, e.g. matrix matched blanks, solvent, background, drift pseudo-spectra, or pure component spectra of new artifacts. In this case, $\mathbf{y_L} = \mathbf{0}$. Similarly, key eigenvectors from the singular value decomposition (SVD) of $\mathbf{L}$ with samples containing constant or no analyte, such as spectra from repeatedly measuring the spectrum of a single sample [15,16], could be used for $\mathbf{L}$ with $\mathbf{y_L} = \mathbf{0}$.

It should be noted that while solution of Equation (2) by PLS or PCR requires two meta-parameters, so does the TR form of Equation (2) and Expression (3) which must be modified to

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \\ \lambda\mathbf{y_L} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \alpha\mathbf{I} \\ \lambda\mathbf{L} \end{pmatrix}\mathbf{b} \tag{4}$$

and

$$\min\left(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \alpha^2\|\mathbf{b}\|_2^2 + \lambda^2\|\mathbf{Lb} - \mathbf{y_L}\|_2^2\right) \tag{5}$$

in order to provide stable solutions [26]. In Equation (4) and Expression (5), $\mathbf{I}$ denotes the identity matrix. The first two terms in Expression (5) formulate ridge regression and thus Expression (5) can be considered ridge regression with updating. Using Expression (5) will be henceforth termed TR in 2-norm or equivalently 2-norm TR.

3.2. TR in 1-norm

TR in Expression (3) can be modified to

$$\min\left(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \tau\|\mathbf{Eb} - \mathbf{y_E}\|_1^1\right) \tag{6}$$

where $\mathbf{E}$ represents a matrix that can be the standardization set spectra ($\mathbf{E} = \mathbf{L}$) and in this case, $\mathbf{y_E} = \mathbf{y_L}$, the 1 indicates the vector 1-norm and τ symbolizes the penalty meta-parameter on the model vector 1-norm. Using the 1-norm under certain conditions ($\mathbf{E} = \mathbf{I}$ or a diagonal matrix) causes wavelengths to be simultaneously selected as the model forms and hence, a sparse solution is obtained. When $\mathbf{E}$ is rectangular, then such a sparse solution is not guaranteed. The special case of TR in Expression (6) with $\mathbf{E} = \mathbf{I}$ and $\mathbf{y_E} = \mathbf{0}$ is more commonly known as the least absolute shrinkage and selection operator (LASSO) [45–48]. A version known as the adaptive LASSO results when $\mathbf{E}$ is a diagonal array and $\mathbf{y_E}$ is the zero vector [45,49]. The purpose of setting $\mathbf{E}$ to a diagonal array is to weight model coefficients by using *a priori* information such as a regression vector previously obtained by PLS. Such an approach was recently studied [50] and was found to provide improved predictions.

When $\mathbf{E}$ contains spectra for calibration maintenance and transfer, it is necessary to add a second regularizing meta-parameter α to Expression (6) [27]. This alters Expression (6) to

$$\min\left(\|\mathbf{Xb} - \mathbf{y}\|_2^2 + \alpha^2\|\mathbf{b}\|_2^2 + \tau\|\mathbf{Eb} - \mathbf{y_E}\|_1^1\right) \tag{7}$$

As with Expression (5), the first two terms in Expression (7) formulate ridge regression except now the model updating is with a 1-norm penalty instead of the 2-norm penalty. The minimization expressed in Expression (7) is studied in this paper using random standardization samples for $\mathbf{E}$. Because $\mathbf{E}$ is not diagonal, then a sparse solution is not guaranteed as shown in Reference [27] that also describes the algorithm. Hereafter, using Expression (7) shall be termed TR in 1-norm or equivalently, 1-norm TR.

3.3. TR in 2-/1-norm

The TR variants in Expressions (3) and (6) can be combined to form

$$\min\left(\|\mathbf{Xb}-\mathbf{y}\|_2^2+\lambda^2\|\mathbf{Lb}-\mathbf{y_L}\|_2^2+\tau\|\mathbf{Eb}-\mathbf{y_E}\|_1^1\right) \quad (8)$$

Note that Expression (7) with $\mathbf{L}=\mathbf{I}$ and $\mathbf{y_L}=\mathbf{0}$ is a special case of Expression (8). A modification of Expression (8) used in the previous work for calibration maintenance and transfer [27] and used in this paper is with $\mathbf{E}=\mathbf{I}$ and $\mathbf{y_E}=\mathbf{0}$ giving

$$\min\left(\|\mathbf{Xb}-\mathbf{y}\|_2^2+\lambda^2\|\mathbf{Lb}-\mathbf{y_L}\|_2^2+\tau\|\mathbf{b}\|_1^1\right) \quad (9)$$

In this 1-norm situation, a sparse solution is guaranteed. Henceforth, Expression (9) shall be referred to as TR in 2-/1-norm or, equivalently, 2-/1-norm TR.

Another modification to Expression (8) that is similar to Expression (9) is known as the elastic net [45,51] where $\mathbf{L}=\mathbf{E}=\mathbf{I}$ and $\mathbf{y_L}=\mathbf{y_E}=\mathbf{0}$ in Expression (8). Not studied to date is Expression (8) as a third meta-parameter, such as α, would probably be needed.

3.4. Previous TR calibration and transfer work

Using Expression (9) for calibration maintenance and transfer with a Kennard Stone selection of the standardization set provided improved predictions for the secondary and primary conditions compared to using TR in Expression (5) where no wavelengths are selected [26,27]. Previous work also evaluated Expression (7) for calibration maintenance and transfer, and improved results were also obtained compared to results from Expression (5) [27]. Presented in this paper are results from a study using Expressions (5), (7) and (9) for calibration maintenance and transfer to ascertain if the 1-norm improves the ability of TR to be less sensitive to the composition of **L** and **E** and hence, more applicable to those situations when the user does not have the luxury to select the standardization subset for **L** and **E** from a larger preexisting set where the Kennard Stone approach could be used to ensure an **L** and **E** representative of the secondary condition.

4. CONSENSUS MODELING

If a large number of samples have been measured in the secondary conditions, then there exists the flexibility of selecting standardization samples with an algorithm such as the Kennard Stone. Alternatively, the process of randomly selecting standardization samples followed by forming the respective updated model could be repeated to form numerous models. Each new sample would be predicted with the collection of models and a composite (fused) prediction would be formed. Such approaches are known as consensus or ensemble modeling. This is different from that performed in Reference [52] where multiple models were formed and each was tested for ruggedness using artificial spectral perturbations and the most rugged model was selected as best for predicting future samples.

The usual format to form multiple models is by random sampling across samples (bagging), variables (random subspace method) or both [53–55]. Once a set of models is formed, filtering for model quality is needed. To effectively leverage consensus modeling, a high degree of prediction accuracy is needed in conjunction with small but significant divergences between models (model diversity), i.e. there are no gains from consensus modeling if all the models are similar (same model vector $\hat{\mathbf{b}}$) and hence, give the same prediction [56,57]. The gain comes when individual models are different and hence, provide different but accurate predictions. Reference [58] describes a study on whether it is best to combine model predictions or select one model for the sample prediction. Based on the prediction combination process and model selection criterion, it was found that combining predictions from different models is not necessarily the most accurate prediction, but that it is less risky to combine predictions than to select one model for the prediction.

In recent work it was shown that indeed, for a given data set, different regression vectors can produce acceptable predictions and hence, interpretations of regression vector shape in terms of good or bad (model validation) are difficult if not impossible [59]. A distance measure has been developed to measure the similarity and diversity of models thereby providing a mechanism to cluster similar models [60].

Various methods exist to form the composite prediction from the acceptable models including weighted linear combinations, majority vote and fuzzy logic. In Reference [61], the final prediction is a weighted mean of the predictions where weights are determined by prediction errors of neighborhood samples. Neighborhood samples are those similar to the new sample where similarity can be spectroscopic similarity. Based on prediction errors of neighborhood samples, a final prediction correction error is made to the new sample. Alternatively, an average of true reference values for neighborhood samples can be used as the predicted value of the new sample. Such an approach is part of the comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D) method [62]. The simple approach of using the mean predictions from all models without any determination of predictability of the models is used in this paper.

5. EXPERIMENTAL

5.1. Software

MatLab 7 (The MathWorks, Natick, MA) programs for 2-norm TR were written by the authors and 1-norm and 2-/1-norm TRs are that previously described [27]. The 1-norm and 2-/1-norm TR algorithms are based on the least angle regression (LAR) algorithm [45,63] and the reader is referred to References [27] and [50] for details.

5.2. Data centering

Matrices **X**, **y**, **L**, **E**, $\mathbf{y_L}$ and $\mathbf{y_E}$ are mean centered to respective means. This local centering approach has been shown to provide improved modeling performance [64,65]. Validation samples measured under the same conditions as **L** or **E** are centered to the mean of **L** or **E** prior to prediction.

5.3. Data sets

5.3.1. 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 [66]. Spectra from 850 to 1049 nm were used. Temperature-specific calibration and validation sets described in

Reference [66] 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 70°C. In a separate study to exemplify the new process of determining meta-parameters α and λ in Expression (5), the same temperature conditions as used in Reference [26] were used here, namely primary calibration at 30°C for ethanol to predict ethanol at 50°C.

5.3.2. Corn

Eighty samples of corn were measured from 1100 to 2498 nm at 2 nm intervals on three near infrared (NIR) spectrometers designated as m5, mp5 and mp6 [67]. 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 were selected from the 80 samples measured on instrument m5 for the primary calibration set using the Kennard Stone algorithm selecting the first sample closest to the mean and successive samples are furthest from this sample. The validation samples are the remaining samples but measured on instrument mp5.

5.3.3. Standardization set

For both data sets, one of the nine sample standardization sets studied is that determined by Kennard Stone algorithm. For the temperature data, this consisted of selecting four samples by applying the Kennard Stone algorithm to the 13 sample calibration set at 30°C, but measured at 70°C. For the corn data set, the standardization set consists of four samples selected from the same 30 calibration samples but measured on instrument mp5. Eight other randomly selected standardization sample sets with four samples were used. For the two data sets, this consisted selecting four samples from the respective calibration sample number but measured in the secondary conditions. The same sets of random four samples were used for TR in Expressions (5), (7) and (9), i.e. $\mathbf{L} = \mathbf{E}$. In other work [26,27], four samples were found to be adequate for the temperature and corn data sets.

5.3.4. Model evaluation

Model evaluation criteria tabulated in Tables include the root mean square error of validation (RMSEV), R^2, and the 2- and 1-norms of the model vectors. Reported for the all nine models formed on each data set using each of the three methods are respective means, standard deviations and relative standard deviation as a per cent (RSD).

6. RESULTS AND DISCUSSION

This paper investigates nine spectral situations for the standardization set in **L** or **E**. One consists of spectra of samples selected by the Kennard Stone algorithm and the other eight are randomly selected. These nine standardization sets are used to compare the subset composition impact on TR in 2-norm, 1-norm and 2-/1-norm.

Final TR models for Expressions (5), (7) and (9) are based on two meta-parameters. The τ parameter for TR in 1-norm and 2-/1-norm are used to penalize the 1-norm prediction error of the standardization set or guide wavelength selection, respectively. Note that the LAR algorithm used in 1-norm and 2-/1-norm does not directly use a τ value. Instead, each iteration corresponds to a non-zero model vector coefficient in **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 (7) and (9) with the modification that the columns of **X** are not scaled to mean zero and standard deviation one (autoscaled) as originally suggested for LAR [45,63]. Details of the exact LAR algorithms implemented are provided in References [27] and [50]. The meta-parameter λ for TR in 2-norm and in 2-/1-norm weights the quality of prediction in the secondary conditions. The α meta-parameter in 2-norm TR and 1-norm TR weights the size of the model vector and is necessary to stabilize the algorithm as with ridge regression. Restating, 2-norm TR is ridge regression with a third penalty on the predicting standardization set in 2-norm; 1-norm TR is ridge regression with a third penalty on predicting the standardization set in 1-norm; and 2-/1-norm TR is LASSO with a third penalty on predicting the standardizations set in 2-norm.

Determining optimal meta-parameters is based on evaluating tradeoffs in *L*-curves as dictated in Expressions (5), (7) and (9). The *L*-curve approach was used in the previous 2-norm, 1-norm and 2-/1-norm TR calibration maintenance and transfer studies as well other studies selecting wavelengths, number of basis vectors for PLS and PCR, and the ridge parameter in ridge regression [27,50,68,69]. Briefly, for a range of meta-parameter combinations, a model prediction variance indicator (such as the 2- or 1-norm of the model vector) is plotted against a bias indicator such as the prediction error for the calibration set, standardization set, a monitoring set and/or the corresponding R^2 values [24,25,69,70]. In these plots, an *L* shaped curve is formed and the better model is assumed to reside in the corner of the *L*-curve with an acceptable tradeoff in plotted criteria. Sometimes a log–log plot is used to accentuate the *L* curve shape. Respective *L*-curve processes are described in the following section and used throughout the study. For the corresponding three TR cases, α and λ values were varied to generate reasonable *L*-curves. The maximum model number for the 1-norm and 2-/1-norm is a default of the LAR algorithm.

6.1. Calibration maintenance

The temperature data represent the situation where a primary model is formed under one condition and has to be updated to handle a new condition. For the nine standardizations sets, results are first presented for 2-norm TR followed by 1-norm TR and then 2-/1-norm TR. Following this is the result of consensus modeling.

6.1.1. TR in 2-norm

As previously noted, one goal of this paper is to present an easier and more clear approach to select an appropriate α value in Expression (5). Thus, for equivalency with Reference [26], calibrating at 30°C and predicting at the new temperature 50°C is first described. Plotted in Figure 1 are prediction error criteria root mean square error of calibration (RMSEC), the RMSE of the standardization set (RMSEL), and for follow-up analysis, the RMSE of validation (RMSEV) trends. From Figures 1a and b, $\alpha = 0.0012$ in the regions of common agreement is selected. Inspecting these plots for common agreement in α between

RMSEC and RMSEL is much simpler than the process described in Reference [26] to determine α. As expected, the validation set also requires this α for optimal prediction as the standardization set was selected by the Kennard Stone algorithm and appropriately spans the new temperature. Note that in the *L*-curves plotted in Figure 1d, the area of agreement in the variance/bias tradeoff is further established and also assists in selecting the α value. Once the α value is established, the λ value must be determined. The same process used in Reference [26] is used here where the tradeoff between RMSEC and RMSEL is ascertained.

This procedure just outlined for calibrating at 30°C and predicting at 50°C was used to determine α for predicting at 70°C with the Kennard Stone selected standardization set. Values in Table I show that some of the randomly selected **L**s provide slightly improved results over the Kennard Stone selected samples while others are slightly worse. This seems reasonable as the secondary calibration space is small and with randomly selected samples, it is not difficult to span or nearly-span the secondary condition. The correlations between the nine model vectors range from 0.889 to 0.998 demonstrating little differences and hence, similar predictions. All nine models perform better than using the primary calibration model at 30°C without updating. While biases are larger than using a 2-norm TR model using calibration samples at 70°C, the biases are an order of magnitude smaller than that with no updating.

6.1.2. *TR in 1-norm*

As outlined in Reference [27] for TR in 1-norm (Expression 7), the first meta-parameter to determine is τ (or actually, LAR model number) from the *L*-curves formed by plotting model 1-norm against the RMSEC for each LAR model over the α values, i.e. points on each LAR model *L*-curve are α values and the most Pareto *L*-curve (LAR model) is selected. Respective *L*-curves with RMSEL are not informative and do not assist in selecting the meta-parameter. After the LAR model selection, α is determined from the corner region of the *L*-curve.

Results shown in Table II reveal that for the temperature data, the prediction error using the Kennard Stone selected samples is slightly better than the TR in 2-norm model in Table I with the

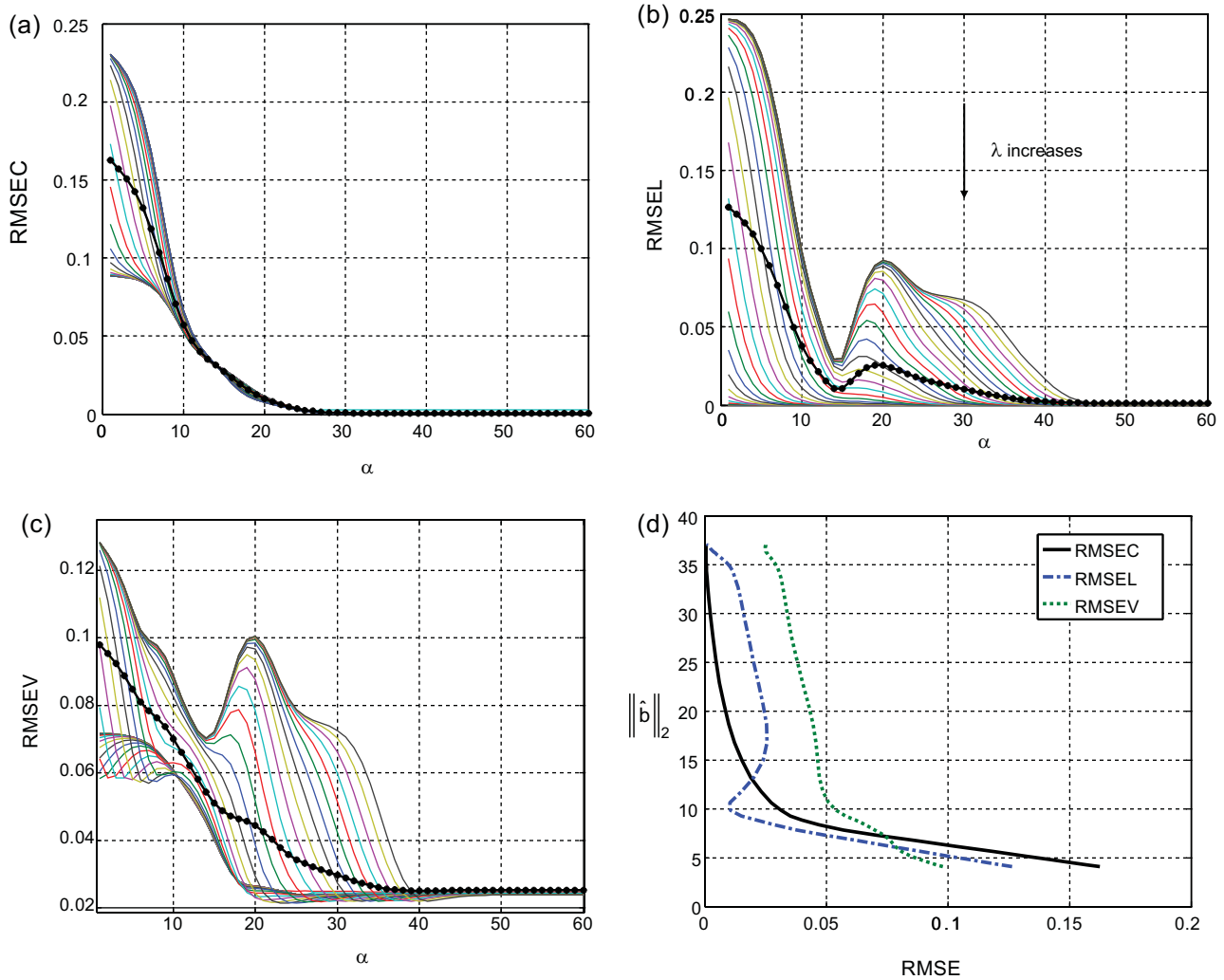

Figure 1. Temperature data for calibrating at 30°C and predicting at 50°C for TR in 2-norm prediction error criteria and mean *L*-curve. For mean plots, closed circle line in (a), (b) and (c), the mean is taken across λ values (50 squared values ranging from 1×10^{-5} to 1×10^{9}) for each α value (60 squared values ranging from 1×10^{-15} to 0.5). Standardization set is that selected by Kennard Stone algorithm and the simplicity of plot can be compared to that in Reference [26]. Note consistency in agreement of α for the calibration, standardization and validation samples a, b and c, respectively. Shown in (d) are the same mean curves in (a), (b) and (c) plotted against the model 2-norm.

Table I. TR in 2-norm, Expression (5), temperature data prediction results for 70°C using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias[a]	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_1$	α	λ
KS	0.0599	0.868	−0.0339	10.28	123.0	0.0117	0.1863
Set 1	0.0581	0.918	−0.0519	9.24	104.0	0.0140	0.4603
Set 2	0.0988	0.948	0.0964	9.21	112.6	0.0140	0.4603
Set 3	0.0616	0.860	−0.0347	10.23	122.0	0.0140	0.1271
Set 4	0.0534	0.855	0.0117	10.44	126.2	0.0117	0.0610
Set 5	0.0392	0.920	0.0046	10.00	122.7	0.0140	0.5532
Set 6	0.0463	0.895	−0.0132	10.02	120.1	0.0140	0.1863
Set 7	0.0323	0.951	0.0110	9.35	109.1	0.0168	0.0880
Set 8	0.0340	0.965	−0.0200	9.42	108.4	0.0140	0.3830
Mean[b]	0.0537	0.909	0.0308	9.80	116.4	0.0138	0.2784
Std Dev[b]	0.0202	0.041	0.0288	0.49	8.0	0.0015	0.1857
RSD %[c]	37.60	4.57	93.49	5.00	6.86	10.89	66.72
Cal at 30°C	0.2956	0.932	−0.2782	9.995	114.2	0.0140	—
Cal at 70°C	0.0357	0.938	−0.0038	9.328	97.0	0.0168	—

[a] $\text{Bias} = \sum_{i=1}^{m} (\hat{y}_i - y_i) \Big/ m.$

[b] Mean and standard deviation values are based on absolute values.

[c] RSD %: relative standard deviation per cent = (Std Dev/Mean) × 100.

Kennard Stone selected samples. Random selected samples for **E** form some models with slightly improved predictions as well as some models with slightly degraded predictions; similar to the 2-norm TR models. As characterized by the relative standard deviation RSD, the degree of variation in the RMSEV values is less with 1-norm TR compared to 2-norm TR while the R^2 variation is slightly better suggesting that TR in 1-norm is more robust to the standardization set composition. That is, regardless of the standardization set composition, the correlation between the 1-norm TR models ranges from 0.917 to 0.993, which is a smaller range than the 2-norm TR models. Because of this tighter correlation range, the models are more similar and hence, similar predictions. All nine models perform better than primary calibration at 30°C without any updating.

6.1.3. *TR in 2-/1-norm*

The first task with TR in 2-/1-norm is to determine a good τ value for Expression (9) (LAR model) followed by determining λ. The approach is described in Reference [27]. Briefly, *L*-curves are formed by plotting model 1-norm against RMSEC and RMSEL for each λ, i.e. points on each *L*-curve for a particular λ are the LAR models (τ values). Once the LAR model is

Table II. TR in 1-norm, Expression (7), temperature data prediction results for 70°C using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_1$	α
KS	0.0537	0.889	−0.0289	10.66	120.7	0.0117
Set 1	0.0550	0.962	−0.0461	11.22	118.1	0.0117
Set 2	0.0587	0.887	−0.0367	12.34	128.2	0.0081
Set 3	0.0535	0.883	−0.0266	11.05	120.6	0.0117
Set 4	0.0736	0.851	0.0488	13.54	135.2	0.0097
Set 5	0.0342	0.963	0.0220	9.28	105.8	0.0168
Set 6	0.0371	0.927	−0.0023	11.10	121.1	0.0097
Set 7	0.0445	0.908	−0.0146	13.03	139.8	0.0067
Set 8	0.0557	0.945	−0.0425	11.28	119.6	0.0140
Mean	0.0518	0.913	0.0298	11.50	123.2	0.0111
Std Dev	0.0119	0.039	0.0154	1.29	10.0	0.0030
RSD %	22.99	4.27	51.59	11.22	8.13	27.39

Table III. TR in 2-/1-norm, Expression (9), temperature data prediction results for 70°C using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_1$	λ
KS	0.0361	0.979	0.0034	48.95	77.45	6.100
Set 1	0.1031	0.938	−0.0927	37.98	85.04	67.00
Set 2	0.0317	0.960	0.0093	49.26	78.87	24.00
Set 3	0.0446	0.925	−0.0234	53.19	89.63	13.00
Set 4	0.0356	0.964	0.0208	43.23	95.80	70.00
Set 5	0.0270	0.966	−0.0090	48.09	116.5	46.00
Set 6	0.0227	0.986	0.0067	50.18	92.38	21.00
Set 7	0.0481	0.886	−0.0129	46.85	86.64	48.00
Set 8	0.0554	0.958	−0.0437	40.06	85.63	60.00
Mean	0.0449	0.951	0.0247	46.21	89.77	39.46
Std Dev	0.0241	0.031	0.0283	4.99	11.61	24.04
RSD %	53.72	3.24	114.7	10.75	12.93	60.92

determined from the corner region, λ is selected at the point of model vector convergence across all LAR models in the corner region. Using this approach forms the model evaluation criteria tabulated in Table III. The listed values show that except for the first random set (Set 1), TR in 2-/1-norm is also essentially unaffected by the composition of the standardization set **L**. The correlations between the nine models range from -6.74×10^{-5} to 0.931 indicating the lack of total agreement in selected wavelengths between each model. Shown in Figure 2 are the nine model vectors demonstrating the variances in selected wavelengths. Apparently, several wavelengths sets can be selected with small variations and still provide acceptable updated models.

In comparison to 2-norm TR and 1-norm TR, the 2-/1-norm TR approach provides several models with lower prediction errors and performs best with the Kennard Stone selection. As with the other two TR modes, TR in 2-/1-norm substantially improves over primary calibration at 30°C without any calibration maintenance.

6.1.4. *Consensus modeling*

The simple approach of using the mean prediction from all nine models without any prior model quality assessment is used in this paper with the results presented Tables I–III. Of the three TR approaches, 2-/1-norm TR does best with respect to mean RMSEV and R^2 values. Due to the larger prediction error for one of the random samples with 2-/1-norm TR, the approach of TR in 1-norm has smaller RSD values implying better consistency in predictions. However, an approach that could be taken (not studied in this paper) is to use an *a priori* cutoff of the similarities between models such as the reported correlations. For example, if all

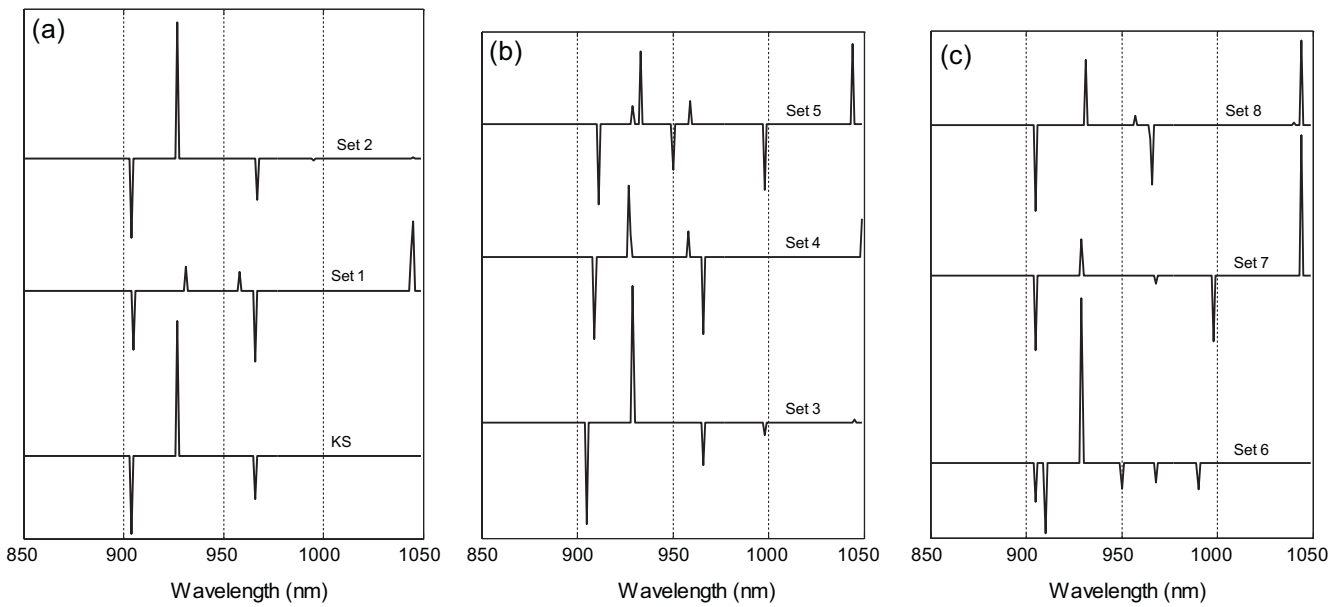

Figure 2. Nine model vectors for temperature data for calibrating at 30°C and predicting at 70°C using TR in 2-/1-norm. Standardization sets correspond to (a) Kennard Stone (KS), Sets 1 and 2; (b) Sets 3, 4 and 5; (c) Sets 6, 7 and 8. Model vectors are directly comparable as only an offset has been added for clarity and hence, there is no *y*-axis label.

intermodal correlations are greater than some cutoff, then essentially, there is not enough diversity in the models and no gains will be obtained from using a consensus approach, i.e. there is no risk in selecting one model over another model. This observation would suggest that the 1-norm TR may not be appropriate for consensus modeling due to the high degree of model similarity as noted by the inter-correlations reported in Section 6.1.2. Conversely, there is a large degree of diversity between the 2-/1-norm TR models (see inter-correlations in Section 6.1.3) and simultaneously, good accuracy, the two requirements needed for effective consensus modeling.

6.2. Calibration transfer

The corn data represent the situation where a primary model is formed on one instrument (m5) and has to be updated to handle a new instrument (mp5).

6.2.1. *TR in 2-norm*

With the corn data, results listed in Table IV demonstrate that similar to temperature data, 2-norm TR is able to form acceptable models with the random samples. Model vectors range in correlation from −0.561 to 0.701.

6.2.2. *TR in 1-norm*

Using the same random samples as with 2-norm TR, Table V shows that switching the 2-norm penalty in Expression (5) with the 1-norm penalty in Expression (7) also generates good predictive models. Additionally, the 1-norm TR models are much more consistent in shape as the correlation between models ranges from 0.875 to 1.00. As with temperature data, the model vectors are not sparse (no specific wavelengths are selected). The α meta-parameter is also much less varied than 2-norm TR. For 1-norm TR, the Kennard Stone selected standardization set is not the best and all nine standardization samples are better than no calibration transfer. With the temperature data, only 2-/1-norm TR was able to form a model with better predictive performance than calibration at the primary condition to predict at the same primary condition. With the corn data, all nine of the 1-norm TR models outperform calibration on the mp5 instrument to predict validation samples measured on the same mp5 instrument.

6.2.3. *TR in 2-/1-norm*

Results listed in Table VI show that the 2-/1-norm TR functions similar to 1-norm TR. One exception is that while 1-norm TR is consistent in the prediction errors due to model similarities, 2-/1-norm has one standardization set (Set 8) that does not form as good as a predicting model relative to the other eight. However, this model essentially predicts equivalently to the primary model built for mp5 to predict the secondary samples also measured on instrument mp5 (see Table IV for values). Except for this one standardization set, most of the 2-/1-norm TR models outperform the 1-norm TR models, including the Kennard Stone selected samples.

Similar to the temperature data, the 2-/1-norm TR models are varied in the wavelengths selected. The correlations between the models range from -5.50×10^{-5} to 0.930. Thus, as observed for the temperature data, several wavelength combinations are possible that allow model updating from one instrument to another.

6.2.4. *Consensus modeling*

Consensuses modeling approaches for the three TR variants operate equivalently in spite of the one poor 2-/1-norm TR model. As with the temperature data, the corn 1-norm TR models show poor diversity and hence, 1-norm TR is again probably not useful for consensus modeling. It appears that if a large set of samples is in fact available for selecting the standardization set from, using consensus modeling approach with 2-/1-norm TR can reduce the prediction risk compared to trying to select one updated model.

7. CONCLUSION

This study has shown that the three TR variants are not significantly sensitive to the standardization sample subset

Table IV. TR in 2-norm, Expression (5), corn data prediction results for mp5 using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias	$\|\hat{\mathbf{b}}\|_2$	$\|\hat{\mathbf{b}}\|_1$	α	λ
KS	0.1520	0.896	0.0267	88.95	1242	0.0057	13.34
Set 1	0.1865	0.888	0.1035	105.8	1455	0.0057	51.09
Set 2	0.1923	0.889	0.1090	79.03	1137	0.0101	3.48
Set 3	0.1390	0.920	0.0440	103.6	1546	0.0032	3.48
Set 4	0.1371	0.913	−0.0010	101.0	1445	0.0032	3.48
Set 5	0.1870	0.895	0.1093	95.29	1406	0.0057	6.81
Set 6	0.1463	0.901	0.0040	87.22	1250	0.0057	13.33
Set 7	0.1845	0.861	−0.0556	67.32	1003	0.0101	6.81
Set 8	0.1844	0.847	−0.0220	121.4	1781	0.0057	6.81
Mean	0.1677	0.890	0.0528	94.40	1363	0.0061	12.07
Std Dev	0.0233	0.0232	0.0443	15.96	232.7	0.0025	0.1513
RSD %	13.91	2.61	83.92	16.91	17.08	40.68	125.31
Cal on m5	0.7396	0.773	−1.6776	37.98	594.9	0.0265	—
Cal on mp5	0.1971	0.824	−0.0446	46.71	529.9	0.0144	—

Table V. TR in 1-norm, Expression (7), corn data prediction results for mp5 using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_1$	α
KS	0.1413	0.912	0.0238	75.96	1069	0.0064
Set 1	0.1691	0.913	0.0972	71.92	1010	0.0078
Set 2	0.1282	0.926	−0.0091	103.0	1517	0.0028
Set 3	0.1490	0.905	0.0357	70.39	1004	0.0078
Set 4	0.1528	0.895	−0.0162	64.39	927.3	0.0096
Set 5	0.1514	0.922	0.0757	88.55	1284	0.0043
Set 6	0.1617	0.882	−0.0071	58.47	858.5	0.0118
Set 7	0.1311	0.912	−0.0100	76.38	1090	0.0064
Set 8	0.1408	0.912	0.0231	70.39	1090	0.0064
Mean	0.1472	0.909	0.0331	75.49	1094	0.0070
Std Dev	0.0134	0.013	0.032	13.25	198	0.0026
RSD %	9.10	1.47	96.65	17.56	18.10	37.98

Table VI. TR in 2-/1-norm, Expression (9), corn data prediction results for mp5 using the Kennard Stone (KS) selection and eight random selections

Method	RMSEV	R^2	Bias	$\|\|\hat{\mathbf{b}}\|\|_2$	$\|\|\hat{\mathbf{b}}\|\|_1$	λ
KS	0.1261	0.936	0.0459	391.2	1032	17.59
Set 1	0.1483	0.925	0.0761	439.0	1049	12.76
Set 2	0.1598	0.926	0.0909	434.5	1080	17.59
Set 3	0.1186	0.936	0.0155	384.8	933.3	13.57
Set 4	0.1345	0.916	−0.0005	395.0	909.8	16.58
Set 5	0.1564	0.924	0.0891	412.0	1066	9.150
Set 6	0.1218	0.931	0.0043	466.2	1104	9.350
Set 7	0.1252	0.931	−0.0218	419.9	1053	19.60
Set 8	0.2078	0.817	−0.0509	312.5	921.5	15.38
Mean	0.1443	0.916	0.0439	406.1	1016	14.62
Std Dev	0.0283	0.038	0.0355	43.7	74.2	3.69
RSD %	19.59	4.12	80.91	10.76	7.30	25.26

composition. Prediction results and models were comparatively consistent between the Kennard Stone selected samples and random samples. Even in the case with the largest prediction error for the corn validation set, the R^2 and bias values are suitable. This work also reported on an easier methodology to select an α and λ for TR in 2-norm in Expression (5) compared to that presented in a previous publication [26].

For the data sets studied, once a 'golden' set of primary calibration samples has been measured, it is relatively easy to update the golden set to the new condition. Expected fundamental variances can be spanned in the primary golden set and any unexpected new variances can be included with only a few new samples. For example, under current investigation is using virgin olive oil samples from one geographic region to form a golden model for predicting per cent adulteration and then applying 1-norm TR and 2-/1-norm TR to form an updated model to predict samples for a different geographic region.

In this study, standardization samples are from the prediction sample space of the secondary condition and instrument. While literature suggests prediction sample-like samples perform better compared to generic, sealed reference samples, or glass standards for the standardization set [29–33], it may be that this observation is less critical when TR is used. Studies of this nature are currently under investigation.

An approach that may also prove to be robust to which samples make up the standardization set is to weight each standardization sample relative to a similarity measure to the particular current prediction sample, i.e. a different set of weights, and hence model, for each prediction sample. With each sample *a priori* weighted, then Expression (7) or Expressions (9) would be used. Weighting standardization samples in calibration transfer and maintenance has proven effective in other studies [71,72].

A consensus modeling approach was also briefly evaluated and it was found that acceptable results could be formed from averaging respective predictions. In the truest form of consensus modeling desiring model prediction accuracy and simultaneously, diversity in model structure, the 2-/1-norm best meets these goals. However, using a consensus modeling approach for multivariate calibration maintenance and transfer would require

a pool of sample spectra measured in the secondary condition to select from.

From the 2-/1-norm TR results, it appears that different sets of selected wavelengths provide acceptable prediction results. Methods of wavelength selection are usually one of the two modes. One is to select individual wavelengths (as with 2-/1-norm TR) and the other is determining wavelength intervals (bands) [73]. There is probably no unique best set of wavelengths and the method of TR 2-/1-norm appears to determine a suitable subset of wavelengths. It could be possible to use TR in 2-/1-norm for wavelength intervals by using a bootstrapping approach that randomly selects samples for **X** and a histogram of preferred selected wavelengths (acceptable RMSEC and RMSEL) is used with an appropriate frequency cutoff to determine wavelength intervals instead of a mean (consensus) prediction from the collection of models or a mean model vector from the collection of model vectors. Alternatively, wavelengths could be randomly selected and TR in 2-norm would be used to form models and the resultant histogram of preferred wavelengths would be used to determine wavelength bands [73].

Acknowledgements

This study is based upon work supported by the National Science Foundation under Grant No. CHE-0715179 (co-funded by the MPS Chemistry and DMS Statistics Divisions, and by the MSPA Program) and is gratefully acknowledged by the authors.

REFERENCES

1. de Noord OE. Multivariate calibration standardization. *Chemom. Intell. Lab. Syst.* 1994; **25**: 85–97.
2. Fearn T. Standardization and calibration transfer for near infrared instruments: a review. *J. Near Infrared Spectrosc.* 2001; **9**: 229–244.
3. Feudale RN, Woody NA, Tan H, Myles AJ, Brown SD, Ferré J. Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.* 2002; **64**: 181–192.
4. Cogdill RP, Anderson CA, Drennen JK, III. Process analytical technology case study, part III: calibration monitoring and transfer. *AAPS Pharm. Sci. Tech.* 2005; **6**: E284–E297.
5. Brown SD. Transfer of multivariate calibration models. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis Vol. 3*, Brown SD, Tauler R, Walczak B (eds-in-chief). Kalivas J, (section ed.). Elsevier: Amsterdam, 2009; 345–377.
6. Zeaiter M, Roger JM, Bellon-Maurel V. Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. *Chemom. Intell. Lab. Syst.* 2006; **80**: 227–235.
7. Haaland DM. Synthetic multivariate models to accommodate unmodeled interfering spectral components during quantitative spectral analysis. *Appl. Spectrosc.* 2000; **54**: 246–254.
8. Maruo K, Oota T, Tsurugi M, Nakagawa T, Arimoto H, Hayakawa M, Tamura M, Ozaki Y, Yamada Y. Noninvasive near-infrared blood glucose monitoring using a calibration model built by a numerical simulation method: trial application to patients in an intensive care unit. *Appl. Spectrosc.* 2006; **60**: 1423–1431.
9. Westerhaus MO. Improving repeatability of NIR calibrations across instruments. In *Proceedings of the Third International Near Infrared Spectroscopy Conference*, Biston R, Bartiaux-Thill N (eds). Agriculture Research Centre Publishing: Gembloux, Belgium, 1991; 671–674.
10. Wang Y, Veltkamp DJ, Kowalski BR. Multivariate instrument standardization. *Anal. Chem.* 1991; **63**: 2750–2756.
11. Capron X, Walczak B, de Noord OE, Massart DL. Selection and weighting of samples in multivariate regression model updating. *Chemom. Intell. Lab. Syst.* 2005; **76**: 205–214.
12. Kramer KE, Small GW. Blank Augmentation protocol for improving the robustness of multivariate calibrations. *Appl. Spectrosc.* 2007; **61**: 497–506.
13. Riley MR, Arnold MA, Murhammer DW. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. *Appl. Spectrosc.* 1998; **52**: 1339–1347.
14. Sulub Y, Small GW. Spectral simulation methodology for calibration transfer of near-infrared spectra. *Appl. Spectrosc.* 2007; **61**: 406–413.
15. Wehlburg CM, Haaland DM, Melgaard DK, Martin LE. New hybrid algorithm for maintaining multivariate quantitative calibrations of a near-infrared spectrometer. *Appl. Spectrosc.* 2002; **56**: 605–614.
16. Wehlburg CM, Haaland DM, Melgaard DK. New hybrid for transferring multivariate quantitative calibrations of intra-vendor near-infrared spectrometers. *Appl. Spectrosc.* 2002; **56**: 877–886.
17. Stork CL, Kowalski BR. Weighting schemes for updating regression models – a theoretical approach. *Chemom. Intell. Lab. Syst.* 1999; **48**: 151–166.
18. Sáiz-Abajo MJ, Mevik BH, Segtnan VH, Næs T. Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. *Anal. Chim. Acta* 2005; **533**: 147–159.
19. Zhu M. Kernels and ensembles. *Am. Statist.* 2008; **62**: 97–109.
20. Mevik BH, Segtnan VH, Næs T. Ensemble methods and partial least squares regression. *J. Chemom.* 2004; **18**: 498–507.
21. Wold S. Exponentially weighted moving principal components analysis and projections to latent structures. *Chemom. Intell. Lab. Syst.* 1984; **23**: 149–161.
22. Helland K, Berntsen HA, Borgen OS, Martens H. Recursive algorithm for partial least squares. *Chemom. Intell. Lab. Syst.* 1991; **14**: 129–137.
23. Condolfi A, Massart DL. Model updating for the identification of NIR spectra from a pharmaceutical excipient. *Appl. Spectrosc.* 2000; **54**: 48–53.
24. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM: Philadelphia, PA, 1998.
25. Aster RC, Borchers B, Thurber CH. *Parameter Estimation and Inverse Problems*. Elsevier: Amsterdam, 2005.
26. Kalivas JH, Siano GS, Andries E, Goicoechea HC. Calibration Maintenance and transfer using Tikhonov regularization approaches. *Appl. Spectrosc.* 2009; **63**: 800–809.
27. Kunz MR, Andries E, Kalivas JH. Model updating for spectral calibration maintenance and transfer using 1-norm variants of Tikhonov regularization. Submitted to Anal. Chem. 2009.
28. Wang Y, Kowalski BR. Calibration transfer and measurement stability of near-infrared spectrometers. *Appl. Spectrosc.* 1992; **46**: 764–771.
29. Sohn M, Barton FE II, Himmelsbach DS. Transfer of near-infrared calibration model for determining fiber content in flax: effects of transfer samples and standardization procedure. *Appl. Spectrosc.* 2007; **61**: 414–418.
30. Sohn M, Himmelsbach DS, Barton FE, II, De Haseth JA. Transfer of calibrations for barley quality from dispersive instrument to Fourier transform near-infrared instrument. *Appl. Spectrosc.* 2009; **63**: 1190–1196.
31. Bouveresse E, Massart DL, Dardenne P. Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm: effects of different standardization samples. *Anal. Chim. Acta* 1994; **297**: 405–416.
32. Garcia-Olmo J, Garrido-Varo A, De Pedro E. The transfer of fatty acid calibration equations using four sets of unsealed liquid standardization samples. *J. Near Infrared Spectrosc.* 2001; **9**: 49–62.
33. Osborne BG, Kotwal Z, Wesley IJ, Saunders L, Dardenne P, Shenk JS. Optical matching of near infrared reflectance monochromator instruments for the analysis of ground whole wheat. *J. Near Infrared Spectrosc.* 1999; **7**: 167–178.
34. Park KS, Ko YH, Lee H, Jun CH, Chung H, Ku MS. Near-infrared spectral data transfer using independent standardization samples; as case study on the trans-alkylation process. *Chemom. Intell. Lab. Syst.* 2001; **55**: 53–65.
35. Shenk JS, Westerhaus MO. New standardization and calibration procedures for NIRS analytical systems. *Crop Sci.* 1991; **31**: 1694–1696.
36. Zeaiter M, Rutledge D. Preprocessing methods. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis Vol. 3*, Brown SD, Tauler R, Walczak B (eds-in-chief). Kalivas J, (section ed.). Elsevier: Amsterdam, 2009; 121–231.
37. Guenard RD, Wehlburg CM, Pell RJ, Haaland DM. Importance of prediction outlier diagnostics in determining a successful inter-vendor multivariate calibration model transfer. *Appl. Spectrosc.* 2007; **61**: 747–754.
38. Bates JM, Granger CWJ. The combination of forecasts. *Oper. Res. Quart.* 1969; **20**: 451–468.

39. Bunn DW. *Expert Use of Forecasts: Bootstrapping and Linear Models*, Wright G, Ayton P (eds). Wiley: New York, 1987; 229–241.
40. Bunn DW. Combing forecasts. *Eur. J. Operat. Res.* 1988; **33**: 223–229.
41. Clemen RT. Combining forecasts: a review and annotated bibliography. *Int. J. Forecast.* 1989; **5**: 559–583.
42. Breiman L. Stacked regressions. *Mach. Learn.* 1996; **24**: 49–64.
43. Breiman L. Bagging predictors. *Mach. Learn.* 1996; **24**: 123–140.
44. Næs T, Isaksson T, Fern T, Davies T. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications: Chichester, 2002.
45. Hastie TJ, Tibshirani RJ, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, (2nd edn). Springer-Verlag: New York, 2009.
46. Claerbout JF, Muir F. Robust modeling with erratic data. *Geophysics* 1973; **38**: 826–844.
47. Tibshirani R. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. B* 1996; **58**: 267–288.
48. Stout F, Kalivas JH, Héberger K. Wavelength selection for multivariate calibration using Tikhonov regularization. *Appl. Spectrosc.* 2007; **61**: 85–95.
49. Zou H. The adaptive lasso ad its oracle properties. *J. Am. Statist. Assoc.* 2006; **101**: 1418–1429.
50. Ottaway J, Andries E, Kalivas JH. Spectral multivariate calibration with wavelength selection using variants of Tikhonov regularization. Submitted to Anal. Chim. Acta 2010.
51. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J. Royal Statist. Soc. B* 2005; **67**: 301–320.
52. Gemperline PJ. Rugged spectroscopic calibration for process control. *Chemom. Intell. Lab. Syst.* 1997; **39**: 29–40.
53. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998; **20**: 832–844.
54. Ho TK. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* 1994; **16**: 66–75.
55. Kittler J, Hatef M, Duin RPW, Matas J. On combining classifers. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998; **20**: 226–239.
56. Tong W, Hong H, Fang H, Xie Q, Perkins R. Decision forests: combining the predictions of multiple independent decision tree models. *J. Chem. Inf. Comput. Sci.* 2003; **43**: 525–531.
57. Van Rhee AM. Use of recursion forests in the sequential screening process: consensus selection by multiple recursion trees. *J. Chem. Inf. Comput. Sci.* 2003; **43**: 941–948.
58. Hibbon M, Evgenoiu T. To combine or not to combine: selecting among forecasts and their combinations. *Int. J. Forecast.* 2005; **21**: 15–24.
59. Brown CD, Green RL. Critical factors limiting the interpretation of regression vectors in multivariate calibration. *Trend. Anal. Chem.* 2009; **28**: 506–514.
60. Todeschini R, Consonni V, Pavan M. A distance measure between models: a tool for similarity/diversity analysis of model populations. *Chemom. Intll. Lab. Syst.* 2004; **70**: 55–61.
61. Hansen K, Rathke R, Schroeter T, Rast G, Fox T, Kriegl JM. Bias-correction of regression models: a case study on hERG inhibition. *J. Chem. Inf. Model.* 2009; **49**: 1486–1496.
62. Davies AMC, Fearn T. Quantitative analysis via near infrared databases: comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D). *J. Near Infrared Spectrosc.* 2006; **14**: 403–411.
63. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression (with discussion). *Anna. Statist.* 2004; **32**: 407–499.
64. Du YP, Kasemsumran S, Maruo K, Nakagawa T, Ozaki Y. Improvement of the partial least squares model performance for oral glucose intake experiments by inside mean centering and inside multiplicative signal correction. *Anal. Sci.* 2005; **21**: 979–984.
65. Kalivas JH. Learning from Procrustes analysis to improve multivariate calibration. *J. Chemom.* 2008; **22**: 227–234.
66. Wülfert F, Kok TW, Smilde AK. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal. Chem.* 1998; **70**: 1761–1767.
67. Wise BM, Gallagher NB, Bro R, Shaver JM. *PLS_Toolbox 3.0 for use with MATLAB*. Eigenvector Research: Manson, WA, 2003.
68. Forrester JB, Kalivas JH. Ridge regression optimization using a harmonious approach. *J. Chemom.* 2004; **18**: 372–384.
69. Kalivas JH. Calibration methodologies. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Brown SD, Tauler R, Walczak B (eds-in-chief). Kalivas J, (section ed.). Elsevier: Amsterdam, 2009; 1–32.
70. Lawson CL, Hanson RJ. *Solving Least Square Problems*. SIAM: Philadelphia, PA, 1995.
71. Bouveresse E, Massart DL, Dardenne P. Modified algorithm for standardization of near-infrared spectrometric instruments. *Anal. Chem.* 1995; **67**: 1381–1389.
72. Cole SB, Kalivas JH. Calibration maintenance and transfer using weighted spectral differences. *In preparation for J. Chemom.*
73. Brenchley JM, Hörchner U, Kalivas JH. Wavelength selection characterization for NIR spectra. *Appl. Spectrosc.* 1997; **51**: 689–699.