Special issue article

Journal of CHEMOMETRICS

Received: 19 December 2015, Revised: 10 January 2016, Accepted: 17 January 2016, Published online in Wiley Online Library: 14 March 2016

(wileyonlinelibrary.com) DOI: 10.1002/cem.2784

Penalty processes for combining roughness and smoothness in spectral multivariate calibration

Alister Tencate[a], John H. Kalivas[a]* and Erik Andries[b]

Tikhonov regularization (TR) has been successfully applied to form spectral multivariate calibration models by augmenting spectroscopic data with a regulation operator matrix. This matrix can be set to the identity matrix I (ridge regression), yielding what shall be considered rough regression vectors. It can also be set to the *i*-th derivative operator matrix $_i$L to form smoothed regression vectors. Two new penalty (regularization) methods are proposed that concurrently factor both roughness and smoothness in forming the model vector. This combination occurs by augmenting calibration spectra simultaneously with independently weighted I and $_i$L matrices. The results of these two new methods are presented and compared with results using ridge regression forming rough model vectors and only using the smoothing TR processes. Partial least squares regression is also used to combine roughness and smoothness, and these results are compared with the TR variants. The sum of ranking differences algorithm and the two fusion rules sum and median are used for automatic model selection, that is, the appropriate tuning parameters for I and $_i$L and partial least squares latent vectors. The approaches are evaluated using near-infrared and ultraviolet-visible spectral data sets. The near-infrared set consists of corn samples for the analysis of protein and moisture content. The ultraviolet-visible set consists of a three-component system of inorganic elements. The general trends found are that when spectra are originally generally smooth, then using the smoothing methods provides no improvement in prediction errors. However, when spectra are considered noisy, then smoothing methods can assist in reducing prediction errors. This is especially true when the spectroscopic noise is more widespread across the wavelength regions. There was no difference in the results between the different smoothing methods. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: smooth regression vectors; penalty smoothing; partial least squares; Tikhonov regularization; sum of ranking differences

1. INTRODUCTION

Characteristics of model vectors generated through spectral multivariate calibration methods can have significant impacts on prediction accuracy (and precision) as well as model robustness. Rough (or jagged) model vectors can be found to predict accurately under the given calibration conditions. However, such model vectors can be overfitted to the specific calibration conditions used and can fail to predict new samples accurately. A smooth model vector on the other hand can be more robust to prediction than its rough counterpart [1–3]. For example, small changes in temperature or pressure have been found to cause shifts in measured sample spectra. Smoothness in the model vector may be useful, permitting robustness to spectral perturbations.

There are a number of calibration smoothing approaches reported in the literature [1–14]. Many of these approaches are penalty based such as the efficient Whittaker smoother [10], and some methods incorporate flexibility with B-splines allowing vectors to be unimodal, convex, or concave [11]. The focus of these processes is to smooth spectra as data pre-processing. Also used is Savitzky–Golay smoothing. While the goal of these methods is to increase the spectral signal-to-noise ratio, there are concerns that through the smoothing process, some useful spectral information may be lost and not carried forward to the calibration model [15].

An alternative to using penalty smoothing processes directly on spectra is to smooth basis vectors (such as partial least squares (PLS) latent vectors (LVs)) prior to forming calibration model vectors [12–14]. Penalty-based approaches have been used as well a Savitzky–Golay process. However, as with smoothing spectra, no significant advantages have been found unless the original spectra contain large measurement noise.

Rather than using penalties on spectra or the basis vectors, the approach used in this study, and found to be successful in many situations [2–9], is to form smooth regression vectors directly via in-process regression vector penalties. New here are Tikhonov regularization (TR) variants that simultaneously optimize the

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

a *A. Tencate, J. H. Kalivas*
Department of Chemistry, Idaho State University, Pocatello, Idaho, 83209, USA

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

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

degree of model smoothness and roughness by applying separate smoothing and roughness penalties in the minimization. This is similar to the TR penalty variant known as the elastic net that balances full variables models with sparseness [16,17].

The method of PLS regression is a common method to generate model vectors. Thus, PLS variants analogous to the new TR variants are also studied and compared.

2. MODELING PROCESSES

2.1. Tikhonov regularization and variants

The fundamental TR approach applied to spectral multivariate calibration relies on a regulation operator matrix (**L**) that augments the spectral data [17–22]. The **L** matrix is adjusted by a regularization (tuning or penalty) parameter (λ). This tuning parameter has values from 0 to ∞, and the value determines the effectiveness of the generated model. Hence, the model characteristics are determined by the choice of **L** and λ. Setting $\mathbf{L}=\mathbf{I}$ (identity matrix) yields the standard form of TR (referred to hereafter as ridge regression (RR) [23]) expressed by

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{I}\end{pmatrix}\mathbf{r} \tag{1}$$

with solutions from the minimization expression

$$\min\left(\|\mathbf{Xr}-\mathbf{y}\|^2+\lambda^2\|\mathbf{r}\|^2\right) \tag{2}$$

where **X** denotes the $m\times n$ matrix of calibration spectra, **y** represents the corresponding vector of reference values for the prediction property to be model, **r** signifies the model vector to be estimated, and $\|\bullet\|$ symbolizes Euclidean vector norm (L_2 norm). The model vector is denoted **r** because the RR modeling process tends to generate rough model vectors with spectroscopic data. The degree of roughness depends on the value chosen for λ. A small λ value produces rough models, while larger λ values generate smoother models. The method of RR in Eq. 1 is labeled Method 1 in Table I and shall be referred to as either RR or Method 1, and the TR variants defined in the following are referred to as TR variants or the respective Method number. A similar statement is true for the PLS variants introduced in Section 2.2.

Table I. TR and PLS method names

Method name	Equation
TR Method 1 (RR)	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{I}\end{pmatrix}\mathbf{r}$
TR Method 2	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \eta\mathbf{L}\end{pmatrix}\mathbf{s}$
TR Method 3	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{I}\\ \eta\mathbf{L}\end{pmatrix}\mathbf{b}$
TR Method 4	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X} & \mathbf{X}\\ \lambda\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \eta\mathbf{L}\end{pmatrix}\begin{pmatrix}\mathbf{r}\\ \mathbf{s}\end{pmatrix}$
PLS Method 1 (PLS)	$\mathbf{y}=\mathbf{Xr}$
PLS Method 3	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \eta\mathbf{L}\end{pmatrix}\mathbf{s}$ $\mathbf{y}_a=\mathbf{X}_a\mathbf{s}$
PLS Method 4	$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X} & \mathbf{X}\\ \mathbf{0} & \eta\mathbf{L}\end{pmatrix}\begin{pmatrix}\mathbf{r}\\ \mathbf{s}\end{pmatrix}$ $\mathbf{y}_a=\mathbf{X}_a\mathbf{b}_a$

Setting $\mathbf{L}={}_2\mathbf{L}$, the second derivative operator defined by the discrete matrix approximation of the second derivative operator

$${}_2\mathbf{L}=\begin{bmatrix}1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1\end{bmatrix}\in\mathbb{R}^{(n-2)\times n} \tag{3}$$

and using

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \eta\mathbf{L}\end{pmatrix}\mathbf{s} \tag{4}$$

with solutions for the minimization

$$\min\left(\|\mathbf{Xs}-\mathbf{y}\|^2+\eta^2\|\mathbf{Ls}\|^2\right) \tag{5}$$

forms smoothed model vectors **s** subject to the η tuning parameter value that can vary from 0 to ∞. As with RR, small η tuning parameter values for this TR variant [18–22] produce rougher models than those with larger η values. Essentially, the derivative operator in the second minimization term of expression 5 penalizes (constrains) the model vector roughness. Higher order derivatives place a stronger constraint on the model vector roughness, leading to smoother and broader regression vectors [2]. For the remainder of the paper, when the symbol **L** is used, it is assumed to represent the second derivative operator. The approach noted by Eq. 4 is labeled Method 2 in Table I.

A new variant of TR, termed Method 3 in Table I, can be formed by combining Eqs. 1 and 4, obtaining

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \lambda\mathbf{I}\\ \eta\mathbf{L}\end{pmatrix}\mathbf{b} \tag{6}$$

with the minimization expression

$$\min\left(\|\mathbf{Xb}-\mathbf{y}\|^2+\lambda^2\|\mathbf{b}\|^2+\eta^2\|\mathbf{Lb}\|^2\right) \tag{7}$$

By tuning the identity and the derivative penalties independently, the degrees of roughness and smoothness can be simultaneously varied in forming regression vectors. Because this TR modeling process concurrently penalizes roughness and smoothness, the model vector in Eq. 6 is designated **b**. As noted previously, these dual penalties are similar to the concept of the elastic net where two separate penalties on the regression vector are also used, one favoring full variables and the other preferring sparseness (selected variables) [16].

A second new TR variant studied in this paper and previously proposed in general terms [24] is expressed as

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{X} & \mathbf{X}\\ \lambda\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \eta\mathbf{L}\end{pmatrix}\begin{pmatrix}\mathbf{r}\\ \mathbf{s}\end{pmatrix} \tag{8}$$

with minimization

$$\min\left(\|\mathbf{X}(\mathbf{r}+\mathbf{s})-\mathbf{y}\|^2+\lambda^2\|\mathbf{r}\|^2+\eta^2\|\mathbf{Ls}\|^2\right) \tag{9}$$

This TR variant, Method 4 in Table I, can also factor in a smoothing penalty simultaneously with a roughness penalty. However, this method produces two regression vectors. One is

formed through the rough penalty, and as such, is designated by **r**. The other vector incorporates the smoothing penalty constraint and is designated by **s**. The analyte present in a new sample spectrum is predicted by the sum of the two regression vectors; that is, the final model prediction vector is $\mathbf{b} = \mathbf{r} + \mathbf{s}$.

2.2. Partial least squares solutions to augmented arrays

The PLS regression method (also referred to as PLSR) [25,26] has been widely used in the literature for a number of different applications. For this reason, the PLS approach has been adapted to the aforementioned TR variants for regression purposes. It is assumed that the reader is familiar with PLS regression and the variants of PLS relative to the aforementioned TR methods are only briefly noted.

Essentially, a PLS algorithm is used on each of the aforementioned augmented arrays in Eqs. 1, 4, and 8 with the corresponding **0** and λ**I** arrays removed. The augmented arrays are treated as respective **y** and **X** single arrays for the PLS algorithm. The PLS LVs replace the λ tuning parameter. The TR and PLS variants are identified in Table I with regard to specific method numbers as discussed in this paper. The PLS methods are not sequentially numbered in order to better correlate the TR and PLS methods that operate on obtaining similar model vectors.

2.3. Model quality measures and tuning parameter selection

Three fusion processes known as sum, median, and sum of ranking differences (SRD), sum, and median [27,28] are used with numerous model quality measures in order to select the best calibration tuning parameter values for λ, η, and/or LVs as the case may be. From fusion of model quality measures favoring biased models (leading toward overfitted models), low variance models (preferring underfitted models), and combinations of bias and variance measures, tuning parameters are selected, balancing the bias/variance trade-off needed with inverse regression models [29]. The SRD, mean, and median fusion process have been shown to be effective in selecting up to two calibration tuning parameters [30,31].

Specific model measures used are the root-mean-squared error (RMSE) of calibration (RMSEC) and cross validation (RMSECV), both of which assess the deviation of predicted values from the known reference values, respective statistical measures R^2, slope, and *y*-intercept from plotting the corresponding predicted values against reference values, model vector L_2 norms, model jaggedness J where $J_i = \sqrt{\sum_{j=2}^{p} \left(\hat{b}_{ij} - \hat{b}_{i(j-1)}\right)^2}$ for the *i*-th model, U-shaped curves [29] formed by

Table II. Corn protein results

Modeling process	Fusion rule	Tuning parameter[1] λ or PLS LVs, η	RMSECV	R^2_{CV}	$\|\hat{\mathbf{b}}\|$
TR Method 1	SRD	0.004 (48)	0.153	0.905	92.4
	Sum	0.005 (47)	0.152	0.906	79.3
	Median	0.005 (47)	0.152	0.906	79.3
	Minimum[2]	0.005 (47)	0.152	0.906	79.3
TR Method 2	SRD	0.048 (48)	0.199	0.851	322
	Sum	0.494 (37)	0.166	0.890	106
	Median	2.17 (30)	0.180	0.870	71.0
	Minimum[2]	0.324 (39)	0.165	0.892	126
TR Method 3	SRD	0.002 (50), 0.060 (47)	0.158	0.900	121
	Sum	0.004 (48), 0.074 (46)	0.154	0.904	88.4
	Median	0.004 (48), 0.009 (56)	0.154	0.905	91.5
	Minimum[2]	0.005 (47), 0 (100)	0.152	0.906	79.3
TR Method 4	SRD	5.88e-4 (55), 0.007 (57)	0.295	0.718	1.02e3
	Sum	0.004 (48), 149 (10)	0.154	0.905	92.4
	Median	0.004 (48), 530 (4)	0.154	0.905	92.4
	Minimum[2]	0.005 (47), 1000 (1)	0.152	0.906	79.3
PLS Method 1	SRD	17	0.186	0.864	277
	Sum	16	0.182	0.868	234
	Median	10	0.153	0.905	75.6
	Minimum[2]	10	0.153	0.905	75.6
PLS Method 3	SRD	80, 0.001 (67)	0.369	0.628	1.87e3
	Sum	34, 0.060 (47)	0.155	0.903	79.9
	Median	31, 0.060 (47)	0.154	0.904	77.2
	Minimum[2]	15, 0.032 (50)	0.152	0.906	73.5
PLS Method 4	SRD	80, 0.009 (56)	0.251	0.778	811
	Sum	70, 0.172 (42)	0.155	0.903	81.3
	Median	20, 0.048 (48)	0.154	0.904	79.4
	Minimum[2]	15, 0.048 (48)	0.152	0.906	72.6

[1]Values in parenthesis are respective tuning parameter index.
[2]Not fusion minimum rule but instead, model at actual minimum RMSECV.

$$C1_i = \left(\frac{\|\mathbf{b}\|_i - \|\mathbf{b}\|_{\min}}{\|\mathbf{b}\|_{\max} - \|\mathbf{b}\|_{\min}}\right) + \left(\frac{\text{RMSEC}_i - \text{RMSEC}_{\min}}{\text{RMSEC}_{\max} - \text{RMSEC}_{\min}}\right) \tag{10}$$

$$C2_i = \frac{\text{RMSEC}_i + \text{RMSECV}_i}{\left(\text{RMSEC}_i / \text{RMSECV}_i\right)} \tag{11}$$

with three other variations of Eq. 10 by replacing the model vector L_2 norms with the model vector jaggedness J and/or the RMSEC with the RMSECV term, and lastly, measures formed from modifying Eq. 11 by exchanging the RMSEC and RMSECV terms in the numerator with $\left(1 - R^2_{\text{cal}}\right)$ and $\left(1 - R^2_{\text{CV}}\right)$. These combinations of measures were found to be useful with SRD in selecting RR and PLS tuning parameters and intermodal comparison, but others can be added or used instead of those presented [30].

2.4. Fusion processes

Previous work involving selection of single tuning parameter values for RR and PLS (respective Methods 1) used only SRD and evaluated all possible tuning parameter values [30]. Work presented here is the same for RR and PLS Method 1 and TR Method 2. As noted previously, the fusion rules sum and median are used in addition to SRD. Previous work on selecting two tuning parameters (as required for Methods 3 and 4) considered three fusion methods (SRD, sum, and median) using different approaches of dealing with the large number of models from all the possible combinations of tuning parameter values [31]. While no one approach was better than another, the process of using thresholds to preliminarily filter all the possible models before fusion was used in this study. This procedure consists of using R^2 values for the calibration samples (R^2_{cal}) where the lower 50% and top 10% of all models are removed from further consideration for selecting the pair of tuning parameter values to/from the model.

Regardless of whether one or two tuning parameters are being selected, the matrix of model quality measures for column-wise fusion is composed of a row for each model quality measure and a column for each tuning parameter pair. All values are row-wised normalized to unit length before applying the fusion rules. The SRD process is then used to rank the columns (models). The sum fusion rule sums the values in each column and then ranks the column. The median rule identifies the median value in each column and then ranks the columns.

In order to validate rankings, an internal cross validation is used on the matrix of model quality measures. In this case, some model quality measures are removed, and the fusion rules are applied on the remaining model quality measures to obtain rankings. The process is repeated depending on the type of cross validation being used, that is, *n*-fold, leave multiple out, and so on. From the internal cross validation, statistical difference testing is possible [27,32] if desired. Additionally, a rank reliability

Table III. Corn moisture results with simulated noise added to spectra

Modeling process	Fusion rule	Tuning parameter[1] λ or PLS LVs, η	RMSECV	R^2_{CV}	$\|\hat{\mathbf{b}}\|$
TR Method 1	SRD	0.119 (35)	0.192	0.763	7.12
	Sum	0.346 (31)	0.249	0.603	2.07
	Median	8.38 (19)	0.331	0.434	0.093
	Minimum[2]	6.42e-6 (72)	0.108	0.926	17.7
TR Method 2	SRD	1.15 (33)	0.072	0.966	25.5
	Sum	6.25 (25)	0.076	0.962	22.7
	Median	4.10 (27)	0.074	0.964	22.6
	Minimum[2]	1.42 (32)	0.071	0.966	24.6
TR Method 3	SRD	0.003 (49), 0.007 (57)	0.073	0.964	22.2
	Sum	0.019 (42), 1.76 (31)	0.070	0.967	19.8
	Median	0.001 (52), 7.72 (24)	0.080	0.958	18.6
	Minimum[2]	0.011 (44), 0.754 (35)	0.068	0.969	22.3
TR Method 4	SRD	0.001 (52), 0.014 (54)	0.075	0.962	29.2
	Sum	0.006 (46), 0.039 (49)	0.081	0.956	31.8
	Median	0.005 (47), 0.032 (50)	0.081	0.956	31.5
	Minimum[2]	0.451 (30), 1.42 (32)	0.071	0.966	24.6
PLS Method 1	SRD	5	0.109	0.925	16.8
	Sum	2	0.261	0.570	2.27
	Median	2	0.261	0.570	2.27
	Minimum[2]	7	0.108	0.926	17.6
PLS Method 3	SRD	45, 6.87e-5 (79)	0.075	0.962	21.3
	Sum	21, 0.048 (48)	0.072	0.965	23.4
	Median	56, 0.091 (45)	0.072	0.964	26.4
	Minimum[2]	80, 0.400 (38)	0.066	0.970	21.6
PLS Method 4	SRD	71, 0.003 (62)	0.089	0.950	18.3
	Sum	15, 0.021 (52)	0.091	0.947	18.2
	Median	25, 0.001 (65)	0.094	0.944	18.1
	Minimum[2]	80, 0.932 (34)	0.087	0.952	17.8

[1]Values in parenthesis are respective tuning parameter index.
[2]Not fusion minimum rule but instead, model at actual minimum RMSECV.

measure (RRM) was developed to use with the results from the internal cross validation [31]. The RRM for each model is the sum of the range scaled mean and standard deviations of model rankings. This measure favors low overall rank and consistency in that rank. A potential concern is some models could result in low RRM values where overall model rank is not low but are extremely closely ranked over the cross validations (very low standard deviation). Such models could have artificially low RRM values, but not the lowest RRM values. For this reason, only the 100 lowest mean ranked models are included in the computation of RRM, thereby placing a greater weight on the actual data fusion ranks than on the standard deviation of those ranks.

3. EXPERIMENTAL

3.1. Algorithms

All algorithms (RR, PLS, TR and PLS variants, SRD, RRM, and all measures of model quality) were written by the authors using MATLAB 8.1 (The MathWorks, Natick, MA, USA). The MATLAB SRD algorithm is available for download at [33]. The PLS regression algorithm is that based on reference [25].

3.2. Near-infrared corn data

Spectra were measured for 80 samples of corn over 700 wavelengths from 1100 to 2498 nm at 2 nm intervals [34]. Included in the data set are spectra for the same 80 samples measured on three near-infrared spectrometers designated m5, mp5, and mp6. Reference values for moisture, oil, protein, and starch content for each sample are available. For the purpose of this study, the focus was protein using spectra from mp6 and moisture using m5. The η and λ tuning parameters ranged exponentially from 1000 to 10^{-6} for 80 and 100 values, respectively, for RR and the other TR variants. For PLS models, the number of PLS LVs was limited to 80 in order to facilitate easier comparison between the TR (including RR) and PLS model measures.

In a separate study, noise based on a random normal distribution with mean zero and standard deviation one was added to the corn spectra to further characterize the rough and smoothing processes. Noise added was homoscedastic at 0.25% of the maximum peak amplitude for each spectrum and heteroscedastic at 0.25% of each spectral wavelength intensity. No correlated noise was used, and other distributional noise structures were not studied. Other than the altered spectra, the modeling procedures were the same as those described previously. The analyte studied was moisture on the instrument designated m5.

3.3. Ultraviolet-visible inorganic data

Three-component mixtures of three metal ions, cobalt II, chromium III, nickel II, with varying concentrations for each ion were measured at 176 wavelengths from 300 to 650 nm at 2 nm intervals on a diode array spectrophotometer. Reference concentration values (mM) are available for a total of 128 samples [35]. Analytes nickel and cobalt were studied. For RR and the TR variants, η tuning parameter values ranged exponentially from 1000 to 10^{-6} for 80 values, and the λ ranged exponentially from 10^{5} to 10^{-4} for 100 values. For PLS models, the number of PLS LVs was limited to 80 in order to facilitate easier comparison between the TR (including RR) and PLS model measures.

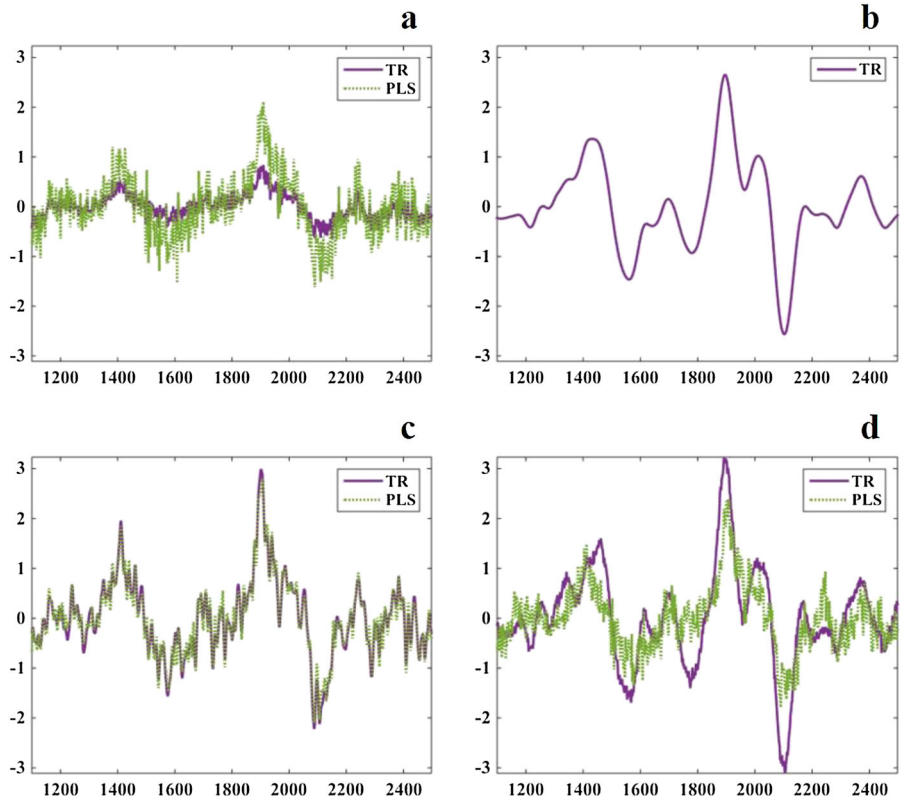

Figure 1. Corn moisture model vectors based on adding simulated noise to spectra. Model vectors are selected by SRD. Modeling processes are (a) Method 1, (b) Method 2, (c) Method 3, and (d) Method 4 for TR (purple line) and PLS (green dots).

3.4. Cross validation

Confidence in the significance of the model rankings and selection is essential in order to form any conclusions. To this end, cross validation is performed. Twenty leave-multiple-out cross validation was used for both data sets. In this case, 60% of the full data are randomly removed, forming the calibration set, and the remaining 40% make up the validation set. The process is repeated 20 times. In each case, the data were mean centered by column based on the calibration set. All of the model measures previously noted are computed for each of the cross-validation splits. Thus, for the fusion rules, there are 20 rows for measures of model quality. Thus, instead of the usual practice of using the mean model quality measures to select tuning parameters, the model quality measures generated for each CV split are used. This process guards against overfitting, and the consensus of the tuning parameters is assessed. Further characterization of the SRD fusion rule relative to data splitting is available [36]. For the internal cross validations of the model measures, a 10-fold cross validation was performed on each model quality measure block of 20 rows.

4. RESULTS AND DISCUSSION

Results from the TR (including RR) and PLS variants are discussed together as this best facilitates a comparison between the variety of methods. Two elements are included in the comparison of the modeling methods. The first compares the model with the absolute minimum cross validation error for each method. The second utilizes the three automatic tuning parameter selection methods in order to compare models selected for each modeling method in a realistic setting (rather than the best-case scenario). The results are similar for the two data sets studied, and the general trends for each data set as a whole are discussed.

4.1. Corn data

The following sections discuss results observed for the corn data set predicting protein and moisture contents on two different near-infrared instruments. Results from the seven penalty methods listed in Table I are first presented for protein as the analyte. Modeling moisture resulted in similar trends to protein, and as such, those results are only compared qualitatively. The results for the artificially noisy corn spectra are discussed independently as they represent a different experimental situation.

4.1.1. *Protein*

Table II shows the results for the seven methods applied to the mp6 spectra using protein as the analyte. The models selected at the minimum RMSECV values in Table II indicate that no discernable improvement in prediction is achieved by incorporating

Table IV. Inorganic nickel results

Modeling process	Fusion rule	Tuning parameter[1] λ or PLS LVs, η	RMSECV	R^2_{CV}	$\|\hat{\mathbf{b}}\|_2$
TR Method 1	SRD	0.001 (52)	9.64e-4	0.999	0.144
	Sum	0.054 (38)	8.35e-4	0.999	0.080
	Median	0.070 (37)	8.42e-4	0.999	0.078
	Minimum[2]	0.041 (39)	8.31e-4	0.999	0.083
TR Method 2	SRD	0.003 (83)	5.97e-4	1.00	0.476
	Sum	0.164 (64)	3.28e-4	1.00	0.172
	Median	7.37 (46)	6.64e-4	1.00	0.083
	Minimum[2]	0.250 (62)	3.24e-4	1.00	0.155
TR Method 3	SRD	0.024 (41), 0.133 (65)	4.65e-4	1.00	0.093
	Sum	0.070 (37), 2.07 (52)	4.87e-4	1.00	0.079
	Median	3.16e-5 (66), 0.250 (62)	3.24e-4	1.00	0.155
	Minimum[2]	0 (80), 0.250 (62)	3.24e-4	1.00	0.155
TR Method 4	SRD	0.001 (52), 0.020 (74)	7.59e-4	0.999	0.182
	Sum	0.032 (40), 0.070 (68)	4.53e-4	1.00	0.196
	Median	0.092 (36), 0.024 (73)	4.26e-4	1.00	0.248
	Minimum[2]	0.005 (47), 1000 (1)	3.24e-4	1.00	0.155
PLS Method 1	SRD	13	9.81e-4	0.999	0.074
	Sum	11	1.00e-3	0.999	0.074
	Median	15	9.83e-4	0.999	0.075
	Minimum[2]	30	9.11e-4	0.999	0.082
PLS Method 3	SRD	68, 2.07 (52)	4.38e-4	1.00	0.159
	Sum	74	4.49e-4	1.00	0.083
	Median	74	3.77e-4	1.00	0.098
	Minimum[2]	80	3.54e-4	1.00	0.108
PLS Method 4	SRD	74, 32.4 (39)	8.11e-4	1.00	0.101
	Sum	30, 40.0 (38)	8.15e-4	1.00	0.081
	Median	25, 49.4 (37)	8.40e-4	1.00	0.078
	Minimum[2]	59, 75.4 (35)	7.96e-4	0.999	0.085

[1]Values in parenthesis are respective tuning parameter index relative to figures.
[2]Not fusion minimum rule but instead, model at actual minimum RMSECV.

a smoothing penalty into the modeling process (Methods 2, 3, and 4). Additionally, there is little difference between the new methods (Methods 3 and 4) and Methods 1 and 2. The results for the m5 spectra predicting moisture similarly showed no discernable improvement between the seven methods.

Tabulated results in Table II for the models selected by the fusion processes show a similar lack of variation between the smoothing and unsmoothed methods. Models selected by the SRD fusion rule always had a slightly greater prediction error than the other selected models. However, this trend is not present for the m5 corn moisture analysis; that is, the three fusion rules select models with equivalent measures of model quality.

This clear lack of any consistent variation for predicting protein and moisture can be attributed to the nature of the data itself. There is a high level of collinearity and very low noise across the corn spectra for all instruments. Specifically, the spectra themselves are already fairly smooth. Hence, nothing further is gained by incorporating a smoothing penalty in the modeling process.

4.1.2. *Moisture with simulated noisy spectra*

Noise was added to the corn spectra in order to distort the collinearity and cause modeling difficulties. Table III shows the results for the seven methods, and the smoothing penalty methods all have lower minimum prediction errors and higher R^2_{CV} values than Methods 1. This trend is accentuated when the models selected by the fusion processes are compared.

The improvements observed do appear to be significant, with an increase in R^2_{CV} values from below 0.65 (in general) to approximately 0.95 for the models selected with a smoothing constraint. Coupled with the decrease in the prediction error, these markers clearly show that incorporating a smoothing penalty in the modeling process is beneficial when spectra contain noise across the full wavelength range. There are, however, no prediction improvements by incorporating two tuning parameters (Methods 3 and 4) compared with the prediction by Method 2 with only the smoothing tuning parameter.

Figure 1 shows the regression vectors selected by the SRD fusion rule for each of the modeling methods studied. It is clear that, with the exception of the PLS solution to Method 4, incorporating a smoothing condition in the modeling process does yield regression vectors that are less noisy (compared with Methods 1). Another interesting feature is that, although the regression vectors selected for Methods 2, 3, and 4 have different shapes and degrees of smoothness, all of these models have nearly identical prediction errors, R^2_{CV}, and L_2 norm values. Only the one noise structure was studied (described in Section 3), and it may be that different noise structures could generate different results. Such a full noise study has not been performed.

4.2. Inorganic data

Both nickel and cobalt were studied, but because the results for cobalt are similar to nickel, only the nickel results are detailed. The inorganic spectra show less collinearity than the corn spectra, and there are two highly noisy regions in the spectra. Because adding artificial noise to the corn spectra did show marked improvements in prediction using smoothing regression methods, it is of interest to see whether this trend is consistent for a data set that includes noise. Listed in Table IV are the results for the seven methods with nickel as the analyte. In this case, there is a clear difference in the minimum error of prediction

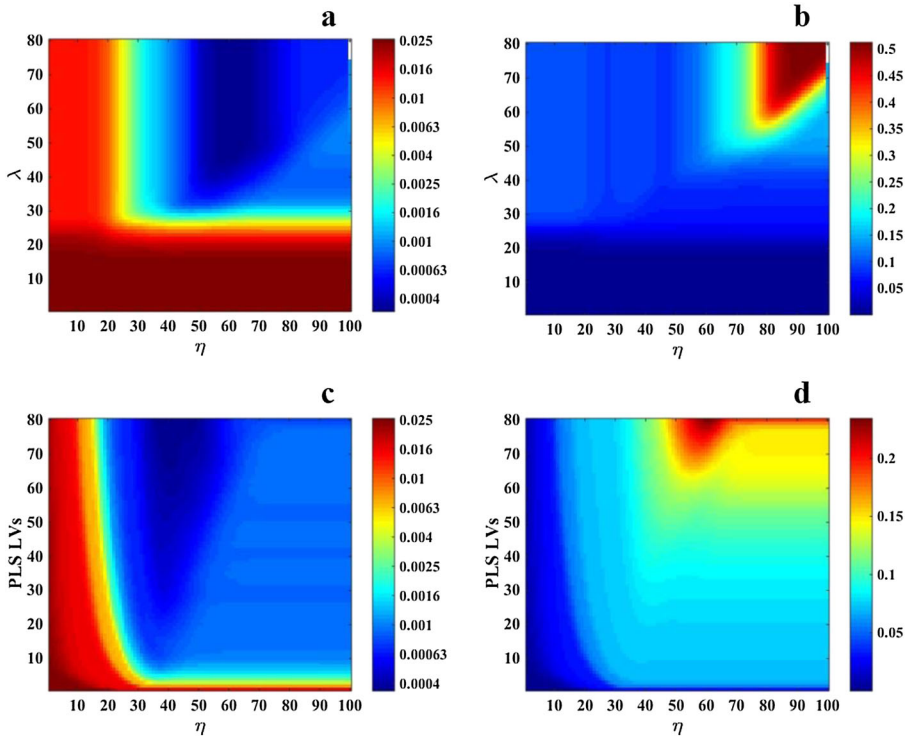

Figure 2. Nickel inorganic data landscape images for TR and PLS Method 3 with (a) RMSECV from TR, (b) model vector L_2 norm from TR, (c) RMSECV from PLS, and (d) model vector L_2 norm from PLS. Log scales are used for (a) and (c), and all color bar values correspond to actual calculated values. In each case, landscapes represent mean values of the 20 cross-validation splits. In all landscapes, the tuning parameter values decrease from left to right and bottom to top (the number of PLS latent vectors used increases where applicable from bottom to top).

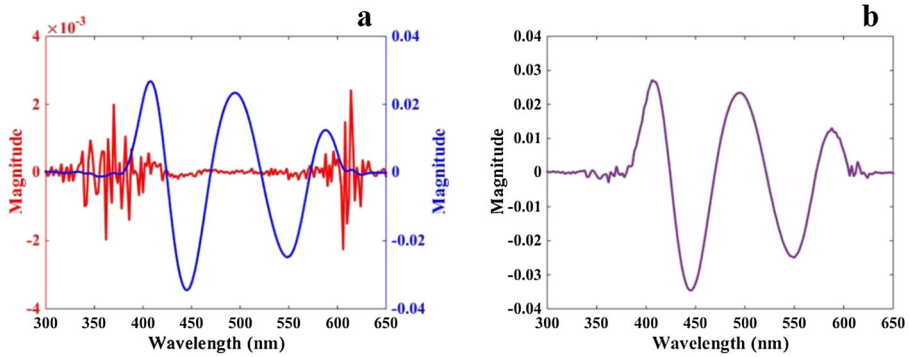

Figure 3. Nickel regression vectors selected by the sum fusion rule from the TR solution to Method 4. (a) Rough component **r** (red) and smooth component **b** (blue) of the regression vector and (b) composite regression vector **r** + **s** used for prediction (purple).

between some of the smoothing methods (Methods 2, 3, and 4 for TR), and Method 1 with no smoothing constraints. These trends are consistent with the results for prediction with cobalt as the analyte.

The models selected by the three fusion processes exhibit similar trends. In general, the modeling methods using a smoothing penalty (except for Method 4 by PLS) have models with lower error of prediction and higher R^2_{CV} values than the standard methods incorporating no smoothing, albeit there are greater differences in prediction errors than the R^2_{CV} values. This result may be attributed to the fact that the spectra are fairly noise free apart from the two noisy wavelength regions on each end of the spectra. Hence, analyte spectral information can be well modeled by the non-noisy spectral region. Specifically, the modeling methods generated model vectors with nearly zero coefficients at points corresponding to the isolated noisy spectral regions (Figure 3b). The near-zero model coefficients negate any influence on predictions by these regions. These result trends are also the case when cobalt is used as the analyte, with the exception of Method 3 by TR, where two of the fusion rules selected models with very poor prediction as compared with the other methods.

Figure 2 contains images of the two model quality measures RMSECV and L_2 norm for Method 3 evaluated by the respective TR and PLS variants. As previously observed [31], the PLS processes can lead to more discrete results, as seen in Figure 2c. The transition zone for the bias/variance trade-off occurs in the first five LVs for η indices greater than 30. However, this discreteness does not appear to produce a difficulty in tuning parameter selection as was the situation observed for model updating [31]. This could be attributed to the fact that the experimental configuration in this study is easier than model updating, leading to a larger minimum zone in prediction error. Additionally, the larger sample size of this data set allows LVs to better decompose the space.

Figure 3 illustrates the formation of the final nickel model vector for TR Method 4 using the sum fusion rule. Similar regression vectors are identified by the other two fusion rules and PLS Method 4. From Figure 3a, the two model vector components are as expected from the design; one component is smooth (blue), while the other is rougher (red). Varying the relative values of the two tuning parameters dictates the relative magnitudes of the two modeling components. The final model plotted in Figure 3b is a sum of the two components in Figure 3a. From Figure 3b, it is noticeable that there is more significant weight on the smoothing penalty parameter, leading to the overall fairly smooth regression vector.

5. CONCLUSION

This study showed that directly forming smoothed model vectors has some targeted benefits. Specifically, advantages are observed when the calibration spectra contain noise, especially when that noise is spread across the entire wavelength range and not localized. This observation was evident when the models were selected based on the minima RMSECV as well as models selected by the fusion processes. When improvement was observed, it was generally characteristic of all of the methods that include a smoothing process. Specifically, throughout the cases studied, no noticeable decrease in prediction error was gained for the smoothing methods by including a second tuning parameter (Methods 3 and 4). The improvement of Methods 3 and 4 over Method 1 can be attributed to the benefit of the smoothing penalty, because Methods 3 and 4 show no improvement over Method 2. That is, when provided the opportunity, smoothing takes precedence over roughness even if roughness can be simultaneously included in the procedure. For this reason, Method 2 is recommended when dealing with noisy spectra as the method uses a single tuning parameter.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant no. CHE-1111053 (co-funded by MPS Chemistry and the OCI Venture Fund) and is gratefully acknowledged by the authors. This research was also partially supported by NSF Award 1156879 through the Department of Physics Research Experience for Undergraduates summer program.

REFERENCES

1. Marx BD, Eilers PHC. Multivariate calibration stability: a comparison of methods. *J. Chemom.* 2002; **16**: 129–140.
2. Stout F, Kalivas JH. Tikhonov regularization in standardized and general form for multivariate calibration with application towards removing unwanted spectral artifacts. *J. Chemom.* 2006; **20**: 22–33.
3. Camacho J, Lauri D, Lennox B, Excabias M. Evaluation of smoothing techniques in the run to run optimization of fed-batch processes with u-PLS. *J. Chemom.* 2015; **29**: 338–348.
4. Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. *Technometrics* 1989; **31**: 3–39.
5. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 1996; **11**: 89–121.
6. Marx BD, Eilers PHC. Direct generalized additive modeling with penalized likelihood. *Comput. Stat. Data Anal.* 1998; **28**: 193–209.

7. Marx BD, Eilers PHC. Generalized linear regression on sampled signal and curves: a P-spline approach. *Technometrics* 1999; **41**: 1–13.
8. Marx BD, Eilers PHC. Multidimensional penalized signal regression. *Technometrics* 2005; **47**: 13–22.
9. Eilers PHC, Li B, Marx BD. Multivariate calibration with single-index signal regression. *Chemom. Intell. Lab. Syst.* 2009; **96**: 196–202.
10. Eilers PHC. A perfect smoother. *Anal. Chem.* 2003; **75**: 3631–3636.
11. Eilers PHC. Unimodal smoothing. *J. Chemom.* 2005; **19**: 317–328.
12. Krämer N, Boulesteix AL, Tutz G. Penalized partial least squares with application to B-splines transformations and functional data. *Chemom. Intell. Lab. Syst.* 2008; **94**: 60–69.
13. Goutis C, Fearn T. Partial least squares regression on smooth factors. *J. Am. Stat. Assoc.* 1996; **91**: 627–632.
14. Rutledge DN, Barros A, Delgadillo I. PoLiSh – smoothed partial least-squares regression. *Anal. Chim. Acta* 2001; **446**: 281–296.
15. Brown CD, Wentzell PD. Hazards of digital smoothing filters as pre-processing tool in multivariate calibration. *J. Chemom.* 1999; **13**: 133–152.
16. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J. Royal Stat. Soc. Series B* 2005; **67**: 301–320.
17. Kalivas JH. Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance. *J. Chemom.* 2012; **26**: 218–230.
18. Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* 1992; **34**: 561–580.
19. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM: Philadelphia, PA, 1998.
20. Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* 1963; **4**: 1035–1038.
21. Dax A. On regularized least norm problems. *SIAM J. Optimization* 1992; **2**: 602–618.
22. Aster RC, Borchers B, Thurber CH. *Parameter Estimation and Inverse Problems*. Elsevier: Amsterdam, 2005.
23. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 1970; **12**: 55–67.
24. Naumova V, Pereverzyev SV. Multi-penalty regularization with a component-wise penalization. Johann Radon Institute for Computational and Applied Mathematics Austrian Academy of Sciences Report 2013-03, https://www.ricam.oeaw.ac.at/publications/reports/13/rep13-03.pdf (assessed 12/2015).
25. Manne R. Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemom. Intell. Lab. Syst.* 1987; **2**: 187–197.
26. Næs T, Isaksson T, Fern T, Davies T. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications: Chichester, UK, 2002.
27. Héberger K. Sum of ranking differences compares methods or models fairly. *Trends Anal. Chem.* 2010; **29**: 101–109.
28. Willett P. Combination of similarity rankings using data fusion. *J. Chem. Inf. Model.* 2013; **53**: 1–10.
29. Kalivas JH, Palmer J. Characterization multivariate calibration tradeoffs (bias, variance, selectivity, and sensitivity) to select model tuning parameters. *J. Chemom.* 2014; **28**: 347–357.
30. Kalivas JH, Héberger K, Andries E. Sum of ranking differences (SRD) to ensemble multivariate calibration model merits for tuning parameter selection and comparing calibration methods. *Anal. Chim. Acta* 2015; **869**: 21–33.
31. Tencate AJ, Kalivas JH, White AJ. *Fusion Processes for Selecting Multiple Tuning Parameters for Multivariate Calibration and Other Penalty Based Processes: A Model Updating Application for Pharmaceutical Analysis*. Submitted to Anal. Chim: Acta December, 2015.
32. Héberger K, Kollár-Hunek K. Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers. *J. Chemom.* 2011; **25**: 151–158.
33. Download address: http://www.isu.edu/chem/people/faculty/kalijohn/ (assessed December 2015).
34. Eigenvector Research, Inc, Wenatchee, Washington, http://www.eigenvector.com/index.htm (assessed December 2015).
35. Wentzel PD, Andrews DT, Kowalski BR. Maximum likelihood multivariate calibration. *Anal. Chem.* 1997; **69**: 2299–2311. http://groupwentzell.chemistry.dal.ca/downloads.html (assessed December 2015).
36. Rácz A, Bajusz D, Héberger K. Consistency of QSAR models: correct split of training and test sets, ranking of models and performance parameters. *SAR QSAR Environ. Res.* 2015; **26**: 683–700.