

Multivariate calibration stability: a comparison of methods

Brian D. Marx^{1*} and Paul H. C. Eilers²

¹Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA

²Department of Medical Statistics, Leiden University Medical Center, NL-2300 RA Leiden, The Netherlands

Received 25 October 2000; Revised 16 August 2001; Accepted 16 August 2001; Published online 18 January 2002

In the multivariate calibration framework we revisit and investigate the prediction performance of three high-dimensional modeling strategies: partial least squares, principal component regression and P-spline signal regression. Specifically we are interested in comparing the stability and robustness of prediction under differing conditions, e.g. training the model under one temperature and using it to predict under differing temperatures. An example illustrates stability comparisons.
Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: multivariate calibration; partial least squares; principal component regression; P-splines; signal regression; transfer

1. INTRODUCTION

Multivariate calibration (MVC) is a fascinating subject. If MVC succeeds for the chemist, then it promises fast, inexpensive and reliable on-line measurements based on spectroscopic methods. Moreover, owing to the ill-posed nature of the problem, there is no shortage of interesting challenges for the statistician. For instance, the spectral regressors are highly correlated and constitute a signal. It is no wonder that MVC has received a great deal of attention in the chemometric and statistical literature. Martens and Næs [1] devoted an entire book to it, and Frank and Friedman [2] gave an in-depth statistical treatment.

Essentially MVC leads to a regression model, providing a (high-dimensional) vector of estimated coefficients. These coefficients weigh the spectrum (or more generally the signal) and provide the estimated response, e.g. a component concentration of interest. To develop a model, a set of calibrated concentrations (the response) and the accompanying spectra are collected at discrete intervals (the regressors). As there are generally many more regressors than training observations, classical linear regression breaks down and special methods have to be developed. The most important ones are the following.

- Ridge regression. A ridge penalty shrinks all coefficients toward zero and reduces the effect of collinearity [3].
- Projection of the spectra on an externally given low-dimensional basis, e.g. B-splines [4,5].
- Principal component regression (PCR). A limited number of singular vectors (the ones with the highest singular values, or alternatively choosing the ones with

the most significant *t*-statistics) of the spectra are used to build a low-dimensional orthogonal basis [6].

- Partial least squares (PLS). A low-dimensional basis is also built, but weights for the orthogonally constructed variables are chosen based on the (partial) correlation between the (residual) regressors and the response [7].
- Penalized signal regression (PSR). PSR introduces a roughness penalty to force the vector of regression coefficients to vary smoothly with the ordered wavelengths, thereby eliminating collinearity. Originally proposed by Hastie and Mallows [8], it was developed into a practical tool by Marx and Eilers [9].

A problem of great practical interest is reliable prediction and the robustness of the MVC training model. The calibration data are generally collected with one instrument under stable conditions. When the model is going to be used on-line, another instrument may be used and/or the conditions may change, which may lead to shifting, warping and scaling of the spectral regressors. This is the problem we term (multivariate) calibration stability (MCS), and it can be even more interesting than MVC itself.

A major goal of this paper is to compare prediction performance among PLS, PCR and PSR methods. Since presenting PSR, we have had a special interest in MCS. One conjecture we had was that the smoothness of the regression coefficients might lead to more stable prediction under a variety of conditions. That is, perhaps the combination of relatively erratic estimated coefficients in PLS/PCR, together with small deformations of spectra, could have a significant impact on prediction performance.

The paper is structured as follows. First we give a short review of PSR. We pay extra attention to the penalty, because we learned that it is more critical than we assumed beforehand. We next present and analyze the performance of PCR, PLS and PSR with different penalties on a data set

*Correspondence to: B. D. Marx, Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA.
E-mail: bmarx@lsu.edu

comprising spectra collected at various temperatures [10]. We end with a discussion of the results.

2. AN INTRODUCTION TO P-SPLINE SIGNAL REGRESSION

We assume that the reader is familiar with PCR and PLS techniques. These techniques have in common that the natural order of the wavelengths is unimportant: the same results will be obtained when these are scrambled arbitrarily. Brown (Reference [11], Chapter 6) took into account contiguous regressors in a Bayesian context. PSR tries to exploit order by forcing the regression coefficients to be smooth. Such an assumption can be quite natural with smooth spectral regressors. We will see below that PSR does not necessarily require smoothness of the spectra, but only that the coefficient vector can be smoothed without doing much harm. Consider modeling the mean response as

$$\mu = \alpha_0 + X_{m \times p} \alpha_{p \times 1} \quad (1)$$

where X is severely ill-conditioned, α_0 is the unknown intercept and α is the unknown spectral coefficient vector. Typically the number of regressors (p) far exceeds the number of observations (m), i.e. $p \gg m$. What is essential to PSR is that it achieves smoothness in α . At first thought one might find it practical to place a roughness penalty directly on α , provided that dimension is not too unwieldy. However, it can also be a good idea to considerably reduce dimension, while preserving smoothness, by first projecting α onto a known (n -dimensional) basis of smooth functions, i.e. $\alpha_{p \times 1} = B_{p \times n} \beta_{n \times 1}$. We prefer a B-spline basis, in part because B-splines are easy to compute and have excellent numerical properties (for details see References [12,13]). The vector β is the unknown vector of basis coefficients of modest dimension. As seen, α is now re-expressed linearly, in fact much like a multiple regression with regressors constrained in a smooth subspace.

Traditionally the major obstacle with B-spline smoothers is the choice of the number and placement of *knots*—that is, the places where the smooth polynomial segments of B-splines join, as well as specify their limited support. Too many (few) knots will lead to overfitting (underfitting), and optimization schemes are complicated non-linear problems. We eliminated the knot placement problem by using P-splines [14], a combination of (a) projecting α onto a B-spline basis using a moderate number of equally spaced knots and (b) further increasing smoothness by imposing a difference penalty on adjacent coefficients in the β vector. Notice that Equation (1) can be rewritten as

$$\mu = \alpha_0 + U_{m \times n} \beta_{n \times 1} \quad (2)$$

where $U = XB$. We now have a moderately sized regression problem, which is by design more flexible than needed. Additional smoothness and regularization come from a difference penalty on adjacent B-spline coefficients:

$$P = \lambda \sum_{k=d+1}^n (\Delta_k^d \beta)^2 \quad (3)$$

where Δ_k^d indicates the k th difference operator of order d . The

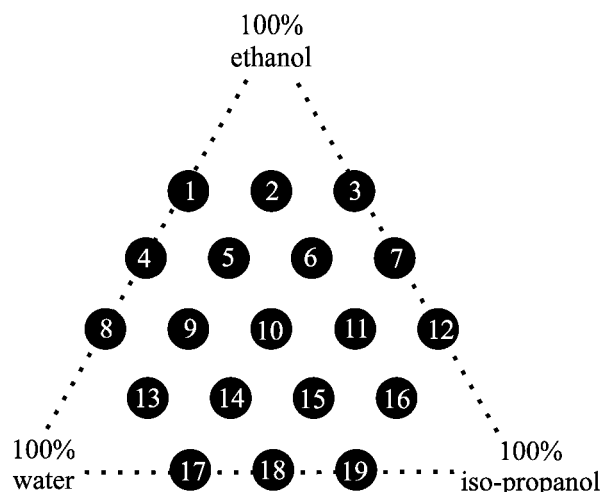


Figure 1. Mixture design for ethanol, water and isopropanol mole fractions.

order of the difference penalty can also moderate smoothing, i.e. increasing d translates into more neighboring B-spline coefficients *holding hands*.

Our preliminary experiments frequently indicated that the PSR-estimated α s were too enthusiastic, with magnitude too large for successful stability. Thus we consider an additional ridge penalty on β , which shrinks the coefficients toward zero. This doubly penalized least squares solution simplifies to, and is convenient to express as,

$$\hat{\beta} = (U'U + \lambda D_d' D_d + \kappa I_n)^{-1} U' y \quad (4)$$

where D_d is an $(n-d) \times n$ banded matrix of contrasts resulting from differencing adjacent rows of the identity matrix (I_n) d times. Note that when $\lambda = \kappa = 0$ we have the unpenalized least squares solution. The non-negative parameters λ and κ regularize and tune the penalties and can be chosen, for example, via a two-dimensional grid search.

Table I. Mole fractions (%) of the samples

Mixture	Ethanol	Water	Isopropanol
1	66.4	33.6	0
2	67.2	16.3	16.5
3	66.6	0	33.4
4	50.0	50.0	0
5	50.0	33.3	16.7
6	49.9	16.7	33.3
7	50.0	0	50.0
8	33.3	66.7	0
9	33.2	50.0	16.7
10	33.3	33.4	33.3
11	32.2	16.6	51.2
12	33.5	0	66.5
13	16.6	66.7	16.7
14	16.7	50.0	33.3
15	16.6	33.3	50.1
16	16.2	16.3	67.5
17	0	66.7	33.3
18	0	50.0	50.0
19	0	33.4	66.6

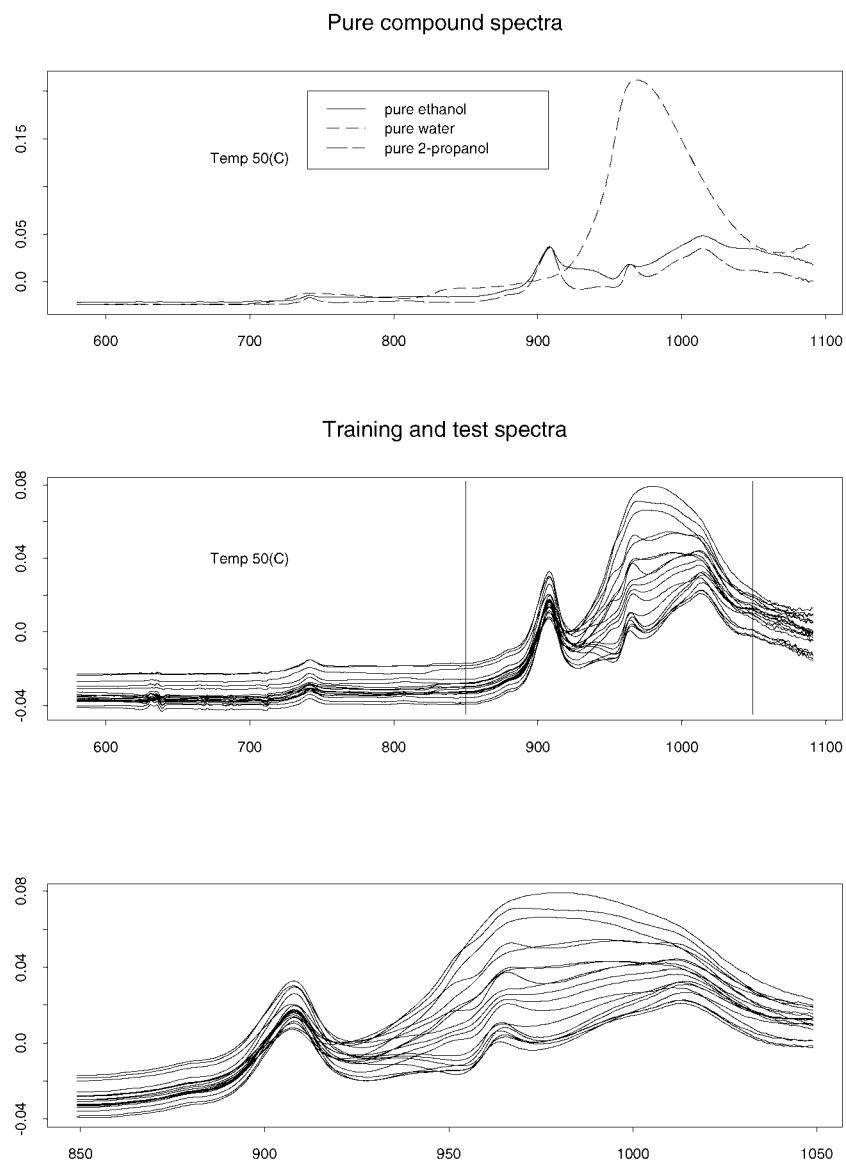


Figure 2. Spectra measured at 50°C. The top panel provides the spectra for the pure components. The middle panel displays the spectra for the 19 mixtures, where the vertical lines indicate the interior region of the spectra used in data analyses. The bottom panel is the same as the middle panel, but just focusing on the interior 200 wavelengths.

A *general recipe* for PSR is given by Marx and Eilers (Reference [9], Section 4). To give an idea of default design parameters, we typically use between 10 and 40 equally spaced cubic B-splines. In fact, in this previous paper we had not highlighted the importance of the choice of the order of the penalty, but the stability results have taught us otherwise. We vary the difference penalty from order $d = 3$ (previous default) to $d = 0$. A difference penalty of order $d = 0$ simplifies to another ridge penalty (on β , not on α), and thus κ can be set to zero. It is important to note that the penalty terms introduce very little extra computational effort, since $U^T U$ and $U^T y$ do not have to be recomputed when the smoothness parameter λ or shrinkage parameter κ is changed. For fixed d and κ , increasing λ makes β smoother (toward a polynomial of degree $d - 1$). For fixed d , optimal λ and κ are searched for systematically by monitoring, for instance, cross-validation prediction error. Results of these

optima can be directly compared over the various $d = 0, 1, 2, 3$. Given the choice of d , λ and κ , then the p -dimensional signal coefficient vector can be constructed, $\hat{\alpha} = B\hat{\beta}$. Since PSR is grounded in classical regression, diagnostics involving hat diagonal information (such as the PRESS statistic), standard error bands for coefficients, and extensions to the generalized linear model are available.

3. THE MIXTURE EXAMPLE

We revisit the experiment presented by Wulfert *et al.* [10], which involves mixtures of ethanol, water and isopropanol prepared according to the design given in Figure 1 and Table I. Details of the experimental equipment and procedures can be found in Reference [10], and the data are available at www-its.chem.uva.nl. The 19 mixtures and three pure

Ethanol: 66.4%, Water: 33.6%, 2-propanol: 0%

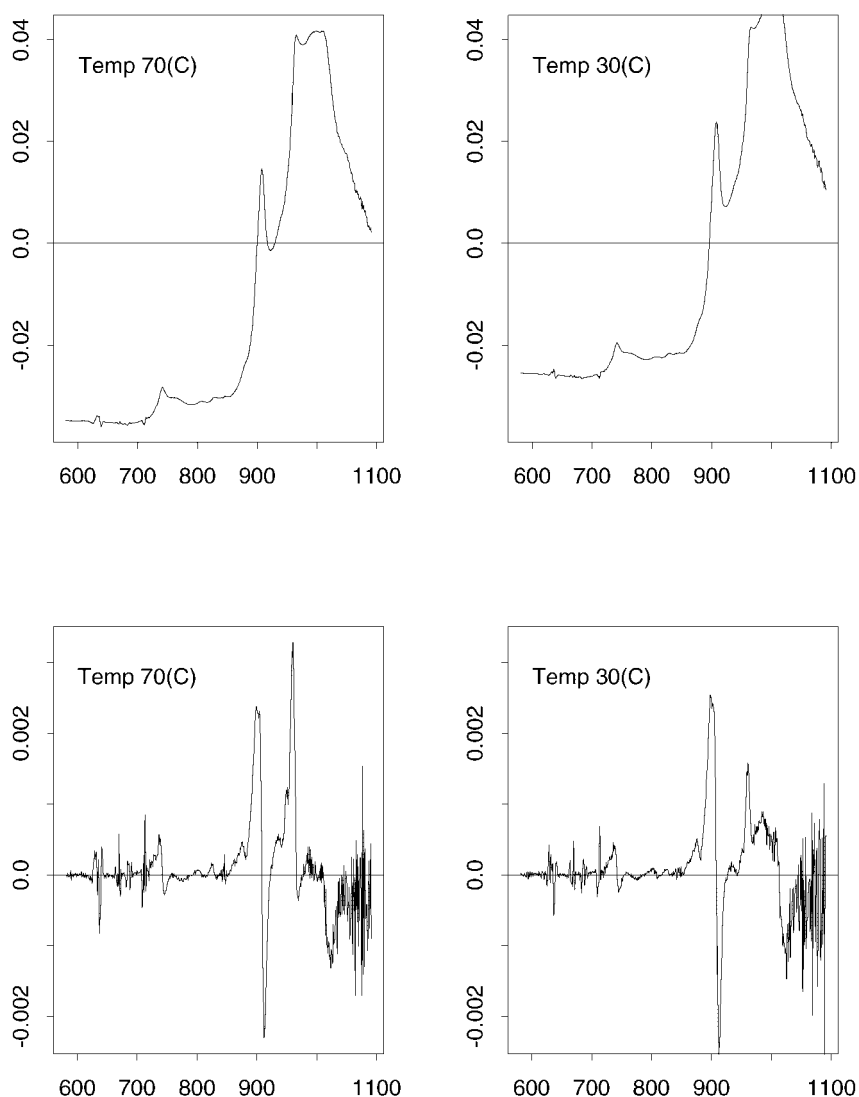


Figure 3. Spectrum of mixture 1 at temperatures of 70 and 30°C (top panels). The bottom panels correspond to the same spectrum as above, but using the first derivative (difference) to remove shifts.

compounds were each measured under several temperature conditions: 30, 40, 50, 60 and 70°C ($\pm 0.2^\circ\text{C}$).

Figure 2 displays the 22 spectra measured at 50°C, which are the discrete (512) representation of measured short-wave near-infrared spectra ranging from 580 to 1091 nm in 1 nm intervals. There is a separate spectrum for each mixture and temperature combination (in total $22 \times 5 = 110$). For absorption and noise reasons, all data analyses were performed using only an interior region of 200 wavelengths, i.e. 850–1049 nm. The spectra were pretreated to remove instrumental baseline drift.

Figure 3 (top panels) displays the spectrum for mixture 1 (ethanol 66.4%, water 33.6%, isopropanol 0%) at the two extreme temperatures (30 and 70°C). We clearly see some shifting of the spectrum across temperature. Although it is not presented, this is in fact the general case as temperature varies within any one of the 19 mixtures or even within the three pure components. The bottom panels of Figure 3

display the first-derivative spectrum (199 digitizations), which can effectively remove shifts that may not be important for modeling.

For a moment, consider situations where the instrumental precision improves, e.g. the number of columns in X grows by an order of magnitude within a fixed band of wavelengths. Despite such an increase in dimension, the increase in the relevant detail within a constructed signal may be marginal, i.e. Figures 2 and 3 could look approximately the same. PSR estimation is not burdened by such increased precision, since the dimension of U does not change. However, PLS and, especially, PCR computations are significantly taxed.

3.1. Training and testing the mixture example

To start, we consider models to predict % ethanol that are trained using only spectra at 50°C, and then see how stable prediction is using spectra at different temperatures. This is

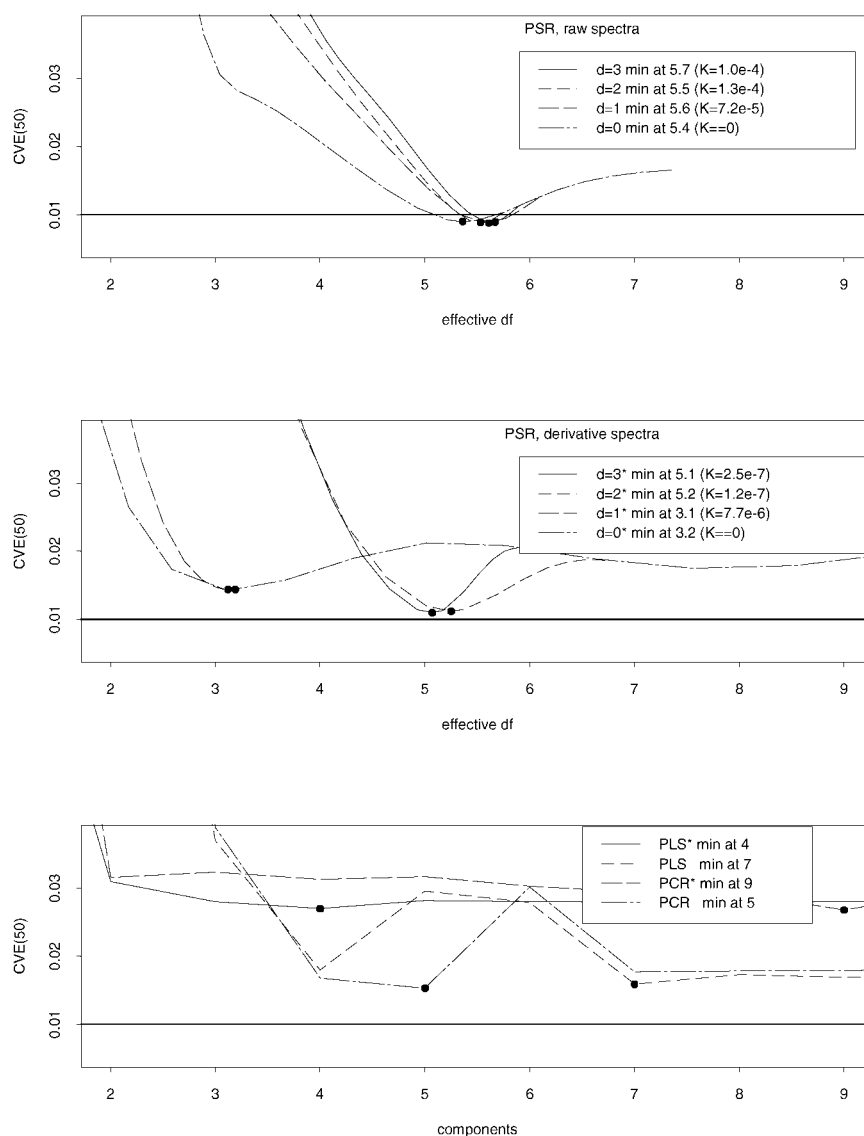


Figure 4. Optimization of dimension. The top and middle panels display CVE_{50} as a function of PSR effective degrees of freedom when using the raw and first-derivative (*) spectra respectively. The λ parameter is varied here at the optimal κ . The bottom panel is a similar plot, but now with the abscissa corresponding to the number of components in PLS or PCR.

how we define the MCS scenario. PLS, PCR and PSR models are built using the 200 (199) raw (first-derivative) spectral regressors (from 850 to 1049 nm) at 50°C. As in Reference [10], models are first fit for optimization purposes using the mixtures on the edge of the experimental design (mixtures 1–4, 7, 8, 12, 13, 16–19) and in the center (mixture 10); see Figure 1. The number of components in PLS and PCR or (λ , κ) in PSR is optimized by using a validation set consisting of the remaining mixtures in the inner ring (mixtures 5, 6, 9, 11, 14, 15). As seen, no extrapolation is done during this optimization. Specifically we search for model components or (λ , κ) that minimize the cross-validation error.

$$CVE = \left(\frac{1}{m_C} \sum_{i \in C} (y_i - \hat{y}_i)^2 \right)^{1/2} \quad (5)$$

where C is the calibration validation set of size m_C , and \hat{y}_i is

the predicted response using a validation spectrum in the fitted model described above. The final training measure can be defined in this context as model is the one fit to all 19 mixtures using the optimal components or (λ , κ), based on minimizing Equation 5. In Section 3.2 we will evaluate how this training model does in terms of prediction stability when conditions vary.

Some behind-the-scenes details now follow. The PLS/PCR components are centered and of unit length, so as to have estimation that is orthogonal to the intercept term. For PLS the response vector is also normalized. Given the number of components, (α_0 , α) are estimated. The six validation observations are standardized in an obvious way using the standardization information from the fitted model with 13 observations. Thus, given estimated ($\hat{\alpha}_0$, $\hat{\alpha}$), new predicted values are produced. Under transfer conditions we use 19 new spectral regressors, say from 30°C, and these new

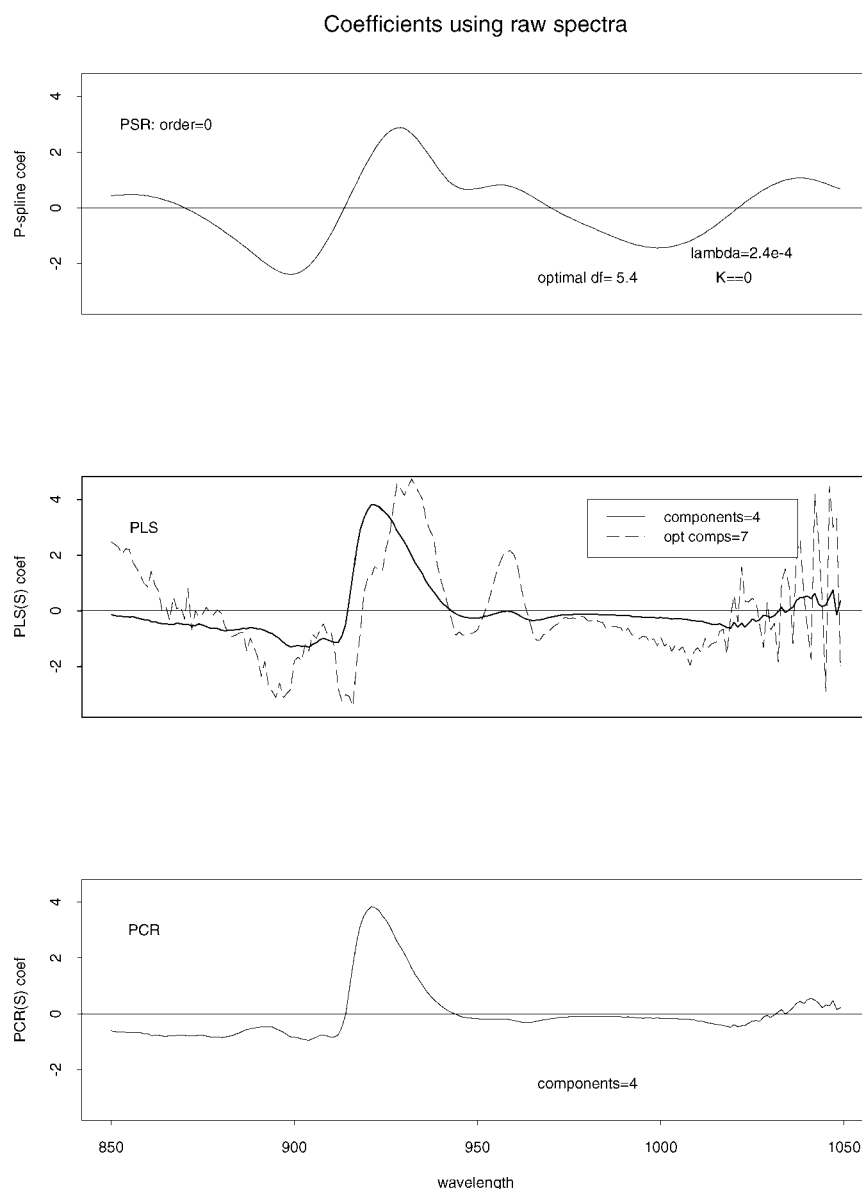


Figure 5. Estimated spectral coefficients using raw spectra. The top panel corresponds to PSR estimates for difference penalty $d=0$. The middle and bottom panels correspond to PLS and PCR estimates respectively. The effective df or number of components was optimized using CVE_{50} . All panels have the same scale.

spectral regressors are again standardized in an obvious way using the optimally trained standardization information. Exact PLS/PCR/PSR cross-validation measures are presented. For PSR, similar strategies were taken, except standardization is not an issue, since the raw (or first-derivative) spectra are used directly. All PSR models were fit using cubic B-splines on 15 equally spaced interior knots; this produced amply flexibility, since the effective df (degrees of freedom) was under 10 for all PSR models.

Figure 4 (top and middle panels) displays profiles of CVE_{50} (using the optimal κ slice) as a function of effective df for PSR. The PSR models were trained using either the raw or first-derivative spectral regressors (varying the penalty order from $d=0$ to 3). In the top panel, PSR using the raw spectra shows that approximately the same CVE_{50} (0.0089 in units of % ethanol) is achieved for all $d=0, 1, 2, 3$ with rather

stable effective df (5.6). In the central panel, d is affecting optimization a bit more with the first-derivative spectra, with roughly the same CVE_{50} (0.011) for $d=3$ and 2. Again, a similar CVE_{50} (0.014) is found for $d=1$ and 0. The first-derivative spectral PSR optimal effective df ranges from approximately 3.1 to 5.2. The effective degrees-of-freedom measure can be defined in this context as

$$\text{effective } df(\lambda, \kappa, d) = \text{trace}[U'U(U'U + \lambda D'_d D_d + \kappa I_n)^{-1}]$$

and is an idea borrowed from Hastie and Tibshirani [15].

The bottom panel of Figure 4 shows CVE_{50} as a function of the number of components in PLS/PCR models. Comparing with the middle and top panels, note that we are beginning to see some evidence that PSR may be a reasonable alternative strategy, with an approximately 30% reduction

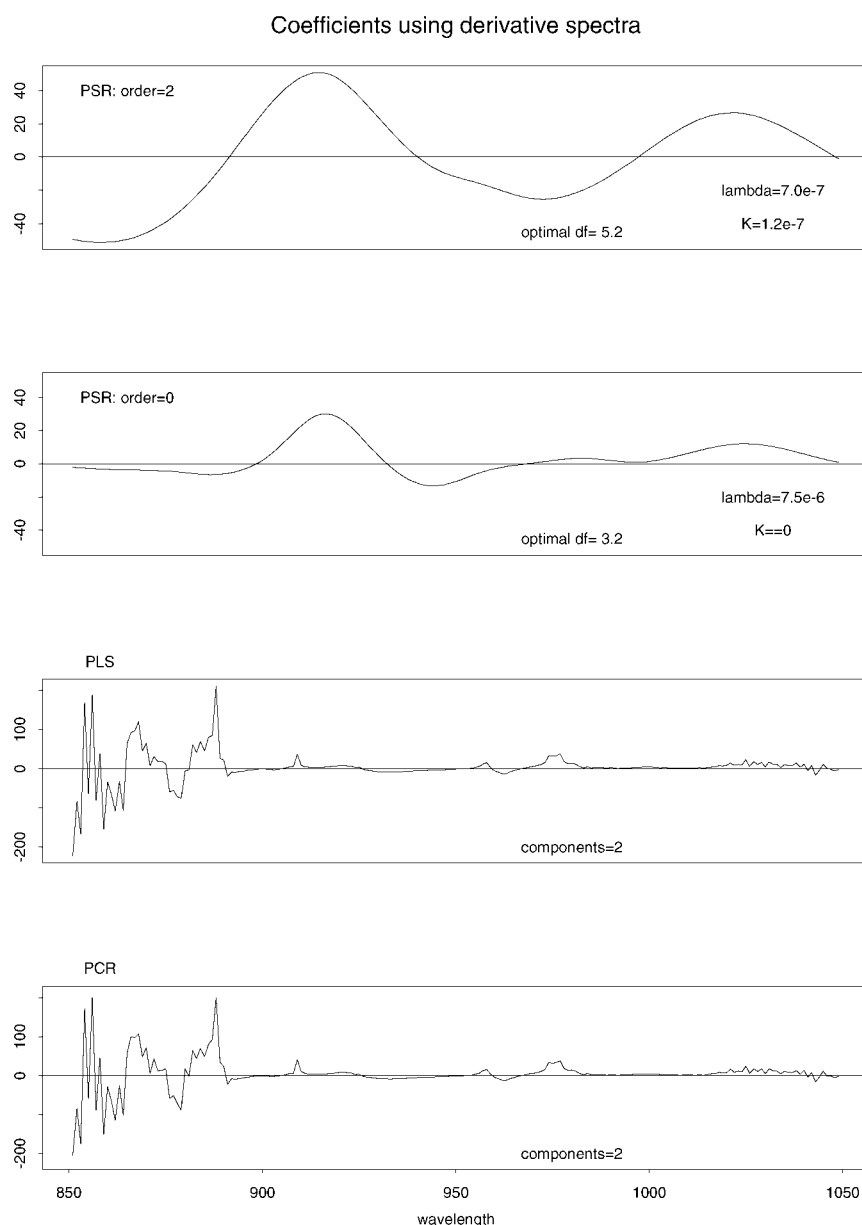


Figure 6. Estimated spectral coefficients using first-derivative spectra. The top two panels correspond to PSR estimates (on the same scale) for penalty orders $d = 2$ and 0 respectively. The bottom two panels correspond to PLS and PCR estimates respectively (on the same scale, but different from the PSR scale). The effective df or number of components was optimized using CVE_{50} .

in CVE_{50} relative to PLS/PCR. The PLS and PCR results with the first-derivative signal regressors show considerably higher optimal CVE_{50} (approximately 0.027) than that achieved using raw signals (approximately 0.015). The optimal dimension ranges from approximately two to nine for PLS/PCR models. We keep in mind that the data reflect ternary mixtures only. Based on Figure 4, we thus chose four components for both PLS and PCR raw spectral training models, but chose two components for both PLS and PCR first-derivative spectra. These decisions were also made in light of the fact that exaggerated numbers of components (e.g. seven) lead to rather rough estimated coefficient vectors and damaged prediction performance; see the middle panel of Figure 5.

We would like to caution that one should not necessarily compare directly the PSR effective dimension to the number of PLS/PCR components. The actual counterpart for the PLS/PCR effective dimension is not easily obtainable. PLS and PCR coefficient plots can sometimes suggest great detail, but some care must be taken, since these estimators are really constructed using a combination of a few low-dimensional, sometimes rough, basis vectors. The truth is that it is impossible to get several hundred meaningful coefficients from only 19 observations; the data themselves have little power to steer such enormous dimensions. We now see that PSR does not allow the coefficient vector to present any more detail than the data permit.

Figure 5 displays the estimated coefficient vectors when

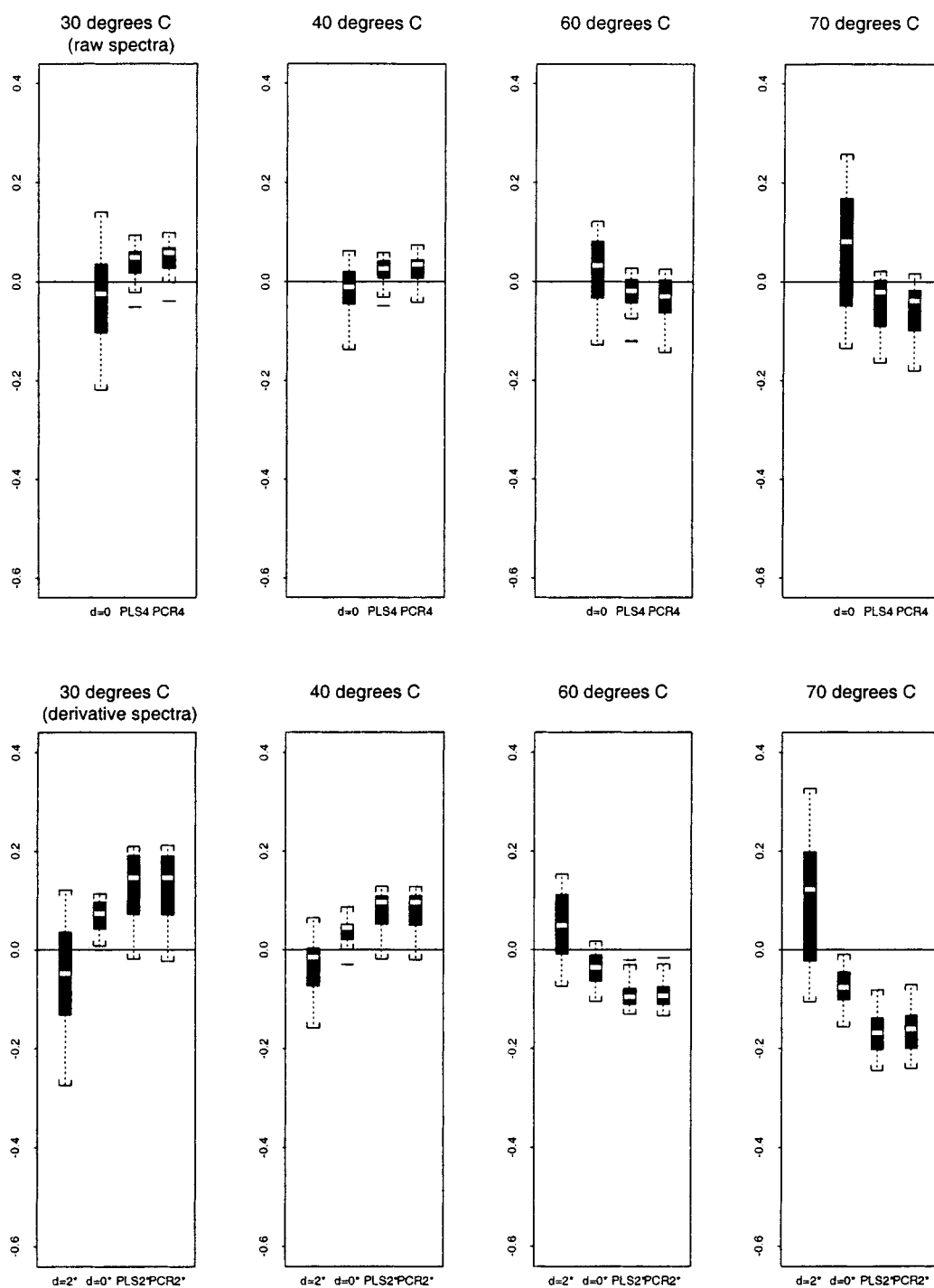


Figure 7. Stability results for models trained at 50°C (optimized using the original six-observation test set). The box plots summarize the 19 values of $\hat{\mu} - y$, by method and by temperature. An asterisk (*) denotes that the first-derivative spectra are used.

the raw spectra are used as regressors. The top panel corresponds to the estimated smooth PSR signal coefficient vectors for order $d=0$. Higher orders of the difference penalty d resulted in a relatively dominant κ (e.g. $d=2$, $\lambda = 2.2 \times 10^{-5}$, $\kappa = 1.3 \times 10^{-4}$) and thus essentially reproduced the $d=0$ result. The displayed PSR coefficient vector is essentially a (smooth) representation of the 200-dimensional α for the spectra. The middle and bottom panels of Figure 5 correspond to estimated coefficient vectors using PLS and

PCR respectively, which are surprisingly smooth along the wavelength index and have similar features to the PSR coefficients. This is particularly noteworthy since, as mentioned, both PLS and PCR are essentially built using a (rough) low-dimensional basis. The middle panel also provides the estimated PLS coefficient vector with seven components (optimal in CVE, but overfit). In all three panels we find some evidence of a positive effect near 920 nm.

Figure 6 is similar in presentation to Figure 5, but

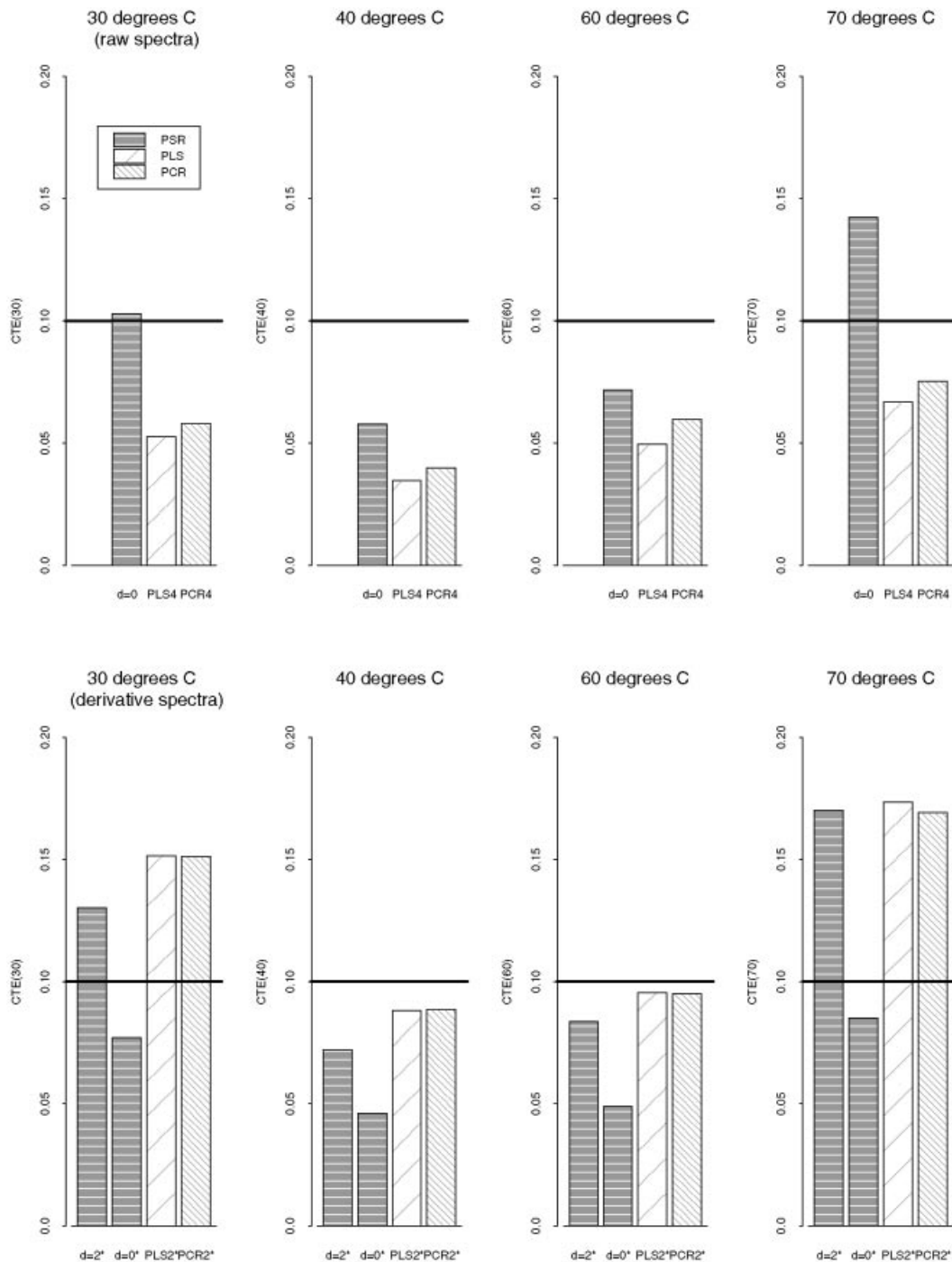


Figure 8. CTE results for models trained at 50°C (optimized using the original six-observation test set), by method and by temperature. An asterisk (*) denotes that the first-derivative spectra are used.

Table II. Summary of the optimal models trained at 30°C using the alternative seven-observation test set. An asterisk (*) denotes that the first-derivative spectra were used

Method	Effective <i>df</i> or components	Minimum CVE_{30}	(λ, κ)
PSR ($d = 2$)	5.88	0.005	$(3.6 \times 10^{-5}, 1.0 \times 10^{-8})$
PSR ($d = 1$)	5.95	0.006	$(2.8 \times 10^{-5}, 1.0 \times 10^{-16})$
PLS	5 (4)	0.009 (0.015)	
PCR	10 (4)	0.010 (0.015)	
PSR* ($d = 2$)	4.86	0.006	$(1.0 \times 10^{-6}, 4.6 \times 10^{-8})$
PSR* ($d = 1$)	4.89	0.008	$(7.7 \times 10^{-7}, 2.7 \times 10^{-8})$
PLS*	4 (2)	0.023 (0.031)	
PCR*	10 (2)	0.022 (0.034)	

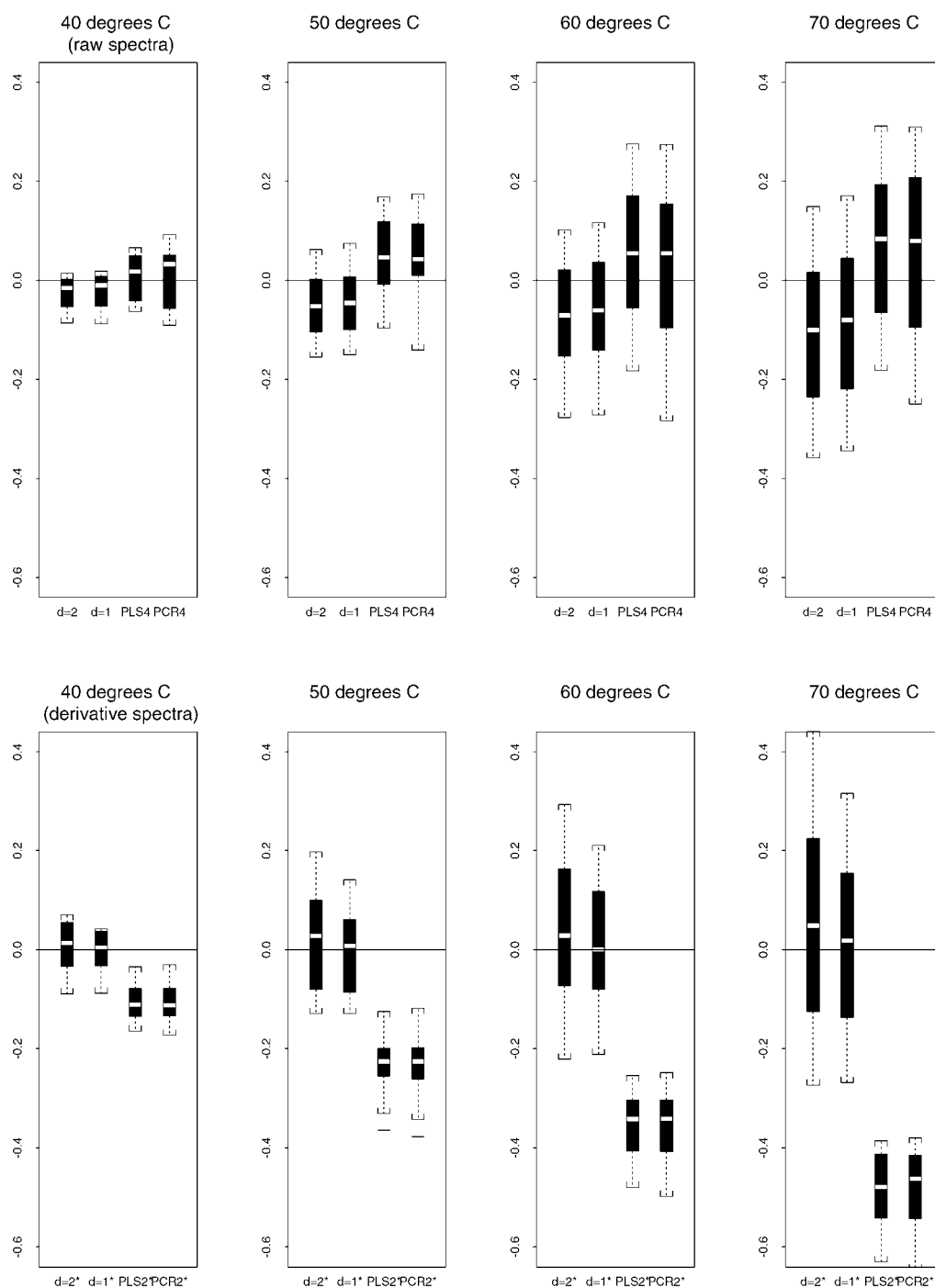


Figure 9. Stability results for models trained at 30°C (optimized using the alternative seven-observation test set). The box plots summarize the 19 values of $\hat{\mu} - y$, by method and by temperature. An asterisk (*) denotes that the first-derivative spectra are used.

corresponds to the estimated coefficients when the first-derivative spectra are used. The top two panels correspond to PSR estimates for varying penalty, $d = 2$ and 0. To save space, the PSR estimate using $d = 3$ is not presented, but is very similar to the one using $d = 2$. There was a striking similarity between $d = 1$ ($\lambda = 3.6 \times 10^{-7}$, $\kappa = 7.7 \times 10^{-6}$) and $d = 0$ ($\lambda = 7.5 \times 10^{-6}$, $\kappa = 0$), again due to the relatively dominant influence of the ridge penalty. Hence $d = 1$ is also suppressed from the figure. In the case of $d = 2$, again we find some evidence of positive effects at approximately 920 and

1025 nm, and a possible negative effect is seen in the region of 970 nm. The $d = 0$ panel shows considerable dampening of the estimated coefficients. Notice that PSR coefficients emphasize approximately the same regions of wavelengths, regardless of using raw or first-derivative spectra. This is not the case for PLS/PCR. In fact, some PLS/PCR estimates are now of much higher magnitude when compared with PSR estimates, and thus are presented on a different scale. We now find the PLS/PCR estimates to be very rough, but, unlike the previous Figure 5, now in the lower-wavelength region.

3.2. Calibration stability results for the mixture experiment

We now have several optimal models trained with spectral regressors at 50°C: PSR (raw spectra at $d = 0$, first-derivative spectra at $d = 0$ and 2), PLS (raw and first-derivative spectra) and PCR (raw and first-derivative spectra). We are next interested in seeing how well prediction compares across these particular models when using spectral regressors generated under different temperature conditions, e.g. 30, 40, 60 and 70°C. Figure 7 displays box plots of the 19 transfer residuals $\hat{\mu} - y$, by method and by transfer temperature. We additionally define the calibration transfer error

$$\text{CTE} = \left(\frac{1}{m_T} \sum_{i \in T} (y_i - \hat{\mu}_i)^2 \right)^{1/2} \quad (6)$$

where T is the transfer set of size m_T , and $\hat{\mu}$ is a predicted response using the trained model (at 50°C) with spectra generated at either 30, 40, 60 or 70°C. The top (bottom) panels of Figure 8 display the CTE using the raw (first-derivative) spectra, by temperature. Fixing any one method in Figure 7 or 8, notice that the more extreme temperatures (e.g. 70°C) tend to have more variability, are further off target and have a higher CTE than the ones closer to the trained conditions (e.g. 40°C).

When focusing on the raw spectral results (top panels), PLS appears to be highly competitive among methods with respect to CTE and at all temperatures. It is also surprising how well PCR performs given that it does not take into account the response information when constructing variables. Recall from Figure 5 that the both the PLS and PCR coefficient vectors are rather smooth (using four components). In Section 3.3 a limited sensitivity analysis is provided.

A somewhat different situation arises when the first-derivative spectra are used (indicated by an asterisk in the bottom panels of Figures 7 and 8). It is interesting to note that first-derivative spectra boil down to very rough signal regressors, and we observe that PLS and PCR performances are now not as competitive compared with PSR. It is also true that now both PLS and PCR have very rough estimated coefficient vectors, even using a relatively low number of components (refer to Figures 3 and 6). We observe that PLS and PCR have a strong bias that seems to systematically change with temperature. Perhaps there may be some opportunity for a remedy using a correction factor, but we do not pursue this further. As for PSR, we find that varying the order of the penalty can significantly tune prediction. Based on CTE, we also find that PSR is clearly a strong competitor to both PLS/PCR (in the case of $d = 0$ with approximately 50% reduction).

3.3. A snapshot of sensitivity

The previous analyses were trained using the 50°C spectra. The 13 training mixtures were the ones comprising the outer ring plus the center mixture in Figure 1. Hence the six-observation validation set consisted of the inner ring of mixtures. In the interest of sensitivity, models were alternatively built using the 30°C spectra (the lowest temperature), further using different fitting and test sets. The new set

to build models for optimization purposes has 12 mixtures (1, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 19), and thus the remaining seven mixtures (2, 4, 7, 10, 13, 16, 18) comprise the validation set. Notice that this alternative validation set is more of a compromise between interpolation and extrapolation than the original one. PSR (15 equally spaced interior knots), PLS and PCR models were constructed using both the raw and first-derivative spectra. Using Equation 5, the optimal dimension was determined by minimizing CVE_{30} , now based on $m_C = 7$. Table II provides a summary of optimal models, as well as other PLS/PCR candidate models that do not overfit the ternary nature of the experiment. Note that PSR with $d = 0$ is not presented, since in this case the search for optimal λ went to the boundary of zero (for both the raw and first-derivative spectra).

Figure 9 displays box plots of the 19 transfer residuals $\hat{\mu} - y$, by method and by transfer temperature. In general, the results are consistent with Figures 7 and 8. For a fixed method the CTE increases in monotone fashion when going from 40 toward 70°C. We now see some evidence that stability worsens at 70°C here, since evaluation of the model is 40°C (not 20°C) away from the training conditions. When the raw spectra are used (top panels), notice that PSR, PLS and PCR all have generally comparable and good performance. Perhaps the shifts inherent in the raw spectra provide useful predictive information for the PLS and PCR methods. For the first-derivative spectra we again find notable bias for PLS and PCR. However, the PSR results do not differ considerably with either the raw or first-derivative spectra, at all temperatures, and especially when d is small.

4. DISCUSSION

We realize that the results presented only provide some empirical evidence for stability comparisons during transfer; certainly further comparisons need to be done. However, some consistent and interesting insights have surfaced for PSR from these mixture data: (a) PSR validates very well under the training conditions; (b) PSR appears to transfer better with lower than with higher d ; and (c) PSR appears somewhat more satisfactory than PLS/PCR with rough spectra. In all fairness, there are some subjective procedures in all these methods. For example, in PLS/PCR we did not strictly pick the number of components based on minimizing CVE, but rather chose four or less components owing to the ternary nature of the mixtures. Such natural dimension restriction may be more difficult to implement in PSR, and perhaps this is the reason for the relatively large influence of the ridge penalty. Further, in PSR the choice of d can possibly mislead the researcher, i.e. the optimal trained model might suggest high d , but transfer results may be better for low d , as indicated in the middle panel of Figure 4 and the bottom panels of Figure 8. We hope that readers will consider PSR not only as a competitor for high-dimensional prediction models, but also as practical tool when calibration transfer scenarios arise. PSR S-plus functions, which also accommodate the generalized linear model setting, are available at www.stat.lsu.edu/bmarx.

Acknowledgements

The authors would like to thank the editor Age Smilde and two anonymous referees for their helpful suggestions to improve this paper, and Elizabeth Swoope for her help with Figure 1.

REFERENCES

1. Martens H, Næs T. *Multivariate Calibration*. Wiley: New York, 1989.
2. Frank IE, Friedman JH. A statistical view of some chemometric regression tools (with discussion). *Technometrics* 1993; **35**: 109–148.
3. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 1970; **12**: 55–67.
4. Alsberg BK. Representation of spectra by continuous functions. *J. Chemometrics* 1993; **7**: 177–193.
5. Denham MC, Brown PJ. Calibration with many variables. *Appl. Statist.* 1993; **42**: 515–528.
6. Webster JT, Gunst RF, Mason RL. Latent root regression analysis. *Technometrics* 1974; **16**: 513–522.
7. Wold H. Soft modeling by latent variables: the nonlinear iterative partial least squares approach. In *Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett*. Gani J (ed.). Academic Press: London, 1975; 117–142.
8. Hastie T, Mallows C. A discussion of 'A statistical view of some chemometrics regression tools' by I. E. Frank and J. H. Friedman. *Technometrics* 1993; **35**: 140–143.
9. Marx BD, Eilers PHC. Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics* 1999; **41**: 1–13.
10. Wülfert F, Kok W, Smilde A. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal. Chem.* 1998; **70**: 1761–1767.
11. Brown PJ. *Measurement, Regression, and Calibration*. Clarendon: Oxford, 1993.
12. De Boor C. *A Practical Guide to Splines*. Springer: New York, 1978.
13. Dierckx P. *Curve and Surface Fitting with Splines*. Clarendon: Oxford, 1993.
14. Eilers PHC, Marx BD. Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statist. Sci.* 1996; **11**: 89–121.
15. Hastie T, Tibshirani R. *Generalized Additive Models*. Chapman and Hall: London, 1990.