



# Multivariate calibration with temperature interaction using two-dimensional penalized signal regression

Paul H.C. Eilers<sup>a,\*</sup>, Brian D. Marx<sup>b</sup>

<sup>a</sup>*Department of Medical Statistics, Leiden University Medical Center, 2300 RA, Leiden, The Netherlands*

<sup>b</sup>*Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA*

Received 15 October 2002; accepted 12 February 2003

## Abstract

The Penalized Signal Regression (PSR) approach to multivariate calibration (MVC) assumes a smooth vector of coefficients for weighting a spectrum to predict the unknown concentration of a chemical component. B-splines and roughness penalties, based on differences, are used to estimate the coefficients. In this paper, we extend PSR to incorporate a covariate like temperature. A smooth surface on the wavelength–temperature domain is estimated, using tensor products of B-splines and penalties along the two dimensions. A slice of this surface gives the vector of weights at an arbitrary temperature. We present the theory and apply multi-dimensional PSR to a published data set, showing good performance. We also introduce and apply a simplification based on a varying-coefficient model (VCM).

© 2003 Elsevier Science B.V. All rights reserved.

*Keywords:* Calibration transfer; Multivariate calibration; P-splines; Signal regression; Stability; Tensor products; Varying-coefficient models

## 1. Introduction

The typical multivariate calibration (MVC) problem is this: for a number of chemical samples, optical spectra are obtained, as well as concentrations of an analyte; from these data one wishes to derive a vector of coefficients for predicting unknown concentrations given a measured spectra. Generally, the number of samples in the training data is less than 100, while the spectra are measured on several hundreds or even over a thousand wavelengths; thus the problem is inherently ill posed. See Ref. [1] for an extensive presen-

tation and Ref. [2] for a discussion from a statistical point of view.

Two general approaches have been used to make the MVC problem well posed: (a) reduction of the regression bases, and (b) penalized estimation. The first approach can use, for example: principal component regression (PCR), partial least squares regression (PLSR), or projection onto B-splines. Penalized regression comes in many forms, two of which are: (i) ridge regression, which shrinks the regression coefficients towards zero, and (ii) penalized signal regression (PSR), which forces the vector of coefficients to vary smoothly with wavelength [3].

Most applications of MVC have a static flavor: the operating conditions are assumed to be more or less constant. In practice, this might not be the case. Wülfert et al. [4] studied the effect of changing

\* Corresponding author.

*E-mail addresses:* [p.eilers@lumc.nl](mailto:p.eilers@lumc.nl) (P.H.C. Eilers), [bmarx@lsu.edu](mailto:bmarx@lsu.edu) (B.D. Marx).

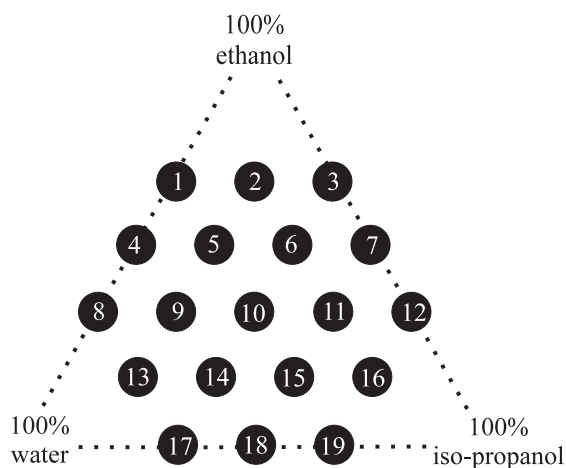


Fig. 1. Mixture design for ethanol, water, and isopropanol mole fractions.

temperature on the predictive ability of an MVC model. Using the same data, Marx and Eilers [5] made a systematic comparison of PCR, PLSR and PSR. They also coined the name *multivariate calibration stability* for the systematic analysis of an MVC model that is trained under one condition, then monitored for prediction performance under changing operating conditions. This might also include calibration transfer: e.g. developing a model on one instrument (in the laboratory) and using it on another instrument (in a production environment).

A logical further development is to extend the model by including additional covariates (like temperature or pressure) in a systematic way, thereby hoping to improve performance. In this paper, we report on a first step in that direction. Using the data of Ref. [4], we extend PSR to include temperature information. Our strategy is to introduce a *coefficient surface*, defined on the two-dimensional wavelength–temperature domain. At a specific temperature, one cuts through this surface to get the “classical” MVC regression coefficient vector with which to weigh a spectrum. We assume the surface to be smooth (in both the directions of wavelength and temperature) and estimate it with a two-dimensional extension of PSR, based on tensor products of B-splines and appropriate roughness penalties. We refer to this extension as TPSR.

The way that we estimate the surface allows complicated interactions of wavelength and temper-

ature. The actual results indicate a less complicated structure, and we develop simpler models that implement the ideas of varying-coefficient models (VCM) [6].

Although we use an example with changing temperature, our approach is also applicable to changes in time, to correct for instrumental drift, assuming that it is possible to analyze calibration samples at regular intervals.

In Section 2, we discuss the data structure of a mixture experiment that has motivated this research. A recap of the one-dimensional PSR approach is given in Section 3. Tensor-product B-splines in a nutshell are presented in Section 4, followed by our proposed two-dimensional PSR extension in Section 5. The results of this extension applied to the example are given in Section 6, and we close with a discussion in Section 7.

## 2. The motivating example

Wülfert et al. [4] presented an experiment that involved mixtures of ethanol, water and isopropanol prepared according to the design given in Fig. 1 and Table 1. Specific details can be found in their article, and the data are available at [www-its.chem.uva.nl](http://www-its.chem.uva.nl).

Table 1  
Mole fractions (%) of the samples

	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

Each of the 19 mixtures, as well as the three pure compounds had measured spectra under several temperature conditions: 30, 40, 50, 60, and 70 °C ( $\pm 0.2$  °C), which were short-wave near-infrared spectra ranging from 580 to 1091 nm, by 1 nm. There is a separate spectrum for each mixture and temperature combination (in total  $(19+3)\times 5=110$ ), and Ref. [7] attempted to take temperature information into account. Due to absorption and noise, all data ana-

lyses are performed only using an interior region of 200 wavelengths, i.e. 850–1049 nm. Fig. 2 provides the spectra for the middle temperature 50 °C. The top panel displays raw spectra for pure components. The second panel displays the raw spectra for the 19 mixtures, where the vertical bars indicate the central 200 wavelength region, which is further displayed in the third panel. Finally, the bottom panel displays the spectra used in the analyzes to follow: first-derivative

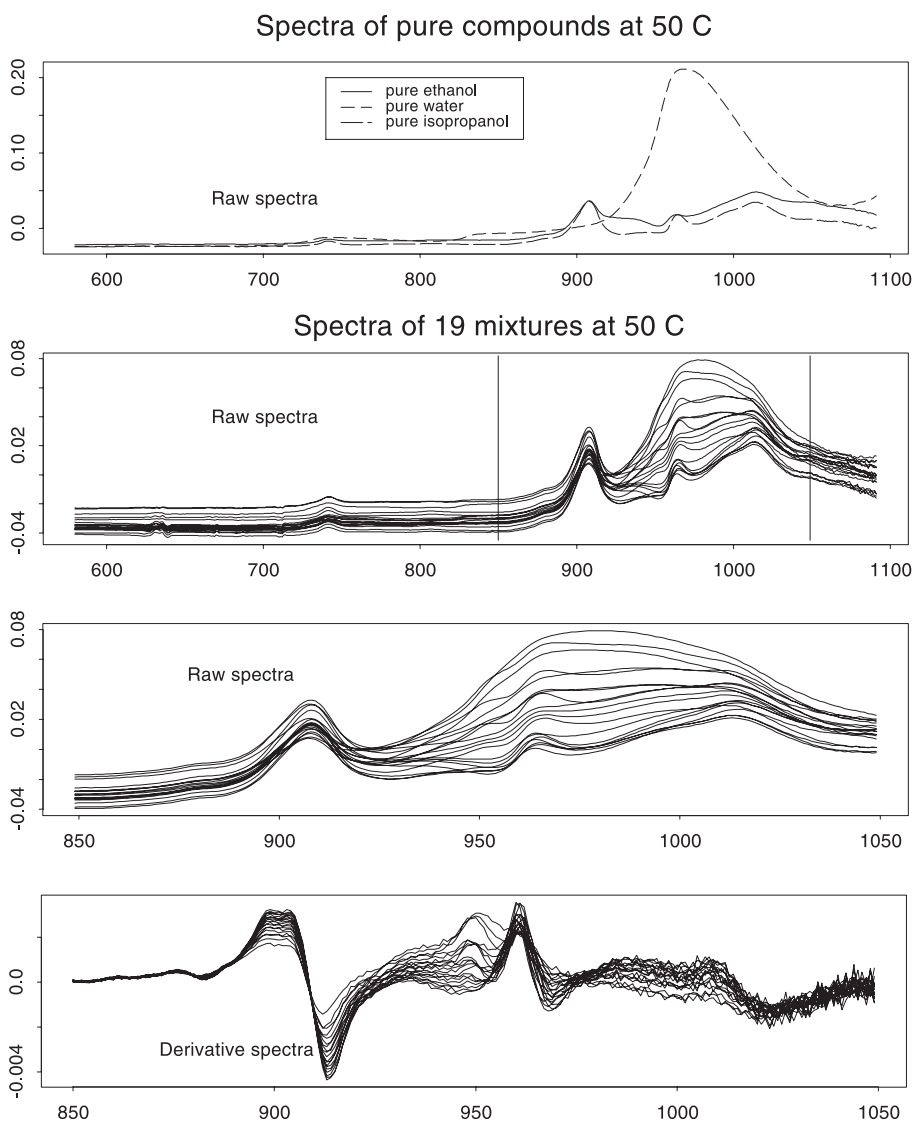


Fig. 2. Measured spectra at 50°. The top panel provides the raw spectra for the pure components. The second and third panels display the raw spectra for the 19 mixtures. The bottom panel gives derivative spectra.

spectrum (199 digitizations), which can effectively remove trends that may not be important to modeling.

Fig. 3 gives an impression of how the spectra vary over mixture and temperature. The four panels correspond to each of the four mixtures: #2, 10, 13, 16, the center, plus the three mixtures that are closest to each of the design corners. Within each panel, spectra for three temperatures 30°, 50° and 70° are given. The raw spectra (rather than the derivative) are presented here since they gave a clearer impression of potential shifting, warping or scaling.

Other approaches have appeared in the literature to correct for the influences of temperature with these mixture data. Wülfert et al. [7,8] investigated linear corrections and continuous piecewise corrections, and Swierenga et al. [9] developed calibration models that are robust to external variations in temperature. Our proposed method is highly competitive to these other approaches and allows for complex interactive relationships between wavelength and temperature, while using data at all temperatures simultaneously.

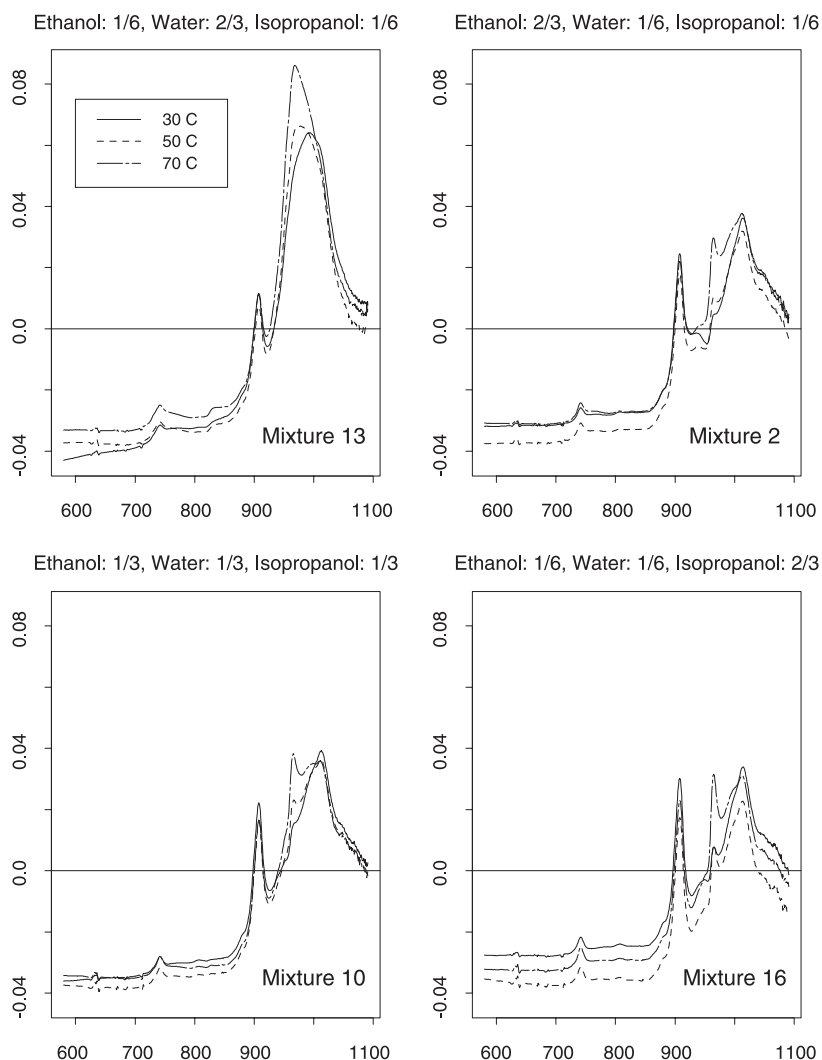


Fig. 3. Spectrum of four mixtures, each at three different temperatures.

### 3. Recap: P-spline signal regression (PSR)

Marx and Eilers solved the standard multivariate calibration problem with penalized signal regression (PSR): forcing the coefficients to be smooth. Consider modeling the mean response  $E(Y)=\mu$  as

$$\mu_{m \times 1} = \alpha_0 \mathbf{1}_m + X_{m \times p} \alpha_{p \times 1}, \tag{1}$$

where  $\alpha_0$  is the intercept,  $X$  is the matrix of digitized spectra, and  $\alpha$  is the unknown coefficient vector. As mentioned, 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  $\alpha$ , by exploiting the natural order of the wavelengths. Dimension can be considerably reduced, while preserving smoothness, by first projecting  $\alpha$  onto a known ( $n$ -dimensional) basis of smooth functions, i.e.  $\alpha_{p \times 1} = B_{p \times n} \beta_{n \times 1}$ . Our choice is a B-splines basis, which has local influence and is stable and efficient computationally. The vector  $\beta$  is the unknown vector of basis coefficients of modest dimension. We re-express  $\alpha$  linearly and constrained in a smooth subspace. We stress that the spectra coefficients, and not the spectra, are smoothed. PSR chooses  $\beta$  to minimize

$$S(\alpha_0, \beta) = |y - \alpha_0 - XB\beta|^2 + \lambda_d |D_d \beta|^2 + \lambda_0 |\beta|^2,$$

with explicit solution

$$\hat{\beta} = (U^*{}'U^* + \lambda_d D_d^*{}'D_d^* + \lambda_0 I^*)^{-1} U^*{}'y, \tag{2}$$

where  $U=XB$ ,  $U^*=(\mathbf{1}_m | U)$  are the effective regressors to estimate  $(\alpha_0, \alpha)$ ,  $D_d^*=(0 | D_d)$ , and  $I^*=\text{diag}(0, I_n)$ . The notation  $|$  denotes column augmentation of matrices and the augmented zeros in  $D_d^*$  and  $I^*$  ensures that the intercept is unpenalized. The nonnegative tuning parameters  $(\lambda_d, \lambda_0)$  regularize the influence of the penalty: the larger the  $\lambda$  the smoother the  $\beta$ . Choice of optimal  $\lambda$  is discussed further below and again in Section 5. PSR uses a “moderate” number of equally spaced knots, and then further increasing smoothness by imposing a difference penalty on adjacent B-spline coefficients  $\beta$ , which provides further smoothness with continuous control. The  $(n-d) \times n$  difference penalty matrix  $D_d$  has a banded structure of  $d$ th order polynomial contrasts. The first difference of  $\beta$ ,  $\Delta^1 \beta$  is the vector with elements  $\beta_{j+1} - \beta_j$ , for  $j=1, \dots, n-1$ . By

repeating this computation on  $\Delta \beta$ , we arrive at higher differences like  $\Delta^2 \beta$  and  $\Delta^3 \beta$ . The  $(n-1) \times n$  matrix  $D_1$  is sparse, with  $d_{j,j}=-1$  and  $d_{j,j+1}=1$  and all other elements zero. Examples of  $D_d$  of small dimension look like

$$D_0 = I_n; \quad D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix};$$

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix}; \quad D_3 = [1 \quad -3 \quad 3 \quad -1],$$

where  $I$  is the identity matrix.

The system of equations leading to Eq. (2) is of the order of what may be encountered in a modest-sized standard multiple regression. Further the spectra matrix  $X$  is no longer needed within the estimation process. Given  $\hat{\beta}$ , then the  $p$ -dimensional  $\hat{\alpha} = B\hat{\beta}$ . Standard error bands for the coefficient vector are tractable; details and caveats are given in Ref. [3]. One main strength of the PSR approach is that it circumvents the choice of the number and placement of B-spline knots, which is a difficult optimization problem. The penalty terms introduce very little extra computational effort, since  $U^*{}'U^*$  and  $U^*{}'y$  do not have to be recomputed when  $\lambda$  is changed. For fixed  $d$ , increasing  $\lambda$  makes  $\beta$  smoother (toward a polynomial of degree  $d-1$ ). Further, since PSR is grounded in classical (penalized) regression, the hat diagonal elements, and thus leave-one-out cross-validation, are very easy to obtain.

A *general recipe* for PSR is given in Ref. [3] and gives an idea of default parameters. Generally, we suggest using between 10 and 40 equally spaced cubic B-splines, but in the example to follow we exceed this recommendation and let the penalty drive the appropriate regularization. Based on new insight and further experience with PSR, we increased the number of knots along the wavelength axis from 18 (used in Ref. [5]) to 83 knots, and such an increase significantly improved cross-validation performance. A double penalty is used in Eq. (2) and in practice it

is useful to compare various penalty orders, say from  $d=3$  to  $d=1$  in combination with the ridge penalty (on  $\beta$ , not  $\alpha$ ). Further, Ref. [5] found some merit in further investigating a polynomial penalty, e.g.  $\sum_{d=0}^3 \lambda_d |D_d \beta|^2$ . For given penalty (combination), optimal  $\lambda(s)$  is (are) searched for systematically by monitoring, for instance, the leave-one-out cross-validation standard error of prediction (CV). The CVs minimum can also be directly compared over several (combinations of)  $d=0, 1, 2, 3$ . The final optimal  $\hat{\beta}$  is a compact representation of the estimated coefficient vector.

Using percent ethanol as the response, Fig. 4 displays the estimated PSR smooth coefficient vectors for a models trained using the derivative spectra ( $p=199$ ), using one temperature at a time. In each case, the  $m=19$  mixtures at each of  $30^\circ, 50^\circ, 60^\circ, 70^\circ$

$^\circ\text{C}$  are used. The top (bottom) panel used cubic B-splines with 18 (83) equally spaced knots along the wavelength index. Marx and Eilers [5] found that a double penalization using  $d=0, 2$  for PSR was competitive in terms of cross-validation standard error for these data. The penalty is optimized through minimizing CV in a two-dimensional grid search and resulted in an effective dimension ranging from 6.4 to 9.2 for the top panel and ranging from 16.2 to 17.3 for the bottom panel, depending on temperature. The effective dimension can be a noninteger value as it is approximated by the trace of the “hat” matrix using Eq. (2),  $\text{trace}(U^*(U^*U^* + \lambda_d D_d^* D_d^* + \lambda_0 I^*)^{-1} U^*)$ . Leave-one-out cross-validation was used, and the minimum standard error of prediction was approximately between 0.0141 and 0.0147 in the 18 knot case, but was considerably reduced with 83 knots,

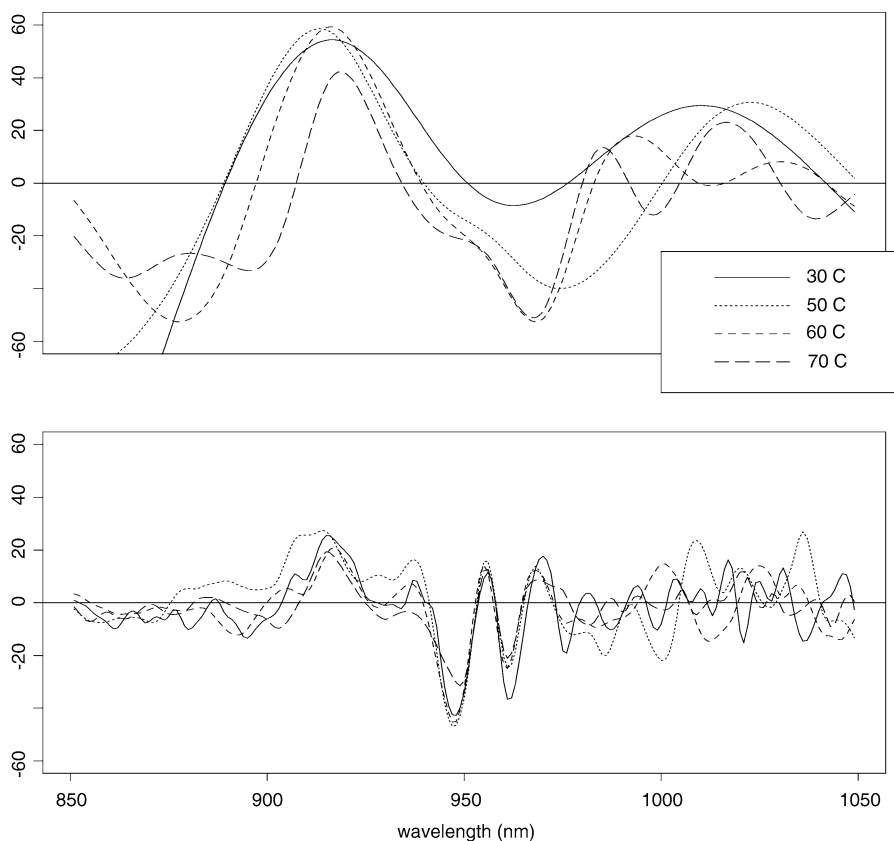


Fig. 4. Optimal smooth  $\hat{\alpha}$  for derivative spectra at several temperatures. Double penalization with  $d=0$  and  $d=2$  is used with 18 knots (top) and 83 knots (bottom).



ranging from 0.00637 (70° data) to 0.0109 (30° data). When compared to the 18-knot results, we find that using 83 knots can result in over a 50% reduction in CV, which leads to even more compelling results than those presented in Ref. [5]. Fig. 4 does not plot the very wiggly estimated PSR coefficient vector at 40°, since the tuning parameter approached zero in the 18-knot case.

In either panel, Fig. 4 gives the impression that the optimal coefficient vector may differ considerably from one temperature to the next, and if information was borrowed from adjacent temperatures then there may be a smoother transition from one  $\hat{\alpha}$  slice to the next. In part due to such a nontrivial temperature effect, Ref. [5] found that transferring a model that was optimized using the data at a single temperature, to predict at another, can lead to unsatisfactory results. Further, one may wonder if a more appropriate and comprehensive model would incorporate temperature in a smooth way using all data simultaneously.

A natural development, and the main contribution of article, is to build a two-dimensional coefficient surface that allows for interaction with the temperature information. Thus the PSR coefficient vectors displayed in Fig. 4 could then each represent a coefficient slice at a specific temperature, which usually will not be an additive adjustment. We assume that the surface is smooth along both wavelength and temperature indices. We will present a rather straight-forward and rich extension of PSR using B-spline tensor products and appropriate difference penalties on the rows and columns of these tensor products. To provide an idea of how a surface may look when using percent ethanol as the response, Fig. 5 provides a range of candidate surfaces using the derivative spectra at the five temperatures simultaneously. The upper, left panel displays a surface constructed from essentially unpenalized tensor products, whereas the lower, right surface displays the limiting plane resulting from large second order penalties on every row and column of tensor products.

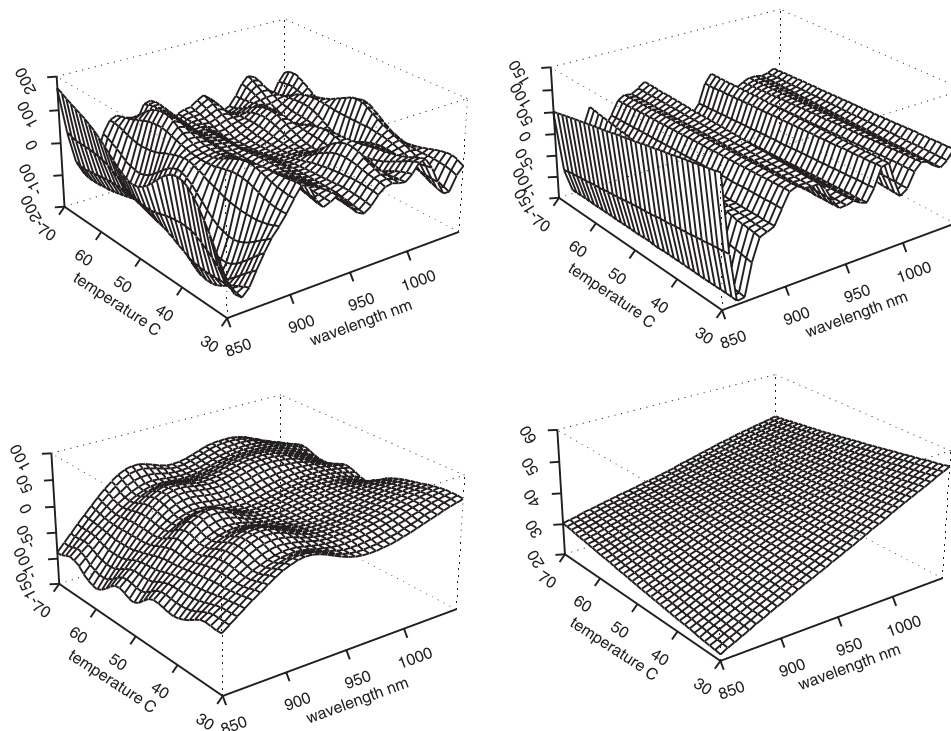


Fig. 5. Perspective plot of candidate coefficient surfaces using the derivative spectra at all temperatures: (upper, left) effective dimension=74, CV=0.0202; (upper, right) effective dimension=34, CV=0.0118; (lower, left) effective dimension=23, CV=0.0470; (lower, right) effective dimension=5, CV=0.0778.

The other two figures have a mixture of a low penalty on one axis and a high penalty on the other. Further details of the surface parameters, performance and advantages of this approach will be given in Section 6. First, we provide some details of tensor products P-splines and introduce the extensions to the model.

#### 4. Tensor product B-splines in a nutshell

Eilers and Marx [10] presented a section *B-splines in a nutshell*. To give the background that is needed for this paper, we extend the nutshell to illustrate the basic simplicity of tensor product B-splines. A more complete and mathematically rigorous presentation of the subject can be found in Ref. [11] (chapters 1 and 2). Fig. 6 displays the essential building block: a bicubic basis function. In short, this figure displays the tensor product of the two univariate (cubic) B-splines,  $B_k$  and  $B_l$  presented along the axes, say  $v$  (wavelength) and  $t$  (temperature), respectively. The tensor product value is zero (nonzero) in the corresponding zero (nonzero) univariate  $v$  and  $t$  ranges. More details follow below, including indexing, but

first we revisit some specifics of a univariate B-spline basis by focusing on the  $v$  axis.

We choose to divide the domain  $v_{\min}$  to  $v_{\max}$  into  $n'$  equal intervals, using  $n'+1$  interior knots. Taking each boundary into consideration, a complete basis needs  $n'+2q+1$  total knots, where  $q$  is the degree of the B-spline. Denote the knots as:  $v_1, \dots, v_{n'+2q+1}$ . The total number of B-splines on the axis is  $n=n'+q$ . For indexing purposes, it is convenient to associate each B-spline,  $B_k(v)$  with exactly one of the (first)  $k=1, \dots, n$  knots. The general properties of a B-spline of degree  $q$  include:

- it consists of  $q+1$  polynomial pieces, each of degree  $q$ ;
- the polynomial pieces join at  $q$  inner knots;
- the joining points have continuous derivatives up to degree  $q-1$ ;
- a B-spline of degree  $q$  is positive on a domain spanned by  $q+2$  knots (zero elsewhere);
- except at the boundaries, it overlaps with  $2q$  polynomial pieces of its neighbors;
- at a given  $v$  (or  $t$ ), exactly  $q+1$  B-splines are nonzero.

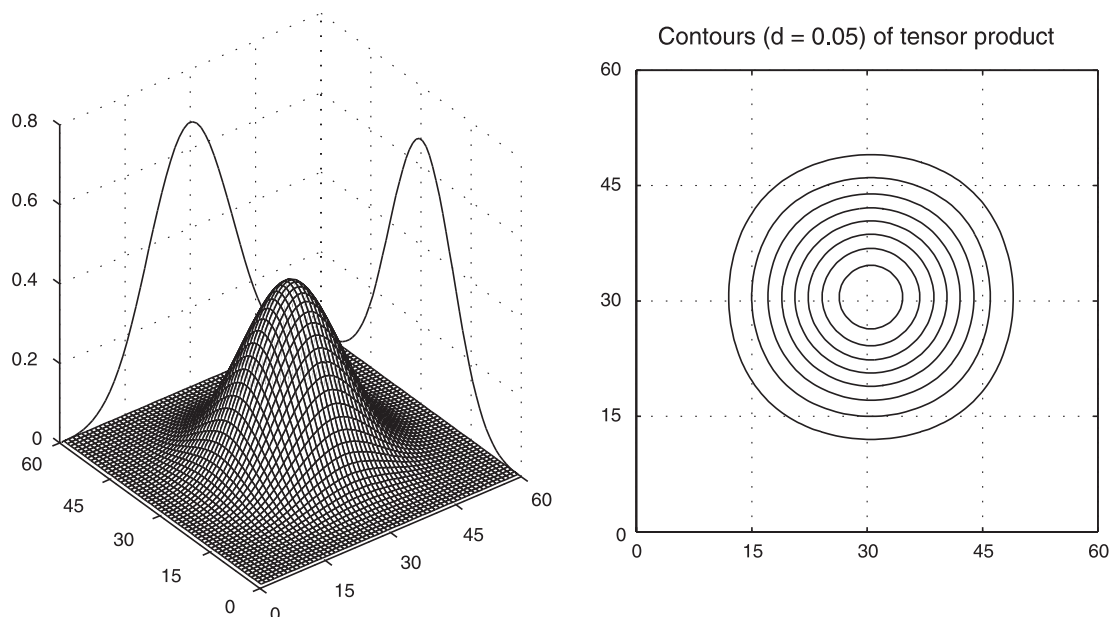


Fig. 6. Tensor product of two cubic B-splines.



A full basis is simply denoted as the  $m \times n$  matrix  $B$ . A similar division of the  $t$  axis is made for  $\check{B}_l(t)$ , also using equally spaced knots, but possibly using a different  $q$  or  $n'$ .

Tensor product B-splines exist in the  $v \times t$  plane. The knots are selected on an equally spaced grid, carving out the plane into subrectangles. A tensor product  $B_k(v)\check{B}_l(t)$  is positive in the rectangular region defined by the knots  $R=[v_k, v_{k+q_v+2}] \times [t_l, t_{l+q_t+2}]$  or on a support of spanned by  $(q_v+2) \times (q_t+2)$  knots. Specifically, each tensor product can be indexed by one of  $(n_v \times n_t)$  knot pairs and

$$B_k(v)\check{B}_l(t) \geq 0 \text{ for all } v,$$

$$t \in R = 0 \text{ for all } v, \quad t \notin R, \tag{3}$$

$k=1, \dots, n_v$  and  $l=1, \dots, n_t$ . Fig. 7 displays nine tensor product B-splines, which represents only a portion of a full basis.

Denote  $\Gamma_{n_v \times n_t} = [\gamma_{kl}]$  as the matrix of unknown tensor product B-spline coefficients. For given knot grid, a surface can be approximated by

$$\alpha(v, t) = \sum_{k=1}^{n_v} \sum_{l=1}^{n_t} B_k(v)\check{B}_l(t)\gamma_{kl}, \tag{4}$$

where  $k=1, \dots, n_v$  and  $l=1, \dots, n_t$ . The system of equations is of order  $n_v n_t$ . It is computationally efficient

(by avoiding looping) to re-express the surface in matrix notation as  $\alpha(v, t) = \mathbf{B}\gamma$ , where  $\gamma = \text{vec}(\Gamma)$ , and

$$\mathbf{B} = B \square \check{B} = (B \otimes 1'_{n_t}) \odot (1'_{n_v} \otimes \check{B}), \tag{5}$$

and is of dimension  $m \times (n_v n_t)$ . The symbols  $\otimes$  and  $\odot$  denote Kronecker product and element-wise multiplication of matrices, respectively. The combination of  $\otimes$  and  $\odot$  matrix operations with vectors of ones, as expressed in Eq. (5) is denoted by  $\square$ . We will see this notation again in the next section. Each column of  $\mathbf{B}$  can be graphically displayed as one “hill” in e.g. Fig. 7; i.e. each “hill” is a unit rank matrix. Penalized estimation of  $\gamma$  and its use with spectra regressors are topics discussed in the next section.

### 5. Two-dimensional Tensor Product PSR (TPSR)

Given spectra matrix  $X=[x_{ij}]$  ( $i=1, \dots, m; j=1, \dots, p$ ) and coefficient surface  $\alpha(v, t)$ , let

$$\mu_i = \alpha_0 + \sum_{j=1}^p x_{ij} \alpha(v_j, t_i). \tag{6}$$

Eq. (6) is akin to Eq. (1), but uses a slice of the coefficient surface that is specific to the value of  $t$ . Recall Fig. 5, which presented several estimated regression coefficient surfaces for the mixture experiment. To give an idea of how the surface is used, Fig. 8 displays various temperature slices of the upper

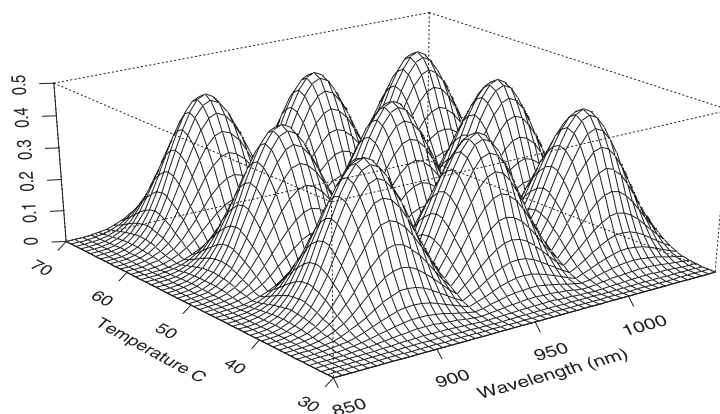


Fig. 7. Landscape of nine cubic B-spline tensor products, a portion of a full basis.

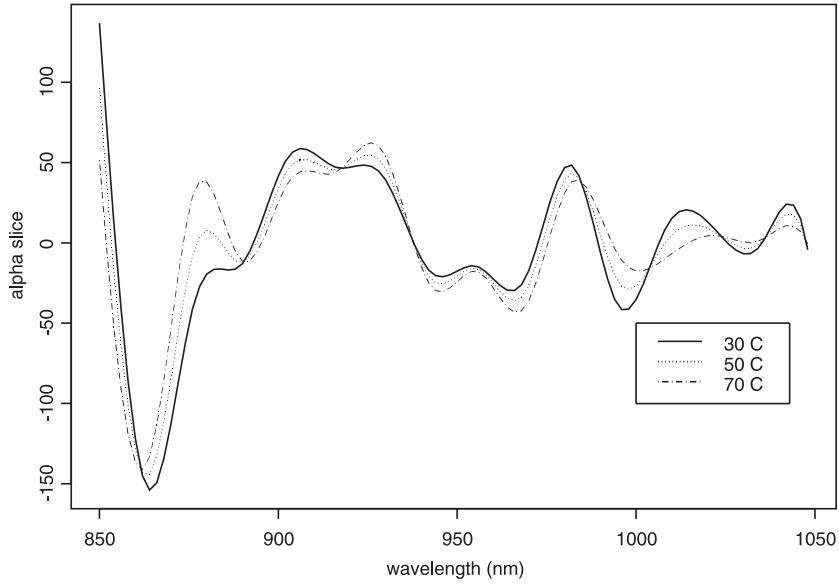


Fig. 8. Slices of  $\hat{\alpha}$  at various temperatures from an estimated tensor product coefficient surface using derivative spectra.

right panel of Fig. 5 that can be used in Eq. (6). If the coefficient surface is approximated by tensor product B-splines, then Eq. (4) can be substituted into Eq. (6) yielding

$$\begin{aligned}
 \mu_i - \alpha_0 &= \sum_{j=1}^p x_{ij} \sum_{k=1}^{n_v} \sum_{l=1}^{n_t} B_k(v_j) \check{B}_l(t_i) \gamma_{kl} \\
 &= \sum_{j=1}^p \sum_{k=1}^{n_v} \sum_{l=1}^{n_t} x_{ij} b_{jk} \check{b}_{il} \gamma_{kl} \\
 &= \sum_{k=1}^{n_v} \sum_{l=1}^{n_t} \left( \sum_{j=1}^p x_{ij} b_{jk} \right) \check{b}_{il} \gamma_{kl} \\
 &= \sum_{k=1}^{n_v} \sum_{l=1}^{n_t} u_{ik} \check{b}_{il} \gamma_{kl}. \tag{7}
 \end{aligned}$$

Borrowing from Eq. (5), we find that Eq. (7) can be expressed in matrix notation as  $\mu = \alpha_0 + \mathbf{U}\gamma$ , where

$$\mathbf{U} = \mathbf{U} \square \check{\mathbf{B}} = (\mathbf{U} \otimes \mathbf{I}_{n_t}) \odot (\mathbf{I}_{n_v} \otimes \check{\mathbf{B}}),$$

is of dimension  $m \times (n_v n_t)$  and  $\mathbf{U} = \mathbf{X}\mathbf{B}$ .

We aim to find a practical solution to minimize  $Q(\alpha_0, \gamma) = |y - \alpha_0 - \mathbf{U}\gamma|^2$ . We see the use of tensor product B-splines can dramatically reduce the dimension of estimation, but there are still  $n_v n_t + 1$  unknown parameters. For even moderately complex surfaces, ill-posed estimation problems can arise, as it may be necessary to increase the number on knots on the grid to allow enough flexibility.

In the spirit of P-splines [10], appropriate penalties can be put on  $\gamma$  and thus regularize estimation. A separate difference penalty is attached to each of the rows and each of the columns of  $\Gamma$ . The objective function is now modified to minimize

$$\begin{aligned}
 Q^*(\alpha_0, \gamma) &= |y - \alpha_0 - \mathbf{U}\gamma|^2 + \lambda_v |P_v \gamma|^2 \\
 &\quad + \lambda_t |P_t \gamma|^2 + \lambda_0 |\gamma|^2. \tag{8}
 \end{aligned}$$

The last term in Eq. (8) is an overall ridge penalty. Indexing can quickly get out of hand and the penalties are most compactly represented in matrix notation as:  $P_v = (D'_d D_d) \otimes \mathbf{I}_{n_t}$ , and  $P_t = \mathbf{I}_{n_v} \otimes (D'_d D_d)$ , where  $\mathbf{I}$  denotes the identity matrix. Although it is not reflected in the notation, the order of the row penalty ( $d_v$ ) can be different from that of the column penalty ( $d_t$ ). In principle, more than one order (or polynomial) penal-

ization can be used for rows or columns. Much like the PSR approach presented in Section 3, the difference penalties ensure that adjacent coefficients within the same row (or a column) do not differ too much from each other. The penalties can continuously regulate roughness through the nonnegative  $\lambda_v$ ,  $\lambda_t$  and  $\lambda_0$ .

The explicit P-spline solution for Eq. (8) is

$$\hat{y} = (\mathbf{U}^* \mathbf{U}^* + \lambda_v P_v^* + \lambda_t P_t^* + \lambda_0 I^*)^{-1} \mathbf{U}^* y,$$

with  $\mathbf{U}^* = (1_m | \mathbf{U})$ ,  $P^* = (0 | P)$ , and  $I^* = \text{diag}(0, I_{n_v n_t})$ . Again, the notation  $|$  denotes matrix column augmentation. The predicted values are  $\hat{y} = \mathbf{U}^* (\hat{\alpha}_0, \hat{y}')'$ . Thus the “hat” matrix is  $H = [h_{ij}] = \mathbf{U}^* (\mathbf{U}^* \mathbf{U}^* + \lambda_v P_v^* + \lambda_t P_t^* + \lambda_0 I^*)^{-1} \mathbf{U}^*$ . Given  $\hat{y}$  and the diagonal of  $H$ , leave-one-out cross-validation standard error of prediction can be quickly calculated:

$$\text{CV}(\lambda_v, \lambda_t, \lambda_0) = \sqrt{\frac{1}{m} \sum_{i=1}^m \left( \frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2}.$$

The optimal  $(\lambda_v, \lambda_t, \lambda_0)$  can be found, e.g. using a grid search to minimize CV. The dimension of the estimated coefficient surface can be approximated by trace ( $H$ ), and the error variance component estimated by

$$\hat{\sigma}^2 = \frac{|y - \hat{y}|^2}{m - \text{trace}(H)}.$$

For computational purposes, it is worth mentioning that the explicit solution for Eq. (8) can be found efficiently through data augmentation tricks, i.e.  $(\hat{\alpha}_0, \hat{y}') = (\mathbf{U}_+^* \mathbf{U}_+^*)^{-1} \mathbf{U}_+^* y_+$ , where

$$\mathbf{U}_+^* = \begin{bmatrix} 1 & \mathbf{U} \\ 0 & \sqrt{\lambda_v} (D_d \otimes I_{n_t}) \\ 0 & \sqrt{\lambda_t} (I_{n_v} \otimes D_d) \\ 0 & \sqrt{\lambda_0} I_{n_v n_t} \end{bmatrix} \quad \text{and} \quad y_+ = \begin{bmatrix} y \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

## 6. Results for the mixture experiment

We model percent ethanol using the derivative spectra (199 channels) and temperature. The previous sections motivated ideas of tensor product coefficient surface estimation (TPSR) and its application to the mixture data. We first construct a surface with penalty orders along wavelength and temperature of  $d_v=2$  and  $d_t=1$ , respectively. The two-dimensional grid search yields a minimum leave-one-out CV at 0.00594 for  $\lambda_v=10^{-14}$ ,  $\lambda_t=10^{-8}$ , and  $\lambda_0=5 \times 10^{10}$  using 83 (8) knots on the  $v(t)$  axis. Fig. 9 displays various slices of the estimated P-spline tensor product coefficient surface at a sequence of 40 (10) various temperatures (wavelengths). The effective dimension of the surface is approximately 89: we see the penalty working as there are 665  $((83 \times 8) + 1)$  parameters. The choice of  $d_t=1$  is based on the fact that the limiting polynomial is constant: thus if a relatively large penalization should have been needed to minimize CV, then the model would have defaulted to standard PSR (Section 3). When compared to the standard PSR (a model for each temperature), we find a reduction in CV between approximately 6% and 46%, depending on temperature. We find that the adjustments to the estimated spectra coefficient vector can vary approximately 10% relative to the coefficient range. One might wonder if these coefficient adjustments are really needed along temperature. They are essential: if a standard PSR (using all 95 observations) is alternatively optimized (ignoring temperature), then this would lead to a 47% increase in CV to 0.00869 (using 83 knots,  $d=0, 2$ ). Cubic B-splines ( $q=3$ ) were used in all cases. Some guidelines for choice of these design parameter are borrowed from previous analyzes of these data in Ref. [5].

Increasing  $d_t=2$ , Fig. 10 presents the optimal estimated coefficient surface. The effective dimension is now decreased to 70. A two-dimensional grid search found a minimum leave-one-out CV at 0.00531 for  $\lambda_v=10^{-15}$ ,  $\lambda_t=100$ , and  $\lambda_0=3 \times 10^{-9}$  again using 83 (8) knots on the  $v(t)$  axis. This is about a 10% decrease over the minimum found above using  $d_t=1$ . The relatively large penalty on the coefficients along temperature leads to essentially a linear trend, as seen in Fig. 11. This is a general feature of P-splines: as the penalty gets large, then the smooth approaches a polynomial of degree  $d-1$ .

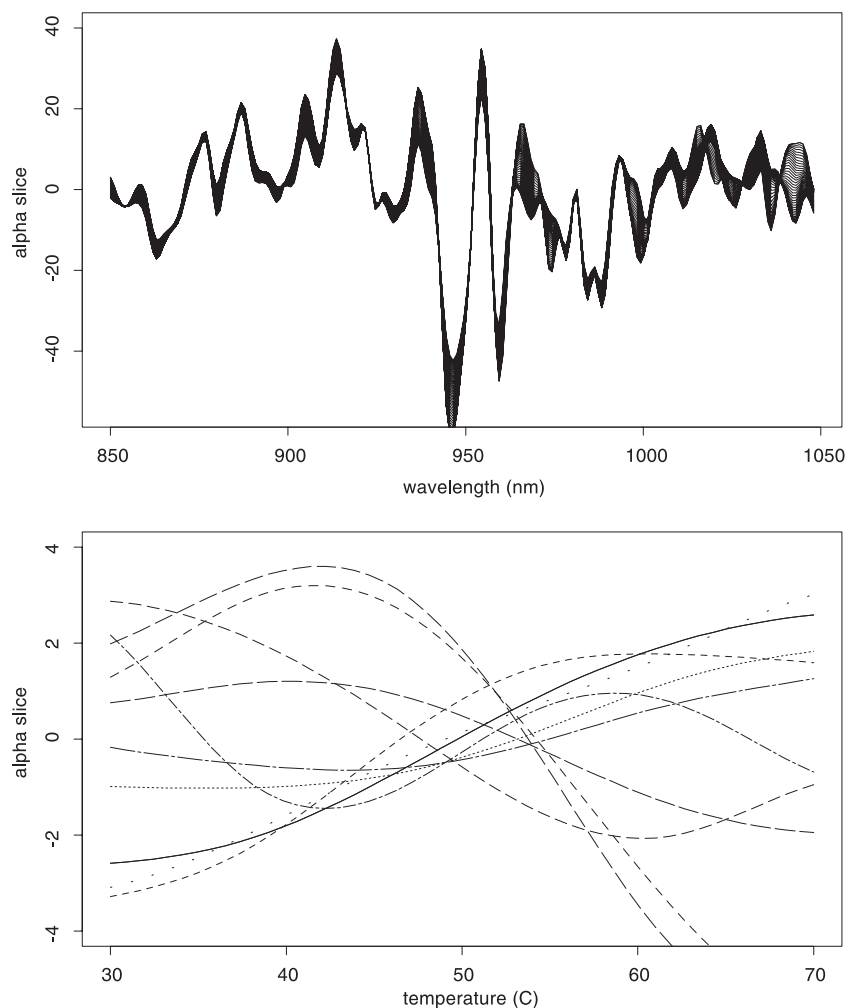


Fig. 9. Slices of the optimal TPSR  $\hat{\alpha}$  surface using  $d_v=2$  and  $d_t=1$ : (top) at various temperatures, (bottom) at various wavelengths (centered with mean 0).

Figs. 10 and 11 display notable interaction in the estimated surface. Although we can fit a general interaction model through tensor products, these figures suggest that these data may be well fit by a simpler model. We next explore an alternative interaction model.

#### 6.1. A simpler varying PSR model for the mixture data (VPSR)

Figs. 9 and 11 provide some evidence that the mixture coefficient surface may have a rather simple

interaction structure. For these data, an attractive alternative is the varying PSR (VPSR) model:

$$\alpha(v, t) = f(v) + (t - 50)g(v), \quad (9)$$

where  $f$  and  $g$  are smooth functions of wavelength. This model has ease of interpretation: when temperature is 50 °C, then the coefficient vector is the slice  $f(v)$ . Further, the smooth  $g(v)$  provides the values of the slope in the temperature direction at a fixed wavelength  $v$ . The upper panels of Fig. 12 provides a comparison of the tensor product and VPSR models

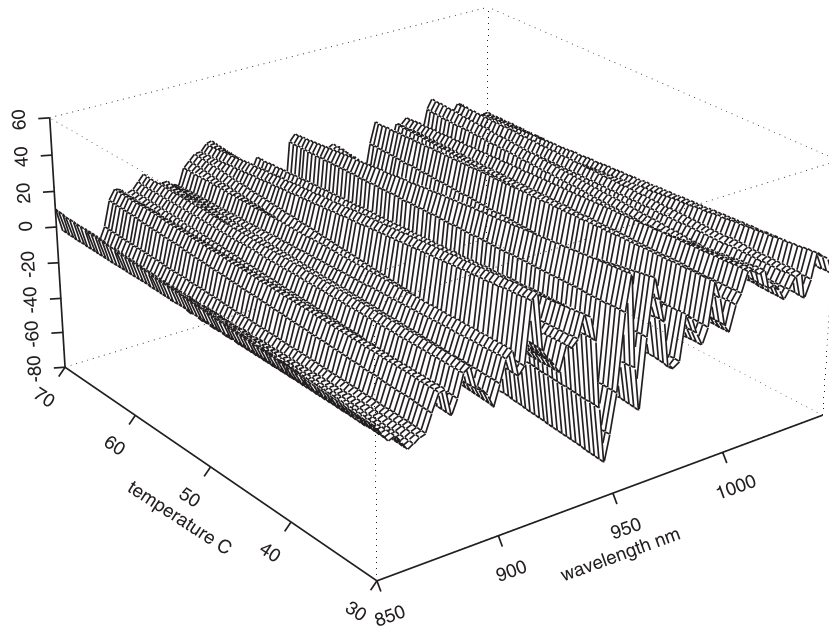


Fig. 10. Optimal perspective plot of tensor product (TPSR) coefficient surface using the derivative spectra over temperature,  $d_v, d_t=2$ .

at two (interpolated) coefficient slices: we find that behave somewhat similarly, but with VPSR coefficients much more enthusiastic near the regions of 870,

910, and 940 nm. The lower, left panel of Fig. 12 displays the estimated smooth function  $g(v)$ , which again demonstrate rather subtle, but important, values

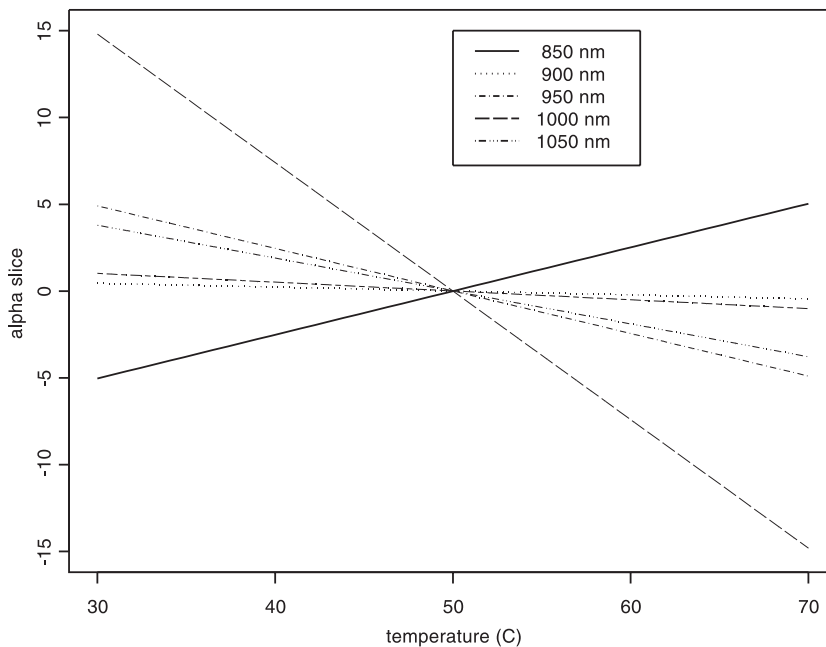


Fig. 11. Slices of  $\hat{\beta}$  at various wavelengths (centered with mean 0) from the estimated tensor product coefficient surface using derivative spectra.

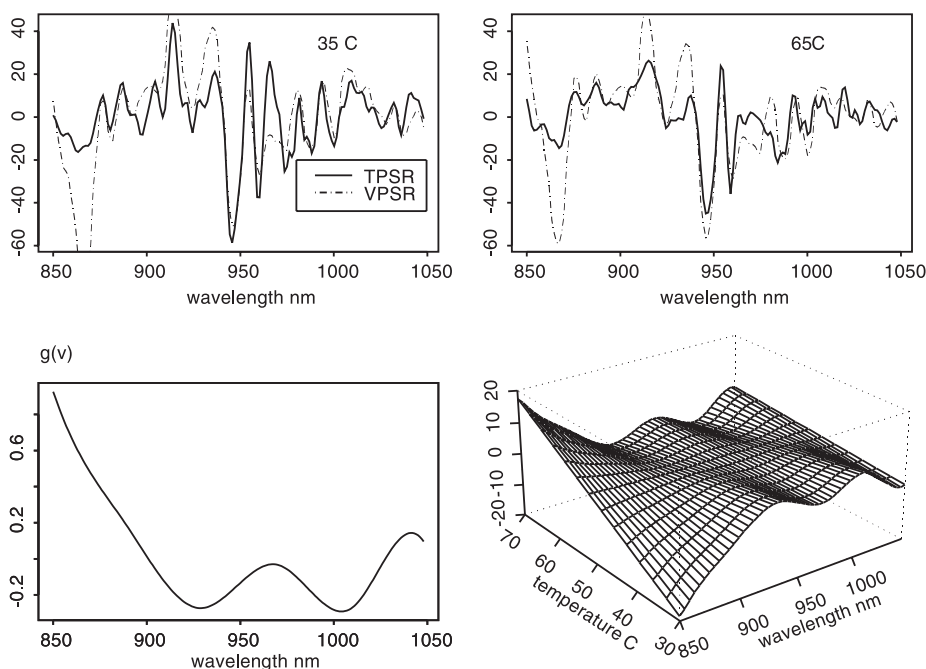


Fig. 12. Upper panels: slices of the optimal tensor product and multiplicative VCM  $\hat{\alpha}$  surface for 35° and 65°; (lower, left): smoothly varying slope function (of wavelength) for the surface; (lower, right) perspective plot of  $\phi(v, t)=(t-50)g(v)$  displaying VPSR interaction structure.

of slopes in the  $t$  direction ranging within  $(-0.3, 0.9)$ . The lower, right panel of the figure displays the VPSR interaction structure  $\phi(v, t)=(t-50)g(v)$ . Using Eq. (9), we can write

$$\begin{aligned} \mu &= \alpha_0 + XB_1\beta_1 + VXB_2\beta_2 \\ &= \alpha_0 + U_1\beta_1 + U_2\beta_2, \end{aligned} \quad (10)$$

$V=\text{diag}\{t-50\}$ . Actually Eq. (10) can be viewed as an additive signal model, with effective signals  $S_1=X$  and  $S_2=VX$ . Such a model is described in detail, under a more general framework, in Ref. [12]. Difference penalties in Eq. (10) were placed on  $\beta_1$  and  $\beta_2$  with orders  $d_1=2$  and  $d_2=2$ , respectively. B-splines were used with  $n_1=83$  and  $n_2=8$  equally spaced knots. This lead to optimal penalty tuning parameters, found by grid search, of  $\lambda_1=1.5\times 10^{-8}$  (with a ridge penalty of  $\lambda_0=10^{-11}$  on  $\beta_1$ ) and  $\lambda_2=10^{-6}$ . The leave-one-out CV was minimized at 0.006810 nearly the optimal value using the tensor product model above. The effective dimension of the VPSR surface is approximately 49

(42 in  $f$  and 7 in  $g$ ), which is about 21 less than the optimal TPSR model.

## 7. Discussion

We have presented a modeling approach that allows the coefficient vector to vary smoothly (interact) with another variable, e.g. temperature, yielding a surface. Denote the triplet  $(y_i, x(v_{ji}), t_i)$  for the response, signal, and covariate, respectively,  $i=1, \dots, m$ ;  $j=1, \dots, p$ . We moved from a smooth PSR vector  $\alpha(v_j)$  to a tensor product smooth surface TPSR  $\alpha(v_j, t_i)$ . We have also presented a simplified varying penalized signal regression VPSR surface:  $\alpha(v_j)+t_i\alpha^*(v_j)$ . Although we did not consider it in this paper, one could also image an additive coefficient (APSR):  $\alpha(v_j)+\alpha^*(t_i)$  similar to Ref. [13].

The CV values presented thus far have been for model optimization purposes, i.e. the choice of the penalty tuning parameters. Some additional care must taken for direct comparison of prediction performance to that found within other research. To this end, we



Table 2  
TPSR cross-validation (RMSEP): training and test set, by mixture component

Component	CV ( $i, -i$ ) training	CV test set
Ethanol	0.0065	0.0071
Water	0.0031	0.0039
Isopropanol	0.0044	0.0063

now optimize the TPSR penalty parameters by minimizing the leave-one-out CV for the training set mixture numbers (see Fig. 1): 1–4, 7, 8, 10, 12, 13, 16–19. The test set for model assessment uses mixture numbers: 5, 6, 9, 11, 14, 15. We find CV values that are highly competitive and an improvement to those found in Refs. [4,5,7–9]. Table 2 summarizes our TPSR results, by mixture ingredient.

Although the model uses a rather large number of B-spline tensor products, in this example over 850 parameters, one should view these as only an initial reduction in dimension toward estimation of the approximately 1400 unknown parameters across wavelength and temperature. The penalty ensures further reduction in (effective) dimension (89 in TPSR and 49 in VPSR). The regularization of the penalty is chosen based on cross-validation (CV): the combination of the penalty and CV prevents over-fitting. At first thought, it may appear that a dimension of 89 or 49 is somewhat large, especially when compared to, for example, a three component PLSR model for this tertiary experiment. Yet, it is not fair to compare values 89 or 49 to 3: each PLSR component is constructed with 200 (rough) weighting coefficients. The effective dimension of such a PLSR model is less clear to quantify, but in between 3 and  $3 \times 200$ .

We believe that the proposed models not only can be easy to use in practice, but have many nice features. For example, rather than monitoring temperature, one can use such coefficient surfaces to model instrumental drift. The methods are not limited to spectra and can be used for more general signal regression problems. Further there is nothing prohibitive from extending this work into higher dimensions that smoothly account for other variables. Since the penalized signal regression approach is grounded in classical regression, outlier diagnostics, standard error bands for the estimated surface, as well as extension to the generalized linear model are all tractable. Due to the projection of coefficients onto smooth bases,

this method will not be burdened if, for example, improved technology increases the number of measured spectra channels. An exciting feature of the proposed approach is that temperatures do not have to be precisely controlled: they just have to be measured. More specifically, the fitting procedure uses measured temperatures and does not care about precise control for each mixture of components. Of course, it is important to aim at a more or less regular spacing of temperatures. This can be a great advantage under process circumstances, outside the laboratory, where temperature control may be quite difficult or even impossible. Lastly, we believe that a smooth surface provides a viable and rather natural interpolation between temperatures, especially when compared to the option of interpolating separate models for each temperature. The proposed method provides a smooth coefficient vector that can be estimated at any temperature slice, not just the measured ones. S-PLUS and Matlab code will be available at [www.stat.lsu.edu/bmarx](http://www.stat.lsu.edu/bmarx).

## Acknowledgements

We thank Age Smilde for valuable discussion regarding this research. Research supported in part for Brian Marx by NSF Grant DMS-0102131.

## References

- [1] H. Martens, T. Næs, *Multivariate Calibration*, Wiley, New York, 1989.
- [2] I.L. Frank, J.H. Friedman, A statistical view of some chemometric regression tools, *Technometrics* 35 (1993) 109–148.
- [3] B.D. Marx, P.H.C. Eilers, Generalized linear regression on sampled signals and curves: a p-spline approach, *Technometrics* 41 (1999) 1–13.
- [4] F. Wülfert, W. Kok, A. Smilde, Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models, *Analytical Chemistry* 70 (1998) 1761–1767.
- [5] B.D. Marx, P.H.C. Eilers, Multivariate calibration transfer: a comparison of methods, *Journal of Chemometrics* 16 (2002) 1–12.
- [6] T. Hastie, R. Tibshirani, Varying-coefficient models (with discussion) *Journal of the Royal Statistical Society, B* 55 (1993) 757–796.
- [7] F. Wülfert, W. Kok, O. de Noord, A. Smilde, Linear techniques to correct for temperature induced spectra variation in

- multivariate calibration, *Chemometrics and Intelligent Laboratory Systems* 51 (2000) 189–200.
- [8] F. Wülfert, W. Kok, O. de Noord, A. Smilde, Correction of temperature-induced spectral variation by continuous piecewise standardization, *Analytical Chemistry* 72 (2000) 1639–1644.
- [9] H. Swierenga, F. Wülfert, O. de Noord, A.K. de Weijer, A. Smilde, L.M.C. Buydens, Development of robust calibration models in near infra-red spectrometric applications, *Analytica Chimica Acta* 411 (2000) 121–135.
- [10] P.H.C. Eilers, B.D. Marx, Flexible smoothing with b-splines and penalties (with comments and rejoinder) *Statistical Science* 11 (1996) 89–121.
- [11] P. Dierckx, *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford, 1995.
- [12] P.H.C. Eilers, B.D. Marx, Generalized linear additive smooth structures, *Journal of Computational and Graphical Statistics* 11 (2003) 758–783.
- [13] T. Hastie, R. Tibshirani, *Generalized Additive Models*, Chapman & Hall, London, 1990.