# Multidimensional Single-Index Signal Regression

Brian D. Marx*
Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
*contact author bmarx@lsu.edu

Paul H.C. Eilers
Department of Biostatistics
Erasmus Medical Centre
3015 GE, Rotterdam, The Netherlands
p.eilers@erasmusmc.nl

Bin Li
Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
bli@lsu.edu

July 27, 2011

**Abstract**

In general, linearity is assumed to hold in multivariate calibration (MVC), but this may not be true. We approach the MVC problem using multidimensional penalized signal regression, which can be extended with an explicit link function between linear prediction and response and in the spirit of single-index models. As the two-dimensional surface of calibration coefficients is smoothly and generally estimated with tensor product P-splines, the unknown link function is estimated using univariate P-splines. The methods presented are grounded in penalized regression, where difference penalties are placed on the rows and columns of the tensor product coefficients, as well on the link function coefficients, each having their own tuning parameter. An application to ternary mixture data shows that a non-linearity is present. Performance comparisons are made to standard penalized signal regression, not only demonstrating the nonlinear effect, but also improvements in external prediction.

Keywords: Multivariate calibration; P-splines; signal regression; single-index; spectra; tensor product; ternary mixtures.

## 1 Introduction

In this paper, we take a novel approach the multivariate calibration problem, in particular where the signal (spectra) regressors have two-dimensional structure. Our application

considers UV-VIS spectra taken over several temperatures. Through simultaneous estimation, we parse out and estimate two separate modeling components: (1) a single *smooth regression coefficient surface* associated with the two-dimensional signal (Marx and Eilers, 2005), and (2) an unknown, possibly nonlinear, *link function* (Eilers, Li and Marx, 2009). Although the first component is linear, the second component explicitly models the nonlinearity, allowing us to learn something about its features, while enhancing insight into the measurement process. We will see that the combination of these components can lead to a systematic and tractable modeling approach, that is statistical in nature, while having improved external prediction performance when compared to standard signal regression approaches and partial least squares.

### 1.1 Multivariate calibration with two-dimensional signals

At the heart of a multivariate calibration problem is rich regressor data, often compactly given as a digitized signal, curve, or spectra. Such regressor information can also be in two or more dimensions, such digitized images. An often ironic consequence of such data is that as more and more precisely regressor information is obtained, the more and more ill-conditioned estimation becomes. Since classical least squares modeling approaches usually fail, there have been numerous competing methods developed to provide tractable and reliable prediction; see Eilers, Li, and Marx (2009) and Eriksson *et al.* (2000) for partial lists. We will see that, unlike most of the other approaches, our proposed method additionally takes advantage of the ordered or array structure among the regressors.

To motivate the problem, Figure 1 displays signal regressors (at two different temperatures) for each of $m = 34$ observations, coming from a ternary mixture experiment using spectroscopy. Each "signal" actually consists of numerous digitizations ($p = 401$) along the wavelength axis (700 to 1100, by 1 nm). The top (bottom) panels present the raw (first differenced) spectra. If such optical regressors are to be related, e.g. to a chemometric response, then some regularization is needed. Generally, not only is $p \gg m$, but the regressors are highly correlated.

Notice that the left and right panels of Figure 1 presents "signals" at *temperature* levels of $30^o$ and $70^o$ C, respectively, and one could imagine even more, forming a sequence of several "extremely narrow images". Thus a natural question to ask is: What if the signal regressors become fully two-dimensional, and we wish to take into account spatial information in both directions? One could view this problem as multivariate calibration with multidimensional spectra, where, e.g., the second dimension is temperature. Figure 2 presents such a two-dimensional spectra structure with 4800 regressors, summarized in a $12 \times 400$ matrix (using first differences), for the center mixture unit, with corresponding scalar responses (water, 1,2-ethanediol, 3-amino-1-propanol, each at 0.33).

### 1.2 Notation and data structure

The data structure is as follows, each observation consists of the data pair: $(y_i, X_i)$, where $i = 1, \ \ldots \ , m$. The response $y_i$ is scalar. We assume independence among the responses, with common variance $\text{var}(y) = \sigma^2$.

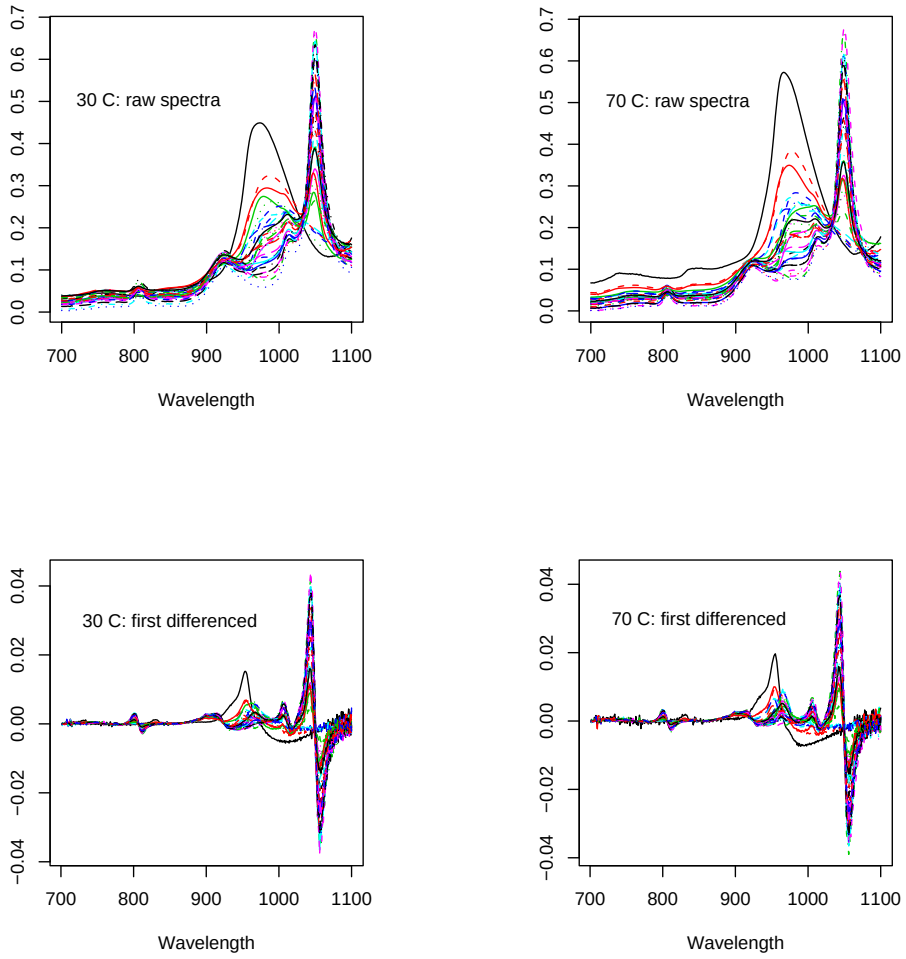


Figure 1: *Signal regressors (raw and first differenced) for mixture experiment, at two different temperatures.*

The two-dimensional signal consists of (often thousands of) digitized regressors, $X_i$, arranged in a $p \times \breve{p}$ array. The indexing axes, i.e. $v$ and $\breve{v}$, that define the support coordinates of $X_i$ are usually on a regular grid, but the only requirement for our method is that the scatter of digitizations are common for all $i$. As suggested by Figure 2, the number of regressors are rich, over one hundred times greater in number than observations. The regressor support is specified as $v^\star$ (wavelength) with $p = 400$ channels (701 to 1100nm, by 1 nm) and $\breve{v}^\star$ with $\breve{p} = 12$ temperature channels (30, 35, 37.5, 40, 45, 47.5, 50, 55, 60, 62.5, 65, 70$^o$ C).

The response $y$ comes from the composition (mole fraction) of a mixture, here consisting of three components (water, 1,2-ethanediol, 3-amino-1-propanol). The ternary plot

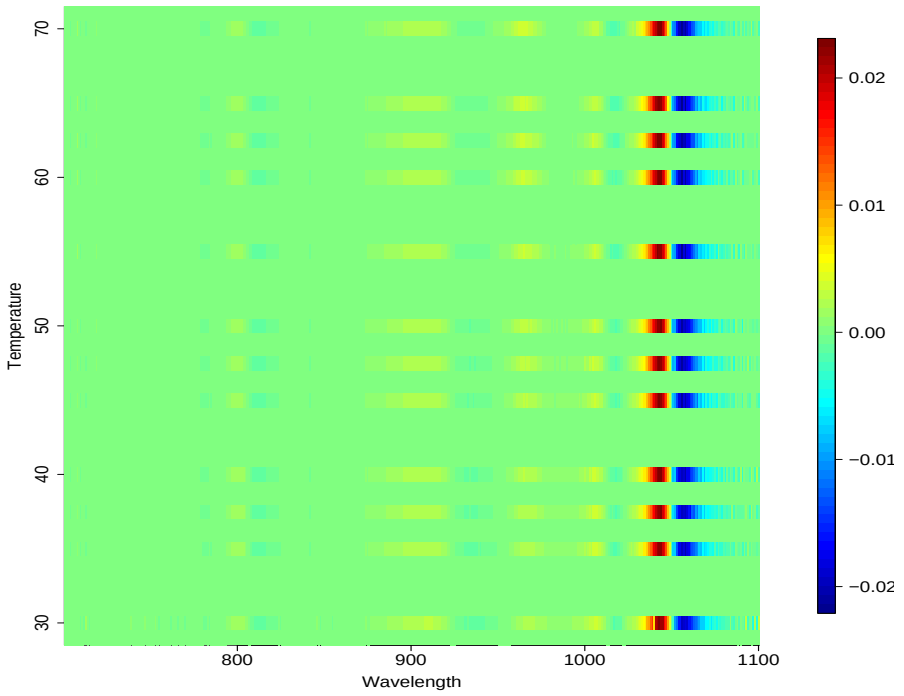


Figure 2: Two-dimensional (first differenced) signal regressor image for center mixture.

for the $m = 34$ mixtures is provided in Figure 3. The center data point in the triangle represents equal concentrations of the three components, the edge points are mixtures containing only two components, and the corners are pure. Note that there are 3 pure, 12 edge, and 19 interior (1 center) mixtures.

### 1.3 First modeling component: MPSR

The multidimensional signal regression (MPSR) model was first presented in Marx and Eilers (2005), initially motivated by both Marx and Eilers (1999) and Eilers and Marx (2003). The model's goal is to provide an extremely practical solution for functional linear models using the entire two-dimensional signal as regressors. Associated with the regressors is a single overarching coefficient surface which serves to smoothly weigh each two-dimensional signal digitization over its support. Regularization is needed, and we choose to impose some sensible constraints: ones that take into account the spatial structure of the regressors, while ensuring smoothness in the coefficient surface. As with any P-spline approach (Eilers and Marx, 1996), we take two steps towards smoothness: (a) The coefficient surface (not the signal) is intentionally overfit using two-dimensional tensor product B-splines, making the surface more flexible than needed. (b) Tensor product coefficient estimates are penalized using difference penalties on each of the rows and columns.

The first step provides an initial reduction in parameter estimation through smoothness, as the higher dimensional two-dimensional signal coefficient surface is projected onto a lower dimensional tensor product basis, where the knots are "richly" chosen on a regular grid, thus circumventing knot selection schemes. The second step ensures further smoothness, as well as regularization, while allowing general surface candidates. Two tuning parameters, associated with the row and column penalties, respectively, are needed

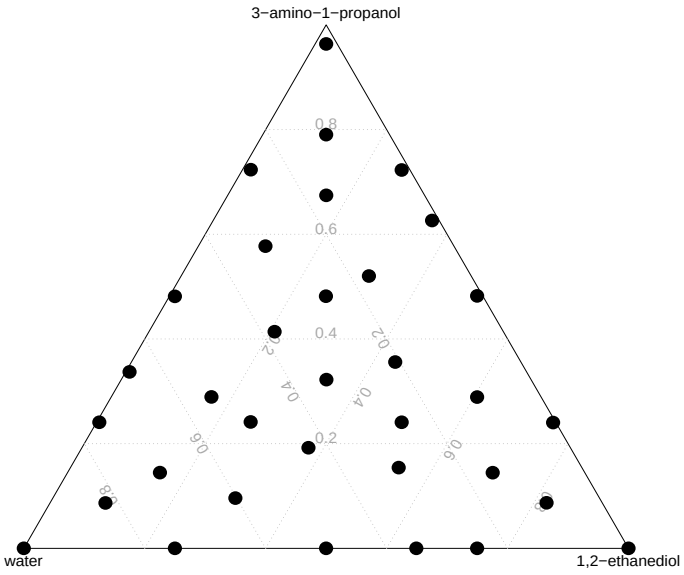


Figure 3: *Ternary plot for mixtures, with* $m = 34$*:* 3 *pure,* 12 *edge,* 19 *interior.*

to allow for continuous control over the surface. Figure 4 displays a variety examples of (coefficient) surfaces using tensor products B-splines. 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. The other two panels have a mixture of a low penalty on one axis and a high penalty on the other.

## 1.4 Second modeling component: SISR

The second modeling component is single-index signal regression (SISR), which was presented in Eilers, Li, and Marx (2009) for one-dimensional signals, and is a method that can provide additional insight through the explicit modeling of any nonlinear behavior that may exist with the response. In fact, one could view the standard multivariate calibration problem as using an identity link function, which in actuality may be (slightly) misspecified. In effect, there may exist a true, but "missing link" function (that is nonlinear and monotone) (Cox, 1984), and SISR serves the purpose of estimating this link while improving external prediction. Many approaches to the multivariate calibration problem, e.g. support vector machines (SVM) (Thissen *et al.*, 2004) or the genetic algorithm (Üstün *et al.*, 2005), do implicitly "capture" nonlinearities to improve prediction performance, but in the end do not necessarily provide the researcher with meaningful information about

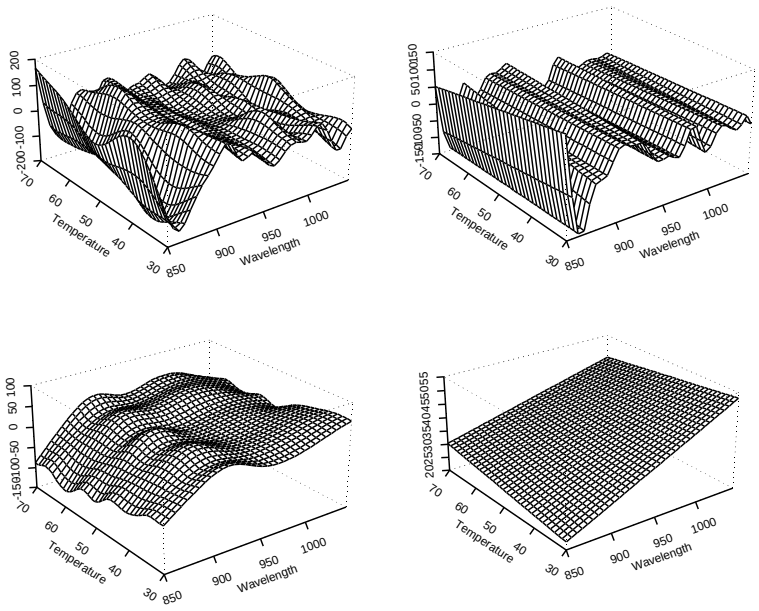


Figure 4: Examples of surfaces that can be generated from tensor products when constraining roughness in two dimensions.

the character of such nonlinearity; we aim to do so.

In multidimensional penalized signal regression, the mean $\mu_i$ of a response is given by $\sum_{jk} x_{ijk}\alpha_{jk}$, where the matrix $X_i$ contains the measured two dimensional spectrum with elements $[x_{jk}]$, with corresponding (smooth) coefficient matrix $[\alpha_{jk}]$. SISR introduces a small modification: $\mu_i = f(\sum_{jk} x_{ijk}\alpha_{jk})$. The function $f(\cdot)$ is assumed to be smooth and is estimated from the data using univariate P-splines, having its own additional tuning parameter. This model is generally related to "projection pursuit" (Friedman and Stuetzle, 1981), with additional smoothness demands on $\alpha$.

While imposing some identifiability constraints and using a linear approximation, we can cycle back and forth between MPSR and SISR, until convergence, to simultaneously estimate the coefficient surface and the nonlinear relationship.

### 1.5 Aims and benefits of the combined MSISR approach

We refer to the combination of MPSR and SISR as *multidimensional single-index signal regression* or MSISR. We will see that the basic appeal of this particular single-index model is its explicit estimation of meaningful components. We use the two-dimensional signal (spectrum) in a natural way: with a two-dimensional coefficient surface. Unlike some other "black box" approaches to the multivariate calibration paradigm that target nonlinear structure, our proposed approach provides explicit estimation of any nonlinearity relationship that may be present. Thus our combined approach provides a systematic means toward estimation of two separate model components: (1) the coefficient surface associated with the signal regressors, and (2) any nonlinear features or "link function"

associated with the response.

As we will see, MSISR is extremely flexible: even minor departures in $f$ from the identity function can lead to relatively dramatic changes in the estimated coefficient surface, while significantly improving (external) prediction. Moreover, although the estimation between $f$ and $\alpha$ is iterative, it is extremely tractable, essentially boiling down to repeated alternate applications of MPSR and P-spline smoothing on "working" responses and regressors.

Other additional features of MSISR that are worthy of note include: (a) Although smooth, $f$ can be assumed to be very general, an explicit function can be estimated. (b) Heavy penalization associated with $f$ typically produces low degree polynomial estimates for $f$, in some cases reducing back to standard MPSR. (c) The entire signal can be used as regressors. (d) The number of highly spatially correlated regressors can far exceed the number of observations. (e) The parameterization (and the effective dimension) of the coefficient surface is dramatically reduced, with a very manageable system of equations. (f) The candidate coefficient surface can be very general (non-additive), with heavy penalization yielding polynomial surfaces. (g) Since the two-dimensional signals and single estimated coefficient surface have a common indexing plane, potentially important regions can be visually identified. (h) Although we do not pursue it here, the MSISR approach can be transplanted to the generalized linear model (e.g. binary response) framework (Marx, Eilers, and Li, 2010).

### 1.6 Outline of this paper

In the next section, we provide a basic overview of tensor product B-splines and motivate the development of the penalty and its implementation to estimate the smooth coefficient surface. Section 3 brings the function $f$ into the MPSR setting providing the combined MSISR approach. We aim for reliable prediction, and in Section 4 we discuss cross-validation measures that can be used to optimally tune the parameters associated with the penalty. We also discuss effective dimension of the estimated coefficient surface. In Section 5, the proposed methodology is applied to the ternary mixture data set and directly compared to both the standard MPSR approach and partial least squares (PLS). We close with a brief discussion.

## 2 Recap: Tensor Product B-splines and MPSR

Eilers and Marx (2003, Section 4; 2005) provided a basic presentation *Tensor product B-splines in a nutshell*, giving essential details to building smooth surfaces; Dierckx (1995, Chapters 1 and 2) provides a much more complete coverage. The essential building block is a bicubic basis function, which is the *tensor product* of the two univariate (cubic) B-splines, say $B$ and $\breve{B}$ (presented on the margins). The axes are $v$ and $\breve{v}$ (e.g. emission and excitation wavelength), respectively. The tensor product has a zero (non-zero) value in the corresponding zero (non-zero) univariate B-spline support along $v$ and $\breve{v}$.

Tensor product B-splines exist in the $v \times \breve{v}$ plane. For our presentation, $n$ ($\breve{n}$) equally-

spaced indexing knots are placed on $v$ ($\breve{v}$) to yield a regularly-spaced grid, carving out the plane into subrectangles. The $r$th-$s$th single tensor product $B_r(v)\breve{B}_s(\breve{v})$ is positive in the rectangular region defined by the knots $R = [\varphi_r,\ \varphi_{r+q+2}] \times [\breve{\varphi}_s,\ \breve{\varphi}_{s+\breve{q}+2}]$ or on a support of spanned by $(q+2) \times (\breve{q}+2)$ knots. Similar to univariate B-splines, it is convenient to index each tensor product by one of the $n \times \breve{n}$ knot pairs and

$$\begin{aligned} B_r(v)\breve{B}_s(\breve{v}) &> 0 \text{ for all } v,\ \breve{v} \in R \\ &= 0 \text{ for all } v,\ \breve{v} \notin R, \end{aligned} \tag{1}$$

$r = 1,\ \ldots,\ n$ and $s = 1,\ \ldots\ , \breve{n}$. Figure 5 sparsely displays nine (scaled) tensor product B-splines, which represents only a portion of a full basis. A graphic of a complete basis would be difficult to appreciate, as the "hills" strongly overlap. Associated with each "hill", there is an unknown coefficient. A complete tensor product B-splines basis thus has an unknown coefficient matrix, denoted by $\Gamma_{n\times\breve{n}} = [\gamma_{rs}]$. For given knot grid, a very flexible surface can be approximated, e.g. at the digitized coordinates. For $j = 1, \cdots,\ p$ and $k = 1, \cdots,\ \breve{p}$,

$$\alpha(v_j^\star, \breve{v}_k^\star) = \sum_{r=1}^{n}\sum_{s=1}^{\breve{n}} B_r(v_j^\star)\breve{B}_s(\breve{v}_k^\star)\gamma_{rs}. \tag{2}$$

The surface is in fact driven by relatively few parameters ($n\breve{n}$), changing $\Gamma$ changes the surface.

## 2.1 Unfolding $\Gamma$ and notation

It is computationally efficient to reexpress the surface in a "unfolded" notation. Before doing so, some further notation is needed. Denote the support coordinate matrix $C = (v^\star \otimes \mathbf{1}_{\breve{p}},\ \mathbf{1}_p \otimes \breve{v}^\star)$ of dimension $p\breve{p} \times 2$. Let the matrix $\mathbf{B}_1^\star$ and $\mathbf{B}_2^\star$ (with respective dimensions $p\breve{p} \times n$ and of $p\breve{p} \times \breve{n}$) be the univariate B-spline basis matrix evaluated at each entry of the first and second column of $C$, respectively. The unfolded expression at the support coordinates then has the standard multiple regression form

$$\text{vec}\{\alpha(v^\star, \breve{v}^\star)\} = \mathbf{T}^\star\gamma, \tag{3}$$

where $\gamma = \text{vec}(\Gamma)$. Define the matrix

$$\mathbf{T}^\star = (\mathbf{B}_1^\star \otimes \mathbf{1}'_{\breve{n}}) \odot (\mathbf{1}'_n \otimes \mathbf{B}_2^\star) \tag{4}$$

of dimension $p\breve{p} \times n\breve{n}$. The symbols $\otimes$ and $\odot$ denote Kronecker product and elementwise multiplication of matrices, respectively. Penalized estimation of $\gamma$ and its use with two-dimensional signal regressors are topics discussed in the next section.

## 2.2 Smooth two-dimensional coefficient surfaces

Given the $i$th regressor matrix $X_i = [x_{ijk}]$ of dimension $p \times \breve{p}$, signal regressor support matrix $C$, and coefficient surface $\alpha(v, \breve{v})$, express the mean

$$\mu_i = \sum_{j=1}^{p}\sum_{k=1}^{\breve{p}} x_{ijk}\alpha(v_j^\star, \breve{v}_k^\star), \tag{5}$$

where $i = 1, \ldots, m;\ j = 1, \ldots, p;\ k = 1, \ldots, \breve{p}$. Using tensor product B-splines, (2) can be substituted into (5) yielding

$$\mu_i = \sum_{j=1}^{p} \sum_{k=1}^{\breve{p}} x_{ijk} \sum_{r=1}^{n} \sum_{s=1}^{\breve{n}} B_r(v_j^\star) \breve{B}_s(\breve{v}_k^\star) \gamma_{rs} = \mathbf{x}_i' \mathbf{T}^\star \gamma, \tag{6}$$

where $\mathbf{x}_i' = \text{vec}(X_i)$. We can further express (6) in matrix form as

$$\mu = \mathbf{X}\mathbf{T}^\star \gamma = \mathbf{M}\gamma,$$

where $\mathbf{X}$ is the $m \times p\breve{p}$ matrix of vectorized signals and $\mathbf{M} = \mathbf{X}\mathbf{T}^\star$.

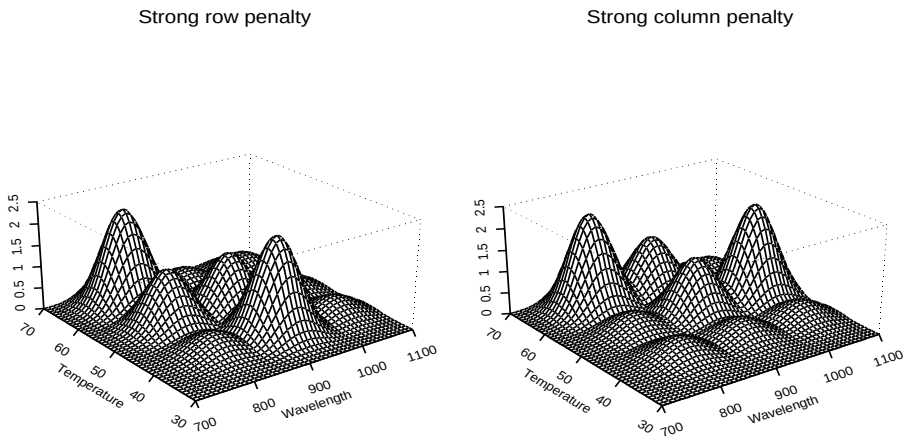


Figure 5: *Nine cubic B-spline tensor products, with a strong linear row penalty (left panel) and a strong linear column penalty (right panel)*

## 2.3 Penalizing the coefficient surface

In the P-spline spirit, a separate difference penalty is assigned to each row and to each column of $\Gamma$. The penalties have structure to effectively break the linkage in the penalty from row to row or from column to column. The objective function is now modified, using penalties, to minimize

$$\begin{aligned} Q_P(\gamma) &= \text{Residual SS} + \text{Row Penalty} + \text{Column Penalty} \\ &= \sum_{i=1}^{m} (y_i - \mathbf{x}_i' \mathbf{T}^\star \gamma)^2 + \lambda \sum_{r=1}^{n} \gamma_{r\bullet} D_d' D_d \gamma_{r\bullet}' + \breve{\lambda} \sum_{s=1}^{\breve{n}} \gamma_{\bullet s}' D_{\breve{d}}' D_{\breve{d}} \gamma_{\bullet s} \\ &= ||y - \mathbf{M}\gamma||^2 + \lambda ||P\gamma||^2 + \breve{\lambda} ||\breve{P}\gamma||^2, \end{aligned} \tag{7}$$

where $\gamma_{r\bullet}$ ($\gamma_{\bullet s}$) denotes the $r$th row (the $s$th column) of $\Gamma$.

The penalties are compactly represented using Kronecker products and matrix notation: $P = (D_d'D_d) \otimes I_{\breve{n}}$ and $\breve{P} = I_n \otimes (D_{\breve{d}}'D_{\breve{d}})$, where $I$ denotes the identity matrix and $d$ denotes the order of the difference penalty. Both $P$ and $\breve{P}$ are square with dimension $n\breve{n}$. The matrices $D_d$ and $D_{\breve{d}}$ are banded in structure, each have rows that consist of polynomial contrasts (see Marx and Eilers, 1996), and have dimension $(n-d)\times n$ and $(\breve{n}-\breve{d})\times\breve{n}$, respectively. The order of the penalties ($d$, $\breve{d}$) are in principle additional hyperparameters, but in practice are usually fixed by the user. The non-negative $\lambda$s essentially provide continuous control over smoothness. The influence or weight of the penalty is the same for each row, the same for each column, but are allowed to differ from rows to columns. Figure 5 displays a possible scenario resulting from strong row (top panel) and strong column (bottom panel) penalization using a second order penalty on each row and column with large $\lambda$ and $\breve{\lambda}$.

The explicit P-spline solution for (7) is

$$\hat{\gamma} = (\mathbf{M}'\mathbf{M} + \lambda P'P + \breve{\lambda}\breve{P}'\breve{P})^{-1}\mathbf{M}'y. \tag{8}$$

We see from (8) that the system of equations remains $n\breve{n}$ even as the resolution of the two-dimensional signal, $p\breve{p}$, increase dramatically. The predicted values are $\hat{y} = \mathbf{M}\hat{\gamma}$.

### 2.4 Modification for an intercept term

The model can also include an intercept term $\alpha_0$ which results in the modified P-spline solution

$$(\hat{\alpha}_0, \hat{\gamma}')' = (\mathbf{M_1'M_1} + \lambda P_\mathbf{1}'P_\mathbf{1} + \breve{\lambda}\breve{P}_\mathbf{1}'\breve{P}_\mathbf{1})^{-1}\mathbf{M_1'}y,$$

with $\mathbf{M_1} = (1_m|\mathbf{M})$, $P_\mathbf{1} = (0|P)$, and $\breve{P}_\mathbf{1} = (0|\breve{P})$. The zero vector in $P_\mathbf{1}$ and $\breve{P}_\mathbf{1}$ ensures an unpenalized intercept.

## 3 Bringing in $f$: MSISR Methodology

As mentioned, one problem with MPSR is that prediction quality is limited to estimated coefficients that are linear in the signal regressors, and this may be one explanation as to why penalized signal regression sometimes has difficulties competing with, e.g., machine learning approaches that can take advantage of nonlinear features of the signals.

The MSISR model has the form $\mu = f(\mathbf{M}\gamma)$, where the function $f$ and the smooth coefficient surface are unspecified and approximated with P-spline coefficients $\alpha$ and $\gamma$. Consequently, a modification of the MPSR objective in (7) can be rewritten as

$$Q_P^\star = ||y - f(\mathbf{M}\gamma)||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 + \lambda_f||D_d\alpha||^2. \tag{9}$$

Given the tensor B-spline coefficient vector $\gamma$, the estimation of function $f$ becomes a one-dimensional smoothing problem, and we can apply any scatter-plot smoother to obtain its estimate, which driven by the basis coefficient estimates $\hat{\alpha}$. As in Eilers, Li, and Marx

(2009), we estimate $f$ using a (cubic) P-spline scatter smoother (Eilers and Marx, 1996). The penalty on $\alpha$ ensures a smooth $f$; recall that $\alpha$ is the vector of B-spline coefficients with equally-spaced knots placed along estimated linear predictor, $\hat{\eta} = M\hat{\gamma}$. P-splines are used for the following reasons: (a) P-splines smoothers are easy to use and optimize. (b) Heavy smoothing (with a second order penalty) leads to approximately monotone linear $f$, which is expected by the thermodynamic properties. (c) The first derivative of $f$ (denoted as $\dot{f}$), which is needed in our algorithm, can be easily computed.

Note that derivatives of smoothers with equally-spaced B-splines have the pleasant property that they are equivalent to $B_{(q-1)}(\Delta\alpha)/b$, where $q$ is the degree, $\Delta$ denotes the first difference operator, and $b$ is the step length on the equally-spaced knots. For simplicity in notation, denote $S(V, W, \lambda_f, d_f, n_f)$ as the operation of fitting a cubic P-spline scatter smoother on $V$ (the input variable) and $W$ (the response) using the penalty tuning parameter $\lambda_f$ and difference order $d_f$ on the $n_f$ equally-spaced knots.

### 3.1 The model fitting algorithm

Once given an estimate of $f$, the coefficient vector $\gamma$ can be estimated using a (first-order) Taylor series approximation of the function $f$ (about the current estimate, $\gamma_0$). Specifically, if $\gamma_0$ is the current estimate for $\gamma$, then the current estimate of $\mu = f(\mathbf{M}\gamma)$ can be approximated by

$$f(\mathbf{M}\gamma) \approx f(\mathbf{M}\gamma_0) + \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}(\gamma - \gamma_0). \tag{10}$$

Using (10), with fixed $f$, we have an approximation of $Q_P^\star$

$$\begin{aligned} Q_P^\star &\approx ||y - f(\mathbf{M}\gamma_0) - \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}(\gamma - \gamma_0)||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \\ &= ||\left[y - f(\mathbf{M}\gamma_0) + \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}\gamma_0\right] - \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}\gamma||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \\ &= ||y^\star - \mathbf{M}^\star\gamma||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2, \end{aligned} \tag{11}$$

where $y^\star = y - f(\mathbf{M}\gamma_0) + \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}\gamma_0$ and $\mathbf{M}^\star = \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}$. Note that (11) implies that given $f$, the optimal $\alpha$ that minimizes the right-hand side of (11) can be obtained through a MPSR$(\mathbf{M}^\star, y^\star, (\lambda, \breve{\lambda}), (D_d, D_{\breve{d}}), (n, \breve{n}))$.

Hence, in our algorithm, we first carry out a MPSR with the response $y$ on $\mathbf{M}$ (Step 1). Then, given $\gamma$, an estimate of $f$ is obtained (Step 2). The two steps, estimation of $f$ and $\gamma$, are iterated until convergence of $\hat{\gamma}$. We set the B-spline basis degree $q = 3$ (cubic splines) as default value for both steps. Again only for simplicity of presentation, the intercept term is suppressed ($\alpha_0 = 0$) in the algorithm.

`Algorithm MSISR`

1. Initializations:
   - Choose the tuning parameter values $(\lambda, \breve{\lambda}, \lambda_f)$ for Steps 1 and 2
   - Choose number of knots $(n, \breve{n}, n_f)$
   - Choose penalty order $(d, \breve{d}, d_f)$

- Set all tuning parameters to $\lambda_0$ for the initial Step 1 (default value is $10^6$)
- Create $\mathbf{M} = X\mathbf{T}^\star$
- Calculate $\hat{\gamma} = \text{MPSR}(\mathbf{M}, y, (\lambda_0, \lambda_0), (d, \breve{d}), (n, \breve{d}))$

2. Cycle until convergence of $\hat{\gamma}$'s
   - Estimate $\hat{f}$ and the estimate of the derivative $\dot{f}$ from $S(\mathbf{M}\hat{\gamma}, y, \lambda_f, d_f, n_f)$
   - Obtain $y^\star$ and $\mathbf{M}^\star$
   - Update $\hat{\gamma} = \text{MPSR}(\mathbf{M}^\star, y^\star, (\lambda, \breve{\lambda}), (d, \breve{d}), (n, \breve{n}))$
   - Constrain $\hat{\gamma}/||\hat{\gamma}||$
3. Prediction: $\hat{y}^{new} = \hat{f}(x^{new}\mathbf{T}^\star\hat{\gamma})$

`end algorithm`

To ensure identifiability, we constrain the $\gamma$ vector to have a unit $L_2$ norm, i.e. $\gamma^{cur}/||\gamma^{cur}||$. To define the convergence criterion, denote $||\gamma||^2 = \sum_{k=1}^{n} \gamma_k^2$. The algorithm terminates when

$$\frac{(\gamma_k^{cur}/||\gamma^{cur}||) - (\gamma_k^{pre}/||\gamma^{pre}||)}{\gamma_k^{cur}/||\gamma^{cur}||} < \epsilon$$

for $k = 1, \ldots, n\breve{n}$, where $\gamma^{cur}$ ($\gamma^{pre}$) is the $\gamma$ vector for the current (previous) iteration, and $\epsilon$ is a prespecified convergence tolerance (default value is $10^{-3}$).

Thus the above algorithm outlines that, in principle, we have a generalized linear model with a least squares objective function. We will see in the next section that, given the link function, we can use the final iteration for some diagnostics, such as effective dimension.

# 4 Optimization of the penalty

Some general guidelines toward MSISR could include using a generous number of tensor product B-splines basis functions (keeping $n\breve{n} < 1000$ for computational efficiency) on the signal plane, as well as a generous number of univariate basis functions for $f$. We suggest using a row, column and $f$ penalty order of $d = 2$ or 3 and to vary $(\lambda, \breve{\lambda}, \lambda_f)$ on a logarithmic grid, while monitoring a performance criterion.

We aim for reliable prediction. The MSISR model is driven by the non-negative penalty regularization parameters $(\lambda, \breve{\lambda}, \lambda_f)$, which drive the continuous control over smoothness. There exists a variety of ways to "optimize" the tuning parameters. We propose the following: the data are split into three groups, denoted as the: *training set, validation set, test set*. Apply MSISR to the training set and choose "optimal" $(\lambda, \breve{\lambda}, \lambda_f)$ to minimize error on the validation set,

$$\text{RMSEV} = \sqrt{\frac{1}{m^{valid}} \sum_{i=1}^{m^{valid}} (y_i - \hat{y}_{vi})^2} \tag{12}$$

where $m^{valid}$ is the number of observations in the validation set and $\hat{y}_{vi}$ is the predicted response for the $i^{th}$ subject in the validation set, using the parameter estimates from the training model.

Given a chosen "optimal" model, evaluation of external predictive performance can be calculated using the root-mean-square error of prediction (RMSEP) on the independent test set:

$$\text{RMSEP} = \sqrt{\frac{1}{m^{test}} \sum_{i=1}^{m^{test}} (y_i - \hat{y}_i)^2} \tag{13}$$

where $m^{test}$ is the number of observations on the test set and $\hat{y}_i$ is the predicted response for the $i^{th}$ subject in the external test set, using the parameter estimates from the combined (training, validation) sets with the "optimal" $(\lambda, \breve{\lambda}, \lambda_f)$. Generally, we perform a full (three dimensional) linear grid search, where each element of (
$lambda,\ \breve{\lambda},\ \lambda_f$) is varied on a logarithm scale.

### Note on effective dimension

For the coefficient surface portion of the algorithm, note that we use penalized least squares. Specifically, we find the estimate

$$\hat{\gamma}(\lambda, \breve{\lambda} \mid \lambda_f) = \text{MPSR}(\mathbf{M}^\star, y^\star, (\lambda, \breve{\lambda}), (d, \breve{d}), (n, \breve{n})),$$

which is a function of the estimated $f$ and $\dot{f}$, with $\mathbf{M}^\star$ and $y^\star$ are defined in (9). An approximation to the effective hat matrix, corresponding to the surface, can then be given by

$$H^\star = [h^\star_{ii'}] = \mathbf{M}^\star(\mathbf{M}^{\star\prime}\mathbf{M}^\star + \lambda P'P + \breve{\lambda}\breve{P}'\breve{P})^{-1}\mathbf{M}^{\star\prime}. \tag{14}$$

It follows that the effective dimension of the estimated coefficient surface, for fixed $f$, can be approximated by

$$\text{ED}_\alpha = \text{trace}(H^\star)$$

(Hastie and Tibshirani, 1990). Using a permuted form of $H^\star$,

$$\text{trace}(H^\star) = \text{trace}\{\mathbf{M}^{\star\prime}\mathbf{M}^\star(\mathbf{M}^{\star\prime}\mathbf{M}^\star + \lambda P'P + \breve{\lambda}\breve{P}'\breve{P})^{-1}\}$$

is computed more efficiently. Additionally, the corresponding $\text{ED}_f$ for $\hat{f}$ can be found in a standard univariate way, as outlined in Eilers and Marx (1996).

## 5 Example: ternary mixture data

As introduced in Section 1.2 and in Figure 3, we apply our MSISR to the ternary mixture data. Recall that the responses are the mole fraction of a mixture, consisting of three components: water, 1,2-ethanediol, and 3-amino-1-propanol. There are 3 pure, 12 edge, and 19 interior (1 center) mixtures. A description of the experimental setting is provided in the Appendix.

The two-dimensional signal is constructed using the $p \times \breve{p} = 4800$ digitized regressors, $X_i$, arranged using the (first) differenced spectra, across the temperature levels. The indexing axes that define the support coordinates of $X_i$ are specified as $v^\star$ (wavelength) with $p = 400$ wavelength channels (701 to 1100nm, by 1 nm) and $\breve{v}^\star$ with $\breve{p} = 12$ temperature levels (30, 35, 37.5, 40, 45, 47.5, 50, 55, 60, 62.5, 65, 70$^o$ C). The data were not preprocessed in any other way.

We focus on a prediction performance study that directly compares the proposed MSISR method to the standard MPSR method (Marx and Eilers, 2005) and to partial least squares (PLS). One goal of this paper is to to exploit and explore the explicit nonlinear effect, which is an added benefit, uniquely provided using MSISR.

We divided the $m = 34$ observation into three subsets as follows. The training set consisted of $m^{train} = 16$ observations using the 3 pure, 12 edge, and 1 center mixtures. The remaining 18 interior observations (apart from the center) were divided into a validation set (to optimize tuning parameters) and a test set (to quantify quality of external prediction): (i) They were first sorted on the response (either on water or 1,2-ethanediol or 3-amino-1-propanol) in increasing order. (ii) The validation (test) set was constructed using the $m^{valid} = 9$ ($m^{test} = 9$) even (odd) rank of observations. Such an approach was taken in an attempt to have a fair and reasonable range of mixture levels for both the estimation of $f$ and the evaluation of external prediction. Additionally, there is no extrapolation for model optimization or model testing.

Optimal tuning parameters were determined by minimizing RMSEV in the trained model. Given these optimal tuning parameters, external prediction was evaluated on the test data using RMSEP using the newly trained model that combined both the training and validation data. We stress that in the determination of every optimal model, all hyper-parameters ($\lambda$s or number of PLS components) are optimized using a validation set which is independent of the external test set. We perform a full (three dimensional) grid search, where each $\log(\lambda), \log(\breve{\lambda}), \log(\lambda_f)$ was varied from $-5$ to 5 in twenty steps. For PLS, we varied the number of components from 1 to 7. The RMSEV optimization for PLS is provided in Figure 6. All reported prediction performance refers to external prediction on the test data. The number of equally-spaced knots were set to $n = 40$, $\breve{n} = n_f = 10$. Second order difference penalties ($d = \breve{d} = d_f = 2$) were used.

*Table 1. MSISR, MPSR, PLS external prediction RMSEP using optimal models.*

| Response | MSISR | MPSR | PLS |
|---|---|---|---|
| Water | 0.0214 | 0.0365 | 0.0465 |
| 1,2-ethanediol | 0.0241 | 0.0338 | 0.0382 |
| 3-amino-1-propanol | 0.0306 | 0.0251 | 0.0359 |

Table 1 presents the root mean square error of prediction (RMSEP) for the external prediction set, using optimal MSISR, MPSR, and PLS models. For responses water and 1,2-ethanediol, we find an improvement in external prediction for MSISR over both MPSR and PLS, leading to RMSEP reductions that range from 30% to 55%. For MSISR, the

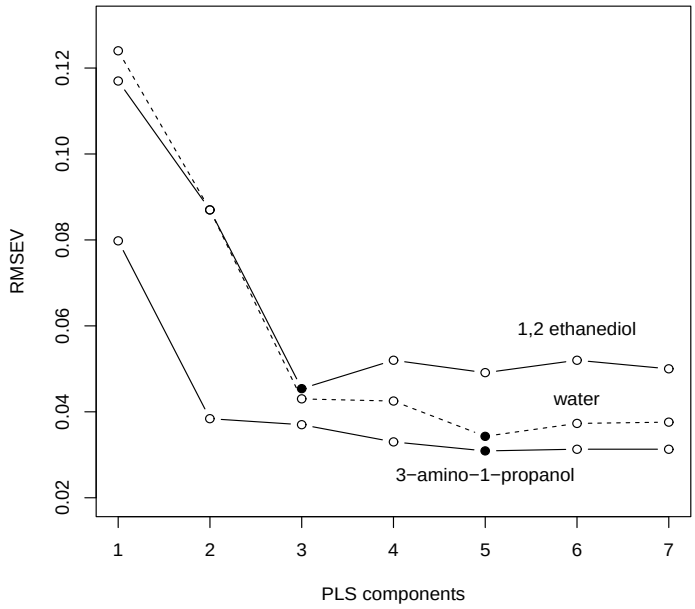


Figure 6: *Optimization the number of PLS components using RMSEV.*

external RMSEP values are between 0.0214 and 0.0241, which when multiplied by 100 gives units of percent mixture. The RMSEV values were optimized, achieving values of (0.0155, 0.0372, 0.0343) for water and (0.0173, 0.0508, 0.0454) for 1,2-ethanediol, for MSISR, MPSR and PLS, respectively.

The response 3-amino-1-propanol also showed very promising results for MSISR, with a 15% reduction in RMSEP when compared to PLS. Although the standard MPSR performed best of all three methods (with RMSEP of 0.0251), this particular MPSR response model model is problematic in that extremely light (boundary) smoothing was required for the estimated two-dimensional coefficient surface. In some respect, the MSISR model further regularizes the standard MPSR model by accounting for the nonlinear effect, yielding stronger prediction results with reasonable tuning parameters. The optimal RMSEV values were (0.0148, 0.0282, 0.0359) for MSISR, MPSR and PLS, respectively. Table 2 provides the optimal tuning parameters for each method and mixture response.

*Table 2. Optimal tuning parameters for MSISR, MPSR, and PLS models.*

| Response | MSISR ( $\lambda$,$\breve{\lambda}$,$\lambda_f$) | MPSR ($\lambda$,$\breve{\lambda}$) | PLS Components |
|---|---|---|---|
| Water | (4e-5, 9e-4, 2e-3) | (3e-5, 5e-4) | 5 |
| 1,2-ethanediol | (4e-5, 2e-1, 5e-1) | (8e-2, 6e-1) | 3 |
| 3-amino-1-propanol | (1e-5, 1e-3, 1e-3) | (1e-7*, 1e-7*) | 5 |

* light (boundary) smoothing required for MPSR

Figures 7−9 provide the estimated $\hat{f}$ function and $(\hat{f} - \hat{\eta})$ (left panels) for the thee mixture components, respectively. The upper, left panel shows $\hat{f}$ relative to the dashed identity line. The plotted points represent the nine observations in the external test data set. The lower, left panel provides a more focused view of the nature of the $\hat{f}$ function, by plotting $(\hat{f} - \hat{\eta})$. These left panels highlight the unique contribution of MSISR, i.e. the explicit estimation of the nonlinear $f$. In addition, the estimated smooth MSISR coefficient surface (top, right panel) is provided. The lower image panel further provides the *difference* in the coefficient surfaces, i.e. MSISR − MPSR. Note for the standard MPSR model for the water component, there are particular large swings in the estimated coefficient surface, ranging from −30 to 40 in the wavelength region 700 − 800nm. Such large values are moderated by a factor of 10 when using MSISR. These right panels further highlight a very nice feature of MPSR, i.e. that the spatial information of the regressors is built into the model and further the user can actually see the relative magnitudes for various regions of the coefficient indexing plane, as well as its interactive features.

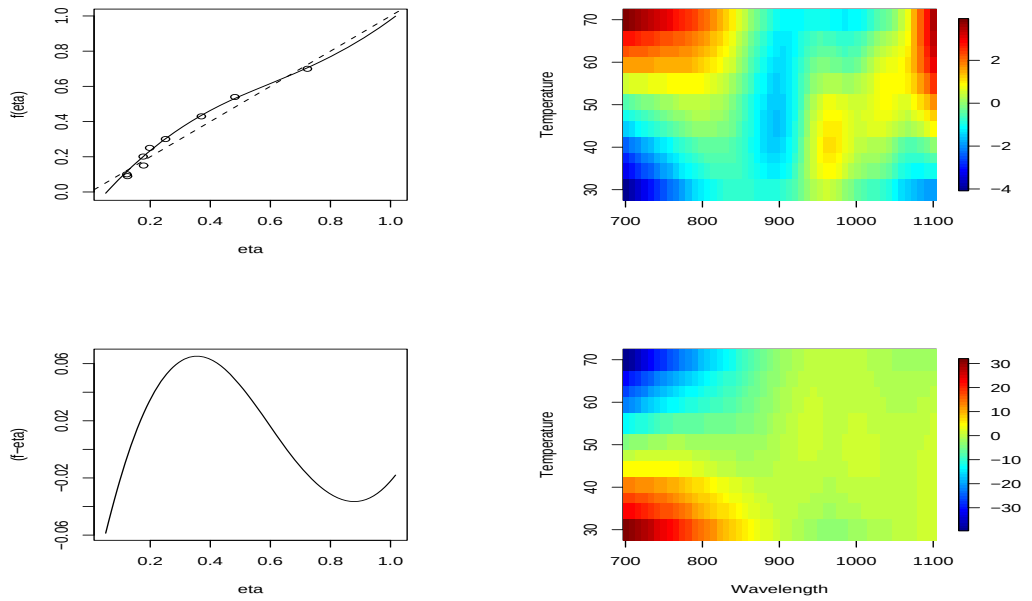


Figure 7: *Water: The estimated $\hat{f}$ function is given (upper, left), along with $(\hat{f} - \hat{\eta})$ (lower, left). The plotted points represent the nine observations in the external test data set. The right panels provide the "optimal" image plots for the estimated coefficient surface (upper) and the coefficient surface difference, MSISR−MPSR (lower).*

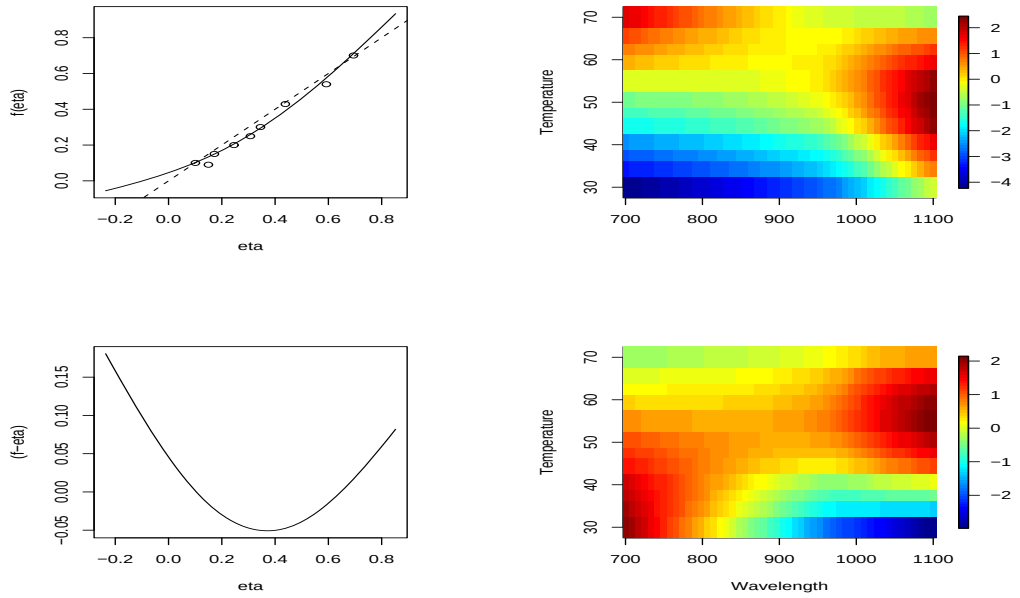


Figure 8: *1,2-ethanediol: The estimated $\hat{f}$ function is given (upper, left), along with $(\hat{f}-\hat{\eta})$ (lower, left). The plotted points represent the nine observations in the external test data set. The right panels provide the "optimal" image plots for the estimated coefficient surface (upper) and the coefficient surface difference, MSISR−MPSR (lower).*

### 5.1 Further inspection of the nonlinear components

The estimated link functions, as shown in Figures 7− 9, are clearly nonlinear. Unfortunately, by splitting our data in training, validation and test sets, no data points with extreme mole fractions are displayed, and thus do not have an idea of how well the functions fit to the data points. To get a better picture, we retrained the model using all data while using the optimal penalty parameters$(\lambda;\ \breve{\lambda};\ \lambda_f)$. Figure 10 shows the results. They suggest a quite sharp bend near zero, in an otherwise quite smooth function, for 1,2-ethanediol and 3-amino-1-propanol. Strong local changes in the the smoothness of a curve cannot be captured by our present model.

To get a better understanding of the behavior of the link function at small concentrations of 1,2-ethanediol and 3-amino-1-propanol, the model has to be improved, but also more detailed measurements at low mole fractions are needed. The data now jump from 0 to 0.15, with nothing in between.

We are not experts in chemical mixtures, so we will not try to give a fundamental explanation of the observed behavior. However, from the data, we find that at low concentrations of one organic compound in water, the spectrum of water behaves in a remarkable way. When the mole fraction of water increases form 0.75 to 1, the spectrum changes much more in size and shape than would be expected from a change by one third.

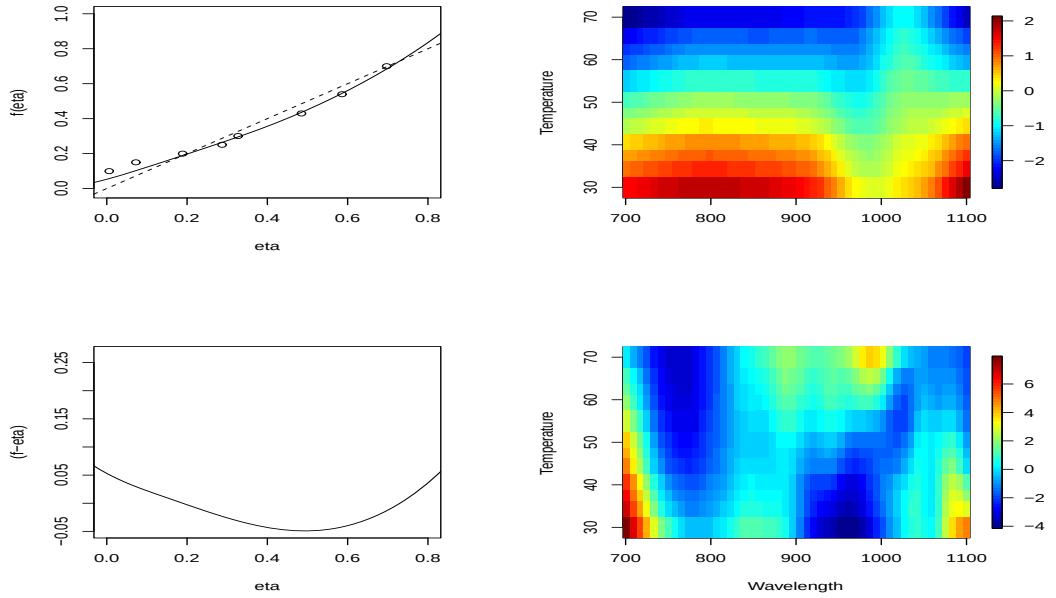


Figure 9: *3-amino-1-propanol: The estimated $\hat{f}$ function is given (upper, left), along with $(\hat{f}-\hat{\eta})$ (lower, left). The plotted points represent the nine observations in the external test data set. The right panels provide the "optimal" image plots for the estimated coefficient surface (upper) and the coefficient surface difference, MSISR−MPSR (lower).*

*Table 3. Summary of average external prediction for PLS, MPSR, and MSISR using optimized models, based on 50 replications. Standard deviations are in parentheses.*

| Method | Water | 1,2-ethanediol | 3-amino-1-propanol |
|---|---|---|---|
| PLS | 0.0531 (0.0092) | 0.0961 (0.0564) | 0.0818 (0.0487) |
| MPSR | 0.0594 (0.0184) | 0.0767 (0.0572) | 0.0819 (0.0619) |
| MSISR | 0.0411 (0.0208) | 0.0582 (0.0252) | 0.0726 (0.0543) |

## 5.2 Prediction performance: PLS, MPSR, and MSISR

In the experiment, we compared the prediction performance of each PLS, MPSR, and MSISR against each other through the *comparative test error*, defined by

$$c_{i,j} = \frac{RMSEP_{ij}}{min\{RMSEP_{i,l}\}_{l=1,2,3}},\ i = 1,\ldots,50,\ j = 1,2,3 \tag{15}$$

over 50 replications for each method. This quantity facilitates individual comparisons by using the test error of the best method for each data set to calibrate the difficulty of the problem. Within each replication, the entire data set is randomly split into training (16), validation (9) and testing (9) sets. Figure 11 shows the boxplots of comparative external

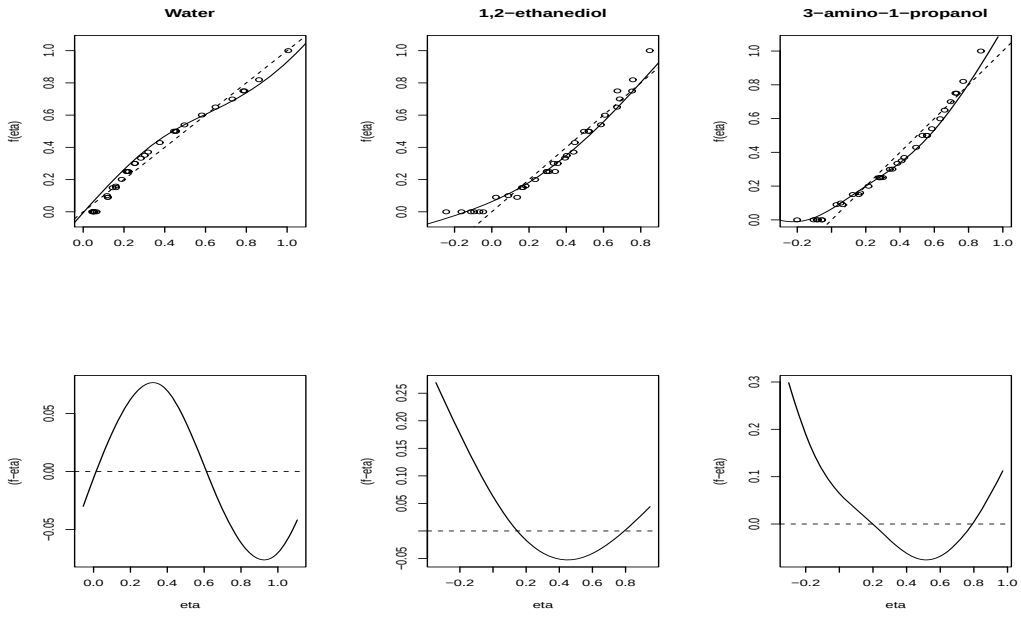


Figure 10: *The $\hat{f}$ functions trained with all $m = 34$ mixtures at optimal $(\lambda, \breve{\lambda}, \lambda_f)$. The corresponding $(\hat{f} - \hat{\eta})$ functions are provided in the lower panels.*

test errors for PLS, MPSR, and MSISR. For each replication, the tuning parameters for each method were optimized using RMSEV with the respective validation set. The optimum number of PLS components ranged from 3 to 12. Table 3 further shows the average RMSEP for PLS, MPSR, and MSISR for the external test sets based on the 50 replications, with corresponding standard deviations in parentheses. Figure 11 and Table 3 provides convincing evidence that MSISR achieves better external prediction performance than both MPSR or PLS, for all three components. For MPSR and MSISR, the number of knots and order of the penalty were set as above.

# 6 Evaluation of simulated nonlinear effect

We simulated a nonlinear effect on the response of the form $f(y) = y + 3(y - \frac{1}{2})^2$. We applied both the proposed MSISR, as well as the standard MPSR algorithms to these data. We expect that MSISR can capture explicit nonlinearity through the estimation of $f$, while have a smaller RMSEV than the (linear) MPSR counterpart.

Figure 12 (left) shows the nonlinear structure of the simulated response with an additive shock $\epsilon = z/30$ of scaled standard normals. Figure 12 (right) also provides the (optimal) MSISR estimated $f$, which clearly exhibit the recovery of some of the true underlying nonlinear response features. Although the coefficient surfaces are not displayed for either approach, the standard (linear) MPSR coeffient values took on a much wider

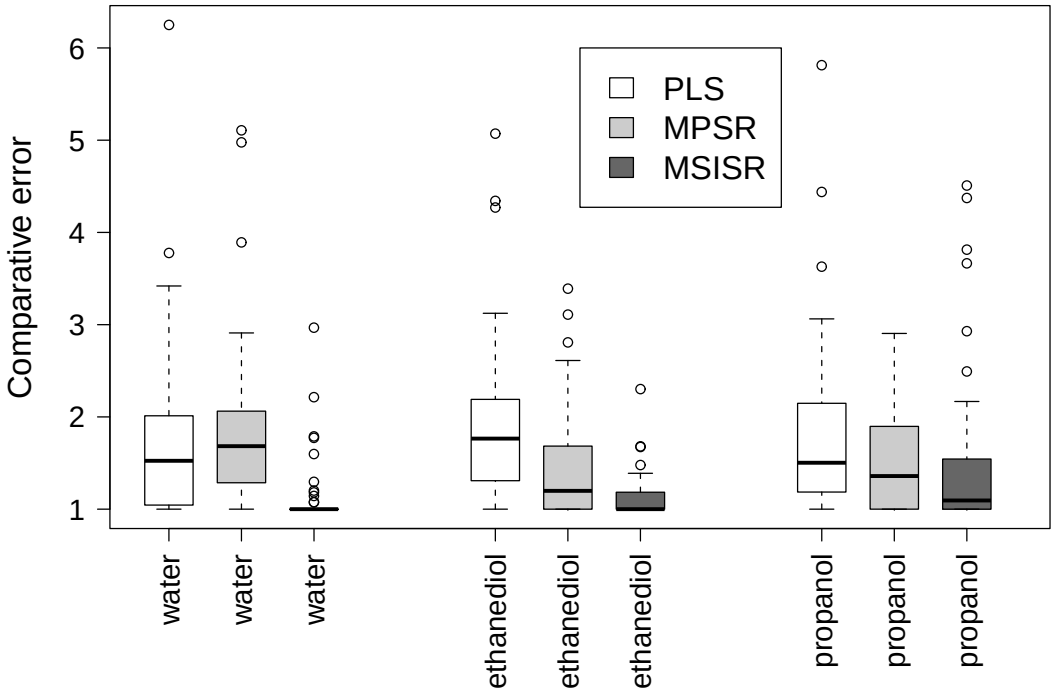


Figure 11: *Boxplot of comparative test errors for PLS, MPSR, and MSISR, based on 50 random splits of the data.*

range of values (approximately $\pm 30$) than MSISR (approximately $\pm 6$). We suspect that the more erratic nature MPSR coefficient surface is perhaps due to the linear estimators inability to capture true nonlinear response effects, which is also consistent with the corresponding increase in optimal RMSEV of approximately 0.05 (MSPR) compared to 0.02 (MSISR).

Some technical details follow regarding the construction of the response and optimization. The nonlinear $y^{\star\star}$ was obtained by first generating a linear response ($y^{\star}$). The linear $y^{\star}$ was constructed using the derivative spectra from the mixture experiment and a fixed linearly estimated coefficient surface. The fixed coefficient surface was determined using standard MPSR with fixed tuning parameters $(1,\ 0.1)$ and $(10,\ 40)$ equally-spaced knots for temperature and wavelength, respectively. The model was trained with all observations, apart from the validation set, and minimization of RMSEV determined the optimal tuning parameters.

# 7 Discussion

We have shown how to estimate nonlinear relationships in multivariate calibration, by combining the single index model with multidimensional penalized signal regression.

We stress that our MSISR approach, is not only a competitor, but has some clear

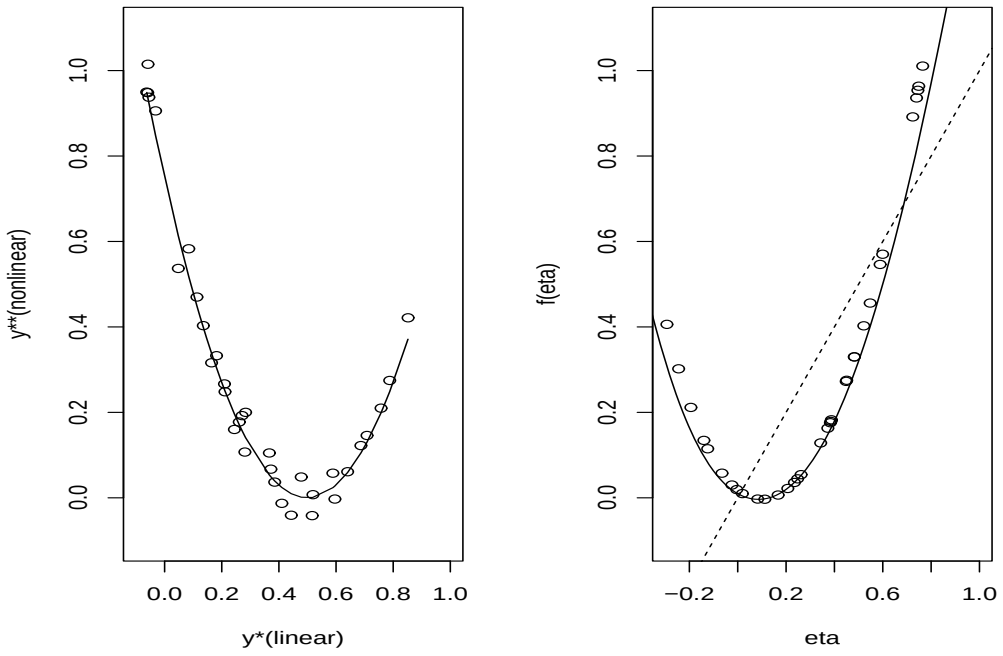


Figure 12: *Simulated nonlinear response (left)and MSISR fitted $f$, with all $m = 34$ observations (right).*

advantages. Unlike some other methods, MSISR takes full advantage of the natural spatial information of the signals. MPSR is straight-forward to use: it uses the entire ("raw") signal and works "right out of the box" without any data preprocessing. The method is intuitive in that you can actually see what is going relative to the spatial indexing plane, i.e. how the coefficient surface is used to contrast the two-dimensional signals for prediction. Further the nonlinearity is clearly estimated with a smooth function.

We found that the explicit estimation of the nonlinearity can provide some insights into the physical and chemical process underlying the measurements, which we view as a contribution over some of the other more "black box" approaches, while modestly improving external prediction.

As pointed out in Marx and Eilers (2005), our applications showed that the physical background of the measurements dictated the wavelength and temperature axes, and it was natural to allow different amounts of smoothing. When working with real images, the axes might be relatively arbitrary, only determined by the orientation of the imaging instrument. Then it might be natural to have an isotropic penalty, forcing both penalty parameters to be equal. Unfortunately, isotropic doe not imply rotation invariant: rotation of the axes might lead to different estimates of the coefficient surface. We note that it is possible to extend the penalties with mixed differences. Presently we use penalties that either work exclusively on the rows (columns) of the matrix of coefficients of tensor

products. A mixed penalty would work on both rows and columns.

The methods developed can be used beyond spectra; they can also be used for medical/grey-scale, or other images. Future research could constrain estimation so that the sum of the mixtures is one. Also future research could model the the two dimensional coefficient surface but explicitly create a varying signal coefficient vector by using slices at an arbitrary temperature, while possibly additionally modeling the link functions smoothly in two-dimensions. One could also imagine modeling two-dimensional image regressors while controlling for other (smooth) covariates or factors, along the lines of Eilers and Marx (2002). Other future research could investigate prediction stability during calibration transfer, i.e. investigating the robustness of prediction quality as an additional covariate changes, e.g. temperature (Marx and Eilers, 2002). In the present case the response is assumed to have a normal distribution. MSISR could also be generalized for binary classification or Poisson counts (Marx and Eilers, 2005, Section 6; Marx, Eilers, and Li, 2010), e.g. a Bernoulli response with probability $\pi_i$ could be modeled with $\log(\pi/(1-\pi)) = f(X\beta)$.

## Acknowledgement

The data we use were obtained in an unpublished experiment in 2001 by Zhenyu Wang at University of Amsterdam, under the supervision of Age Smilde. We are grateful for the permission to use the data.

## Appendix: Description of the experimental setting

According to the documentation provided by Zhenyu Wang, the following instruments and chemicals were used in the experiment:

- HP 8453 spectrophotometer (Hewlett-Packard, Palo Alto, CA)
- 2cm closed quartz cuvette with glass thermostatable jacket
- Pt-100 temperature sensor
- Neslab microprocessor EX-111 circulator bath
- UV-visible Chemstation software (Rev A.02.04) on a Hewlett-Packard Vectra XM2 PC
- Water: subboiled demi water (self made)
- 1,2ethanediol: 99.8% Sigma-Aldrich Germany
- 3amino1propanol: 99% Merk Schuchardt Germany

# References


**Cox, C.** (1984). Generalized linear models- the missing link. *Journal of the Royal Statistical Society, Series C*, **33**, 18-24.

**Dierckx, P.** (1995). *Curve and Surface Fitting with Splines*. Clarendon Press, Oxford.

**Eilers, P.H.C., Li, B. and Marx, B.D.** (2009). Multivariate calibration with single-index signal regression. *Chemometrics and Intelligent Laboratory Systems*, **96**, 196-202.

**Eilers, P.H.C. and Marx, B.D.** (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89-121.

**Eilers, P.H.C. and Marx, B.D.** (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics*, **11**(4), 758-783.

**Eilers, P.H.C. and Marx, B.D.** (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, **66**, 159-174.

**Eriksson, L., Trygg, J., Johansson, E., Bro, R., and Wold, S.** (2000). Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated fluorescence data. *Analytica Chimica Acta*, **420**, 181-195.

**Hastie, T. and Tibshirani, R. (1990).** *Generalized Additive Models*. Chapman and Hall, London.

**Friedman, J. H. and Stuetzle, W.** (1981). Projection pursuit regression, *Journal of the American Statistical Association* **76**, 817823.

**Marx, B.D. and Eilers, P.H.C.** (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, **41**, 1-13.

**Marx, B.D. and Eilers, P.H.C.** (2002). Multivariate calibration: a comparison of methods. *Journal of Chemometrics*, **16**, 129-140.

**Marx, B.D. and Eilers, P.H.C.** (2005). Multidimensional penalized signal regression, *Technometrics* **47**, 13-22.

**Marx, B.D., Eilers, P.H.C., and Li, B.** (2010). Generalized single-index signal regression, Technical Report 2010-1, Department of Experimental Statistics, Louisiana State University.

**Thissen, U., Üstün, B., Melssen, W.J. and Buydens, L.M.C.** (2004). Multivariate calibration with least-squares support vector machines, *Analytical Chemistry* **76**, 3099-3105.

**Üstün, B., Melssen, W.J., Oudenhuijzen, M. and Buydens, L.M.C.** (2005). Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization, *Analytica Chimica Acta* **544**, 292-305.