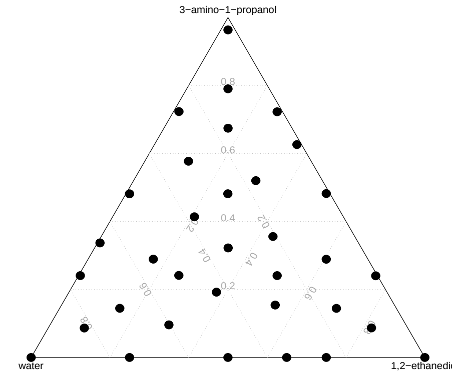
Varying-Coefficient Single-Index Signal Regression

Brian D. Marx
Department of Experimental Statistics
Louisiana State University

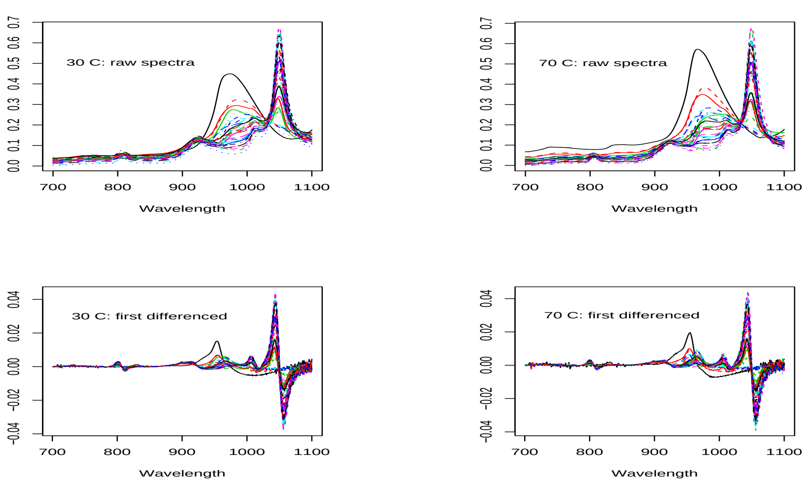
Ternary mixture experiment

- Composition based on mole fraction $(m = 34)$
- Ternary plot: 3 pure $+$ 12 edge $+$ 19 interior



Signal regressors for mixture experiment

- Two (2) temperatures presented $(30^o$, $70^o)$
- Raw and first differenced signals



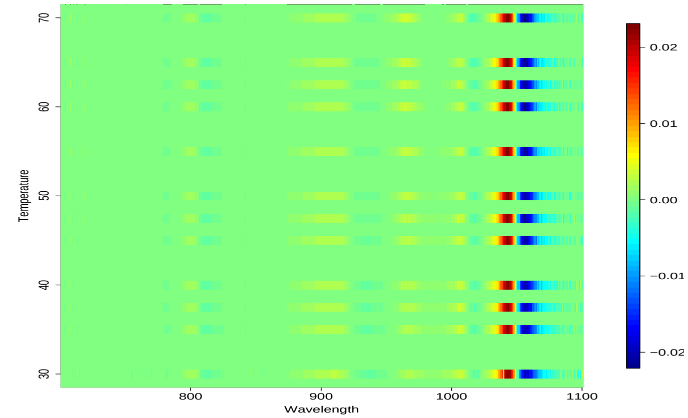
Unique structure and challenges

- *Primary goal*:
  Quality *external* prediction
- *Curious*:
  y is measured *exactly* (at molar level, by design)
- *Small* m:
  Only have $m = 34$ mixtures
- *Large* p:
  Rich signal regressors $(p = 400)$
- *No interest*:
  Internal prediction (perfectly attainable)
- *Oddly*:
  As t changes, then x (signals) change [but *not* y]

Realistic data structure

- Triplet (y_i, x_i, t_i)
  - Concentration (y) Signal digitizations (x) Covariate (t)
- Digitized optical NIR-spectra, $p = 400$
  - Ordered structure: 701 to 1100, by 1 nm
- For each mixture y:
  - Separate spectra at $\breve{p} = 12$ temperatures (t)
  - 30, 35, 37.5, 40, 45, 47.5, 50, 55, 60, 62.5, 65, 70^o C
- Consequence: 408 effective observations: $(N = m\breve{p})$
  - The 34 (m) mixtures have 34×12 recordings
- Responses y:
  - "Independent," with common σ^2

Regressor structure: in a "thousand words"

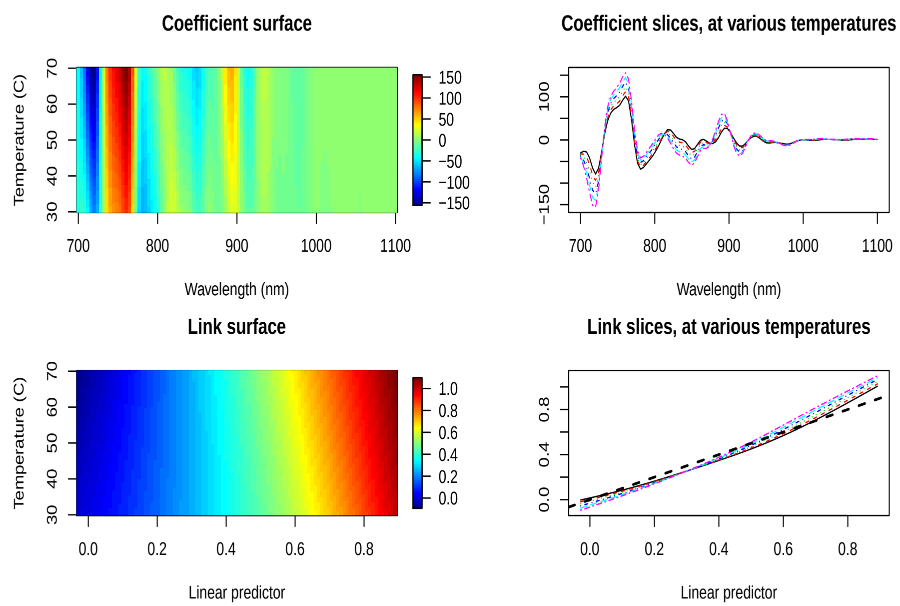


- Unique regressor "fingerprint": center mixture $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
- "Unfold": signal regressor image, by companion temperature (t)

Simultaneous goals

- Identify and estimate two separate modeling surfaces:
  1. Varying-signal coefficient surface (across covariate t)
  2. A *nonlinear* varying-link surface (across $t'(= t)$)
- Slice each surface at t
- Linear predictor: signal regressors with $\alpha(t)$
- Further *bend* mean: use link slice
- Combination: systematic, tractable, competitive
- Aim: reliable prediction + interpretability

Deliverable: Ethanol



Model formulation

Varying (penalized) signal regression

$$\mu_i = \sum_{j=1}^{p} x_{ij}\alpha(v_j, \underbrace{t_i}) = \eta_i$$

Estimate α as smooth surface [Note: t with index i]

1D signal regressors: x_i

Complete set of signals X $(N \times p)$

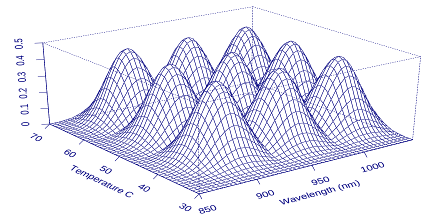
Varying single-index signal regression

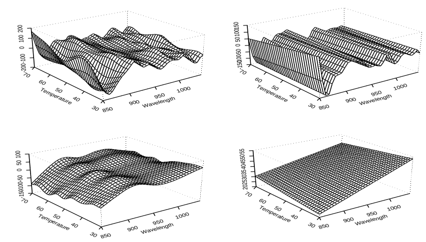
$$\mu_i = f(\eta_i, t_i)$$

Unknown 2D f surface, with own smoothness penalty

Smooth surfaces: Tensor product B-spline

Basis dimension: $n \times \breve{n}$ Coefficients: $\Gamma = [\gamma_{rs}]$





Varying-coefficient signal regression portion

$$\begin{aligned}
\mu_i &= \sum_{j=1}^{p} x_{ij} \overbrace{\sum_{r=1}^{n}\sum_{s=1}^{\breve{n}} B_r(v_j)\breve{B}_s(t_i)\gamma_{rs}}^{\alpha(v_j,t_i)} \qquad (1)\\
&= \sum_{j=1}^{p}\sum_{r=1}^{n}\sum_{s=1}^{\breve{n}} x_{ij}b_{jr}\breve{b}_{is}\gamma_{rs}\\
&= \sum_{r=1}^{n}\sum_{s=1}^{\breve{n}} \left(\sum_{j=1}^{p} x_{ij}b_{jr}\right)\breve{b}_{is}\gamma_{rs}\\
&= \sum_{r=1}^{n}\sum_{s=1}^{\breve{n}} u_{ir}\breve{b}_{is}\gamma_{rs}
\end{aligned}$$

- *Modified* tensor product expression, now using a basis $U = XB$.

Multiple regression format (unfold mountains)

- “Unfold” coefficient surface with tensor product B-splines

$$\text{vec}(\mu) = \mathbf{U}\gamma$$

$$\mathbf{U} = U\Box\breve{B} = (U \otimes \mathbf{1}'_{\breve{n}}) \odot (\mathbf{1}'_{n} \otimes \breve{B})$$

  - B, $\breve{B}$ bases built on: (wavelength, temperature)
  - **U** has dimension $N \times n\breve{n}$

- Similar approach for **link** surface

$$\text{vec}(f) = \mathbf{T}\theta$$

$$\mathbf{T} = B\Box\breve{B} = (B \otimes \mathbf{1}'_{\breve{n}_f}) \odot (\mathbf{1}'_{n_f} \otimes \breve{B})$$

  - B, $\breve{B}$ bases built on: (η, temperature)
  - **T** has dimension $N\breve{p} \times n_f\breve{n}_f$

Penalizing the coefficient surface

- Minimize:

$$\begin{aligned} Q_P(\gamma) &= \text{Residual SS} + \text{Row Penalty} + \text{Column Penalty} \\ &= \|y - \mathbf{U}\gamma\|^2 + \lambda \|P\gamma\|^2 + \breve{\lambda}\|\breve{P}\gamma\|^2, \end{aligned}$$

- Penalties directly on γ (tensor product coefficients)
- Two positive tuning parameters: $\lambda, \breve{\lambda}$
- Compact representation of penalties:

$$P = (D_d' D_d) \otimes I_{\breve{n}} \quad \text{and} \quad \breve{P} = I_n \otimes (D_{\breve{d}}' D_{\breve{d}})$$

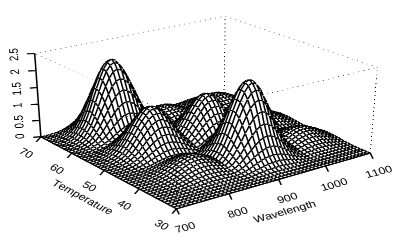
The penalties, more specifically

- Spirit of P-splines
- Must also carefully arrange ("stack") penalties
- Block diagonal to break (e.g. row to row) linkages:
  - $P = D \otimes I_{\breve{n}}$
  - $\breve{P} = I_n \otimes D$
  - Low dimensional example:

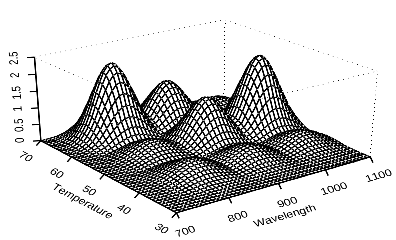
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes D = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Penalties in "action" (coefficient surface)

Strong row penalty



Strong column penalty



VPSR explicit solution (use rich basis):

$$\hat{\gamma} = (\mathbf{U}'\mathbf{U} + \lambda P'P + \breve{\lambda}\breve{P}'\breve{P})^{-1}\mathbf{U}'y$$

Bringing in f

Modified objective:

$$Q_P^\star = \|y - f(\overbrace{\mathbf{U}\gamma}^{\eta}, t)\|^2 + \lambda\|P\gamma\|^2 + \breve{\lambda}\|\breve{P}\gamma\|^2 + \lambda_f\|P_f\theta\|^2 + \breve{\lambda}_f\|\breve{P}_f\theta\|^2$$

- f two-dimensional (P-spline) smoothing of y on (η, t)
- 2D surface imbedded within another 2D surface
- f has own tuning parameters: λ_f, $\breve{\lambda}_f$
- Bonus: Surface (partial) derivative $\dot{f}_\partial$ easy to compute

Estimation, for fixed f [given t]

Approximate:

$$f(\mathbf{U}\gamma, t) \approx f(\mathbf{U}\gamma_0, t) + \dot{f}_\partial(\mathbf{U}\gamma_0, t)\mathbf{U}(\gamma - \gamma_0)$$

Objective:

$$\begin{aligned} Q_P^\star &\approx ||y - f(\mathbf{U}\gamma_0, t) - \dot{f}(\mathbf{U}\gamma_0, t)\mathbf{U}(\gamma - \gamma_0)||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \\ &= ||y^\star - \mathbf{U}^\star\gamma||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \end{aligned}$$

- Boils down to: VPSR$(\mathbf{U}^\star, y^\star, (\lambda, \breve{\lambda}), (D_d, D_{\breve{d}}), (n, \breve{n}))$
- Modified response and regressors:

$$y^\star = y - f(\mathbf{U}\gamma_0, t) + \dot{f}_\partial(\mathbf{U}\gamma_0, t)\mathbf{U}\gamma_0 \quad \text{and} \quad \mathbf{U}^\star = \text{diag}\{\dot{f}_\partial(\mathbf{U}\gamma_0, t)\}\mathbf{U}$$

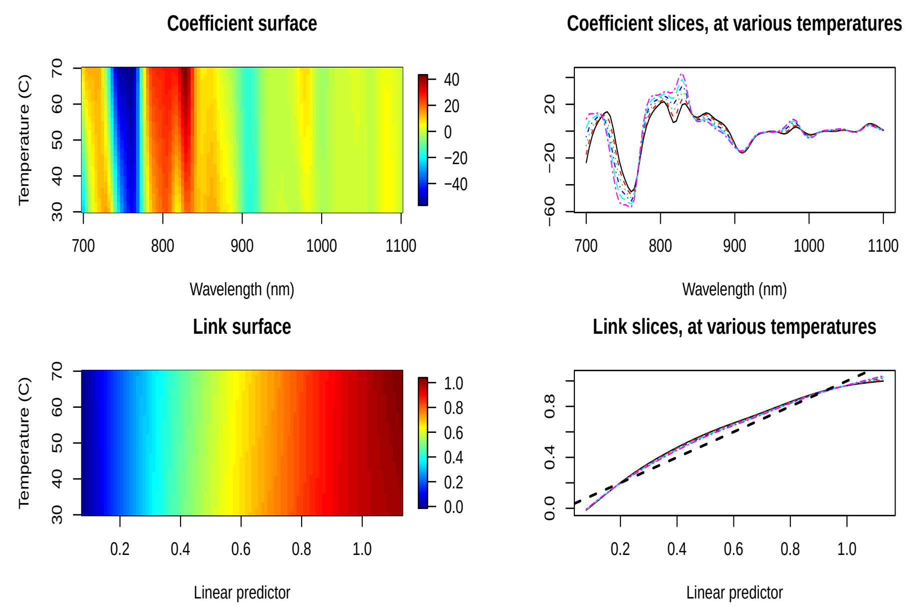
Basics of fitting algorithm (VSISR)

- Initialize with VPSR$(\mathbf{U}, y)$: get $\hat{\gamma}$
- Given $\hat{\gamma}$: get $\hat{f}$ with $T((\hat{\eta}, t), y)$
- Given $\hat{f}$: get partial derivative and modified $y^\star$, $\mathbf{U}^\star$
- Update $\hat{\gamma}$: VPSR$(\mathbf{U}^\star, y^\star)$
- Regularize norm of $\hat{\gamma}$
- Construct $\hat{\eta} = \mathbf{U}\hat{\gamma}$
- Cycle until convergence

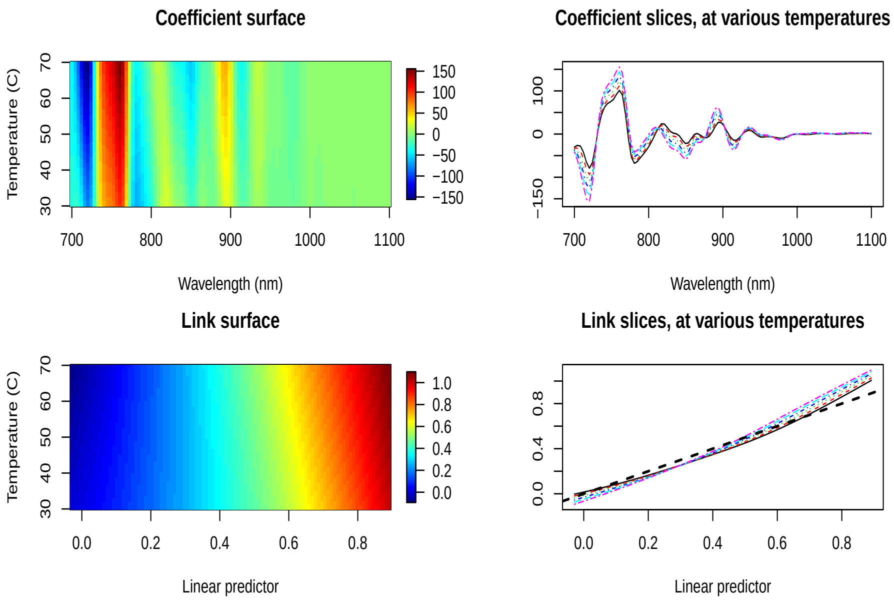
Single-index signal regression (VSISR) algorithm

1. Initializations:
   - Choose the tuning parameter values $(\lambda, \breve{\lambda}, \lambda_f, \breve{\lambda}_f)$ for Steps 1 and 2
   - Choose number of knots $(n, \breve{n}, n_f, \breve{n}_f)$
   - Choose penalty order $(d, \breve{d}, d_f, \breve{d}_f)$
   - Create $\mathbf{U} = U\Box\breve{B}$
   - Calculate $\hat{\gamma} =$ VPSR$(\mathbf{U}, y, (\lambda, \breve{\lambda}), (d, \breve{d}), (n, \breve{n}))$
2. Cycle until convergence of $\hat{\gamma}$
   - Estimate $\hat{f}$ surface and the partial derivative $\dot{f}_\partial$ from $T((\mathbf{U}\hat{\gamma}, t), y, (\lambda_f, \breve{\lambda}_f), (d_f, \breve{d}_f), (n_f, \breve{n}_f))$
   - Obtain $y^\star$ and $\mathbf{U}^\star$
   - Update $\hat{\gamma} =$ VPSR$(\mathbf{U}^\star, y^\star, (\lambda, \breve{\lambda}), (d, \breve{d}), (n, \breve{n}))$
   - Constrain $\hat{\gamma}/||\hat{\gamma}||$
3. Prediction: $\hat{y}^{new} = \hat{f}(\mathbf{u}^{new}\hat{\gamma},\ t^{new})$

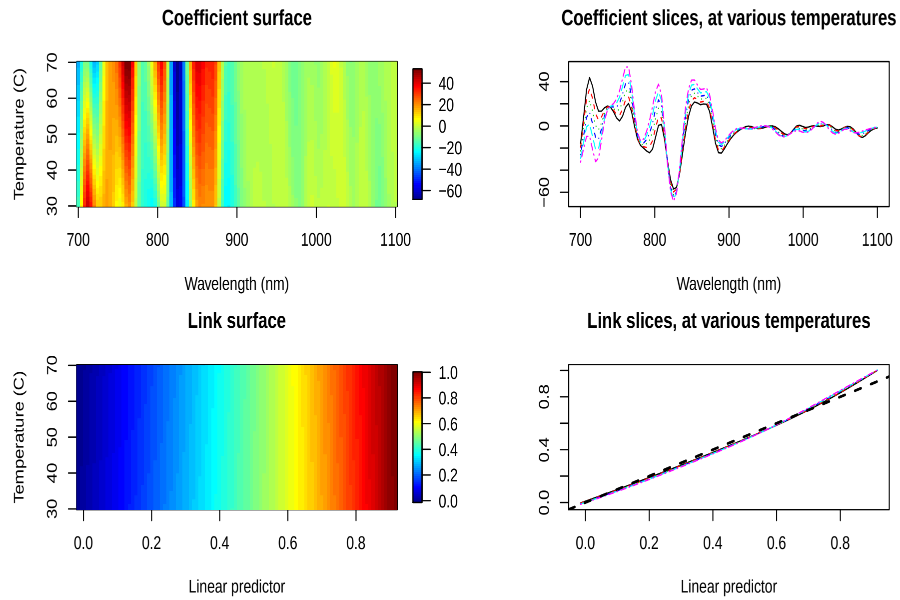
Deliverable: Water



Deliverable: Ethanol (again)



Deliverable: Isopropanol



Findings and design details

- Signal coefficients vary dramatically across temperature
  - Light smoothing in temperature direction ($\breve{\lambda}$)
- Link surfaces display moderate torsion
  - Water: nonlinearity that non-varies (perhaps 1D f)
  - Ethanol: nonlinearity that varies (inverted from Water)
  - Isopropanol: identity link perhaps sufficient
- Design parameters
  - Bicubic tensor products ($q = 3$)
  - Second order penalties ($d = 2$)
  - Coefficient grid: $n \times \breve{n} = 40 \times 20$
  - Link grid: $n_f \times \breve{n}_f = 10 \times 10$

(Optimal) tuning parameters

- Four (4) tuning parameters (λ, $\breve{\lambda}$, λ_f, $\breve{\lambda}_f$): (v, t), $(\eta,\ t)$
- Data splitting: training, validation, test sets
- Choose λs to minimize:

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

- $\hat{y}_v$: validation prediction, using training parameter estimates

(External) prediction performance

- Given "optimal" model, evaluate

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

- $\hat{y}$: test prediction, using combined (train, valid) estimates
- RMSEP is truly external

Idea of prediction performance: RMSEP

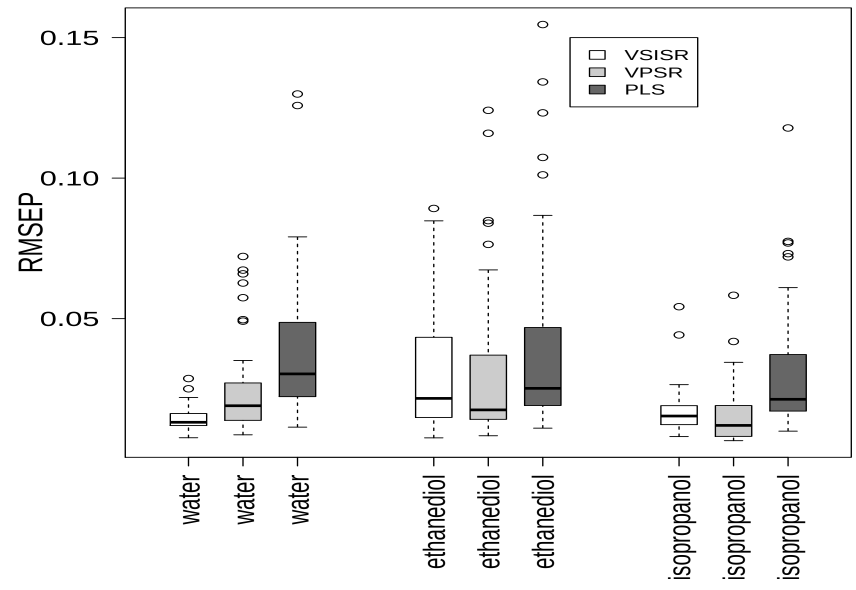
Response	VSISR	VPSR	PLS
Water	**0.0087**	0.0129	0.0367
1,2-ethanediol	**0.0094**	0.0104	0.0134
3-amino-1-propanol	0.0146	**0.0063**	0.0099

- Using optimal models
- RMSEP$\times 100$: units percent mixture (in stdevs)
- VSISR finds reductions of 33% to 77% (Water)
- VSISR finds reductions of 10% to 30% (Ethanol)

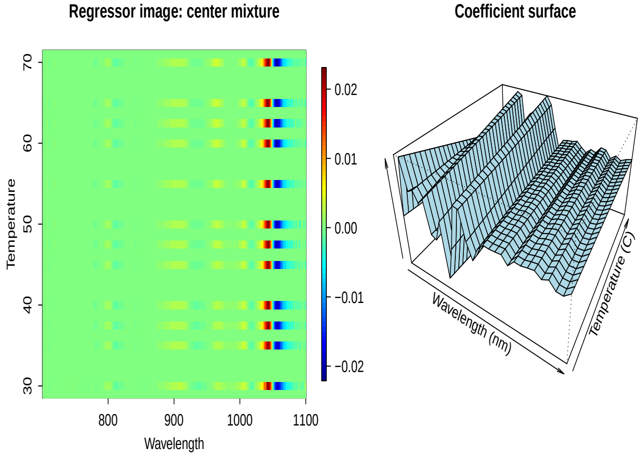
Details: data splitting

- $m = 34$ combinations (16 + 9 +9)
- 16=m^{train} (3 pure + 12 edge + 1 center)
- $9 = m^{valid} = m^{test}$ (each) interior
- Rank 18 interior: use odd (even) for valid (test)
- Use all $\breve{p}$ temperatures at each split
- Fair and reasonable range of mixture levels
- No extrapolation in external prediction

External prediction with random splitting



Alternative view: full 2D image regressors



- MPSR approach. No slices (multidimensional)
- Use entire image and common coefficient surface
- $\mu_i = \sum_{j=1}^{p} \sum_{k=1}^{\breve{p}} x_{ijk} \alpha(v_j, t_k)$ $\quad m = 34$ images (each mixture)

Delineation between VSISR and MSISR

	VPSR/VSISR	MPSR/MSISR
μ_i:	$\sum_{j=1}^{p} x_{ij} \alpha(v_j, t_i)$	$\sum_{j=1}^{p} \sum_{k=1}^{\breve{p}} x_{ijk} \alpha(v_j, t_k)$
Data:	(y, X, t)	(y, X)
length(y):	$m\breve{p}$	m
Regressor type:	signal	image
Dimension X:	$m\breve{p} \times p$	$m \times p\breve{p}$
Dimension B:	$p \times n$	$p\breve{p} \times n$
Dimension $\breve{B}$:	$m\breve{p} \times \breve{n}$	$p\breve{p} \times \breve{n}$
η:	$(XB)\Box\breve{B}\gamma$	$X(B\Box\breve{B})\gamma$

- Not directly comparable
- Number of "observations" differs by factor of $\breve{p}$

VSISR: Discussion and basic appeal

- Simplicity: doubly-varying coefficient and link surfaces
- Surfaces indexing can identify potentially "important" regions
- Nonlinear structure is targeted, gaining insight
- Surfaces with very general (non-additive) structure
- Highly competitive external prediction ability
- No "black box" algorithm
- No data preprocessing: the entire signals used (with t)
- Heavy penalization defaults to polynomial structure
- Managing of severe ill-conditioned model (and collinear data)
- Precision welcomed: the system of equations remains $n\breve{n}$

Future VSISR research

- Models that constrains the sum of mixture concentrations to be one
- Extensions allowing other (smooth) covariates or factors
- Prediction stability during calibration transfer/ robustness
- Generalized linear model approach to VSISR (binary or counts)