Correction of systematic disturbances in latent-variable calibration models

Paman Gujral, Michael Amrhein, Barry M. Wise, Enrique Guzman, Davyd Chivala and Dominique Bonvin

Laboratoire d'Automatique
Ecole Polytechnique Fédérale de Lausanne

Agenda

- Introduction
    - Calibration and prediction
    - Constituents of prediction error
- Unifying framework for different correction methodologies
- Illustrative examples
    - Simulation example
    - Two real data examples

Introduction

Background

Calibration: $\{\mathbf{X}_c, \mathbf{y}_c\} \rightarrow \hat{\mathbf{b}}$ Prediction: $\hat{\mathbf{y}}_p = \mathbf{X}_p\hat{\mathbf{b}}$

Introduction

Background

Calibration: $\{\mathbf{X}_c, \mathbf{y}_c\} \rightarrow \hat{\mathbf{b}}$ Prediction: $\hat{\mathbf{y}}_p = \mathbf{X}_p\hat{\mathbf{b}}$

$$\mathbf{x}_c^{\mathrm{T}} = y_c\mathbf{s}^{\mathrm{T}} + (\text{spectra from other species}) + \text{noise}_c$$

Introduction

Background

Calibration: $\{\mathbf{X}_c, \mathbf{y}_c\} \rightarrow \hat{\mathbf{b}}$ Prediction: $\hat{\mathbf{y}}_p = \mathbf{X}_p\hat{\mathbf{b}}$

$$\mathbf{x}_c^{\mathrm{T}} = y_c\mathbf{s}^{\mathrm{T}} + (\text{spectra from other species}) + \text{noise}_c$$

but

$$\mathbf{x}_p^{\mathrm{T}} = y_p\mathbf{s}^{\mathrm{T}} + (\text{spectra from other species}) + \mathbf{d}^{\mathrm{T}} + \text{noise}_p$$

Introduction

Constituents of prediction error

$$\begin{aligned}\hat{y}_p &= \mathbf{x}_p^{\mathrm{T}}\hat{\mathbf{b}} \\ &= (y_p\mathbf{s}^{\mathrm{T}} + \text{spectra from other species})\hat{\mathbf{b}} + \mathbf{d}^{\mathrm{T}}\hat{\mathbf{b}} + (\text{noise})\,\hat{\mathbf{b}}\end{aligned}$$

Prediction error $(y_p - \hat{y}_p)$ has three constituents:

1. due to noise (variance)
2. due to systematic disturbance (bias)
3. due to the PCR/PLSR modeling error (bias)

Unifying framework

EXPLICIT CORRECTION METHODS USING ADDITIONAL MEASUREMENTS

- **CC**: component correction, 2000
- **IIR**: independent interference reduction, 2001
- **GLSW**: generalized least squares weighting, 2003
- **EPO**: external parameter orthogonalization, 2003
- **TOP**: calibration transfer by orthogonal projection, 2004
- **DCPS**: difference correction of prediction samples, 2005
- **DOP**: dynamic orthogonal projection, 2006
- **EROS**: error removal by orthogonal subtraction, 2008

Unifying framework

STEP 1: ESTIMATION OF DRIFT SPACE

n_τ replicate measurements	n_τ reference measurements
Matched y-values	Non-matched y-values (e.g. uncontrolled online measurements)
D approximated as $\hat{\mathbf{D}} = \underbrace{\mathbf{X}_{\tau,2}}_{\textit{slave}} - \underbrace{\mathbf{X}_{\tau,1}}_{\textit{master}}$	**D** approximated as $\hat{\mathbf{D}} = \underbrace{\mathbf{X}_{\tau}}_{\textit{slave}} - \underbrace{\mathbf{A}\,\mathbf{X}_{c}}_{\textit{master}}$
GLSW & TOP (calibration transfer), EPO, DCPS & EROS (temperature changes), IIR (unknown variation), CC (unknown drift)	DOP (unknown drift)

Unifying framework

STEP 2: DRIFT-CORRECTION

Shrinking

GLSW

Calibrate with $\{\mathbf{X}_c\hat{\mathbf{W}}, \mathbf{y}_c\}$, where

$$\hat{\mathbf{W}} = \left(\frac{\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}}}{n_\tau - 1} + \alpha^2\mathbf{I}\right)^{-\frac{1}{2}}$$

Orthogonal projection

CC, IIR, EPO, DOP, TOP, EROS

Calibrate with $\{\mathbf{X}_c\hat{\mathbf{N}}, \mathbf{y}_c\}$

$$\begin{aligned}\hat{\mathbf{D}} &= \overset{\curvearrowright r}{\boxed{\mathbf{T}\,\mathbf{P}^{\mathrm{T}}}} + \mathbf{E}\\ \hat{\mathbf{N}} &= (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\end{aligned}$$

Subtraction

DCPS

Calibrate with $\{\mathbf{X}_c, \mathbf{y}_c\}$
Assuming one drift factor, $\hat{\mathbf{d}}$, correct the prediction sample:

$$\mathbf{x}_{p*} = \mathbf{x}_p - \hat{\beta}\,\hat{\mathbf{d}}$$

β optimized to minimize the 2-norm $||\mathbf{x}_{p*}||$.

Unifying framework

STEP 2: DRIFT-CORRECTION

Shrinking	Orthogonal projection	Subtraction
GLSW	**CC, IIR, EPO, DOP, TOP, EROS**	**DCPS**
Calibrate with $\{\mathbf{X}_c\hat{\mathbf{W}}, \mathbf{y}_c\}$, where $\hat{\mathbf{W}} = \left(\frac{\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}}}{n_\tau - 1} + \alpha^2 \mathbf{I} \right)^{-\frac{1}{2}}$	Calibrate with $\{\mathbf{X}_c\hat{\mathbf{N}}, \mathbf{y}_c\}$ $\hat{\mathbf{D}} = \overset{\curvearrowright r}{\boxed{\mathbf{T}\,\mathbf{P}^{\mathrm{T}}}} + \mathbf{E}$ $\hat{\mathbf{N}} = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})$	Calibrate with $\{\mathbf{X}_c, \mathbf{y}_c\}$ Assuming one drift factor, $\hat{\mathbf{d}}$, correct the prediction sample: $\mathbf{x}_{p*} = \mathbf{x}_p - \hat{\beta}\,\hat{\mathbf{d}}$ β optimized to minimize the 2-norm $\|\|\mathbf{x}_{p*}\|\|$.

ANALYTICAL RESULTS

$\underbrace{\qquad\qquad\qquad}$

(i) Equivalent for one drift component

Unifying framework

STEP 2: DRIFT-CORRECTION

Shrinking	Orthogonal projection	Subtraction
GLSW	**CC, IIR, EPO, DOP, TOP, EROS**	**DCPS**
Calibrate with $\{\mathbf{X}_c\hat{\mathbf{W}}, \mathbf{y}_c\}$, where $\hat{\mathbf{W}} = \left(\frac{\hat{\mathbf{D}}^{\mathrm{T}}\hat{\mathbf{D}}}{n_\tau - 1} + \alpha^2\mathbf{I}\right)^{-\frac{1}{2}}$	Calibrate with $\{\mathbf{X}_c\hat{\mathbf{N}}, \mathbf{y}_c\}$ $\hat{\mathbf{D}} = \overset{r}{\boxed{\mathbf{T}\,\mathbf{P}^{\mathrm{T}}}} + \mathbf{E}$ $\hat{\mathbf{N}} = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})$	Calibrate with $\{\mathbf{X}_c, \mathbf{y}_c\}$ Assuming one drift factor, $\hat{\mathbf{d}}$, correct the prediction sample: $\mathbf{x}_{p*} = \mathbf{x}_p - \hat{\beta}\,\hat{\mathbf{d}}$ β optimized to minimize the 2-norm $\|\|\mathbf{x}_{p*}\|\|$.

ANALYTICAL RESULTS

(ii) Equivalent when

$r = n_\tau$ and $\alpha \rightarrow 0$

(i) Equivalent for

one drift component

Step 2: Drift correction

CHOICE OF META-PARAMETERS (α, r)

- More complex than determining pseudo-rank of $\hat{\mathbf{D}}$
  - Wilks' $\curlywedge$ test, Malinowski's F-test, Faber-Kowalski F-test
  - Results based on random matrix theory and perturbation theory (Nadler et al.)

Step 2: Drift correction

CHOICE OF META-PARAMETERS (α, r)

- More complex than determining pseudo-rank of $\hat{\mathbf{D}}$
  - Wilks' $\curlywedge$ test, Malinowski's F-test, Faber-Kowalski F-test
  - Results based on random matrix theory and perturbation theory (Nadler et al.)

Constituents of prediction error

$$\begin{aligned}\hat{y}_p &= \mathbf{x}_p^{\mathrm{T}}\hat{\mathbf{b}} \\ &= (y_p\mathbf{s}^{\mathrm{T}} + \text{spectra from other species})\hat{\mathbf{b}} + \mathbf{d}^{\mathrm{T}}\hat{\mathbf{b}} + (\text{noise})_p\,\hat{\mathbf{b}}\end{aligned}$$

Step 2: Drift correction

CHOICE OF META-PARAMETERS (α, r)

- More complex than determining pseudo-rank of $\hat{\mathbf{D}}$
  - Wilks' ⅄ test, Malinowski's F-test, Faber-Kowalski F-test
  - Results based on random matrix theory and perturbation theory (Nadler et al.)

Constituents of prediction error

$$\begin{aligned}\hat{y}_p &= \mathbf{x}_p^{\mathrm{T}}\hat{\mathbf{b}} \\ &= (y_p\mathbf{s}^{\mathrm{T}} + \text{spectra from other species})\hat{\mathbf{b}} + \mathbf{d}^{\mathrm{T}}\hat{\mathbf{b}} + (\text{noise})_p\,\hat{\mathbf{b}}\end{aligned}$$

- Prediction error constituents after drift correction
  - $\mathbf{b} \perp$ estimated drift-space bias due to drift $|\mathbf{d}^{\mathrm{T}}\mathbf{b}| \downarrow$ ✔

Step 2: Drift correction

CHOICE OF META-PARAMETERS (α, r)

- More complex than determining pseudo-rank of $\hat{\mathbf{D}}$
  - Wilks' $\curlywedge$ test, Malinowski's F-test, Faber-Kowalski F-test
  - Results based on random matrix theory and perturbation theory (Nadler et al.)

Constituents of prediction error

$$\begin{aligned}\hat{y}_p &= \mathbf{x}_p^{\mathrm{T}}\hat{\mathbf{b}} \\ &= (y_p\mathbf{s}^{\mathrm{T}} + \text{spectra from other species})\hat{\mathbf{b}} + \mathbf{d}^{\mathrm{T}}\hat{\mathbf{b}} + (\text{noise})_p\,\hat{\mathbf{b}}\end{aligned}$$

- Prediction error constituents after drift correction
  - $\mathbf{b} \perp$ estimated drift-space — bias due to drift $|\mathbf{d}^{\mathrm{T}}\mathbf{b}| \downarrow$ ✓
  - RMSECV ↑ — bias in PCR/PLSR ↑ ✗

Step 2: Drift correction

CHOICE OF META-PARAMETERS (α, r)

- More complex than determining pseudo-rank of $\hat{\mathbf{D}}$
  - Wilks' ⅄ test, Malinowski's F-test, Faber-Kowalski F-test
  - Results based on random matrix theory and perturbation theory (Nadler et al.)

Constituents of prediction error

$$\begin{aligned}\hat{y}_p &= \mathbf{x}_p^{\mathrm{T}}\hat{\mathbf{b}} \\ &= (y_p\mathbf{s}^{\mathrm{T}} + \text{spectra from other species})\hat{\mathbf{b}} + \mathbf{d}^{\mathrm{T}}\hat{\mathbf{b}} + (\text{noise})_p\,\hat{\mathbf{b}}\end{aligned}$$

- Prediction error constituents after drift correction

$\mathbf{b} \perp$ estimated drift-space	bias due to drift $\lvert\mathbf{d}^{\mathrm{T}}\mathbf{b}\rvert \downarrow$	✔
RMSECV ↑	bias in PCR/PLSR ↑	✘
$\lvert\lvert\mathbf{b}\rvert\rvert_2 \uparrow$	variance due to noise ↑	✘

Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

- Data generation
  - Using Beer's law and known pure component spectra of 4 species
  - Drift in 7-dimensional loading space $\mathcal{S}(\mathbf{P}_d)$, overlapping with signal loading space $\mathcal{S}(\mathbf{P}_s)$

Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

- Data generation
  - Using Beer's law and known pure component spectra of 4 species
  - Drift in 7-dimensional loading space $\mathcal{S}(\mathbf{P}_d)$, overlapping with signal loading space $\mathcal{S}(\mathbf{P}_s)$
- Estimated drift-space for one (out of 5) reference sample

$$\hat{\mathbf{d}}^{\mathrm{T}} = {\sigma_{\mathbf{d}}}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{d,1}^{\mathrm{T}} \\ \mathbf{p}_{d,2}^{\mathrm{T}} \\ \mathbf{p}_{d,3}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{d,7}^{\mathrm{T}} \end{bmatrix} +$$

Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

- Data generation
  - Using Beer's law and known pure component spectra of 4 species
  - Drift in 7-dimensional loading space $\mathcal{S}(\mathbf{P}_d)$, overlapping with signal loading space $\mathcal{S}(\mathbf{P}_s)$
- Estimated drift-space for one (out of 5) reference sample

$$\hat{\mathbf{d}}^{\mathrm{T}} = {\sigma_{\mathbf{d}}}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{d,1}^{\mathrm{T}} \\ \mathbf{p}_{d,2}^{\mathrm{T}} \\ \mathbf{p}_{d,3}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{d,7}^{\mathrm{T}} \end{bmatrix} + {\sigma_{\mathbf{s}}}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{s,1}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{s,4}^{\mathrm{T}} \end{bmatrix} +$$

Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

- Data generation
  - Using Beer's law and known pure component spectra of 4 species
  - Drift in 7-dimensional loading space $\mathcal{S}(\mathbf{P}_d)$, overlapping with signal loading space $\mathcal{S}(\mathbf{P}_s)$
- Estimated drift-space for one (out of 5) reference sample

$$\hat{\mathbf{d}}^{\mathrm{T}} = \sigma_{\mathbf{d}}{}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{d,1}^{\mathrm{T}} \\ \mathbf{p}_{d,2}^{\mathrm{T}} \\ \mathbf{p}_{d,3}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{d,7}^{\mathrm{T}} \end{bmatrix} + \sigma_{\mathbf{s}}{}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{s,1}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{s,4}^{\mathrm{T}} \end{bmatrix} + \sigma_n \, randn(1, L)$$

Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

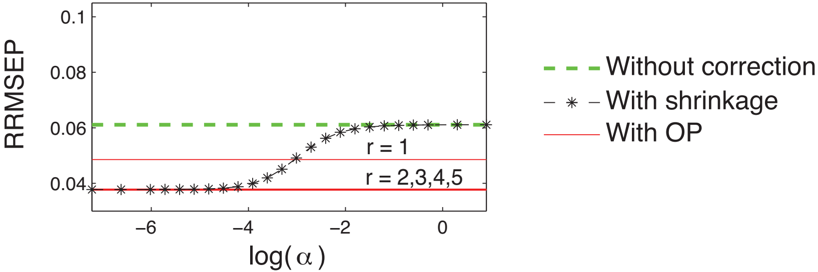
- Data generation
  - Using Beer's law and known pure component spectra of 4 species
  - Drift in 7-dimensional loading space $\mathcal{S}(\mathbf{P}_d)$, overlapping with signal loading space $\mathcal{S}(\mathbf{P}_s)$
- Estimated drift-space for one (out of 5) reference sample

$$\hat{\mathbf{d}}^{\mathrm{T}} = \sigma_{\mathbf{d}}{}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{d,1}^{\mathrm{T}} \\ \mathbf{p}_{d,2}^{\mathrm{T}} \\ \mathbf{p}_{d,3}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{d,7}^{\mathrm{T}} \end{bmatrix} + \sigma_{\mathbf{s}}{}^{\mathrm{T}} \begin{bmatrix} \mathbf{p}_{s,1}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{s,4}^{\mathrm{T}} \end{bmatrix} + \sigma_n \, randn(1, L)$$

$$\begin{aligned} \sigma_{\mathbf{d}}{}^{\mathrm{T}} &= [1 \times randn(1,2) \qquad 0.1 \times randn(1,5)] \\ \sigma_{\mathbf{s}}{}^{\mathrm{T}} &= [0.3 \times randn(1,4)] \\ \sigma_n &= 0.1 \end{aligned}$$

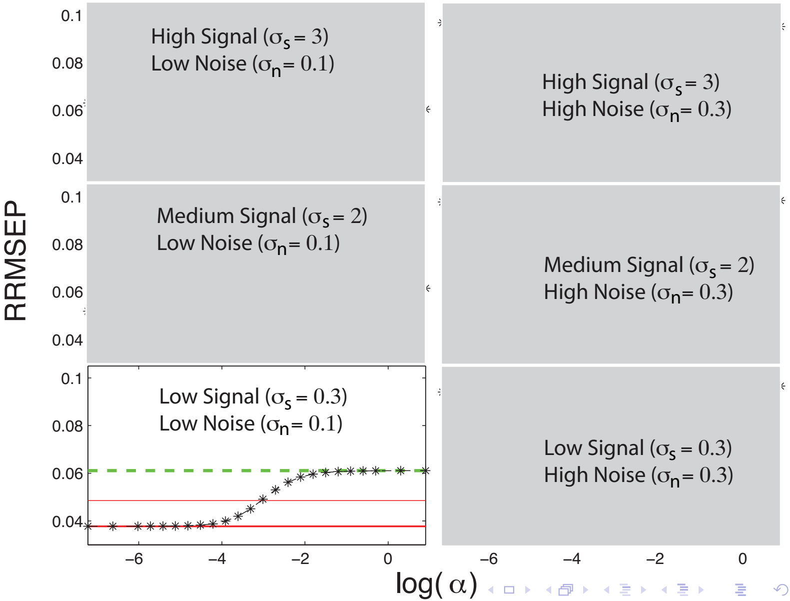
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION



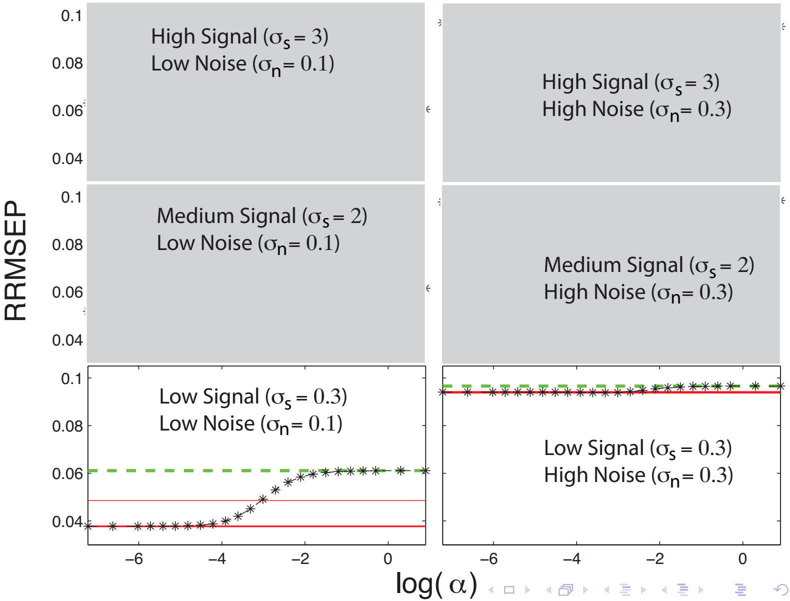
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION



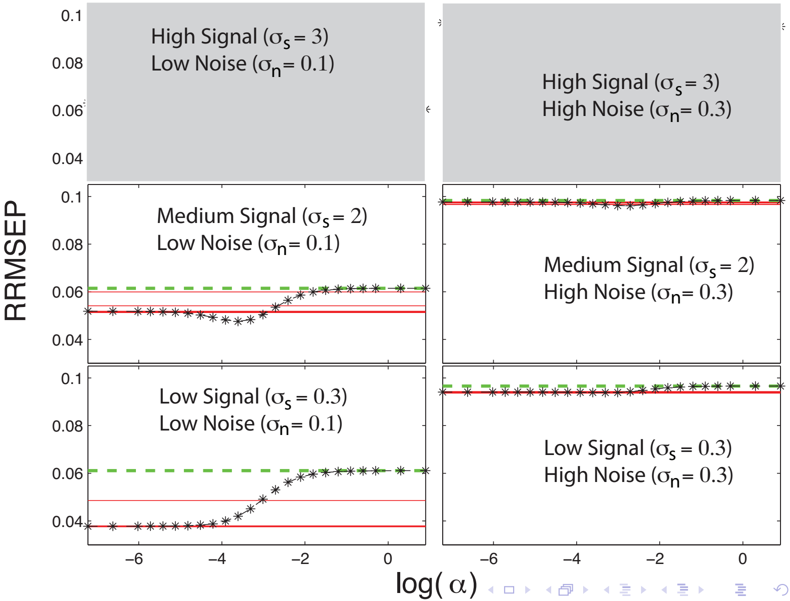
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION



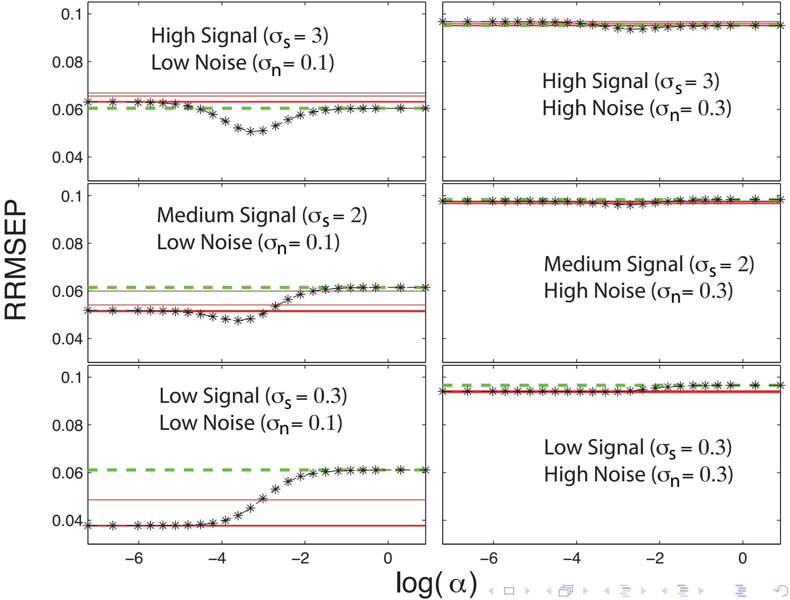
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION



Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION



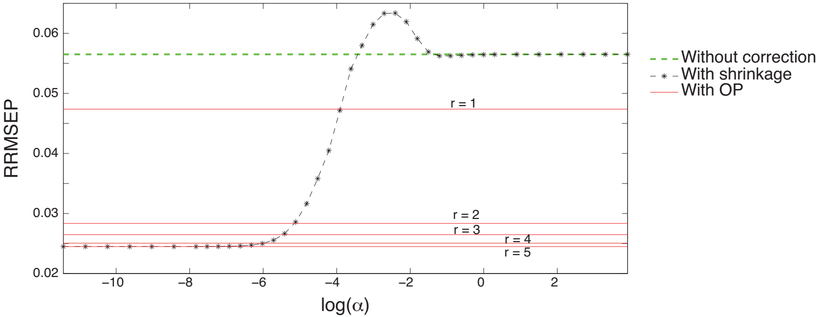
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 2: REAL DATA WITH TEMPERATURE EFFECTS

- NIR, 3 species

Calibration 11 x 2 = 22 @ 30, 40 °C

Prediction 11 x 3 = 33 @ 50, 60, 70 °C $\xrightarrow{\text{Monte Carlo}}$ 5 reference samples



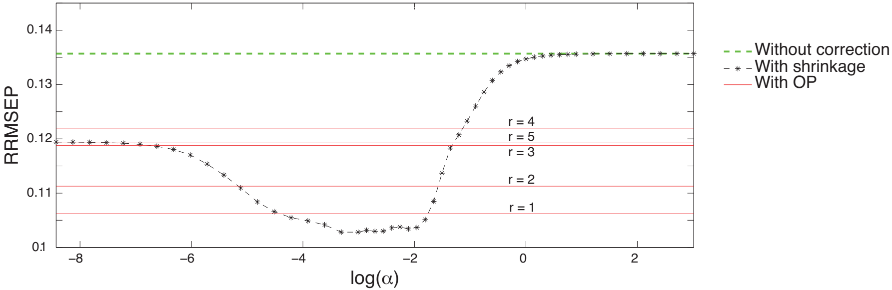
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 3: REAL DATA FOR CALIBRATION TRANSFER

- NIR, 4 measured properties

Calibration 40 on Instrument 1

Prediction 40 on Instrument 2 $\xrightarrow{\text{Monte Carlo}}$ 5 reference samples



Summary

- Unifying framework
  - Many drift-correction methods proceed in two steps: (i) drift estimation, (ii) drift correction

Summary

- Unifying framework
  - Many drift-correction methods proceed in two steps: (i) drift estimation, (ii) drift correction
- Illustrative examples
  - Shrinkage performs better than orthogonal projection typically when significant signal present in estimated drift

Summary

- Unifying framework
    - Many drift-correction methods proceed in two steps: (i) drift estimation, (ii) drift correction
- Illustrative examples
    - Shrinkage performs better than orthogonal projection typically when significant signal present in estimated drift
- Open issues
    - Statistically sound approach to choose meta-parameters

Summary

- Unifying framework
  - Many drift-correction methods proceed in two steps: (i) drift estimation, (ii) drift correction
- Illustrative examples
  - Shrinkage performs better than orthogonal projection typically when significant signal present in estimated drift
- Open issues
  - Statistically sound approach to choose meta-parameters
  - Multi shrinkage parameters $\{\alpha_1, \alpha_2, \dots\}$