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
Introduction
- Calibration and prediction
- Constituents of prediction error

Unifying framework for different correction methodologies

Illustrative examples
- Simulation example
- Two real data examples
Background

Calibration: \( \{X_c, y_c\} \rightarrow \hat{b} \) 

Prediction: \( \hat{y}_p = X_p \hat{b} \)
Background

Calibration: \( \{X_c, y_c\} \rightarrow \hat{b} \)

Prediction: \( \hat{y}_p = X_p \hat{b} \)

\[
x_c^T = y_c s^T + (\text{spectra from other species}) + \text{noise}_c
\]
Introduction

Background

Calibration: \( \{X_c, y_c\} \rightarrow \hat{b} \)  
Prediction: \( \hat{y}_p = X_p \hat{b} \)

\[
x_c^T = y_c s^T + (\text{spectra from other species}) + \text{noise}_c
\]

but

\[
x_p^T = y_p s^T + (\text{spectra from other species}) + d^T + \text{noise}_p
\]
Constituents of prediction error

\[ \hat{y}_p = x_p^T \hat{b} \]
\[ = (y_p s^T + \text{spectra from other species}) \hat{b} + d^T \hat{b} + (\text{noise}) \hat{b} \]

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

<table>
<thead>
<tr>
<th>$n_\tau$ replicate measurements</th>
<th>$n_\tau$ reference measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matched $y$-values</strong></td>
<td><strong>Non-matched $y$-values (e.g. uncontrolled online measurements)</strong></td>
</tr>
</tbody>
</table>

**D** approximated as

$$\hat{D} = X_{\tau,2} - X_{\tau,1}$$

- **slave**
- **master**

GLSW & TOP (calibration transfer), EPO, DCPS & EROS (temperature changes), IIR (unknown variation), CC (unknown drift)

**D** approximated as

$$\hat{D} = X_{\tau} - A X_c$$

- **slave**
- **master**

DOP (unknown drift)
## Unifying framework

### STEP 2: DRIFT-CORRECTION

<table>
<thead>
<tr>
<th>Shrinking</th>
<th>Orthogonal projection</th>
<th>Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLSW</td>
<td>CC, IIR, EPO, DOP, TOP, EROS</td>
<td>DCPS</td>
</tr>
</tbody>
</table>

Calibrate with \( \{X_c \hat{W}, y_c\} \), where

\[
\hat{W} = \left( \frac{\hat{D}^T \hat{D}}{n_T - 1} + \alpha^2 I \right)^{-\frac{1}{2}}
\]

Calibrate with \( \{X_c \hat{N}, y_c\} \)

\[
\hat{D} = TP^T + E
\]

\[
\hat{N} = (I - PP^T)
\]

Calibrate with \( \{X_c, y_c\} \)

Assuming one drift factor, \( \hat{d} \), correct the prediction sample:

\[
x_{p*} = x_p - \hat{\beta} \hat{d}
\]

\( \beta \) optimized to minimize the 2-norm \( ||x_{p*}|| \).
## Unifying framework

### STEP 2: DRIFT-CORRECTION

<table>
<thead>
<tr>
<th>Shrinking</th>
<th>Orthogonal projection</th>
<th>Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLSW</td>
<td>CC, IIR, EPO, DOP, TOP, EROS</td>
<td>DCPS</td>
</tr>
</tbody>
</table>

**Shrinking**

Calibrate with \(\{X_c \hat{W}, y_c\}\), where

\[
\hat{W} = \left( \frac{\hat{D}^T \hat{D}}{n_T - 1} + \alpha^2 I \right)^{-\frac{1}{2}}
\]

**Orthogonal projection**

Calibrate with \(\{X_c \hat{N}, y_c\}\)

\[
\begin{align*}
\hat{D} &= TP^T + E \\
\hat{N} &= (I - PP^T)
\end{align*}
\]

**Subtraction**

Calibrate with \(\{X_c, y_c\}\)

Assuming one drift factor, \(\hat{d}\), correct the prediction sample:

\[
x_{p*} = x_p - \beta \hat{d}
\]

\(\beta\) optimized to minimize the 2-norm \(||x_{p*}||\).

### ANALYTICAL RESULTS

(i) Equivalent for one drift component
## Unifying framework
### STEP 2: DRIFT-CORRECTION

<table>
<thead>
<tr>
<th>Shrinking</th>
<th>Orthogonal projection</th>
<th>Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLSW</td>
<td>CC, IIR, EPO, DOP, TOP, EROS</td>
<td>DCPS</td>
</tr>
</tbody>
</table>

Calibrate with \( \{X_c \hat{W}, y_c\} \), where

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

Calibrate with \( \{X_c \hat{N}, y_c\} \)

\[
\hat{D} = TP^T + E \\
\hat{N} = (I - PP^T)
\]

Calibrate with \( \{X_c, y_c\} \)

Assuming one drift factor, \( \hat{d} \), correct the prediction sample:

\[
x_{p*} = x_p - \hat{\beta} \hat{d}
\]

\( \beta \) optimized to minimize the 2-norm \( ||x_{p*}|| \).

### ANALYTICAL RESULTS

- (i) Equivalent for one drift component
- (ii) Equivalent when \( r = n_\tau \) and \( \alpha \to 0 \)
Step 2: Drift correction

CHOICE OF META-PARAMETERS ($\alpha, r$)

- More complex than determining pseudo-rank of $\hat{D}$
  - Wilks’ $\lambda$ 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** $(\alpha, r)$

- More complex than determining pseudo-rank of $\hat{D}$
  - 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

\[
\hat{y}_p = x_p^T \hat{b} = (y_p s^T + \text{spectra from other species}) \hat{b} + d^T \hat{b} + (\text{noise})_p \hat{b}
\]
Step 2: Drift correction

CHOICE OF META-PARAMETERS ($\alpha, r$)

- More complex than determining pseudo-rank of $\hat{D}$
  - Wilks’ $\lambda$ 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

\[
\hat{y}_p = x_p^T \hat{b} = (y_p s^T + \text{spectra from other species}) \hat{b} + d^T \hat{b} + (\text{noise})_p \hat{b}
\]

Prediction error constituents after drift correction

- $\hat{b} \perp$ estimated drift-space
  - Bias due to drift $|d^T \hat{b}| \downarrow$
Step 2: Drift correction
CHOICE OF META-PARAMETERS ($\alpha, r$)

- More complex than determining pseudo-rank of $\hat{D}$
  - 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

$$\hat{y}_p = x_p^T \hat{b}$$
$$= (y_p s^T + \text{spectra from other species}) \hat{b} + d^T \hat{b} + \text{(noise)}_p \hat{b}$$

- Prediction error constituents after drift correction
  - $\hat{b} \perp \text{estimated drift-space}$
  - $\text{bias due to drift } |d^T \hat{b}| \downarrow$ ✔
  - $\text{bias in PCR/PLSR } \uparrow$ ✗
  - $\text{RMSECV } \uparrow$

Gujral (LA-EPFL)
Step 2: Drift correction

CHOICE OF META-PARAMETERS ($\alpha, r$)

- More complex than determining pseudo-rank of $\hat{D}$
  - 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

$$\hat{y}_p = x_p^T \hat{b}$$

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

- Prediction error constituents after drift correction
  - $\hat{b} \perp$ estimated drift-space
  - RMSECV $\uparrow$
  - $\|\hat{b}\|_2 \uparrow$
  - Bias due to drift $|d^T \hat{b}| \downarrow \checkmark$
  - Bias in PCR/PLSR $\uparrow \times$
  - Variance due to noise $\uparrow \times$
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 $S(P_d)$, overlapping with signal loading space $S(P_s)$
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 $S(P_d)$, overlapping with signal loading space $S(P_s)$

- **Estimated drift-space for one (out of 5) reference sample**

\[
\hat{d}^T = \sigma_d^T \begin{bmatrix}
p_{d,1}^T \\
p_{d,2}^T \\
p_{d,3}^T \\
\vdots \\
p_{d,7}^T 
\end{bmatrix} +
\]
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 $S(P_d)$, overlapping with signal loading space $S(P_s)$

Estimated drift-space for one (out of 5) reference sample

$$\hat{d}^T = \sigma_d^T \begin{bmatrix} p_{d,1}^T \\ p_{d,2}^T \\ p_{d,3}^T \\ \vdots \\ p_{d,7}^T \end{bmatrix} + \sigma_s^T \begin{bmatrix} p_{s,1}^T \\ \vdots \\ p_{s,4}^T \end{bmatrix}$$
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 $S(P_d)$, overlapping with signal loading space $S(P_s)$

- **Estimated drift-space for one (out of 5) reference sample**

\[
\hat{d}^T = \sigma_{d}^T \begin{bmatrix} p_{d,1}^T \\ p_{d,2}^T \\ p_{d,3}^T \\ \vdots \\ p_{d,7}^T \end{bmatrix} + \sigma_{s}^T \begin{bmatrix} p_{s,1}^T \\ \vdots \\ p_{s,4}^T \end{bmatrix} + \sigma_n \text{randn}(1, L)
\]
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 $S(P_d)$, overlapping with signal loading space $S(P_s)$

- Estimated drift-space for one (out of 5) reference sample

\[
\hat{d}^T = \sigma_d^T \begin{bmatrix}
    p_{d,1}^T \\
p_{d,2}^T \\
p_{d,3}^T \\
\vdots \\
p_{d,7}^T
\end{bmatrix} + \sigma_s^T \begin{bmatrix}
p_{s,1}^T \\
\vdots \\
p_{s,4}^T
\end{bmatrix} + \sigma_n \ randn(1, L)
\]

\[
\sigma_d^T = \begin{bmatrix} 1 \times \ randn(1, 2) & 0.1 \times \ randn(1, 5) \end{bmatrix}
\]

\[
\sigma_s^T = \begin{bmatrix} 0.3 \times \ randn(1, 4) \end{bmatrix}
\]

\[
\sigma_n = 0.1
\]
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

![Graph showing RRMSEP vs log(α) with different correction methods.]

- **Without correction**
- **With shrinkage** (r = 1, 2, 3, 4, 5)
- **With OP**

---

Gujral (LA-EPFL)  
SSC11 2009  
08/06/2009  
10 / 15
Step 2: Shrinkage vs orthogonal projection

Example 1: Simulation

- High Signal ($\sigma_s = 3$)
  - Low Noise ($\sigma_n = 0.1$)

- High Signal ($\sigma_s = 3$)
  - High Noise ($\sigma_n = 0.3$)

- Medium Signal ($\sigma_s = 2$)
  - Low Noise ($\sigma_n = 0.1$)

- Medium Signal ($\sigma_s = 2$)
  - High Noise ($\sigma_n = 0.3$)

- Low Signal ($\sigma_s = 0.3$)
  - Low Noise ($\sigma_n = 0.1$)

- Low Signal ($\sigma_s = 0.3$)
  - High Noise ($\sigma_n = 0.3$)
Step 2: Shrinkage vs orthogonal projection

**EXAMPLE 1: SIMULATION**

- **High Signal** ($\sigma_s = 3$)
  - Low Noise ($\sigma_n = 0.1$)

- **Medium Signal** ($\sigma_s = 2$)
  - Low Noise ($\sigma_n = 0.1$)

- **Low Signal** ($\sigma_s = 0.3$)
  - Low Noise ($\sigma_n = 0.1$)

- **High Signal** ($\sigma_s = 3$)
  - High Noise ($\sigma_n = 0.3$)

- **Medium Signal** ($\sigma_s = 2$)
  - High Noise ($\sigma_n = 0.3$)

- **Low Signal** ($\sigma_s = 0.3$)
  - High Noise ($\sigma_n = 0.3$)
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

- High Signal ($\sigma_s = 3$)
  - Low Noise ($\sigma_n = 0.1$)

- Medium Signal ($\sigma_s = 2$)
  - Low Noise ($\sigma_n = 0.1$)

- Low Signal ($\sigma_s = 0.3$)
  - Low Noise ($\sigma_n = 0.1$)

- High Signal ($\sigma_s = 3$)
  - High Noise ($\sigma_n = 0.3$)

- Medium Signal ($\sigma_s = 2$)
  - High Noise ($\sigma_n = 0.3$)

- Low Signal ($\sigma_s = 0.3$)
  - High Noise ($\sigma_n = 0.3$)
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 1: SIMULATION

<table>
<thead>
<tr>
<th>Condition</th>
<th>Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Signal ($\sigma_s = 3$)</td>
<td><img src="#" alt="Graph for High Signal" /></td>
</tr>
<tr>
<td>Low Noise ($\sigma_n = 0.1$)</td>
<td><img src="#" alt="Graph for Low Noise" /></td>
</tr>
<tr>
<td>Low Signal ($\sigma_s = 0.3$)</td>
<td><img src="#" alt="Graph for Low Signal" /></td>
</tr>
<tr>
<td>Low Noise ($\sigma_n = 0.1$)</td>
<td><img src="#" alt="Graph for Low Noise" /></td>
</tr>
<tr>
<td>Medium Signal ($\sigma_s = 2$)</td>
<td><img src="#" alt="Graph for Medium Signal" /></td>
</tr>
<tr>
<td>Low Noise ($\sigma_n = 0.1$)</td>
<td><img src="#" alt="Graph for Low Noise" /></td>
</tr>
<tr>
<td>Medium Signal ($\sigma_s = 2$)</td>
<td><img src="#" alt="Graph for Medium Signal" /></td>
</tr>
<tr>
<td>High Noise ($\sigma_n = 0.3$)</td>
<td><img src="#" alt="Graph for High Noise" /></td>
</tr>
<tr>
<td>Low Signal ($\sigma_s = 0.3$)</td>
<td><img src="#" alt="Graph for Low Signal" /></td>
</tr>
<tr>
<td>Low Noise ($\sigma_n = 0.1$)</td>
<td><img src="#" alt="Graph for Low Noise" /></td>
</tr>
<tr>
<td>Medium Signal ($\sigma_s = 2$)</td>
<td><img src="#" alt="Graph for Medium Signal" /></td>
</tr>
<tr>
<td>High Noise ($\sigma_n = 0.3$)</td>
<td><img src="#" alt="Graph for High Noise" /></td>
</tr>
<tr>
<td>Low Signal ($\sigma_s = 0.3$)</td>
<td><img src="#" alt="Graph for Low Signal" /></td>
</tr>
<tr>
<td>Low Noise ($\sigma_n = 0.1$)</td>
<td><img src="#" alt="Graph for Low Noise" /></td>
</tr>
<tr>
<td>Medium Signal ($\sigma_s = 2$)</td>
<td><img src="#" alt="Graph for Medium Signal" /></td>
</tr>
<tr>
<td>High Noise ($\sigma_n = 0.3$)</td>
<td><img src="#" alt="Graph for High Noise" /></td>
</tr>
</tbody>
</table>
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 2: REAL DATA WITH TEMPERATURE EFFECTS

- NIR, 3 species

<table>
<thead>
<tr>
<th>Calibration</th>
<th>$11 \times 2 = 22$</th>
<th>@ 30, 40 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>$11 \times 3 = 33$</td>
<td>@ 50, 60, 70 °C</td>
</tr>
</tbody>
</table>

Monte Carlo 5 reference samples

**Graph**

- $r = 1$
- $r = 2$
- $r = 3$
- $r = 4$
- $r = 5$

- Without correction
- With shrinkage
- With OP

**Equation**

$$\log(\alpha)$$
Step 2: Shrinkage vs orthogonal projection

EXAMPLE 3: REAL DATA FOR CALIBRATION TRANSFER

NIR, 4 measured properties

- Calibration: 40 samples on Instrument 1
- Prediction: 40 samples on Instrument 2

Monte Carlo: 5 reference samples

Graph showing RRMSEP vs log(α)
- Without correction
- With shrinkage
- With OP

Gujral (LA-EPFL)
Unifying framework

Many drift-correction methods proceed in two steps: (i) drift estimation, (ii) drift correction
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 \( \{\alpha_1, \alpha_2, \ldots\} \)