Chemometric Methods for the Kinetic Hard-modelling of Spectroscopic Data

ETH Zurich, March 23rd 2009

Julien Billeter

julien.billeter@chem.ethz.ch

ETH Zürich, Institute for Chemical and Bioengineering, Safety and Environmental Technology Group, Zürich, Switzerland.
Content

Part 1
Chemometrics and kinetic hard-modelling

Part 2
Uncertainties and error propagation

Part 3
Rank deficiency and spectral validation
Part 1
Chemometrics and kinetic hard-modelling

Part 2
Uncertainties and error propagation

Part 3
Rank deficiency and spectral validation
What is Chemometrics?

- Chemometrics is the chemical discipline that uses mathematical and statistical methods to:
  - (a) design or select optimal measurement procedures and experiments, and
  - (b) provide maximum chemical information by analyzing chemical data

Matthias Otto (2007)
From: Chemometrics – Statistics and Computer Application in Analytical Chemistry
Beer’s law

\[ nt \times nw \times (nt \times ns) \times (ns \times nw) = nt \times nw \times (nt \times ns) + nt \times nw \]

Chemometric Methods for the Kinetic Hard-modelling of Spectroscopic Data

\[ A + B \rightarrow_{k_1} C \]
\[ A + C \rightarrow_{k_2} D \]
Direct fitting

Beer’s law

\[
Y = CA
\]

Modelled part

\[
C_{\text{modelled}} = \text{function (}k\text{)}
\]

\[
A_{\text{modelled}} = \text{function (}\theta\text{)}
\]

Linear counter part

- Implicit calibration
  \[
  A = C^+_{\text{modelled}} Y
  \]

- Explicit calibration
  \[
  A = (C^+Y)_{\text{calibration}}
  
  C = YA^+_{\text{modelled}}
  \]

Least squares fitting

\[
C^+ = (C^T C)^{-1} C^T
\]

\[
A^+ = A^T (A A^T)^{-1}
\]

\[
\min_k \| Y - C_{\text{modelled}} A \|^2
\]

\[
\min_\theta \| Y - C A_{\text{modelled}} \|^2
\]

\[
\min_k \| C - C_{\text{modelled}} \|^2
\]
Indirect (inverse) fitting

Inverse model

\[ C = Y B_C \]

\[ A = B_A Y \]

Modelled part

\[ C_{\text{modelled}} = \text{function (} k \text{)} \]

\[ A_{\text{modelled}} = \text{function (} \theta \text{)} \]

Linear counter part

- Implicit calibration
  \[ B_C = Y^+ C_{\text{modelled}} \]

- Explicit calibration
  \[ B_C = (Y^+ C)_{\text{calibration}} \]

\[ B_A = A_{\text{modelled}} Y^+ \]

\[ B_A = (AY^+)_{\text{calibration}} \]

Least squares fitting

\[ \min_k \| C_{\text{modelled}} - YB_C \|^2 \]

\[ \min_\theta \| A_{\text{modelled}} - A Y \|^2 \]

\[ \min_k \| C - C_{\text{modelled}} \|^2 \]
Modelling concentrations (C) or spectra (A) ?

**Modelling C**

Based on a first-principles model (the kinetic rate law)

Depends on a limited number of kinetic parameters, e.g. rate constants \( k (1 \times nr) \)

**Modelling A**

Based on the modelling of \( A \) using Gaussian functions

Depends on a large number of parameters \( \theta \), which are difficult to determine

Requires a subsequent modelling of \( C \) to determine kinetic parameters, e.g. rate constants \( k (1 \times nr) \)
Fitting method used in Parts 2 and 3

- **Direct fitting by modelling concentrations** \((\mathbf{C})\)

\[
\min_k \| \mathbf{Y} - \mathbf{C}_{\text{modelled}}(\mathbf{k}) \mathbf{A} \|^2
\]

- **using implicit and explicit calibration of the pure component spectra** \((\mathbf{A})\)

  - **Implicit calibration**: (i.e. calibration-free)
    \[
    \mathbf{A} = \mathbf{C}^+_{\text{modelled}}(\mathbf{k}) \mathbf{Y}
    \]
  
  - **Explicit calibration**: (also called Strategy 2)
    \[
    \mathbf{A} = \left( \mathbf{C}^+ \mathbf{Y} \right)_{\text{calibration}}
    \]
Modelling concentration profiles using the rate law

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D \\
\end{align*}
\]

\[k = \begin{bmatrix} k_1 & k_2 \end{bmatrix}\]
Modelling concentration profiles using the rate law

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D \\
\end{align*}
\]

\[
\mathbf{k} = \begin{bmatrix} k_1 & k_2 \end{bmatrix}
\]

\[
\begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0
\end{bmatrix} = \begin{bmatrix}
-1 & -1 & 1 & 0 \\
-1 & 0 & -1 & 1
\end{bmatrix}
\]

Chemometric Methods for the Kinetic Hard-modelling of Spectroscopic Data
Modelling concentration profiles using the rate law

\[\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}\]

\[k = \begin{bmatrix} k_1 & k_2 \end{bmatrix}\]

\[\begin{bmatrix}
A & B & C & D \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0
\end{bmatrix} = \begin{bmatrix}
-1 & -1 & 1 & 0 \\
-1 & 0 & -1 & 1
\end{bmatrix}
\]

\[\text{nr} = 2 \text{ kinetic rate laws}\]

\[\begin{align*}
\frac{dx_{t,1}}{dt} &= k_1 c_{t,A} e_{1,1} c_{t,B} e_{1,2} c_{t,C} e_{1,3} c_{t,D} \\
\frac{dx_{t,2}}{dt} &= k_2 c_{t,A} e_{2,1} c_{t,B} e_{2,2} c_{t,C} e_{2,3} c_{t,D}
\end{align*}\]
Modelling concentration profiles using the rate law

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]

\[
k = \begin{bmatrix} k_1 & k_2 \end{bmatrix}
\]

\[
P = \begin{bmatrix} 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \end{bmatrix} \quad E = \begin{bmatrix} 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \end{bmatrix} \quad N = \begin{bmatrix} -1 & -1 & 1 & 0 \\
-1 & 0 & -1 & 1 \end{bmatrix}
\]

\( ns = 4 \) concentration profiles (System of ODE)

\[
\begin{align*}
\frac{dc_{A,t}}{dt} &= n_{1,1} \frac{dx_{t,1}}{dt} + n_{2,1} \frac{dx_{t,2}}{dt} \\
\frac{dc_{B,t}}{dt} &= n_{1,2} \frac{dx_{t,1}}{dt} + n_{2,2} \frac{dx_{t,2}}{dt} \\
\frac{dc_{C,t}}{dt} &= n_{1,3} \frac{dx_{t,1}}{dt} + n_{2,3} \frac{dx_{t,2}}{dt} \\
\frac{dc_{D,t}}{dt} &= n_{1,4} \frac{dx_{t,1}}{dt} + n_{2,4} \frac{dx_{t,2}}{dt}
\end{align*}
\]

\( nr = 2 \) kinetic rate laws

\[
\begin{align*}
\frac{dx_{t,1}}{dt} &= k_1 c_{t,A} c_{t,B} + k_1 c_{t,A} c_{t,B} \\
\frac{dx_{t,2}}{dt} &= k_2 c_{t,A} c_{t,B} + k_2 c_{t,A} c_{t,B}
\end{align*}
\]
Modelling concentration profiles using the rate law

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]

\[k = \begin{bmatrix} k_1 & k_2 \end{bmatrix}\]

\[\begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0
\end{bmatrix} \times \begin{bmatrix}
1 & -1 & 1 & 0 \\
-1 & 0 & -1 & 1
\end{bmatrix} = \begin{bmatrix}
A & B & C & D
\end{bmatrix}
\]

\(ns = 4\) concentration profiles (System of ODE)

\[
\begin{align*}
\frac{dc_{t,A}}{dt} &= n_{1,1} \frac{dx_{t,1}}{dt} + n_{2,1} \frac{dx_{t,2}}{dt} \\
\frac{dc_{t,B}}{dt} &= n_{1,2} \frac{dx_{t,1}}{dt} + n_{2,2} \frac{dx_{t,2}}{dt} \\
\frac{dc_{t,C}}{dt} &= n_{1,3} \frac{dx_{t,1}}{dt} + n_{2,3} \frac{dx_{t,2}}{dt} \\
\frac{dc_{t,D}}{dt} &= n_{1,4} \frac{dx_{t,1}}{dt} + n_{2,4} \frac{dx_{t,2}}{dt}
\end{align*}
\]

\[
\begin{align*}
\frac{dx_{t,1}}{dt} &= -1 \cdot \frac{dx_{t,1}}{dt} - 1 \cdot \frac{dx_{t,2}}{dt} \\
\frac{dx_{t,2}}{dt} &= -1 \cdot \frac{dx_{t,1}}{dt} + 0 \cdot \frac{dx_{t,2}}{dt} \\
\frac{dx_{t,1}}{dt} &= 1 \cdot \frac{dx_{t,1}}{dt} - 1 \cdot \frac{dx_{t,2}}{dt} \\
\frac{dx_{t,2}}{dt} &= 0 \cdot \frac{dx_{t,1}}{dt} + 1 \cdot \frac{dx_{t,2}}{dt}
\end{align*}
\]

nr = 2 kinetic rate laws

\[
\begin{align*}
\frac{dx_{t,1}}{dt} &= k_1 c_{t,A} c_{t,B} c_{t,C} c_{t,D} = k_1 c_{t,A} c_{t,B} \\
\frac{dx_{t,2}}{dt} &= k_2 c_{t,A} c_{t,B} c_{t,C} c_{t,D} = k_2 c_{t,A} c_{t,C}
\end{align*}
\]

Numerical integration

\[
C = \begin{bmatrix} c_{t,A} & c_{t,B} & c_{t,C} & c_{t,D} \end{bmatrix}
\]

The System of ODE is integrated with initial concentrations \(c_0 = \begin{bmatrix} c_{0,A} & c_{0,B} & c_{0,C} & c_{0,D} \end{bmatrix}\)
Part 1
Chemometrics and kinetic hard-modelling

Part 2
Uncertainties and error propagation

Part 3
Rank deficiency and spectral validation
Newton-Gauss algorithm

$$\min_k \| Y - C_{\text{modelled}} A \|^2$$

Initial guess for $k$

$c_0, f_s$

Numerical integration of the system of ODEs describing the kinetic model

Calculation of $R(k, c_0, f_s)$ and sum of squares ($ssq$)

$sq \approx \text{const}$

yes $\rightarrow k_{\text{opt}}(c_0, f_s)$

no

Calculation of the Jacobian $J$

Calculation of the shift $\Delta k$

and $k = k + \Delta k$

Calculation of $\sigma_k^2$
**Newton-Gauss algorithm**

\[
\min_k \| Y - C_{\text{modelled}} A \|^2
\]

**Newton-Gauss algorithm delivers an estimate of the uncertainties in the optimised parameters (k)**

\[
\sigma_k^2 = \text{diag} \left( H^{-1} \right) \sigma_r^2
\]

- \( \sigma_k^2 \): Variance of the optimised parameters \( k \) (1 x \( nr \))
- \( \text{diag}(\cdot) \): Operator extracting a vector of diagonal elements from a matrix
- \( H = \left( \frac{\partial R}{\partial k} \right)^T \left( \frac{\partial R}{\partial k} \right) \): Sensitivity of the residuals \( R \) with respect to the optimised parameters \( k \)
- \( \sigma_r^2 \): Variance of the residuals
Error propagation in Newton-Gauss algorithm

Classical estimation of uncertainties: \( \sigma_k^2 = \text{diag}(H^{-1})\sigma_r^2 \)

**Problem:** this estimation (\( \sigma_k^2 \)) is lower than the variance calculated by repetition of the experiments.
Error propagation in Newton-Gauss algorithm

Classical estimation of uncertainties: \[ \sigma_k^2 = \text{diag}(H^{-1})\sigma_r^2 \]

Problem: this estimation (\( \sigma_k^2 \)) is lower than the variance calculated by repetition of the experiments.

Uncertainties and error propagation (easily extendable):

\[ \sigma_k^2 = \]

\[ \sigma_k^2 = \]
Error propagation in Newton-Gauss algorithm

Classical estimation of uncertainties: \( \sigma_k^2 = \text{diag}(H^{-1}) \sigma_r^2 \)

Problem: this estimation (\( \sigma_k^2 \)) is lower than the variance calculated by repetition of the experiments.

Uncertainties and error propagation (easily extendable):

\[
\sigma_k^2 = \text{diag}(H^{-1}) \sigma_r^2
\]

\[
\sigma_{k,r}^2 = \sigma_{k,r}^2
\]
Error propagation in Newton-Gauss algorithm

Classical estimation of uncertainties:

\[ \sigma_k^2 = \text{diag} \left(H^{-1}\right) \sigma_r^2 \]

Problem: this estimation (\( \sigma_k^2 \)) is lower than the variance calculated by repetition of the experiments.

Uncertainties and error propagation (easily extendable):

\[ \sigma_k^2 = \text{diag} \left(H^{-1}\right) \sigma_r^2 + \text{diag} \left( \frac{\partial k}{\partial c_0} \right)^T \text{DIAG} \left( \sigma_{c_0}^2 \right) \left( \frac{\partial k}{\partial c_0} \right) + \sigma_{k,r}^2 + \sigma_{k,c_0}^2 \]

DIAG = operator generating a diagonal matrix from the corresponding vector argument
Error propagation in Newton-Gauss algorithm

Classical estimation of uncertainties: 
\[ \sigma_k^2 = \text{diag}(H^{-1}) \sigma_r^2 \]

**Problem:** this estimation \( \sigma_k^2 \) is lower than the variance calculated by repetition of the experiments.

**Uncertainties and error propagation (easily extendable):**

\[ \sigma_k^2 = \text{diag}(H^{-1}) \sigma_r^2 + \text{diag}\left( \left( \frac{\partial k}{\partial c_0} \right)^T \right) \text{DIAG}\left( \sigma_{c_0}^2 \right) \left( \frac{\partial k}{\partial c_0} \right) + \text{diag}\left( \left( \frac{\partial k}{\partial f_s} \right)^T \right) \text{DIAG}\left( \sigma_{f_s}^2 \right) \left( \frac{\partial k}{\partial f_s} \right) \]

\[ \sigma_k^2 = \sigma_{k,r}^2 + \sigma_{k,c_0}^2 + \sigma_{k,f}^2 \]

DIAG = operator generating a diagonal matrix from the corresponding vector argument
Simulation: $A + B \xrightarrow{k_1} C$

\[ \sigma_k^2 = \sigma_{k, r}^2 + \sigma_{k, c_0}^2 \quad \text{with} \quad \sigma_{k, c_0}^2 = \text{diag} \left[ \left( \frac{\partial k}{\partial c_0} \right)^T \text{DIAG} \left( \sigma_{c_0}^2 \right) \left( \frac{\partial k}{\partial c_0} \right) \right] \]

Maxima / Minima in $\sigma_{k, c_0}^2$

\[ \sigma_{c_{0,A}}^2 = (0.2\%)^2 c_{0,A}^2 \]
\[ \sigma_{c_{0,B}}^2 = (0.1\%)^2 c_{0,B}^2 \]

Different uncertainties in the initial concentrations due to sampling

Excess of $A$

Excess of $B$

Stoichiometric ratio

$\alpha = c_{0,A} / c_{0,B}$ with $c_{0,A} + c_{0,B} = 1 \text{ mol} \cdot \text{L}^{-1}$
Simulation: optimal experimental conditions

\[ A + B \xrightarrow{k_1} C \]

\[ \sigma_k^2 = \sigma_{k,r}^2 + \sigma_{k,c_0}^2 \approx \sigma_{k,c_0}^2 = \text{diag}\left( \frac{\partial k}{\partial c_0} \right)^T \text{DIAG}\left( \sigma_{c_0}^2 \right) \left( \frac{\partial k}{\partial c_0} \right) \]

Under exact stoichiometric conditions \((\alpha = 1)\)

Under pseudo-first order conditions \((\alpha \in [\infty, 10] \cup [0.1, 0])\)

Excess in the species with the lowest uncertainty in its initial concentration (here \(B\))

This curve was validated by Monte-Carlo sampling (10 000 points)
Experimental reaction

Overall reaction:

\[ B + P + Aa \overset{k}{\rightarrow} BP + Aa \]

Kinetic mechanism:

Experimental conditions:

25°C, dosing Aa, followed in mid-IR (1200–1650 cm\(^{-1}\)) and UV-vis (240–400 nm).

- \( c_{0,B} = 0.40033 (\pm 0.292\%) \) mol·L\(^{-1}\), \( c_{0,P} = 1.19737 (\pm 0.292\%) \) mol·L\(^{-1}\), \( c_{0,Aa} = 0 \) mol·L\(^{-1}\),
- \( c_{dos,Aa} = 17.48376 (\pm 0\%) \) mol·L\(^{-1}\), dosing rate = 8.17 (\pm 0.14\%) mL·min\(^{-1}\) in 0.6 min.

17 repetitions
## Experimental results

<table>
<thead>
<tr>
<th></th>
<th><strong>UV-vis</strong></th>
<th></th>
<th><strong>mid-IR</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k^a$</td>
<td>$\sigma_k^a$</td>
<td>$k^a$</td>
<td>$\sigma_k^a$</td>
</tr>
<tr>
<td>Experimental</td>
<td>1.76(8)</td>
<td><strong>0.02(8)</strong></td>
<td>1.73(9)</td>
<td><strong>0.05(4)</strong></td>
</tr>
<tr>
<td>(mean and standard deviation over 17 repetitions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted by error propagation</td>
<td>-</td>
<td><strong>0.02(3)</strong></td>
<td>-</td>
<td><strong>0.02(2)</strong></td>
</tr>
<tr>
<td>Literature</td>
<td>1.40 $^b$</td>
<td>-</td>
<td>1.51 $^b$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.65 $^c$</td>
<td>-</td>
<td>1.65 $^c$</td>
<td>-</td>
</tr>
</tbody>
</table>

$a)$ in L$^2$·mol$^{-2}$·s$^{-1}$·10$^{-4}$

$b)$ Carvalho et al., Talanta, 68 (2006), 1190-1200

$c)$ Billeter et al., Chemom. Intell. Lab. Syst., (2009), submitted

\[ \sigma_k^2 = \sigma_{k,r}^2 + \sigma_{k,c_0}^2 + \sigma_{k,f}^2 \]

\[ \text{HPLC pump} \]

92%  

2%  

6%
Part 1
Chemometrics and kinetic hard-modelling

Part 2
Uncertainties and error propagation

Part 3
Rank deficiency and spectral validation
Singular Value Decomposition (SVD)

\[ Y = U S V \]

- **U**: Matrix of orthonormal column eigenvectors
- **S**: Matrix of singular values
- **V**: Matrix of orthonormal row eigenvectors

For \( nw < nt \):

\[
\begin{align*}
Y &= (nt \times nw) \\
U &= (nt \times nw) \\
S &= (nw \times nw) \\
V &= (nw \times nw)
\end{align*}
\]

For \( nt < nw \):

\[
\begin{align*}
Y &= (nt \times nw) \\
U &= (nt \times nt) \\
S &= (nt \times nt) \\
V &= (nt \times nw)
\end{align*}
\]
Principal Component Analysis (PCA)

\[ \bar{Y} = \bar{U} \bar{S} \bar{V} \quad \text{or} \quad \bar{Y} = \bar{U} \bar{S} \bar{V} \]

Reduction of the dimensionality to \( nc \), i.e. the number of significant singular values (or eigenvectors)

\[ R = Y - \bar{Y} = \text{noise} \]

- \( Y (nt \times nw) \)
- \( U (nt \times nw) \)
- \( S (nw \times nw) \)
- \( V (nw \times nw) \)

- \( nt < nw \)
- \( nw < nt \)

\( \bar{U} (nt \times nc) \)
\( \bar{S} (nc \times nc) \)
\( \bar{V} (nc \times nw) \)

- \( nt \) or \( nw \) factors
- \( nc \) factors
Target Factor Analysis (TFA)

$\mathbf{Y} = \mathbf{C} \mathbf{A}$

Where $\mathbf{T}$ is a transformation matrix of dimensions $(nc \times nc)$

$n_c =$ number of significant factors

$ns =$ number of reactive species

TFA: Relationship between PCA and Beer’s law

$\mathbf{C} (nt \times nc) = \mathbf{U} (nt \times nc) \mathbf{T} \neq \mathbf{C} (nt \times ns)$

$\mathbf{A} (nc \times nw) = \mathbf{T}^{-1} \mathbf{S} \mathbf{V} (nc \times nw) \neq \mathbf{A} (ns \times nw)$
Rank deficiency in spectroscopy

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]
Chemometric Methods for the Kinetic Hard-modelling of Spectroscopic Data

Rank deficiency in spectroscopy

Beer’s law ($ns = 4$ species) \( Y = CA \)

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]
Rank deficiency in spectroscopy

Beer’s law \((ns = 4 \text{ species})\) \(Y = CA\)

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]

PCA \((nc = 3 \text{ factors})\) \(\bar{Y} = \bar{U} \bar{S} \bar{V}\)

\[
Y = X + R
\]

\[
Y = \bar{U} \bar{S} \bar{V} + R
\]
Rank deficiency in spectroscopy

Beer’s law ($ns = 4$ species) \[ Y = CA \]

PCA ($nc = 3$ factors) \[ \bar{Y} = \bar{U} \bar{S} \bar{V} \]

Spectroscopic data $Y$ are rank deficient when:

significant factors ($nc$) in PCA < number of reactive species ($ns$) $\Leftrightarrow$ rank($Y$) < $ns$
Sources and problems of rank deficiency

\[ Y = CA \]

Rank deficiencies in \( Y \) is due to:

- Linear dependencies in \( C \) and/or linear dependencies in \( A \)
- Mathematical dilemma in case of implicit calibration
  - \( A \) cannot be computed by \( C^*Y \) as \( A \) is not unique

Example: two species that are consumed or generated at the same rate

\[
\text{rank}(Y) = \min \{ \text{rank}(C), \text{rank}(A) \} = \text{rank}(C)
\]
Strategies to treat rank deficiency in kinetic hard-modelling by implicit calibration
Strategies to treat rank deficiency in kinetic hard-modelling by **implicit calibration**

- **Model (Beer’s law) reduction**
  - **Strategy 1:** define $nu$ uncoloured species
  - **Strategy 2:** include $nks$ known spectra in the analysis (explicit calibration)
Strategies to treat rank deficiency in kinetic hard-modelling by implicit calibration

- **Model (Beer’s law) reduction**
  - **Strategy 1:** define $nu$ uncoloured species
  - **Strategy 2:** include $nks$ known spectra in the analysis (explicit calibration)

- **Rank augmentation**
  - **Strategy 3:** dose one or more species in $nf$ dosing steps
  - **Strategy 4:** perform $ne$ additional experiments by varying the initial concentrations (second order global analysis in global mode)
Strategies to treat rank deficiency in kinetic hard-modelling by implicit calibration

- **Model (Beer’s law) reduction**
  - **Strategy 1:** define \( nu \) uncoloured species
  - **Strategy 2:** include \( nks \) known spectra in the analysis (explicit calibration)

- **Rank augmentation**
  - **Strategy 3:** dose one or more species in \( nf \) dosing steps
  - **Strategy 4:** perform \( ne \) additional experiments by varying the initial concentrations (second order global analysis in global mode)

How to define the species to include in these four Strategies?
Rank and kernel of the concentration matrix (C)

- \( nc = \text{rank} \,(Y) = \text{rank} \,(C) \)
  
  number of linearly independent columns or rows of \( Y \) or \( C \)

The rank of \( Y \) or \( C \) defines the maximum number of columns (species) to keep in Strategy 1 \((ns \,– \,nu)\) and Strategy 2 \((ns \,– \,nks)\)

- \( \text{ker} \, C \)
  
  vector space spanned by the vectors forming the null space \( 0 \) when multiplied by \( C \)

  e.g. \( \text{ker} \, C = \begin{bmatrix} 0.8 & 0.1 \\ -0.3 & -0.7 \\ 0 & 0 \\ 0.5 & -0.6 \end{bmatrix} \)

  (1) \( \text{ker} \, C \) defines a mass balance equation: \( C \,(\text{ker} \, C) = 0 \)

  (2) \( \text{ker} \, C \) defines which columns of \( C \) are:

  (a) linearly dependent (rows comprised by non-zeros entries in \( \text{ker} \, C \)) or

  (b) linearly independent (rows comprised by zeros in \( \text{ker} \, C \))
### Interpretation of the kernel of C

<table>
<thead>
<tr>
<th>Species without zero rows in ker C</th>
<th>Species with zero rows in ker C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>kernel of C</strong></td>
<td><strong>linearly independent from the other species</strong></td>
</tr>
<tr>
<td>Strategy 1:</td>
<td><strong>define this species as uncoloured</strong></td>
</tr>
<tr>
<td>determining its pure spectrum</td>
<td><strong>define this species as coloured</strong></td>
</tr>
<tr>
<td>Strategy 2:</td>
<td><strong>providing its pure spectrum avoids rank deficiency</strong></td>
</tr>
<tr>
<td>Strategy 3:</td>
<td><strong>providing its pure spectrum does not avoid rank deficiency</strong></td>
</tr>
<tr>
<td>Strategy 4:</td>
<td><strong>dosing this species breaks rank deficiency</strong></td>
</tr>
<tr>
<td>varying its initial concentration</td>
<td><strong>dosing this species does not break rank deficiency</strong></td>
</tr>
<tr>
<td>breaks rank deficiency</td>
<td>varying its initial concentration does not break rank deficiency</td>
</tr>
</tbody>
</table>
A time invariant matrix $\Omega$ equivalent to $C$

$$\Omega = \begin{bmatrix} (\mu 1)^{E^T} & \text{DIAG}(k)N \\ C_0 & C_{in} \\ C_0^{ne} \end{bmatrix}$$

$(ns + 1 + nf + ne \times ns - nu - nks)$

$\mu$ an arbitrary positive scalar different from 0 and 1

$1$ $(ns \times nr)$ matrix comprised of ones

$E$ $(nr \times ns)$ matrix of reactant coefficients

$\cdot E^T$ element-wise raise to the power of $E^T$

$\text{DIAG}$ operator generating a diagonal matrix from a vector argument

$k$ $(1 \times nr)$ vector of rate constants

$N$ $(nr \times ns)$ matrix of stoichiometric coefficients

$c_0$ $(1 \times ns)$ vector of initial concentrations

$C_{in}$ $(nf \times ns)$ matrix of the dosing concentrations corresponding to the $nf$ dosing steps

$C_0^{ne}$ $(ne \times ns)$ matrix of the varied initial concentrations corresponding to the $ne$ additional experiments

**Validation of this time invariant matrix:**
- Equivalence can be mathematically proven
- Extensively tested on various mechanisms

**Advantages of the time invariant approach:**
- No numerical integration required
- Analytical (symbolic) relationship between the experimental conditions ($c_0, C_{in}, ...$)
Strategies to treat rank deficiency applied to $\Omega$

**Model reduction**

- **Strategy 1**
  - $nu$ uncoloured species
  - $ns - nu$
  - $ns (\mu^1)^{E\text{T}} \text{DIAG}(k)N$
  - $c_0$

- **Strategy 2**
  - $nks$ known pure spectra
  - $ns - nks$
  - $ns (\mu^1)^{E\text{T}} \text{DIAG}(k)N$
  - $c_0$

**Rank augmentation**

- **Strategy 3**
  - $nf$ dosing steps
  - $ns$
  - $(\mu^1)^{E\text{T}} \text{DIAG}(k)N$
  - $c_0$
  - $C_{in}$
  - $nf$

- **Strategy 4**
  - $ne$ varied initial concentrations
  - $ns$
  - $(\mu^1)^{E\text{T}} \text{DIAG}(k)N$
  - $c_0$
  - $C_{0}^{ne}$
  - $ne$
Spectral consequence of Strategy 1 (defining uncoloured species)

Spectral contribution of the $nu$ uncoloured species is linearly transferred into the fitted pure component spectra of the coloured species.

The fitted component spectra $A_c$ of the $(ns-nu)$ coloured species are comprised of linear combinations of the $ns$ true pure component spectra $A$. 
Spectral consequence of Strategy 1 (defining uncoloured species)

Spectral contribution of the $nu$ uncoloured species is linearly transferred into the fitted pure component spectra of the coloured species.

The fitted component spectra $A_c$ of the ($ns$–$nu$) coloured species are comprised of linear combinations of the $ns$ true pure component spectra $A$.

$$\Delta (ns - nu \times ns) = C_c^+ C = \left(\Omega\right|_{\text{comprised of coloured species}} + \Omega\right|_{\text{comprised of all species}}^+$$
Example: calculation of the kernel

\[ \begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*} \]

\[ \Omega = \left[ \begin{array}{c}
\left( \mu 1 \right)^{\text{E}^T} \text{DIAG}(k) N \\
c_0
\end{array} \right] \]

\[ \ker \Omega = \begin{bmatrix}
-\alpha \\
1 \\
1-\alpha \\
1-2\alpha
\end{bmatrix} \]

Matlab code (8 lines)

```matlab
>> syms c0A alpha mu k1 k2
>> N = [-1 -1 1 0; -1 0 -1 1];
>> E = [1 1 0 0; 1 0 1 0];
>> k = [k1, k2];
>> c0 = [c0A, alpha*c0A, 0, 0];
>> one = ones(size(E'));
>> omega = [(mu*one).^(E')*diag(k)*N; c0];
>> null(omega)
ans =
-\alpha \\
1 \\
1-\alpha \\
1-2\alpha
```

\[ \alpha = \frac{C_{0,B}}{C_{0,A}} \]
**Example: calculation of the kernel**

\[
A + B \xrightarrow{k_1} C \\
A + C \xrightarrow{k_2} D
\]

\[
\Omega = \left[ \left( \mu \mathbf{1} \right)^{\mathbf{E}^\top} \text{DIAG}(\mathbf{k}) \mathbf{N} \right]_{\mathbf{c}_0} = \begin{bmatrix}
-\mu k_1 - \mu k_2 \\
-\mu k_1 - k_2 \\
-k_1 - \mu k_2 \\
-k_1 - k_2 \\
\mu k_1 - k_2 \\
\mu k_2 \\
1 - \alpha \\
1 - 2\alpha
\end{bmatrix}
\]

**Matlab code (8 lines)**

```matlab
>> syms c0A alpha mu k1 k2
>> N = [-1 -1 1 0; -1 0 -1 1];
>> E = [1 1 0 0; 1 0 1 0];
>> k = [k1, k2];
>> c0 = [c0A, alpha*c0A, 0, 0];
>> one = ones(size(E'));
>> omega = [(mu*one).^(E')*diag(k)*N; c0];
>> null(omega)
ans =
   -alpha
    1
  1-alpha
3-2*alpha
```

If \( \alpha = 1 \), species \( C \) is linearly independent from the others.
Example: calculation of the kernel

\[
\begin{align*}
A + B & \xrightarrow{k_1} C \\
A + C & \xrightarrow{k_2} D
\end{align*}
\]

\[
\Omega = \left[ (\mu \mathbf{1})^{\mathbf{E}^\top} \text{DIAG} \left( \mathbf{k} \right) \mathbf{N} \right] \frac{c_0}{c_0}
\]

If \( \alpha = 1 \), species \( C \) is linearly independent from the others.

If \( \alpha = 0.5 \), species \( D \) is linearly independent from the others.

Matlab code (8 lines)

\[
\begin{align*}
&>> \text{syms} \; c0A \; \alpha \; \mu \; k1 \; k2 \\
&>> \text{N} \quad = \quad [ -1 \; -1 \; 1 \; 0; \; -1 \; 0 \; -1 \; 1 ]; \\
&>> \text{E} \quad = \quad [ 1 \; 1 \; 0 \; 0; \; 1 \; 0 \; 1 \; 0 ]; \\
&>> \text{k} \quad = \quad [ k1, \; k2 ]; \\
&>> \text{c0} \quad = \quad [ c0A, \; \alpha \cdot c0A, \; 0, \; 0 ]; \\
&>> \text{one} \quad = \quad \text{ones} (\text{size(E')}); \\
&>> \text{omega} \quad = \quad (\mu \cdot \text{one}) \cdot \text{E}' \cdot \text{diag} \left( \mathbf{k} \right) \cdot \mathbf{N}; \; c0]; \\
&>> \text{null} (\text{omega}) \\
&\text{ans} = \\
&\quad \begin{bmatrix}
-\alpha \\
1 \\
1-\alpha \\
1-2*\alpha
\end{bmatrix}
\]

\[
\alpha = \frac{c_{0,B}}{c_{0,A}}
\]
Example: calculation of the kernel

\[ \begin{align*}
A + B \xrightarrow{k_1} C \\
A + C \xrightarrow{k_2} D
\end{align*} \]

\[ \Omega = \begin{bmatrix}
(\mu \mathbf{1})^\mathbf{T} \cdot \text{DIAG}(\mathbf{k}) \mathbf{N} \\
\mathbf{c}_0
\end{bmatrix} =
\begin{bmatrix}
-\mu k_1 - \mu k_2 & -\mu k_1 & \mu k_1 - \mu k_2 & \mu k_2 \\
-\mu k_1 - k_2 & -\mu k_1 & k_1 - \mu k_2 & k_2 \\
-\mu k_1 - k_2 & -k_1 & k_1 - k_2 & k_2 \\
-c_0, A & \alpha c_0, A & 0 & 0
\end{bmatrix}
\]

\[
\text{ker } \Omega = \begin{bmatrix}
\begin{array}{c}
-\alpha \\
1 \\
1 - \alpha \\
1 - 2\alpha
\end{array}
\end{bmatrix}
\]

- If \( \alpha = 1 \), species \( C \) is linearly independent from the others.
- If \( \alpha = 0.5 \), species \( D \) is linearly independent from the others.
- If \( \alpha \neq 1 \) or 0.5, all species are linearly dependent from the others.

Matlab code (8 lines)

\[
\begin{align*}
\text{>> syms c0A alpha mu k1 k2} \\
\text{>> N} &= [-1 -1 1 0; -1 0 -1 1]; \\
\text{>> E} &= [1 1 0 0; 1 0 1 0]; \\
\text{>> k} &= [k1, k2]; \\
\text{>> c0} &= [c0A, alpha*c0A, 0, 0]; \\
\text{>> one} &= \text{ones(size(E'))}; \\
\text{>> omega} &= ((\mu*one).^\mathbf{T}(\mathbf{E})^\mathbf{T}\text{diag}(\mathbf{k})\mathbf{N}\mathbf{c}_0); \\
\text{>> null(omega)} \\
\text{ans} &=
\begin{bmatrix}
-\text{alpha} \\
1 \\
1-\text{alpha} \\
1-2\text{alpha}
\end{bmatrix}
\end{align*}
\]
**Example: design of experiments**

\[
A + B \xrightarrow{k_1} C \\
A + C \xrightarrow{k_2} D
\]

\[
\ker \Omega \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{bmatrix} -\alpha \\ 1 \\ 1 - \alpha \\ 1 - 2\alpha \end{bmatrix}
\]

\[
\begin{array}{cccc}
\alpha = 1 & & & \\
A & B & C & D \\
0/1 & 0/1 & 0 & 0/1 \\
\end{array}
\quad \begin{array}{cccc}
\alpha = 0.5 & & & \\
A & B & C & D \\
0/1 & 0/1 & 0/1 & 0 \\
\end{array}
\quad \begin{array}{cccc}
\alpha \neq 1 \text{ or } 0.5 & & & \\
A & B & C & D \\
0/1 & 0/1 & 0/1 & 0/1 \\
\end{array}
\]

**Strategy**

1: uncoloured, pure spectrum provided, dosed or initial concentration varied

0: coloured, pure spectrum not provided, not dosed or initial concentration not varied

- **(1): define one uncoloured species**
  - \(a = 1\): [0/1 0/1 0 0/1]
  - \(a = 0.5\): [0/1 0/1 0/1 0]
  - \(a \neq 1 \text{ or } 0.5\): [0/1 0/1 0/1 0/1]

- **(2): provide one known pure spectrum**
  - \(a = 1\): [0/1 0/1 0 0/1]
  - \(a = 0.5\): [0/1 0/1 0/1 0]
  - \(a \neq 1 \text{ or } 0.5\): [0/1 0/1 0/1 0/1]

- **(3): dose one species**
  - \(a = 1\): [0/1 0/1 0 0/1]
  - \(a = 0.5\): [0/1 0/1 0/1 0]
  - \(a \neq 1 \text{ or } 0.5\): [0/1 0/1 0/1 0/1]

- **(4): vary one initial concentration**
  - \(a = 1\): [0/1 0/1 0 0/1]
  - \(a = 0.5\): [0/1 0/1 0/1 0]
  - \(a \neq 1 \text{ or } 0.5\): [0/1 0/1 0/1 0/1]
Example: spectral consequence of Strategy 1 (defining uncoloured species)

Let’s define species A uncoloured

\[ A_c = \Delta A \]  
(Spectral balance)

\[ \Delta (3 \times 4) = \left( \Omega \right|_{\text{coloured species}}^\text{comprised of all species} \left( \Omega \right|_{\text{coloured species}}^\text{comprised of all species} \right)^+ + \left( \Omega \right|_{\text{coloured species}}^\text{comprised of all species} \left( \Omega \right|_{\text{coloured species}}^\text{comprised of all species} \right)^+ \]
Example: spectral consequence of Strategy 1 (defining uncoloured species)

True: $A$ (4 x $nw$)

Fitted: $A_c$ (3 x $nw$)
Example: spectral consequence of Strategy 1 (defining uncoloured species)

True: \( A \ (4 \times nw) \)

Fitted: \( A_c \ (3 \times nw) \)

\[
\begin{bmatrix}
\alpha^{-1} & 1 & 0 & 0 \\
\alpha^{-1} & 1 & 0 & 0 \\
\alpha^{-1} & 2 & 0 & 1 \\
\end{bmatrix}
= \Delta A
= \begin{bmatrix}
0.5 \\
-0.5 \\
-1.5 \\
\end{bmatrix}
\begin{bmatrix}
a_{A_i} \\
a_{B_i} \\
a_{C_i} \\
a_{D_i} \\
\end{bmatrix}
+ \begin{bmatrix}
1 \\
0 \\
0 \\
0 \\
\end{bmatrix}
\begin{bmatrix}
a_{A_i} \\
a_{B_i} \\
a_{C_i} \\
a_{D_i} \\
\end{bmatrix}
+ \begin{bmatrix}
0 \\
1 \\
0 \\
1 \\
\end{bmatrix}
\begin{bmatrix}
a_{A_i} \\
a_{B_i} \\
a_{C_i} \\
a_{D_i} \\
\end{bmatrix}
\text{ when } \alpha = 2
Kinetic and spectral validation

- In kinetic hard-modelling, the validation of a kinetic mechanism is based on:
  - (C) The kinetic consistency and the reproducibility of fitted kinetic parameters under different experimental conditions.
  - (A) The spectral consistency of fitted pure component spectra compared to independently measured ones.

\[ Y = C A \]
Kinetic and spectral validation

- In kinetic hard-modelling, the validation of a kinetic mechanism is based on:
  - (C) The kinetic consistency and the reproducibility of fitted kinetic parameters under different experimental conditions.
  - (A) The spectral consistency of fitted pure component spectra compared to independently measured ones.

The spectral validation is facilitated, when Strategy 1 is used, as linear combinations of true pure component spectra can now be explained (Δ)!

\[
\Delta \left( ns - nu \times ns \right) = C_c^+ C = \left( \Omega \text{ comprised of coloured species} \right)^+ \Omega \text{ comprised of all species}
\]
Experimental reaction

Overall reaction:

\[ B + P + Aa \xrightarrow{k} BP + Aa \]

Kinetic mechanism: \( B + P + Aa \xrightarrow{k} BP + Aa \)

Experimental conditions:
25°C, followed in mid-IR (1200–1650 cm\(^{-1}\)) and UV-vis (240–400 nm).

Batch conditions

- **Dosing Aa**
- **Dosing B**
- **Dosing B + Aa**
- **Dosing P**

Reactors: CRC.v4 with FT-IR and UV-vis
Kinetic validation of the model

<table>
<thead>
<tr>
<th>Experimental conditions</th>
<th>Strategy a)</th>
<th>Mid-IR</th>
<th>UV-vis</th>
<th>UV-vis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>k b)</td>
<td>Published k b)</td>
<td>k b)</td>
</tr>
<tr>
<td>Batch conditions</td>
<td>only (1)</td>
<td>1.58 (± 0.02)</td>
<td>1.74 (± 0.05) c)</td>
<td>1.63 (± 0.02)</td>
</tr>
<tr>
<td></td>
<td>(1) + (2)</td>
<td>1.58 (± 0.02)</td>
<td>1.40 d)</td>
<td>1.63 (± 0.02)</td>
</tr>
<tr>
<td></td>
<td>(1) + (4)</td>
<td>1.57 (± 0.02)</td>
<td></td>
<td>1.62 (± 0.02)</td>
</tr>
<tr>
<td>Dosing Aa</td>
<td>(1) + (3)</td>
<td>1.60 (± 0.04)</td>
<td></td>
<td>1.65 (± 0.04)</td>
</tr>
<tr>
<td>Dosing B</td>
<td>(1) + (3)</td>
<td>1.60 (± 0.03)</td>
<td></td>
<td>1.64 (± 0.03)</td>
</tr>
<tr>
<td></td>
<td>(1) + (3)</td>
<td>1.55 (± 0.02)</td>
<td></td>
<td>1.57 (± 0.02)</td>
</tr>
<tr>
<td>Dosing B + Aa</td>
<td>only (3)</td>
<td>1.59 (± 0.06)</td>
<td></td>
<td>1.64 (± 0.06)</td>
</tr>
<tr>
<td>Dosing P</td>
<td>(1) + (3)</td>
<td>1.62 (± 0.06)</td>
<td></td>
<td>1.67 (± 0.06)</td>
</tr>
</tbody>
</table>

b) in L² mol⁻² s⁻¹ x 10⁻⁴  
c) Billeter et al., Chemom. Intell. Lab. Syst., 93 (2008), 120-131  
c) Carvalho et al., Talanta, 68 (2006), 1190-1200

\[ B + P + Aa \overset{k_1}{\rightarrow} BP + Aa \]
## Kinetic validation of the model

<table>
<thead>
<tr>
<th>Experimental conditions</th>
<th>Strategy a)</th>
<th>Mid-IR</th>
<th>UV-vis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k^b)$</td>
<td>Published $k^b)$</td>
</tr>
<tr>
<td><strong>Batch conditions</strong></td>
<td>only (1)</td>
<td>$1.58$ (± 0.02)</td>
<td>$1.74$ (± 0.05) c)</td>
</tr>
<tr>
<td></td>
<td>(1) + (2)</td>
<td>$1.58$ (± 0.02)</td>
<td>$1.40$ d)</td>
</tr>
<tr>
<td></td>
<td>(1) + (4)</td>
<td>$1.57$ (± 0.02)</td>
<td></td>
</tr>
<tr>
<td><strong>Dosing Aa</strong></td>
<td>(1) + (3)</td>
<td>$1.60$ (± 0.04)</td>
<td></td>
</tr>
<tr>
<td><strong>Dosing B</strong></td>
<td>(1) + (3)</td>
<td>$1.60$ (± 0.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) + (3)</td>
<td>$1.55$ (± 0.02)</td>
<td></td>
</tr>
<tr>
<td><strong>Dosing B + Aa</strong></td>
<td>only (3)</td>
<td>$1.59$ (± 0.06)</td>
<td></td>
</tr>
<tr>
<td><strong>Dosing P</strong></td>
<td>(1) + (3)</td>
<td>$1.62$ (± 0.06)</td>
<td></td>
</tr>
</tbody>
</table>

b) in L^2.mol^{-2}.s^{-1} x 10^{-4}  
c) Billeter et al., Chemom. Intell. Lab. Syst., 93 (2008), 120-131  
c) Carvalho et al., Talanta, 68 (2006), 1190-1200

The model is kinetically validated

$$B + P + Aa \xrightarrow{k_1} BP + Aa$$
Spectral validation of the model

\[ B + P + Aa \xrightarrow{k_i} BP + Aa \]

Strategy (3): dosing

Species B and Aa dosed

Full spectral resolution

- Fitted spectra
- ••• Measured spectra
- XXX Predicted spectra
Spectral validation of the model

$$B + P + Aa \rightarrow B'P' + Aa$$

**Strategy (1): uncoloured species**

- **Strategy (1)+(2):** provided known spectrum

  - **Strategy (3): dosing**
  - **Strategy (1)+(4):** second order global analysis

- **Initial concentration of B varied, Aa set uncoloured**
  - **Full spectral resolution**
  - **Fitted spectra**
  - **Measured spectra**
  - **Predicted spectra**

The model can be spectroscopically validated with and without rank deficiency!

Species *P* and *Aa* set uncoloured

Partial spectral resolution

Pure spectrum of *B* provided

*Aa* set uncoloured

Partial spectral resolution
Conclusion: general methods

- **Chemometrics**
  - *Extraction* of information from complex multivariate signals
  - *Identification* of significant contributions (SVD, PCA)
  - *Support* for the elaboration of models

- **Kinetic hard-modelling of spectroscopic data**
  - *Determination* of kinetic parameters (e.g. rate constants)
  - *Assessment* of kinetic models by comparing fitted and independently measured pure component spectra (direct fitting)
  - *Calibration-free* method (implicit calibration)
Conclusion: specific methods in hard-modelling

- **Method of error propagation**
  - **Determination** of reliable uncertainties in the fitted kinetic parameters
  - **Prediction** of the experimental conditions minimising the uncertainties

- **Method for prediction of rank deficiencies and spectral validation**
  - **Design of rank deficient experiments** when Strategies 1 – 4 have to be used
  - **Interpretation** of the fitted pure component spectra when Strategy 1 (defining uncoloured species) is used
  - **Spectral validation** of rank deficient kinetic models
Acknowledgements

Group of Safety and Environmental Technology
  Prof. Dr. K. Hungerbühler  Head of the group

Subgroup of Reaction Analysis
  Dr. Y.-M. Neuhold  Senior scientist, subgroup leader
  Dr. G. Puxty  Former subgroup leader (now at CSIRO)
  Dr. G. Richner
    S. Cap
    T. Godany
    S. Gianoli
  Post-doctoral scientist
  Doctoral student
  Doctoral student
  Doctorat student

Subgroup of Reaction Process Design and Optimisation
  Dr. S. Papadokonstantakis  Senior scientist, subgroup leader

Subgroup of Separation Technologies
  Dr. L. Simon  Senior scientist, subgroup leader

and all members of the group of Safety and Environmental Technology
Thank you for your attention

Any question or comment?
Chemometric Methods for the Kinetic Hard-modelling of Spectroscopic Data