Kernel-Partial Least Squares regression coupled to pseudo-sample trajectories for the analysis of mixture designs of experiments

Supporting Material

I Relationship between Scheffé and Cox model coefficients

It can be easily demonstrated that the Scheffé model coefficients can be derived from the Cox model ones as:

Linear model: \( \beta_i = \alpha_0 + \alpha_i \)  

Quadratic model: \( \beta_i = \alpha_0 + \alpha_i + \alpha_{i,i} \) 
\[ \beta_{i,j} = \alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j} \]  

Consider the following formulations of the second-order Scheffé and Cox polynomials (Equation S4 and S5, respectively):

\[
y = \sum_{i=1}^{I} \beta_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \beta_{i,j} x_i x_j + \epsilon \quad (S4)
\]
\[
y = \alpha_0 + \sum_{i=1}^{I} \alpha_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \alpha_{i,j} x_i x_j + \sum_{i=1}^{I} \alpha_{i,i} x_i^2 + \epsilon \quad (S5)
\]
\[
\text{s.t.} \begin{cases} 
\sum_{i=1}^{I} \alpha_i s_i = 0 \\
\sum_{j=1}^{I} c_{i,j} \alpha_{i,j} s_j = 0 & \forall i \in [1, 2, \ldots, I]
\end{cases}
\]

By applying to Equation S5 the mixture constraint:

\[
\sum_{i=1}^{I} x_i = 1 \quad (S6)
\]

and reformulating the second-order terms, \( x_i^2 \), as:

\[
x_i^2 = x_i(1 - \sum_{j=1, j\neq i}^{I} x_j) \quad (S7)
\]
it follows:

\[
\hat{y} = \alpha_0 \sum_{i=1}^{L} x_i + \sum_{i=1}^{L} \alpha_i x_i + \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_{i,j} x_i x_j + \sum_{i=1}^{L} \alpha_{i,i} x_i (1 - \sum_{j=1}^{L} x_j)
\]

\[
\hat{y} = \sum_{i=1}^{L} (\alpha_0 + \alpha_i) x_i + \sum_{i=1}^{L} \sum_{j \neq i}^{L} \alpha_{i,j} x_i x_j + \sum_{i=1}^{L} \alpha_{i,i} x_i - \sum_{i=1}^{L} \sum_{j \neq i}^{L} \alpha_{i,i} x_i x_j
\]  

(S8)

\[
\hat{y} = \sum_{i=1}^{L} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{L} \sum_{j \neq i}^{L} (\alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}) x_i x_j
\]  

being \( \hat{y} \) the estimated value of the response property to be predicted. The notation \( \alpha_{i,j}^* \) permits to explicitly differentiate the interaction terms \( x_i x_j \) and \( x_j x_i \). Specifically, \( \alpha_{i,j}^* = \alpha_{j,i}^* \) and \( \alpha_{i,j} = 2\alpha_{i,j}^* = 2\alpha_{j,i}^* \) if \( i \neq j \). Rewriting Equation S8, it is obtained:

\[
\hat{y} = \sum_{i=1}^{L} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} (\alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}) x_i x_j
\]  

(S9)

As:

\[
\hat{y} = \sum_{i=1}^{L} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} (\alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}) x_i x_j
\]  

(S10)

\[
\hat{y} = \sum_{i=1}^{L} \beta_i x_i + \sum_{i=1}^{L} \beta_{i,j} x_i x_j
\]

it is then proved that:

\[
\beta_i = \alpha_0 + \alpha_i + \alpha_{i,i}
\]

\[
\beta_{i,j} = \alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}
\]  

(S11)

In the particular case where \( \alpha_{i,i} = \alpha_{i,j} = \alpha_{j,j} = 0 \) (first-order polynomial), then:

\[
\beta_i = \alpha_0 + \alpha_i
\]  

(S12)

*quod erat demonstrandum.*

On the contrary, if the Scheffé model parameters are given, the corresponding Cox model ones can be calculated by solving the linear equation system encompassing either Equations S1 and 5 or Equations S2, S3 and 7.
Table SM.1 - Tablet data: Scheffé model coefficients estimated by Scheffé polynomial fitting by means of OLS, Cox polynomial fitting by means of PLS and K-PLS

<table>
<thead>
<tr>
<th>Model coefficient</th>
<th>Scheffé model fitting by OLS</th>
<th>Cox model fitting by PLS</th>
<th>K-PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>198.16</td>
<td>198.10</td>
<td>198.16</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>114.06</td>
<td>111.94</td>
<td>114.06</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>328.97</td>
<td>326.21</td>
<td>328.97</td>
</tr>
<tr>
<td>$\beta_{1,2}$</td>
<td>-403.26</td>
<td>-404.98</td>
<td>-403.26</td>
</tr>
<tr>
<td>$\beta_{1,3}$</td>
<td>350.56</td>
<td>347.54</td>
<td>350.56</td>
</tr>
<tr>
<td>$\beta_{2,3}$</td>
<td>330.37</td>
<td>323.24</td>
<td>330.37</td>
</tr>
</tbody>
</table>

Subindices stand for: 1 = cellulose; 2 = lactose; 3 = phosphate

Table SM.2 - Bubbles data: Scheffé model coefficients estimated by Scheffé polynomial fitting by means of OLS, Cox polynomial fitting by means of PLS and K-PLS

<table>
<thead>
<tr>
<th>Model coefficient</th>
<th>Scheffé model fitting by OLS</th>
<th>Cox model fitting by PLS</th>
<th>K-PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-1.49</td>
<td>-1.49</td>
<td>-1.49</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>2.35</td>
<td>2.35</td>
<td>2.35</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-1.35</td>
<td>-1.35</td>
<td>-1.35</td>
</tr>
<tr>
<td>$\beta_{1,2}$</td>
<td>2.08</td>
<td>2.08</td>
<td>2.08</td>
</tr>
<tr>
<td>$\beta_{1,3}$</td>
<td>14.55</td>
<td>14.55</td>
<td>14.55</td>
</tr>
<tr>
<td>$\beta_{1,4}$</td>
<td>7.51</td>
<td>7.51</td>
<td>7.51</td>
</tr>
<tr>
<td>$\beta_{2,3}$</td>
<td>6.70</td>
<td>6.70</td>
<td>6.70</td>
</tr>
<tr>
<td>$\beta_{2,4}$</td>
<td>2.63</td>
<td>2.63</td>
<td>2.63</td>
</tr>
<tr>
<td>$\beta_{3,4}$</td>
<td>7.82</td>
<td>7.82</td>
<td>7.82</td>
</tr>
</tbody>
</table>

Subindices stand for: 1 = dish-washing liquid 1 (DWL1); 2 = dish-washing liquid 2 (DWL2); 3 = water; 4 = glycerol