ON THE USE OF SPARSE-PLS TO EXPLAIN PETROLEUM PROPERTIES AT THE MOLECULAR LEVEL

J-J. Da Costa Soares¹, F. Chainet¹, B. Celse¹, M. Lacoue-Nègre¹, C. Ruckebusch², N. Caillol¹, D. Espinat¹

¹IFP Energies nouvelles, Rond-point de l’échangeur de Solaize, BP3, 69360, Solaize, France
²LASIR, Université de Lille, Sciences et Technologies, 59655 Villeneuve d’Ascq Cedex, France
Process optimization: time consuming and costly experiments → process modelling (feedstock properties, catalyst...)

But some properties are poorly predicted

→ Need to understand and explain the property at the molecular level

Cloud Point (CP)
- ASTM D2500 or NF ISO 23015
- The temperature at which wax crystals begin to form in a petroleum product as it is cooled

Analytical techniques: $^{13}$C NMR and GCxGC
ANALYTICAL METHODS

- GCxGC → molecular families
  - 152 variables
  - Preprocessing: mean centering and autoscaling

- $^{13}$C NMR → carbons molecular environment
  - Up to 65k variables → variable selection: 0 to 60 ppm (13k variables)
  - Preprocessing: binning, normalisation and mean centering

Before binning

After binning
Methodology:
- Development of a model to predict the cloud point from NMR and GCxGC data
- Interpretation of the loadings to extract the relevant information related to the property

1/ PLS
- Estimation of the number of LVs
- Interpretation of the loadings

2/ Sparse-PLS
- Sparsity when constructing the direction vectors
- Variable selection penalty \(\rightarrow\) suppression of the lowest coefficients
RESULTS – GCXGC DATA (1/3)

<table>
<thead>
<tr>
<th>Model</th>
<th>LV</th>
<th>RMSECV (°C)</th>
<th>RMSEC (°C)</th>
<th>τ_{1IC} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>10</td>
<td>1.6</td>
<td>0.56</td>
<td>100</td>
</tr>
<tr>
<td>sparse PLS</td>
<td>10</td>
<td>1.2</td>
<td>0.48</td>
<td>100</td>
</tr>
</tbody>
</table>

\[ \eta = 0.86 \]
RESULTS – GCXGC DATA (2/3)

Example: Loading PLS LV1

Where is the relevant information to explain the cloud point?
RESULTS – GCXGC DATA (3/3)

- Loading sparse-PLS LV1
- Loading sparse-PLS LV2

152 → 24 selected variables

- N-paraffin from C18 to C21 ↑ CP
- Isoparaffin C21 and C22 ↓ CP
RESULTS – $^{13}$C NMR DATA (1/2)

<table>
<thead>
<tr>
<th>Model</th>
<th>LV</th>
<th>RMSECV (°C)</th>
<th>RMSEC (°C)</th>
<th>$\tau_{21C}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>8</td>
<td>1.9</td>
<td>0.9</td>
<td>100</td>
</tr>
<tr>
<td>sparse PLS</td>
<td>8</td>
<td>1.3</td>
<td>0.8</td>
<td>100</td>
</tr>
</tbody>
</table>

$\eta = 0.85$
RESULTS – $^{13}$C NMR DATA (2/2)

29.9 ppm $\leftrightarrow$ N-paraffin: $\uparrow$ CP
19.9 ppm $\leftrightarrow$ isoparaffin $\downarrow$ CP

Consistent with GCxGC

... More to investigate ...
CONCLUSIONS

- **PLS model**
  - Quite easy to develop
  - Loadings interpretation could be very long and complex

- **Sparse-PLS**
  - Quite easy to develop
  - Loadings easily interpretable
  - Identification of the relevant variables to explain the cloud point
Thank you for your attention

Innovating for energy

Find us on:

🌐 www.ifpenergiesnouvelles.com
🐦 @IFPENinnovation