Quantitative resolution of emulsifiers in an agrochemical formulation.

Chimiometrie 2020
Liege, Belgium

S. Mukherjee[1] S. Treacy[2] J. McGrath[2] A. Ayling[2] R. Williams[2] A. A. Gowen[1]

[1]UCD Spectral Imaging Research Group (SIRG), UCD School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin 4, Ireland
[2]Life Scientific Ltd, Block 4, Belfield Office Park, Beech Hill Road, Dublin 4, Ireland.
sindhuraj.Mukherjee@ucd.ie

Context

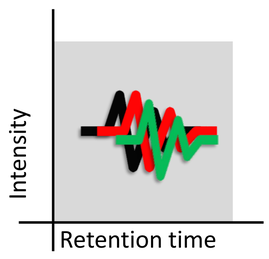
* Emulsifiers stabilise & optimise biological formulations
* Batch to batch variation affects QA, develop model to quantify emulsifiers in blends.
* Quantitation of emulsifiers by LCMS :- co-elution, overlapping peaks, matrix effects; multivariate methods for resolution needed.

Data description & pre-processing

Each formulation is a *mix* of **three emulsifiers/co-formulants (Y1, Y2, Y3)** at the following concentrations

* Low (0.06mg/ml)
* Mid (0.15mg/ml)
* High (0.2399mg/ml)

Objective :- *Create individual models to predict level of each component in any given blend.*

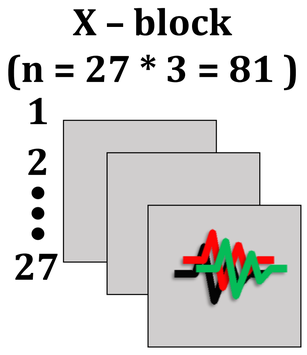


n = 27 unique blends

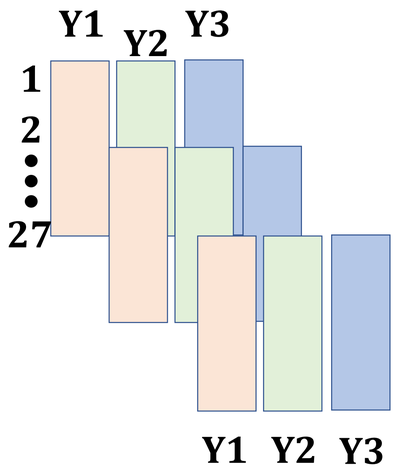
3 replicates

3 concentration values

30-minute LC run (Agilent 1290 Infinity II)

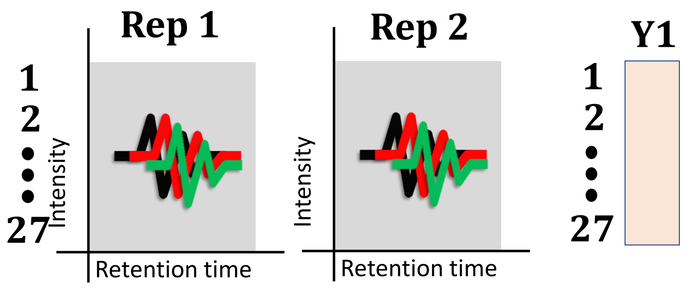


(a) Truncate **X - block** Rt 16 – 27min (*manual varsel*)

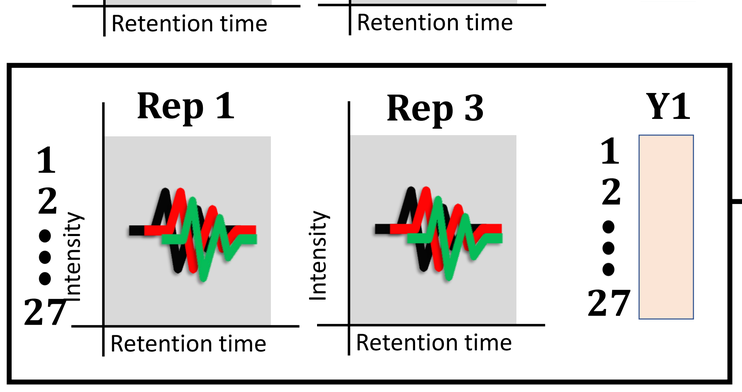


(b) *Autoscale* **X & Y**

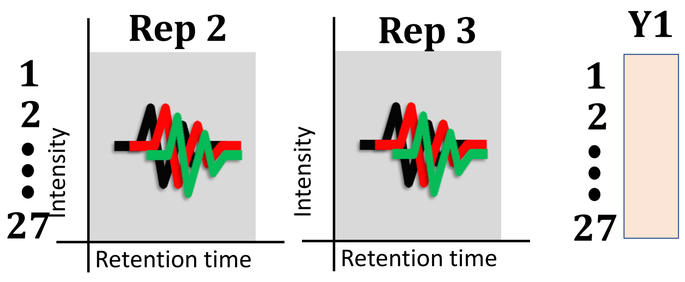
Modelling I (Calibration)



* Create PLSR1 model using n = 27 * 2 as calibration set for **Y1**
* 10-fold cross validation



* Select no. of latent variables based on performance indicators :- **lowest $RMSE_{CV}$, highest R^2_{CV}**



* Check if variable selection needed and/or stable
* Evaluate variability in model performance with respect to cycling of calibration replicates to *select optimal model*

Modelling II (Validation)

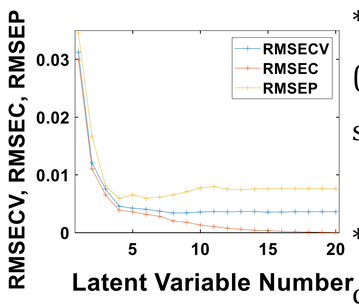
* Apply model on validation set (n = 27),i.e. replicate 2

* Evaluate model performance ;select model with **lowest $RMSE_P$, highest R^2_P**

* Inspect calibration-validation subspace using **Q residuals** to determine if optimal model reached ; consider **exclusion of datapoints** if needed ; consider inclusion of latent variables if needed

* Repeat modelling with **Y2 and Y3** and compare

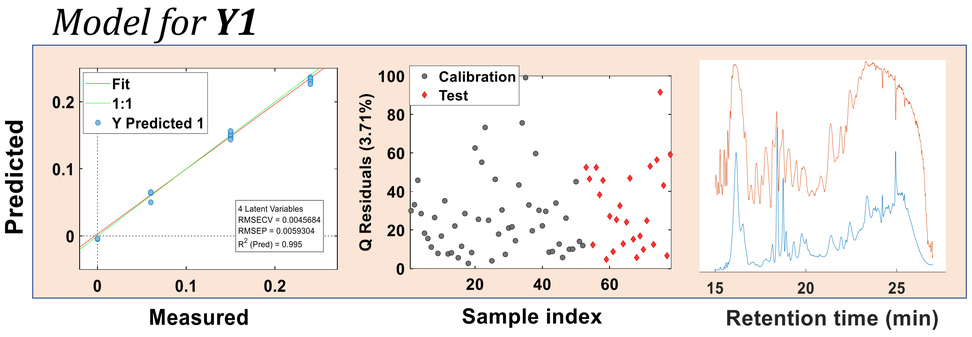
Results & Discussion



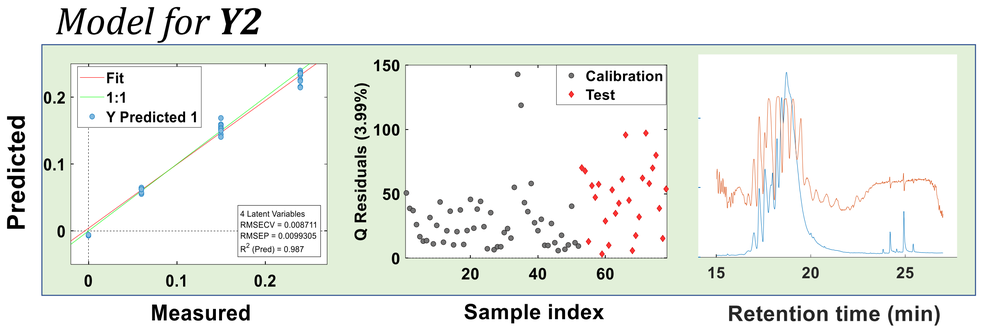
* Lowest $RMSE_{CV}$ for 4 – latent variable model (0.006) and highest **R2** (0.99) for each model separately built for **Y1, Y2 & Y3**

* **Rep 1, Rep 3** selected for calibration :- captures time-based variability, applied to **Rep 2**

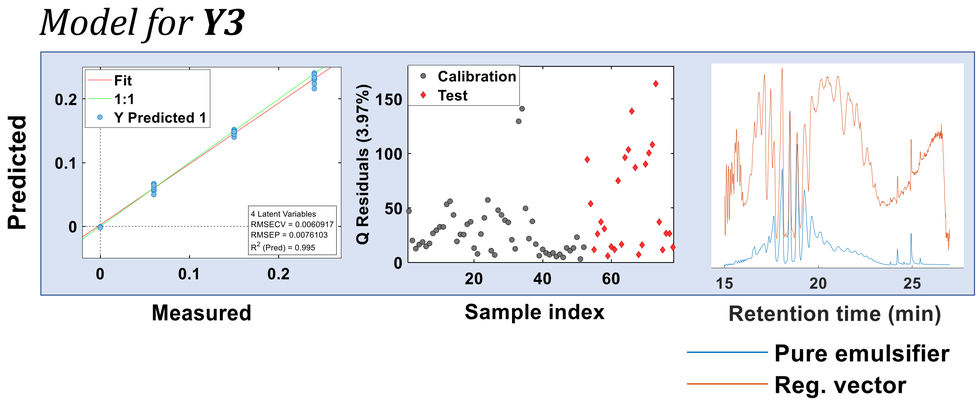
*Model for **Y1***



*Model for **Y2***



*Model for **Y3***



* **Q residual** subspace indicates inclusion of more latent variables could improve model specificity, at the cost of overfitting > more conservative 4 latent variable model preferred to keep predictive performance consistent

* Overlay of 'pure' emulsifier chromatogram with 4-latent variable model regression vector indicates key chromatographic features agree more or less

Acknowledgements

The authors would like to thank Enterprise Ireland for funding support and to SFdS, GC, Univ. Liege for organising Chimiometrie 2020.
All analyses performed using MatLab™ 2019a and PLS Toolbox 8.7.1.