



# La chimiométrie au cours des âges... récents

JM Roger

UMR ITAP, Irstea

[jean-michel.roger@irstea.fr](mailto:jean-michel.roger@irstea.fr)

# Pourquoi la chimiométrie ?



- Parce qu'elle est au cœur des préoccupations d'HélioSPIR :

69/173 = 40% des communications avec ce mot clef

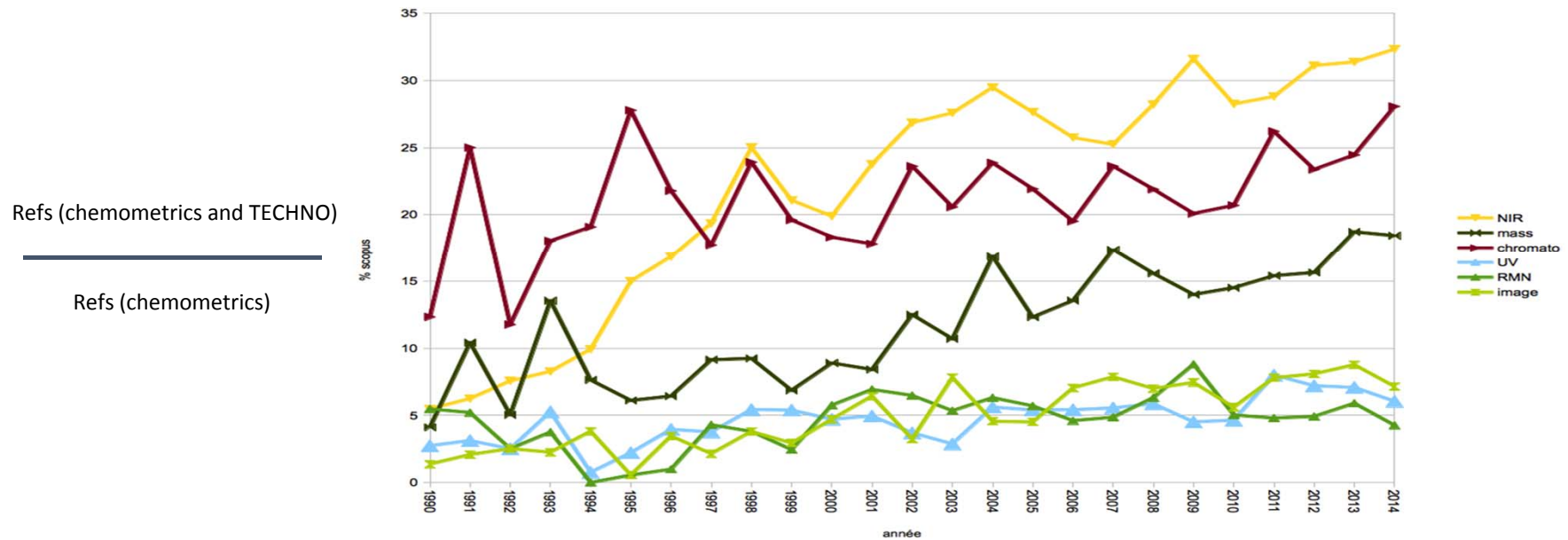
Demande récurrente d'une programmathèque

Sujet principal des demandes de formation

# Pourquoi la chimiométrie ?



- Parce que c'est une préoccupation générale de la SPIR



# Pourquoi la chimiométrie est-elle tellement nécessaire à la SPIR?



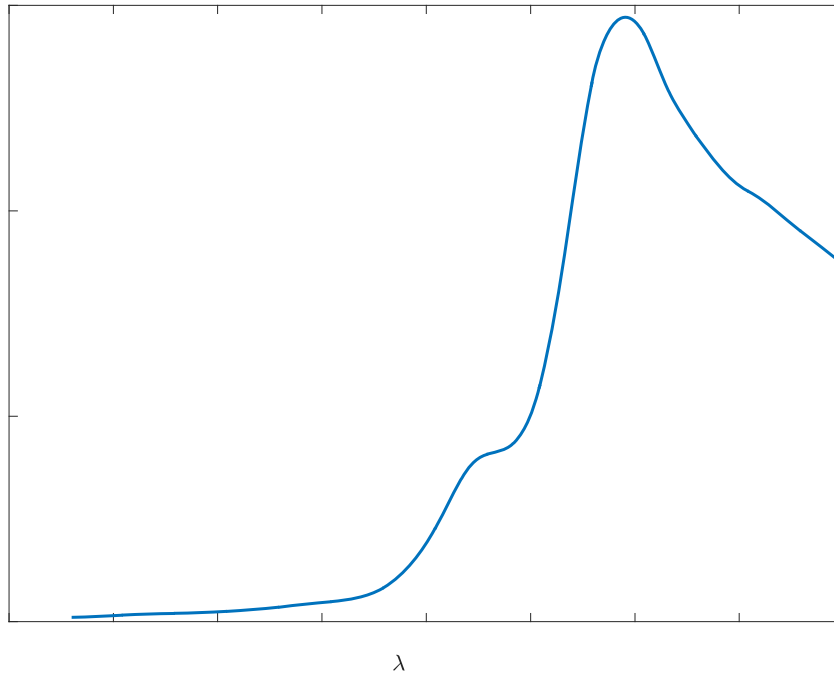
- Un spectre PIR, tout du moins d'un produit biologique, est constitué d'un ensemble de variables peu sélectives.

$$\text{sélectivité } (x_i / y) = \frac{\text{pente } (x_i / y)}{\text{pente } (x_i / \text{interférents})}$$

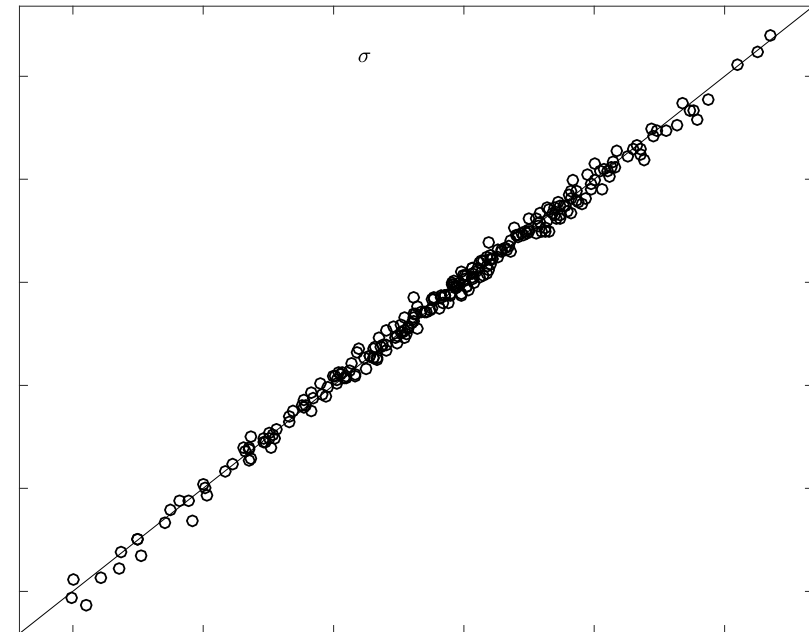
- Or, l'art de la chimiométrie est de trouver la meilleure combinaison de variables peu spécifiques, pour produire de nouvelles variables plus spécifiques :
  - Scores d'une ACP, PLS, MCR, etc

# Exemple : spectres de blé vs protéines

## Challenge IDRC 2016



$R^2$  entre chaque lambda et le taux de protéines  
Maximum 0.15



$R^2$  entre la combinaison (PLS) des lambda et protéines  
= 0.99

# L'âge de pierre

## La PLS et la validation croisée



- Mainstream :
  - Application de la PLS à des bases de données spectrales
  - Erreur souvent estimée par validation croisée
  - Pas ou peu de questions sur la robustesse
  - Pas ou peu de questions sur la conception de capteurs simplifiés
  - Pas ou peu de questions sur les modèles non linéaires

On n'a donc jamais eu d'exposé théorique sur la PLS  
On a supposé que vous connaissiez...

# Et puis, l'âge de fer



- La PLS produit des variables spécifiques vis à vis du jeu d'étalonnage
- Mais aussi plus sensibles aux perturbations non prévues

On vous a donc parlé de robustesse

# L'âge de fer

2004, TOP



2008, EROS



2003, EPO



2006, DOP





# L'adolescence



- Et si on me racontait des bêtises ???
- Et si la PLS ne faisait pas tout ???

On vous a donc parlé de méthodes exotiques

# L'adolescence...suite



- Vous avez eu le plaisir de suivre :
  - S Roussel Méthodes non linéaires pour la détection d'OGM
  - D. Bertrand Les fonctions radiales de base (RBF)
  - D. Rutledge Generalized PLS Cluster
  - R. Leardi Les algorithmes génétiques
  - D. Rutledge L'independent Component Analysis
  - N. Villa L'analyse des données fonctionnelles

Qu'en reste-t-il ?  
L'ICA ?

L'âge de raison

XII<sup>èmes</sup> Rencontres dédiées à la chimiométrie



- On y a vu des thèmes récurrents :
  - Les prétraitements
  - Les transferts d'étalonnage
  - L'incertitude des prédictions
  - La sélection de variables, pour le tri rapide

## Le 4<sup>ème</sup> âge

### Ce que vous verrez dans un futur proche



- La fusion de données : méthodes multi-blocs
- Le deep learning : réseaux CNN
- La discrimination : méthode adaptée à la SPIR?
- ...

# L'âge des apprentissages



3ème saison déjà !!  
1500 inscrits / an

Inscrivez vous sur  
[www.fun-mooc.fr](http://www.fun-mooc.fr)



# L'ère industrielle



Galaxy / ChemFlow Analyze Data Workflow Shared Data Visualization Admin Help User Using 2%

Tools

search tools

- Import Data
- Convert data format
- Utils
- Statistics
- Plots
- Pretreatments
- Exploration
- Regressions
- Clustering
- Discrimination
- Variable selection
- Orthogonal projections
- Calibration transfert
- Unmixing/ Curve Resolution
- Multitable/Multiway Analysis
- Method validation

Workflows

- All workflows

ChemFlow

## Chemometrics without programming

[www.chemproject.fr](http://www.chemproject.fr)

History

search datasets

imported: DATA SCIO 2017-2018 ITB  
17 shown  
23.5 MB

- 32: Spectra\_plot on data 1
- 31: X\_50%2018.tabular
- 30: X\_2017-50%2018.tabular
- 27: Y\_50%2018.tabular
- 19: Y\_2017-50%2018.tabular
- 17: Spectra\_plot on data 16
- 16: X\_2017\_2018\_plot.tabular
- 14: Y\_2017\_2018.tabular

# Conclusion



## Beach safety information



Only swim between the red & yellow flags.  
No flags = no swim

no chemometrics

=

no SPIR