

# Chimiométrie : outils du XX<sup>ème</sup> siècle, méthode du XXI<sup>ème</sup> siècle ?

P. Lantéri et R. Longray

Laboratoire de Chimiométrie, ESCPE, Université Claude Bernard Lyon 1, 43 Bd. du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

The tools and areas connected with the concept of Chemometrics are remembered. Then, gathering together Experimental Design and Data Analysis, we try to demonstrate that the collective term "chemometrics", requires an even greater acceptance by the scientific community. Outside, the simple semantic question, we must see beyond the simple "techniques and tools" aspect and propose a truly experimental methodology for the forthcoming years.

## Introduction

Le premier congrès français de Chimiométrie intitulé « Expériences et Modèles », s'est voulu le premier acte fondateur d'une communauté française trop longtemps dispersée par rapport aux autres communautés internationales qui se réclament de la même discipline. Le terme fédérateur de « Chimiométrie » permet d'afficher une unité thématique regroupant le traitement des données et leur collecte méthodologique dans le cadre des plans d'expériences. Le Groupe Français de Chimiométrie ne s'est constitué que très récemment (1996) sous l'impulsion de membres de la SFC et de la SCI. Cela traduit bien le fait que l'utilisation du label « chimiométrie » ne s'est imposée que lentement au cours de ces deux décennies.

La « chimiométrie » est-elle une discipline ? est-elle encore en émergence ? « Chimie » reflète dans l'imaginaire la notion d'expériences avec de subtils mélanges et « métrie » nous renvoie à la mesure, à la recherche de l'information. Le terme « français » de chimiométrie vient de l'anglais « chemometrics », il peut constituer un paradigme fédérateur et méthodologique. Rappelons que le mot « chemometrics » a une double paternité [1] : celle du Suédois Svante Wold et de l'Américain Bruce Kowalski, avec au départ une acception restrictive associant analyse de données et chimie analytique. Aujourd'hui ce terme recouvre l'ensemble des applications en chimie des méthodes statistiques et mathématiques avec l'appui de l'outil informatique [2].

Le texte présenté ici est extrait de la conférence inaugurale présentée au congrès « Chimiométrie 97 » qui s'est tenu à Lyon en décembre 1997.

Il nous a paru intéressant de préciser ce que ce concept associant des outils mathématiques connus du XX<sup>ème</sup> siècle pouvait avoir de prometteur, sur le plan méthodologique, pour les chimistes du XXI<sup>ème</sup> siècle.

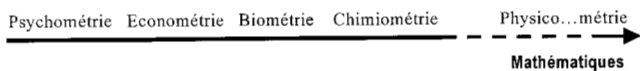
## Les bases historiques de la Chimiométrie

Même si la Chimiométrie est une discipline de création récente, elle n'est pas issue d'une génération spontanée, elle s'inscrit dans un démarche plus ancienne associant expériences et statistiques. Elle pourrait se référer à Descartes et au « Discours de la Méthode » par son souci méthodologique. Notons cependant que le « cartésianisme », poussé à l'extrême, a introduit le dogme, encore tenace, d'une démarche expérimentale reposant sur l'étude « d'un seul facteur à la fois » ce qui s'éloigne fortement, dans les faits et non dans l'esprit, des méthodes de la Chimiométrie. En effet, en raison de la puissance des moyens de calculs, des méthodes chimiométriques ont été élaborées pour conduire et exploiter des processus faisant intervenir simultanément plusieurs facteurs ou pour élaborer des modèles complexes.

L'apport de la statistique dans la démarche scientifique, avec notamment la prise en compte de la notion d'erreur, est relativement récent. Il remonte aux expérimentateurs-statisticiens nés au siècle dernier, dont l'œuvre est pérennisée à travers les outils statistiques qu'ils nous ont légués. Cinq auteurs méritent tout spécialement d'être cités : (i) *Sir Ronald Aylmer Fisher* (1890-1962), diplômé en astronomie de Cambridge (1912), enseignant les mathématiques en 1919 dans le cadre de l'agronomie, de la biologie et de la génétique qui s'intéressa à la méthode expérimentale pour développer des théories statistiques sur les petits échantillons, son nom est définitivement lié à l'analyse de variance ; (ii) *Karl Pearson* (1857-1936), un autre biostatisticien dont les travaux portent sur l'analyse de la régression et l'introduction du test du chi-deux ; (iii) *William Sealey Gosset*, un autre britannique qui se plaçait déjà dans une démarche « qualité » relative à la noble activité de brasseur de bière, allait imprimer sa marque en inventant le test en « t » de « Student » ; (iiii) *Jacques Salomon Hadamard* (1865-1963), un brillant chercheur français sur les séries de Taylor et sur les nombres premiers [3], et bien que n'adhérant pas du tout à cette approche expérimentale, nous a légué les matrices d'Hadamard si chères au cœur de ceux qui pratiquent les plans de pesées (et connues aussi sous le nom de matrices Plackett et Burman [4] ; (iiiiii) pour terminer il faut citer l'américain *George E.P. Box*, qui est un des pères de la méthode expérimentale associant chimie et statistique [5].

Deux ouvrages de Box [6,7], sont devenus des « standards » pour de nombreux expérimentateurs.

### La Chimiométrie et les autres ...« métries »



Une lecture linéaire, un peu élémentaire, consisterait à classer les « métries » selon un axe orienté par la méthode déductive et mathématique, bien illustrée par la physique classique, qui serait la marque du degré de « dureté » de la discipline.

En fait, derrière les différentes « métries », il y a souvent, à l'origine, une science en émergence qui n'était pas encore mathématisée ou mathématisable compte tenu de son degré de complexité. Cette science a été amenée à donner une place importante à l'incertitude dans l'analyse de ses résultats grâce à une méthodologie spécifique utilisant les outils mathématiques et statistiques, mais conservant une approche « expérimentale empirique ». La notion de physico-métrie reste peut être encore une boutade.

La chimie, qui par de nombreux aspects est proche de la physique, a été ainsi une des dernières à être contaminée... mais en n'hésitant pas à faire des emprunts aux métries qui l'avaient précédées (notons par exemple la description du PLS par Horst dans Psychometrika [8]). Il est bon de souligner que ce type de problèmes avait déjà été évoqué dans Biometrika par Hotelling [9] dès 1936.

### Les domaines et outils de la chimiométrie

Pour définir une discipline on peut soit partir d'un concept *a priori* soit d'un constat en recherchant les objets et les domaines qui se réclament de la discipline. C'est cette

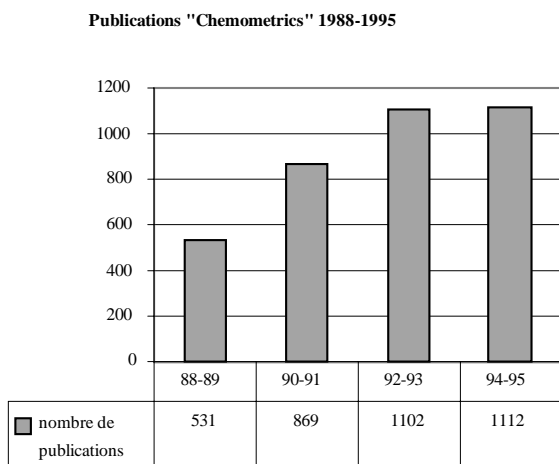


Figure 1. Évolution des publications en chimiométrie.

démarche pragmatique que nous avons adoptée. Une excellente compilation sur la chimiométrie, obtenue par l'exploitation des Chemical Abstracts, est régulièrement proposée, tous les deux ans, par S.D. Brown, depuis deux décennies.

Nous en sommes à la 11<sup>ème</sup> édition et nous n'avons retenu ici que les huit dernières années compilées (plus de 3600 références !) comme indicateur de l'évolution de la discipline [10-13] (Fig. 1).

Outre la progression des publications, cette étude nous renseigne également sur les différents domaines concernés par la chimiométrie et leur évolution au cours de ces dernières années (Fig. 2).

### Chimiométrie : un essai « d'échantillonnage »

S.D. Brown [10] signale déjà la difficulté des mots clés pertinents pour trouver de manière exhaustive les articles ayant trait à la chimiométrie. Le terme de chimiométrie (chemometrics) a d'abord été « rare » (seulement 8 fois utilisé comme mot-clé entre janvier 76 et octobre 79 !).

Afin de situer si ce terme était ou non caractéristique des disciplines qu'il veut représenter, nous avons réalisé une interrogation à partir d'une base de données (Pascal) sur la période 1989-1996.

La règle du jeu que nous nous sommes fixée est une recherche bibliographique avec **le seul mot clé** « Chemometrics » ; nous avons ainsi obtenu un « échantillonnage » de 222 références que nous avons pu analyser : outils utilisés, domaines concernés (Tab. I) et utilisateurs (dans le cadre **volontairement restreint** où nous nous sommes placés). Si l'on peut se permettre de rapprocher ces résultats de ceux obtenus par Brown, on constate :

- que le terme de chimiométrie commence à émerger en tant que mot-clé mais qu'il ne permet d'accéder encore qu'à un faible pourcentage des publications concernées,
- que l'on obtient par contre une bonne image de **l'ensemble des outils** qui se réclament de la chimiométrie.

### Outils du XX<sup>ème</sup> siècle, méthode du XXI<sup>ème</sup> siècle ?

Les outils chimiométriques nécessaires pour permettre la structuration complète de l'expérimentation qui va de l'élaboration des plans d'expérience à l'analyse des données sont en place pour la plupart. La puissance de l'outil informatique laisse espérer de mieux analyser et exploiter la complexité chimique : modèles non linéaires, construction sur mesure et réalisation de systèmes expérimentaux par l'apport des réseaux neuronaux, des algorithmes génétiques.

La chimie de « performance » (souvent chimie de formulation) liée aux propriétés d'usage d'un produit se développe de plus en plus grâce à ces outils.

La chimie combinatoire [14,15], qui semble prometteuse pour la recherche de molécules biologiquement actives, en

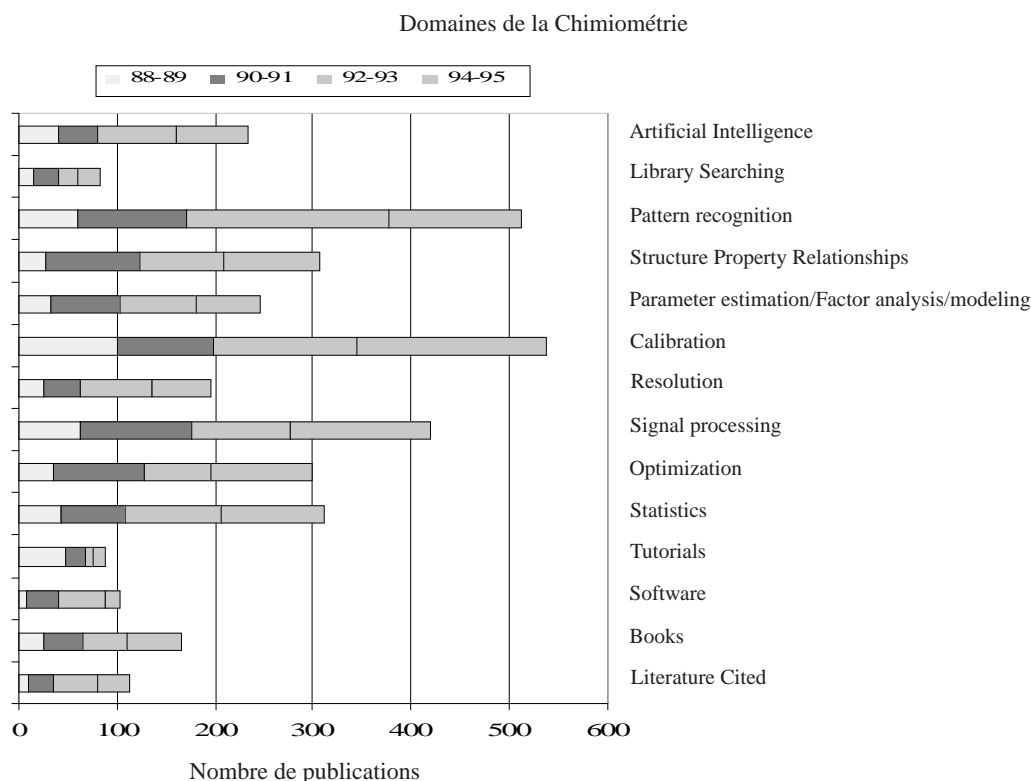


Figure 2. Les domaines se réclamant de la Chimiométrie.

Tableau I. Les outils de la chimiométrie.

| Thème                                   | Méthodes/Outils   | Domaines  |
|---|---|---|
| Analyses de données                     | <ul style="list-style-type: none"> <li>- Analyse Statistique</li> <li>- Analyse factorielle</li> <li>- Analyse en composantes principales (PCA)</li> <li>- Régression PLS-PCR</li> <li>- Réseaux neuronaux</li> <li>- Relation Structure Propriété</li> </ul> | Analyses NIR, IR-FT, UV<br>Chromatographies HPLC<br>Spectrométrie Masse, Fluorescence<br>Cinétique, Propriétés Physicochimiques<br>Analyse des eaux, Archéologie, Biotechnologie,<br>Géochimie<br>Huiles essentielles, Pharmacie, Pollution |
| Calibration                             | <ul style="list-style-type: none"> <li>- Régression MLR, PCR, PLS</li> <li>- Réseaux neuronaux</li> <li>- PCA</li> <li>- Validation croisée</li> </ul>  | UV, IR-FT, NIR<br>Fluorescence<br>Médicament<br>Analyse chimique  |
| Classification                          | <ul style="list-style-type: none"> <li>- Réseaux neuronaux</li> <li>- PCA</li> <li>- Analyse Discriminante</li> <li>- Filtre de Kalman</li> </ul>   | Arôme, Archéologie<br>Dosage des vins<br>Déchets plastiques<br>NIR, HPLC,<br>Contrôle qualité,<br>Analyse chimique  |
| Plans d'expérience<br>Optimisation      | <ul style="list-style-type: none"> <li>- Plan factoriel</li> <li>- Plackett et Burman</li> <li>- Taguchi</li> <li>- Simplexe</li> <li>- Doehlert</li> <li>- Composite</li> <li>- Surfaces de réponses</li> <li>- D-Optimalité</li> </ul>                      | Agroalimentaire, Médicaments<br>Calorimétrie<br>Analyse chimique, Chromatographie<br>HPLC Electrochimie, Electrophorèse<br>Environnement, Ecologie, Biologie<br>Polarographie<br>Mélanges   |
| Algorithmes et<br>Assistance ordinateur | <ul style="list-style-type: none"> <li>- Réseaux neuronaux</li> <li>- Système flou</li> <li>- Algorithme Génétique</li> </ul>   | Analyses Chimiques<br>pH métrie, IR, Spectrométrie, Relations<br>Structures-propriétés, Etalonnage, PLS/calibration   |

chimie pharmaceutique ou en agrochimie, s'intègre dans la démarche chimiométrique à laquelle elle devrait être naturellement associée. Une telle perspective pourrait concerner d'autres domaines d'intérêt que la seule recherche de médicaments dans la mesure où l'on sera capable de relier la description des structures moléculaires à des propriétés pertinentes.

Mais en cette fin de XX<sup>ème</sup> siècle, on peut se poser la question de savoir si la Méthode chimiométrique qui devrait être autre chose qu'une simple juxtaposition d'outils, est vraiment ancrée ? En fait, un des défis pour les prochaines années est de banaliser la discipline pour l'ensemble de la communauté des expérimentateurs universitaires et industriels. Pour cela il faut faire reculer un certain nombre d'idées préconçues qui consistent soit à demander trop à l'outil par rapport à un approche purement phénoménologique, soit à considérer l'idée intéressante mais pas adaptée à son problème.

Ce pari ne sera gagné qu'au prix d'une certaine évolution, voir révolution, des esprits cartésiens.

### Conclusion

La chimiométrie n'est plus seulement un ensemble d'outils statistiques efficaces, elle a gagné son titre de discipline spécifique dans la mesure où elle représente désormais une méthodologie complète pour aborder bien des aspects de la chimie qui vont de la molécule au procédé, de l'aspect fondamental à la propriété d'usage.

Il reste à la faire accepter comme une **méthodologie** de recherche et développement indispensable au XXI<sup>ème</sup> siècle.

### Références

1. Geladi, P.; Esbensen, K. J.; *Chemometrics* **1990**, *4*, 337-354; *ibid* **1990**, 389-412.
2. Lantéri, P.; Longerey, R. *Analisis* **1996**, *24*(2), 17-27.
3. Schwartz, L. « Le petit père Hadamard », « Pour la Science » (mars 97), pp 88-93.
4. Plackett, R. L.; Burman, J. P. *Biometrika* **1946**, *33*, 305-325.
5. Box, G. E. P. Département de statistiques, Université du Wisconsin, Madison, Rapport technique N°455 (1976).
6. Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters: An introduction to design, data analysis and model building*, J. Wiley, NY (1978).
7. Box, G. E. P.; Draper, N. R. *Empirical Model Building and Response Surfaces*, J. Wiley, NY (1987).
8. Horst, P. *Psychometrika* **1961**, *26*, 129-149.
9. Hotelling, H. *Biometrika* **1936**, *28*, 221-277.
10. Stephen, D. Brown « *Chemometrics* », *Anal. Chem.* **1990**, *62*(12), 85-101.
11. Stephen, D. Brown « *Chemometrics* », *Anal. Chem.* **1992**, *64*(12), 23-49.
12. Stephen, D. Brown, « *Chemometrics* », *Anal. Chem.* **1994**, *66*(12), 315-359.
13. Stephen, D. Brown, « *Chemometrics* », *Anal. Chem.* **1996**, *68*(12), 21-61.
14. « Combinatorial, Special Report », *Chemical and Engineering News*, N° 12, 1996; pp 28-73
15. Pierre Laperousse, « La chimie combinatoire pour créer des bibliothèques de molécules ». *L'Usine Nouvelle*, hors série, octobre 1995.