



HAL
open science

Chimométrie et Métabolomique

Jean-Michel Roger, Rémi Servien, Jean-Claude Boulet, Marie Tremblay-Franco, Benoit Jaillais, Julien Bocard, Franck Giacomoni, Marion Brandolini-Bunlon, Estelle Pujos-Guillot, Yann Guitton, et al.

► **To cite this version:**

Jean-Michel Roger, Rémi Servien, Jean-Claude Boulet, Marie Tremblay-Franco, Benoit Jaillais, et al.. Chimométrie et Métabolomique. Annales des falsifications, de l'expertise chimique et toxicologique, 2021, Eco-chimie pour des produits durables. Filières durables, procédés éco-respectueux, impact santé environnement, pp.80-84. hal-03642211

HAL Id: hal-03642211

<https://hal.science/hal-03642211>

Submitted on 2 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chimiométrie et Métabolomique

Jean-Michel Roger^{1,2}, Rémi Servien^{3,2}, Jean-Claude Boulet^{4,2}, Marie Tremblay Franco^{5,6}, Benoit Jaillais⁷, Julien Boccard⁸, Franck Giacomoni^{9,10}, Marion Brandolini-Bunlon⁹, Estelle Pujos-Guillot⁹, Yann Guitton¹¹, Christophe Lebègue¹², Mohamed Hanafi⁷, Douglas N. Rutledge^{13, 14}

¹ INRAE, ITAP, Montpellier SupAgro, University of Montpellier, Montpellier, France

² ChemHouse Research Group, Montpellier, France

³ INRAE, Univ Montpellier, LBE, 102 Av des Etangs, Narbonne F-11100, France

⁴ INRAE, UMR1083, Montpellier SupAgro, University of Montpellier, Montpellier, France

⁵ Toxalim, Toulouse University, INRAE, ENVT, INP-Purpan, UPS, Toulouse, France.

⁶ Metatoul-AXIOM platform, MetaboHUB, Toxalim, INRAE, Toulouse, France

⁷ INRAE/ONIRIS, StatSC, Nantes, France.

⁸ School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland.

⁹ Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000, Clermont-Ferrand, France

¹⁰ CATI INRAE EMPREINTE, département Alimentation Humaine (ALIMH), F-63000, Clermont-Ferrand, France

¹¹ Laboratoire d'Étude des Résidus et Contaminants dans les Aliments (LABERCA), Oniris, INRAE, UMR 1329, 44307, Nantes, France

¹² Formation Permanente, INRAE, Montpellier, France

¹³ Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 75005, Paris, France

¹⁴ National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia

Résumé

La chimiométrie et la métabolomique se sont développées, de manières parallèles, en conséquence de l'apparition des techniques de la chimie analytique moderne. La première peut contribuer à résoudre des problèmes posés par la deuxième. C'est donc tout naturellement que ces deux champs disciplinaires se sont rencontrés, comme en témoignent le nombre croissant de productions scientifiques. Les progrès en cours en matière d'analyses à haute résolution posent de nouveaux problèmes que les chercheurs et utilisateurs visent à résoudre, notamment en se réunissant dans des structures d'animation communes, comme ChemOmics.

Mots clefs : Chimiométrie, Métabolomique, Chimie Analytique, Phénotypage

Abstract

Chemometrics and metabolomics have developed in parallel as a result of the emergence of modern analytical chemistry techniques. The first aims at solving problems posed by the second. It is thus quite naturally that these two disciplinary

fields have met, as shown by the increasing number of scientific productions. The current progress in high-resolution analysis poses new problems that researchers and users aim to solve, notably by meeting in common animation structures, such as ChemOmics.

Keywords: Chemometrics, Metabolomics, Analytical Chemistry, Phenotyping

1. Introduction

Une des premières branches à se développer en chimiométrie fut celle des plans d'expériences, certainement parce qu'elle est peu gourmande en puissance de calcul et parce qu'elle s'appuie sur des outils statistiques classiques tels l'analyse de variance (ANOVA). Mais la chimiométrie telle que nous la connaissons aujourd'hui [1], [2], est née de la rencontre de la chimie analytique quantitative, des méthodes d'analyse multivariées et de la micro-informatique. C'est dans les années 80 que ce mélange a présenté une stœchiométrie favorable au développement de cette jeune discipline. Des conférences, des revues, des logiciels et des livres sont apparus. En Europe, plusieurs groupes de recherche, principalement en chimie analytique, ont posé les bases de la chimiométrie, et Massart et ses collègues ont produit l'un des premiers textes fondateurs [3]. Les premières applications de la chimiométrie ont concerné principalement l'étalonnage des appareils de la chimie analytique quantitative, comme les spectromètres proche infrarouge. Les dernières décennies ont vu naître une autre révolution : la croissance rapide de la puissance de calcul, permettant de traiter les données de nouveaux instruments rapides et automatisés, générant de grands volumes de données. Les progrès techniques ont permis de coupler ces spectrométries à des capteurs matriciels, ce qui a conduit au développement de l'imagerie spectrale, qui associe un spectre à chaque pixel des images.

La métabolomique, qui s'appuie sur des techniques analytiques telles que la spectrométrie de masse (MS) ou la résonance magnétique nucléaire (RMN), vise à offrir une description approfondie des événements biochimiques qui régissent les systèmes vivants par la mesure de l'ensemble des métabolites qui les caractérisent. Les métabolites sont des composés chimiques de faible poids moléculaire (<1500 Da) impliqués dans de multiples voies métaboliques interconnectées, et participant aux processus cellulaires, soit comme intermédiaires soit comme produits finaux de réactions. Ainsi la métabolomique permet une vision intégrée du métabolisme par une détection sensible des changements biochimiques au cours du temps, résultant des interactions de facteurs intrinsèques et extrinsèques des organismes étudiés. A ce

titre, elle constitue un outil de phénotypage puissant pour l'étude des phénomènes biologiques dans une grande diversité de domaines de recherche. Les développements technologiques récents permettent aujourd'hui de mesurer simultanément lors d'une analyse des milliers de métabolites présents dans un biofluide ou un tissu et génèrent des données complexes et massives. L'extraction d'informations pertinentes au sein de ces vastes ensembles de données nécessite donc des outils spécifiques. Dans ce contexte, la chimiométrie propose de nombreuses méthodes adaptées à la fois à la nature des données mesurées et à la structure du dispositif expérimental. Ces outils sont devenus essentiels pour l'extraction de connaissances en métabolomique et c'est donc tout naturellement que ces deux disciplines se sont rencontrées.

2. La chimiométrie

La structure classique des données traitées en chimiométrie est une matrice \mathbf{X} , composée de n lignes (individus) sur lesquels sont mesurées p variables d'intérêt. Dans des cas plus complexes, comme l'imagerie spectrale, les images sont dépliées avant traitement chimiométrique, puis repliées ensuite. Le prétraitement des données est une composante essentielle de la chimiométrie. On y trouve nombre de méthodes adaptées aux spécificités des signaux mesurés en chimie analytique, principalement des données spectrales, utilisant à la fois des outils de traitement du signal et d'algèbre linéaire. D'une manière générale, les prétraitements de données sont utilisés pour corriger les artefacts dus à l'acquisition des données, tels que le bruit, les lignes de base, les effets multiplicatifs, les décalages de pics ou plus simplement pour rétablir la linéarité du signal par rapport aux grandeurs d'intérêt, comme par exemple le calcul d'absorbance. Un très grand nombre de méthodes de prétraitement ont été développées dans le cadre de la chimiométrie [4].

La branche maîtresse de la chimiométrie reste sans aucun doute la très grande famille des méthodes factorielles. La pierre angulaire de cette famille est l'Analyse en Composantes Principales (ACP), qui a été introduite par Pearson en 1901 [5]. L'ACP est devenue un outil classique de l'analyse de données [6] pour finalement être adoptée très rapidement par la communauté de la chimiométrie [7]. L'ACP s'avère particulièrement utile lorsque les variables décrivant les données sont très corrélées, ce qui est le cas par nature des données spectrales. Elle produit de nouvelles variables, en plus petit nombre que celles d'origine, non corrélées et exprimant la plus grande partie de la variance des données. Concrètement, l'ACP est une décomposition

orthogonale de la matrice $\mathbf{X} = \mathbf{TP}' + \mathbf{E}$, où \mathbf{TP}' exprime la partie structurale des données et \mathbf{E} le bruit résiduel. À partir de ce formalisme, un très grand nombre de méthodes ont rejoint l'arsenal de la chimiométrie. La plus populaire est la régression PLS, pour Partial Least Squares ou Projection on Latent Structures [8], et dans laquelle la partie structurée \mathbf{TP}' est reliée à la covariance entre les signaux de \mathbf{X} et une ou plusieurs réponses \mathbf{Y} . La régression PLS (PLSR), permet de réaliser des étalonnages avec moins d'individus que de descripteurs, ce qui se révèle très utile pour l'étalonnage des spectromètres en tous genres. L'analyse discriminante par PLS (PLS-DA) [9], permet de réaliser des modèles de classification supervisée. D'autres méthodes factorielles, dites de démixage, ont été importées depuis le monde des statistiques et du traitement du signal, comme l'Independent Components Analysis (ICA) [10] et la Multi Curve Resolution (MCR-ALS) [11]. Par ailleurs, de nombreuses généralisations de l'ACP d'un à plusieurs tableaux de données, regroupées sous l'appellation méthodes multiblocs, se sont développées et appliquées d'une manière très significative ces dernières décennies.

La communauté française de chimiométrie est structurée autour du Groupe Français de Chimiométrie (GFC) [12]. Un pôle de recherches multi-instituts (ChemHouse) [13] propose des outils de formation (CheMoocs) [14] et des ressources logicielles (ChemFlow) [15].

3. La métabolomique

En biologie, le développement spectaculaire depuis une vingtaine d'années de nouvelles techniques analytiques a conduit à un changement de conception des approches élémentaires pratiquées aux niveaux tissulaires et cellulaires vers des approches globales, orientées systèmes, intégrant les différents niveaux d'organisation du vivant et d'étudier ainsi la biocomplexité. Cependant, ce type d'approche à large échelle requiert une attention particulière du fait de la grande dimensionnalité des données [16], et ce à chaque étape de la démarche : sélection des sujets, collecte et stockage des échantillons, procédures standardisées pour la production des données et leur traitement [17], afin d'être capable d'extraire l'information pertinente relative à la question posée [18]. Même si les données métabolomiques présentent des caractéristiques communes aux jeux de données transcriptomiques et protéomiques, elles possèdent un grand nombre de spécificités. Les stratégies utilisées pour le traitement de ces données se sont donc appuyées sur plusieurs principes d'analyse et de visualisation de données de la génomique fonctionnelle, permettant ainsi de

combiner un contexte biologique avec l'information chimique [19]. Parmi les méthodes et les outils couramment utilisés aujourd'hui par les plateformes et les groupes de recherches, se distinguent les méthodes de prétraitement, les analyses statistiques descriptives et inférentielles des données, et les méthodes *in silico* d'annotation et d'identification des molécules. Ainsi, depuis deux décennies, la métabolomique a su tirer un bénéfice des apports des statistiques, de la bioinformatique et de l'informatique scientifique.

Les données de métabolomique sont caractérisées par un grand nombre de variables semi-quantitatives (plusieurs centaines, voire milliers) largement supérieur au nombre d'observations, de fortes colinéarités entre les variables issues d'une même molécule et la présence de bruit instrumental. Les étapes de prétraitement comprennent de nombreux filtres du bruit de fond, la correction de dérives analytiques ou d'effet de lot d'analyses. Ces étapes sont réalisées grâce à l'analyse d'échantillons qui servent de "contrôles qualité" [20]. Les données peuvent ensuite éventuellement être filtrées ou normalisées par rapport à des données biologiques. Une transformation logarithmique peut aussi être appliquée afin que les distributions des intensités mesurées se rapprochent d'une loi normale. Même en fin de prétraitement, des corrélations entre les variables sont encore présentes. Comme en chimiométrie, les méthodes statistiques multivariées telles que l'ACP, la PLS et ses variantes sont donc beaucoup utilisées en métabolomique pour analyser les données obtenues. Ces méthodes ont toutes pour but la réduction de la dimensionnalité et la création de variables synthétiques "orthogonales" [21]. Les modèles obtenus permettent ainsi de synthétiser et visualiser les données, que ce soit dans le but de détecter des regroupements pertinents d'observations, associés par exemple aux différents groupes expérimentaux, ou d'apprécier la contribution des multiples signaux mesurés à la distribution observée. La méthode PLS et ses dérivées permettent la détermination de variables qui sont prédictives soit de valeurs quantitatives, soit de l'appartenance à un groupe expérimental. Ces méthodes nécessitent l'application d'autres méthodes et l'utilisation de critères pour vérifier la qualité et la robustesse des modèles, et l'utilisation de critères pour effectuer la sélection des variables d'intérêt. Depuis quelques années, des méthodes de machine learning sont utilisées, comme les forêts aléatoires ou réseaux de neurones. Une revue des méthodes statistiques utilisées actuellement en métabolomique est disponible dans [22]. Des méthodes d'analyse multiblocs commencent aussi à être appliquées pour analyser conjointement un jeu de données métabolomiques avec d'autres blocs de données.

Concernant l'annotation et l'identification de molécules, les expérimentations basées sur les techniques de spectrométrie de masse ou de spectroscopie RMN produisent de larges volumes de données disséminées dans les nombreuses publications scientifiques résultantes. La communauté a su s'appuyer sur des ressources bioinformatiques existantes ou développer ses propres banques de données de références, nécessaires aux processus d'annotation, d'identification et d'interprétation biologiques des molécules chimiques. Ainsi, du fait de la multiplicité des formats de données disponibles et de la variabilité des outils et des stratégies de traitement disponibles, la métabolomique a su pleinement tirer profit de l'intégration des principes des flux de travail (ou "workflows") dans la structuration et l'organisation de ses analyses de données. Les principales plateformes en ligne de traitement de données telles que MetaboAnalyste ou XCMS Online [23] ou Workflow4Metabolomics (W4M) [24], sont basées sur cette approche et proposent d'une part des « pipelines experts » d'analyses clés en main, et d'autre part une "boîte à outils" permettant de construire son propre flux de travail relatif à son format de données ou à son design expérimental. Elles permettent également de faciliter l'accès à des infrastructures de calcul haute performance (HPC).

Dans ce contexte, la métabolomique nécessitant un ensemble de compétences et d'expertises très larges, la communauté française du domaine s'est structurée autour du Réseau Francophone de Métabolomique et de Fluxomique (RFMF, <http://www.rfmf.fr/>), d'une communauté active et multidisciplinaire de plateformes et de laboratoires de recherches et d'une infrastructure nationale, MetaboHUB. Cette infrastructure de métabolomique et fluxomique a été créée en 2013 dans le cadre du programme "Investissements d'Avenir", lancé par le Ministère de la Recherche et de l'Enseignement Supérieur et l'Agence Nationale pour la Recherche (ANR). MetaboHUB a pour objectif de fournir des méthodes et outils technologiques de pointe et des services en métabolomique et fluxomique aux équipes de recherche académiques et à des partenaires industriels dans les domaines de la santé, de la nutrition, de l'agriculture, de l'environnement et des biotechnologies.

4. Chimiométrie et métabolomique : une fertilisation croisée

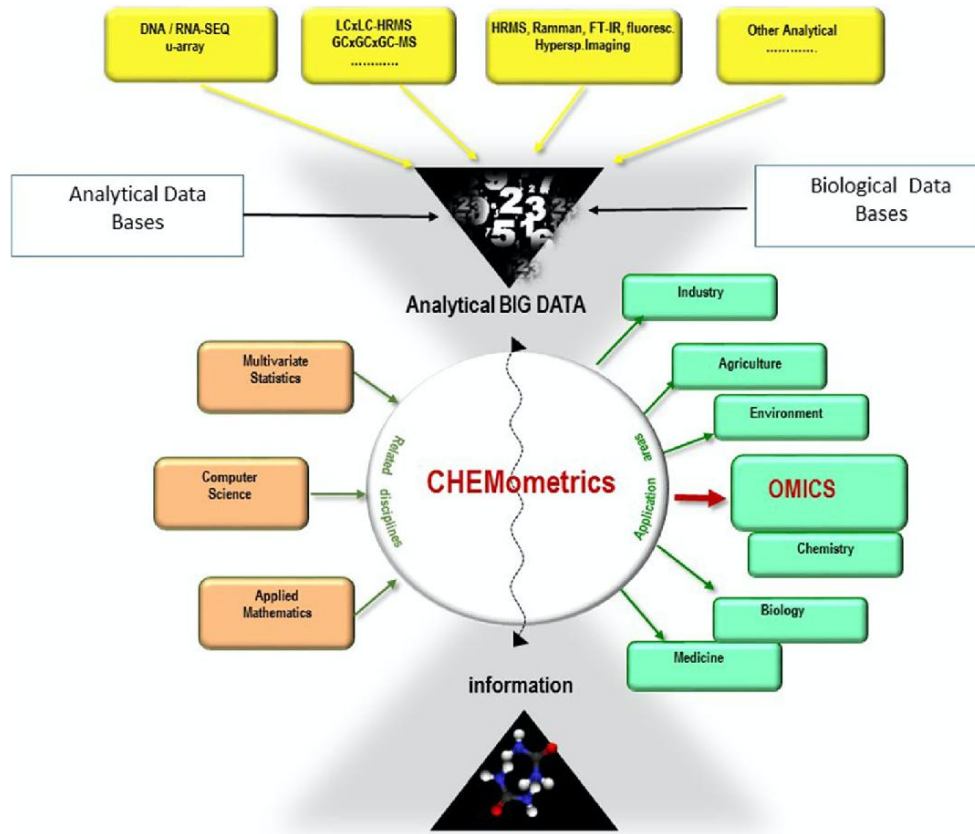


Figure 1 : Le paysage interdisciplinaire de la chimiométrie, tiré de [2]

La fertilisation croisée est un terme qui représente bien les relations entre chimiométrie et métabolomique, des enrichissements mutuels. Ainsi la chimiométrie apporte des outils et méthodes déjà conçus pour d'autres applications, et la métabolomique apporte de nouveaux objets d'étude, de nouvelles questions appelant de nouvelles méthodes.

La chimiométrie se place à l'interface des sciences numériques et des sciences du vivant (Figure 1). Historiquement, la chimiométrie vise à prédire une caractéristique difficile et / ou coûteuse à mesurer de l'échantillon ou à classer des groupes d'échantillons. En métabolomique, de nombreuses études visent à obtenir une vue d'ensemble des métabolites d'un système biologique qui expliquent des combinaisons de facteurs expérimentaux tels que le temps, le régime alimentaire, l'espèce.

Dans les cas de plans d'expériences impliquant plusieurs facteurs, les méthodes de base telles que l'ACP et la PLS peuvent s'avérer limitées. La combinaison de l'ANOVA avec les méthodes factorielles a permis de produire un ensemble de méthodes très utiles pour décrire l'impact de différents facteurs expérimentaux sur le métabolome : ANOVA simultaneous component analysis [25] (ASCA), ANOVA principal component

analysis [26] (ANOVA-PCA), ANOVA- Target Projection [27], AComDim [28], AMOPLS [29] et REP-ASCA [30].

Un aspect très important de l'analyse des données métabolomiques est l'identification des marqueurs liés à un phénomène. Du point de vue de la chimiométrie, cela veut dire être capable d'identifier les variables qui sont responsables d'une discrimination, obtenue par exemple avec un modèle PLS-DA. Les outils chimiométriques dédiés à cette tâche relèvent de la validation de modèles et de la sélection de variables. La chimiométrie offre en effet un arsenal de méthodes adaptées au problème posé par le très grand nombre de variables des jeux de données de métabolomique. Ces méthodes sont généralement couplées au calcul du modèle discriminant, comme la sparse-PLS [31], la méthode CovSel [32] ou l'analyse des VIP [33].

L'analyse longitudinale des altérations métaboliques constitue un des aspects essentiels à une meilleure compréhension de la dynamique des réseaux métaboliques. La collecte de données au cours du temps est donc répandue en métabolomique. Dans ce contexte, des solutions adaptées à l'analyse de données impliquant une composante temporelle sont nécessaires. De plus, les appareils modernes de chimie analytique, utilisés en métabolomique, fournissent très souvent des données composées de plusieurs dimensions analytiques, comme par exemple un temps de rétention, une mobilité ionique, un spectre de masse, ou des spectres RMN à deux dimensions. Le traitement de ces données appelées multivoies nécessite une prise en compte de leur structure et fait ainsi appel à des méthodes chimiométriques dédiées, comme la N-PLS [34], ou PARAFAC [35]. Par ailleurs, la diversité physico-chimique du métabolome constitue un grand défi analytique et il n'est pas possible aujourd'hui de mesurer l'ensemble des constituants de mélanges complexes comme les matrices biologiques dans le cadre d'une seule analyse. Pour faire face à cette limitation, la mise en œuvre de stratégies d'acquisition multi-plateforme est de plus en plus populaire en métabolomique. En conséquence, l'intégration de données collectées à partir de différentes sources constitue aujourd'hui un enjeu majeur pour une meilleure caractérisation des échantillons. Ce type de stratégie fournit plusieurs blocs de données et requiert donc des méthodes d'analyse multiblocs qui ont pour objectif d'offrir une synthèse de l'ensemble des données à disposition [36]. Si le cadre théorique de l'analyse multi-voies et multiblocs est bien posé, l'explosion actuelle de la dimension des données produites en métabolomique confronte la chimiométrie à des problèmes de gestion de données massives, tant horizontalement (données mega-variées) que verticalement (big-data).

Enfin, la possibilité d'effectuer des prédictions *in silico* sur la base de structures moléculaires constitue une approche efficace pour réduire le nombre de candidats et améliorer la robustesse de l'annotation des métabolites. Les outils chimiométriques comme la régression PLS permettent d'étalonner un modèle sur la base de données d'entraînement, pour estimer ensuite des paramètres comme le temps de rétention chromatographique, la mobilité ionique ou encore le profil de fragmentation MS/MS, pour des composés pour lesquels des valeurs expérimentales ne sont pas disponibles.

Du fait de ces proximités et de la complémentarité entre la chimiométrie et la métabolomique, les deux communautés se côtoient et échangent de façon régulière à travers des collaborations bilatérales et des communications ponctuelles lors des événements organisés par leurs sociétés savantes respectives. C'est à l'initiative des porteurs des deux plateformes de traitement de données en chimiométrie (ChemFlow) et en métabolomique (W4M) et suite à des échanges réguliers entre collaborateurs de ces plateformes autour du gestionnaire de flux de travail Galaxy [37], [38], puis du partage de méthodes de traitement statistiques que l'idée de rendez-vous mixtes est née. Un groupe de chercheurs et d'utilisateurs, ChemOmics, s'est alors récemment constitué [39].

5. Conclusions et Perspectives

La chimiométrie a pour but d'aider l'analyste dans le traitement des données et dans l'interprétation des résultats, au moyen de méthodes statistiques. Cet échange a permis l'essor de la chimie analytique, qui en se développant à son tour a enrichi la chimiométrie par l'apport de données toujours plus complexes, et donc le besoin de développement de nouveaux algorithmes. Classiquement appliquée à un unique tableau, la chimiométrie tend actuellement à être appliquée à l'analyse de plusieurs tableaux ou blocs de données. Le challenge consiste donc à extraire conjointement les informations contenues dans chaque bloc, ce qui pose des problèmes conceptuels, en cours de résolution.

La métabolomique est un champ récent de cette bio-chimie analytique, qui pose des problèmes particuliers d'analyse de données, auxquels la chimiométrie pourrait apporter des réponses efficaces, notamment les problèmes liés à l'intégration des données de métabolomique à des données d'autre nature, ou à la fusion de données générées par différentes plateformes analytiques rejoignent le challenge de la chimiométrie. Ainsi, l'intégration des données de métabolomique à d'autres données - omiques, qui mesurent la réponse cellulaire aux différents niveaux fonctionnels de

l'organisme, ou à d'autres données biologiques, a pour objectif d'avoir une vue globale des modifications engendrées par le stimulus. Par ailleurs, la fusion des données générées par différentes plateformes analytiques (RMN, LC-MS, GC-MS...) a pour but d'augmenter la couverture du métabolome et ainsi d'avoir une vue plus globale des "perturbations" métaboliques engendrées par le stimulus. Cette intégration, intra- ou inter-omiques, pose aujourd'hui de nombreux challenges méthodologiques.

Le groupe de chercheurs et d'utilisateurs qui s'est récemment constitué (ChemOmics) a pour objectif de faire vivre la fertilisation croisée entre la métabolomique et la chimiométrie et de jeter les bases d'une recherche méthodologique avancée, afin de lever les verrous posés par les nouvelles données à haute résolution.

6. Références

- [1] Brereton, R. G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., ... & Tauler, R. (2017). Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. *Analytical and bioanalytical chemistry*, 409(25), 5891-5899.
- [2] Brereton, R. G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., ... & Tauler, R. (2018). Chemometrics in analytical chemistry—part II: modeling, validation, and applications. *Analytical and bioanalytical chemistry*, 410(26), 6691-6704.
- [3] D.L. Massart, B.G.M.Vandeginste, S.N.Deming, Y. Michotte and L.Kaufman. *Chemometrics: a textbook.*, Elsevier, *Data Handling in Science and Technology*, Volume 2, Amsterdam 1988.
- [4] Roger, J. M., Boulet, J. C., Zeaiter, M., & Rutledge, D. N. (2020). Pre-processing Methods*. Reference Module in Chemistry, Molecular Sciences and Chemical Engineering.
- [5] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.
- [6] Benzécri, J. P. (1973). *L'analyse des données* (Vol. 2, p. I). Paris: Dunod.
- [7] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.
- [8] Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185, 1-17.
- [9] Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(3), 166-173.
- [10] Monakhova, Y. B., & Rutledge, D. N. (2020). Independent components analysis (ICA) at the "cocktail-party" in analytical chemistry. *Talanta*, 208, 120451.
- [11] Tauler, R. (1995). Multivariate curve resolution applied to second order data. *Chemometrics and intelligent laboratory systems*, 30(1), 133-146.
- [12] Groupe Français de Chimiométrie : <https://www.sfds.asso.fr/fr/chimiometrie/458-groupe-chimiometrie-gfc/>
- [13] ChemHouse : Pôle de recherches inter-instituts en chimiométrie. https://www.chemproject.org/chemhouse_fre/ChemHouse

- [14] CheMoocs : Massive Open Online Courses de chimiométrie. https://www.chemproject.org/chemhouse_fre/CheMoocs
- [15] ChemFlow : Un logiciel gratuit et ouvert, accessible à tous. https://www.chemproject.org/chemhouse_fre/ChemFlow
- [16] Brown, M., Dunn, W. B., Ellis, D. I., Goodacre, R., Handl, J., Knowles, J. D., ... & Kell, D. B. (2005). A metabolome pipeline: from concept to data to knowledge. *Metabolomics*, 1(1), 39-51.
- [17] Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., ... & Nicholls, A. W. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols*, 6(7), 1060-1083.
- [18] Alyass, A., Turcotte, M., & Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC medical genomics*, 8(1), 1-12.
- [19] Pedro, M. (2002). Emerging bioinformatics for the metabolome. *Briefings in bioinformatics*, 3(2), 134-145.
- [20] Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics*, 14(6), 72.
- [21] Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Analytica chimica acta*, 879, 10-23.
- [22] Antonelli, J., Claggett, B. L., Henglin, M., Kim, A., Ovsak, G., Kim, N., ... & Lagerborg, K. A. (2019). Statistical workflow for feature selection in human metabolomics data. *Metabolites*, 9(7), 143.
- [23] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2703878/>
- [24] <https://academic.oup.com/bioinformatics/article/31/9/1493/200011>
- [25] Smilde, A. K., Jansen, J. J., Hoefsloot, H. C., Lamers, R. J. A., Van Der Greef, J., & Timmerman, M. E. (2005). ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13), 3043-3048.
- [26] Harrington, P. D. B., Vieira, N. E., Espinoza, J., Nien, J. K., Romero, R., & Yergey, A. L. (2005). Analysis of variance—principal component analysis: A soft tool for proteomic discovery. *Analytica chimica acta*, 544(1-2), 118-127.
- [27] Marini, F., de Beer, D., Joubert, E., & Walczak, B. (2015). Analysis of variance of designed chromatographic data sets: The analysis of variance-target projection approach. *Journal of Chromatography A*, 1405, 94-102.
- [28] Bouveresse, D. J. R., Pinto, R. C., Schmidtke, L. M., Locquet, N., & Rutledge, D. N. (2011). Identification of significant factors by an extension of ANOVA–PCA based on multi-block analysis. *Chemometrics and Intelligent Laboratory Systems*, 106(2), 173-182.
- [29] Boccard, J., & Rudaz, S. (2016). Exploring Omics data from designed experiments using analysis of variance multiblock Orthogonal Partial Least Squares. *Analytica Chimica Acta*, 920, 18-28.
- [30] Ryckewaert, M., Gorretta, N., Henriot, F., Marini, F., & Roger, J. M. (2020). Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Analytica Chimica Acta*, 1101, 23-31.

- [31] Lê Cao, K. A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics*, 12(1), 253.
- [32] Roger, J. M., Palagos, B., Bertrand, D., & Fernandez-Ahumada, E. (2011). CovSel: Variable selection for highly multivariate and multi-response calibration: Application to IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 106(2), 216-223.
- [33] Favilla, S., Durante, C., Vigni, M. L., & Cocchi, M. (2013). Assessing feature relevance in NPLS models by VIP. *Chemometrics and Intelligent Laboratory Systems*, 129, 76-86.
- [34] Bro, R. (1996). Multiway calibration. multilinear PLS. *Journal of chemometrics*, 10(1), 47-61.
- [35] Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3), 283-319.
- [36] Boccard, J., & Rutledge, D. N. (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica chimica acta*, 769, 30-39.
- [37] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691752/>
- [38] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6030816/>
- [39] ChemOmics, un groupe de chercheurs et d'utilisateurs pour le développement de la chimiométrie en métabolomique : <https://chemproject.org/ChemOmics>