

Offre de stage, 6 mois, 2023, étudiant(e) en Master 2 bio-informatique/statistique appliquée.

Titre du stage : Discrimination en présence de plusieurs blocs multidimensionnels de données. Application à l'identification de marqueurs hépatiques et plasmatiques de l'exposition d'animaux d'élevage aux polychlorobiphényles (PCBs).

Mots-clés : ACP, PLS, MSFA, SCA, PLS-DA, volatolomique, métabolomique

Laboratoire d'accueil : Unité de Statistique, Sensométrie et Chimiométrie (StatSC) Oniris-INRAE

L'unité de Statistique, Sensométrie et Chimiométrie d'Oniris est une unité sous contrat avec l'INRAE. Une des principales thématiques de recherche de l'unité est l'analyse simultanée de plusieurs tableaux de données multidimensionnelles dans un contexte non supervisé ou supervisé. L'objectif général des approches mises en œuvre sont la réduction de la dimensionnalité des données ainsi que la compréhension des relations entre différents types d'informations (intégration ou fusion de données).

Contexte, données et méthodologie envisagée

Le stage s'inscrit dans le cadre d'un projet de recherche ANR nommé SENTINEL [1]. C'est un projet de dimension nationale qui vise, entre autre, à combiner des techniques volatolomiques [2] à des techniques ciblées haut-débit par RMN des marqueurs non volatils d'exposition aux PCBs [3] afin de d'identifier et mesurer rapidement des marqueurs de la contamination de la chaîne alimentaire dans les tissus ou fluides animaux.

Dans cette logique, 88 poulets de chair ont été nourris pendant 42 jours avec de l'aliment contaminé artificiellement ou avec de l'aliment non-contaminé (groupe témoin). Les poulets (individus) répartis en 7 groupes ont reçu les quantités suivantes de PCBs via l'alimentation :

- 1 groupe « témoin » : aliment non-contaminé
- 3 groupes Aroclor 1260 : 2 ; 20 ; 100 µg/kg
- 1 groupe 6PCB-nDL seuls (somme) : 10 µg/kg
- 2 groupes 12 PCB-DL seuls (somme) : 0,15 ; 0,75 ng/kg

L'Aroclor 1260 est un mélange de 209 PCBs avec un effet toxique majoritaire des PCB-nDL. Le mélange des 6 PCB-nDL est utilisé pour évaluer si la réponse à l'exposition à l'Aroclor 1260 est, comme attendu, induite uniquement par les PCB-nDL. Le mélange des 12 PCB-DL est utilisé pour évaluer si la volatolomique permet de distinguer l'impact sur la réponse d'une exposition aux deux types de PCBs : DL et nDL.

Les données, collectées et prétraitées pour l'analyse, se présentent sous la forme de deux blocs de variables (volatolomique et métabolomique) mesurées sur un même ensemble de 88 individus structurés en 7 groupes. L'objectif du stage est double : (i) identifier pour chaque bloc de variables (volatolomique et métabolomique) les biomarqueurs discriminants des 7 groupes et (ii) évaluer la complémentarité informative des deux blocs de variables (volatolomique et métabolomique) dans la discrimination des 7 groupes.

Trois méthodes statistiques sont envisagées pour répondre à l'objectif : (i) SCA (Simultaneous Component Analysis[4]), MSFA (Multi Study Factor Analysis[5]) et PLS-DA (Partial Least Squares Discriminant Analysis[6]). La mission du stagiaire portera sur l'étude de ces trois méthodes, leurs mises en œuvre et l'analyse de leurs performances. La mise en œuvre des méthodes sera réalisée via des packages implémentés en langage R. Par ailleurs, le stagiaire devra assurer la traçabilité de l'ensemble des analyses réalisées avec des programmes facilement réutilisables.

Encadrant du stage : Mohamed Hanafi

Co-encadrant : Benjamin Mahieu

Offre de stage, 6 mois, 2023, étudiant(e) en Master 2 bio-informatique/statistique appliquée.

Localisation : StatSC, ONIRIS, site des Sciences de l'Alimentation, CS 82225, 44322 NANTES cedex 3
Mode de recrutement : Entretien
Candidature : Curriculum vitæ et lettre de motivation à adresser à : <ul style="list-style-type: none">- mohamed.hanafi@oniris-nantes.fr- benjamin.mahieu@oniris-nantes.fr
Compétences attendues : <p>Issu d'une formation en statistique ou bio-informatique (niveau M2), les compétences scientifiques et techniques attendues sont :</p> <ul style="list-style-type: none">- Maîtrise des méthodes statistiques multivariées dans un contexte exploratoire et/ou prédictif- Intérêt pour le caractère aussi bien conceptuel qu'applicatif des méthodes dans le contexte de l'étude- Maîtrise du langage R- Capacité à présenter ses travaux et restituer ses résultats du point de vue triple : méthodologique, logiciel et applicatif- Capacité à interagir dans un contexte pluridisciplinaire
Références. <p>[1] https://anr.fr/Projet-ANR-19-CE21-0011</p> <p>[2] https://tel.archives-ouvertes.fr/tel-01980862/</p> <p>[3] https://new.societechimiquedefrance.fr/wp-content/uploads/2019/12/2012-364-365-juin-juil.-aout-p59-elena-hd.pdf</p> <p>[4] Timmerman ME, Kiers HAL: Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intra individual and inter individual differences. <i>Psychometrika</i> 2003, 68(105–121).</p> <p>[5] De Vito R, Bellio R, Trippa L, Parmigiani G. Multi-study factor analysis. <i>Biometrics</i>. 2019; 75(1):337-346. doi: 10.1111/biom.12974.</p> <p>[6] Svante Wold, Michael Sjöström, Lennart Eriksson. PLS-regression: a basic tool of chemometrics, <i>Chemometrics and Intelligent Laboratory Systems</i>, Volume 58, Issue 2, 2001, Pages 109-130, https://doi.org/10.1016/S0169-7439(01)00155-1.</p>