

PROJET DE FIN D'ETUDES / STAGE M2

Reproductibilité des Résultats Scientifiques en Spectroscopie par Résonance Magnétique In Vivo

Durée

4 à 6 mois

Contacts

Gaël VILA : gael.vila@creatis.insa-lyon.fr

Hélène Ratiney : helene.ratiney@creatis.insa-lyon.fr

Sorina POP : sorina.pop@creatis.insa-lyon.fr

Contexte Scientifique

Un Enjeu Fondamental : la Reproductibilité des Résultats

La majorité des sciences du vivant traverse actuellement une crise de reproductibilité. La réplication des études de l'état de l'art, dans la même configuration et sur les mêmes données expérimentales, échoue fréquemment à produire les mêmes résultats. Du point de vue d'un calcul scientifique, versatilité peut s'observer sur trois niveaux (voir aussi la Figure 1).

- i. Deux **méthodes** d'analyse différentes (ex. : modèles) peuvent générer des conclusions différentes lorsqu'on teste une même hypothèse sur un même jeu de données [1].
- ii. Deux **implémentations** différentes (ex. : codes source) peuvent générer des résultats différents, lorsqu'elles appliquent la même analyse à un même jeu de données [2]. Ce constat est aussi vrai pour deux versions d'un même code.
- iii. Deux **environnements** de calcul différents (ex. : OS, bibliothèques...) peuvent générer des valeurs différentes, lorsqu'on exécute un même code sur un même jeu de données [3]. D'une machine à l'autre, même un algorithme parfaitement déterministe peut ainsi générer deux résultats distincts.

Alors que nous entrons dans l'ère de la science ouverte, la reproductibilité des études sur le long terme est un enjeu fondamental pour la production d'analyses fiables et pérennes.

Un Cas d'Usage : la Spectroscopie par Résonance Magnétique

La spectroscopie par résonance magnétique (SRM) est une technique d'imagerie médicale qui exploite les propriétés magnétiques des tissus biologiques pour en déterminer la composition chimique. En particulier, on peut utiliser le signal de résonance magnétique pour estimer *in vivo* la concentration de certains métabolites (ex. : neurotransmetteurs, acides aminés) à un certain endroit de l'organisme (ex. : une région du cerveau).

Pour cela, on décompose le spectre du signal de résonance en éléments simples :

$$\hat{S} = \sum_i \alpha_i * S_i$$

(modèle simplifié), où \hat{S} est une estimée du spectre original, S_i est le spectre du métabolite i et α_i reflète la concentration de ce métabolite dans la région étudiée. Le processus est illustré dans la Figure 1 ci-dessous, avec les trois niveaux de variabilité mentionnés plus haut.

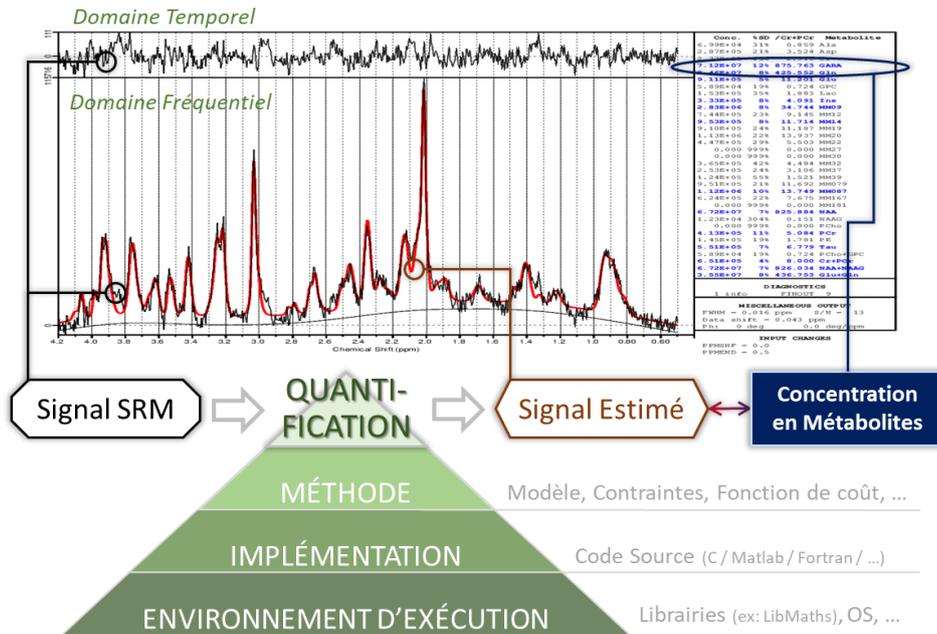


Figure 1. Un résultat scientifique : estimation d'une concentration en métabolites par SRM.

Il n'existe pas de solution unique, ni de méthode idéale pour estimer \hat{S} . Le résultat de la Figure 1 est donc potentiellement variable aux niveaux (i), (ii) et (iii). Dans le cadre du projet ANR ReproVIP ([ANR-21-CE45-0024-01](#)), on cherche à déterminer des intervalles de confiance pour détecter les résultats « faiblement » reproductibles, à l'issue d'un algorithme de quantification. Comme généralement dans un contexte clinique, il existe un grand nombre de sources de variabilité, dont une majeure est la qualité des signaux fournis en entrée. Pour déterminer ce qu'est un « bon » résultat en SRM, il faut donc être capable de diagnostiquer en amont les « bons » signaux par un processus automatique et interprétable.

Sujet de Stage

Objectif – L'objet du stage est de déterminer dans quelle mesure un signal de SRM permet une estimation reproductible de la concentration en métabolites.

Travail existant – Deux algorithmes de quantification (LCModel, CQuest) sont déjà disponibles pour estimer les concentrations en métabolites. Des indicateurs existent également pour *estimer* (avec une incertitude fondée sur la théorie de Cramér-Rao), puis *mesurer* en pratique la reproductibilité des valeurs en sortie de ces algorithmes [4].

Travail à réaliser – Pour évaluer la qualité des signaux en entrée, on pourra utiliser des indicateurs connus (ex. : rapport signal-sur-bruit) ou proposer d'autres indicateurs pertinents. On aura recours à l'analyse de données (statistiques / apprentissage) pour modéliser la variabilité des sorties en fonction de ces indicateurs. On s'intéressera aux variations inter-exécutions (dues aux processus aléatoires), inter-modèles (LCModel, CQuest), inter-environnements (changements de machine). Ce travail s'appuiera sur un ensemble de signaux bruités acquis dans le cadre d'une étude sur le petit animal. On pourra recourir à la simulation pour étendre la base de signaux disponible, ou pour caractériser un phénomène de façon contrôlée (en enlevant notamment les sources de variations expérimentales ou biologiques).

Extension possible – S'il reste du temps, appliquera la même problématique de recherche à d'autres applications médicales, comme la segmentation de tumeurs cérébrales.

Compétences requises

- ❖ Connaissances en traitement du signal / en sciences des données ;
- ❖ Compétences en programmation scientifique (idéalement Python / R) ;
- ❖ Connaissances ou intérêt porté sur l'imagerie médicale ;
- ❖ Capacité à s'intégrer dans une équipe multiculturelle.

Environnement de travail

Le/La stagiaire sera accueilli.e au laboratoire CREATIS, au sein de la plate-forme VIP. VIP (Virtual Imaging Platform) est un portail web pour la simulation et le traitement de données massives en imagerie médicale (<https://vip.creatis.insa-lyon.fr>). Gratuite et open-source, la plateforme compte actuellement plus de 1400 utilisateurs et propose une vingtaine d'applications à destination du monde académique. Elle exploite les ressources de l'*European Grid Infrastructure* (EGI) pour proposer aux utilisateurs des calculs de haute performance.

Le/La stagiaire interagira avec les ingénieur.e.s de VIP et les autres membres du laboratoire. Il/Elle travaillera dans un cadre bienveillant, sur un projet ambitieux visant à accompagner la recherche scientifique en imagerie médicale vers les standards de l'*Open Science*.

Références

- [1] R. Botvinik-Nezer *et al.*, « Variability in the analysis of a single neuroimaging dataset by many teams », *Nature*, vol. 582, n° 7810, p. 84-88, juin 2020, doi: 10.1038/s41586-020-2314-9.
- [2] A. Bowring, C. Maumet, et T. E. Nichols, « Exploring the impact of analysis software on task fMRI results », *Hum. Brain Mapp.*, vol. 40, n° 11, p. 3362-3384, août 2019, doi: 10.1002/hbm.24603.
- [3] T. Glatard *et al.*, « Reproducibility of neuroimaging analyses across operating systems », *Front. Neuroinformatics*, vol. 9, avr. 2015, doi: 10.3389/fninf.2015.00012.
- [4] L. T. Riemann *et al.*, « Assessment of measurement precision in single-voxel spectroscopy at 7 T: Toward minimal detectable changes of metabolite concentrations in the human brain in vivo », *Magn. Reson. Med.*, vol. 87, n° 3, p. 1119-1135, 2022, doi: 10.1002/mrm.29034.