



## Computational Reproducibility

There is no unique solution for the estimate  $\hat{S}$ , because MR signal quantification is a nonlinear problem involving many parameters. Therefore, *the amount of some metabolite ( $\alpha_m$ ) estimated for a given signal ( $S$ ) may vary from one computation to another*, when trying to estimate it multiple times. Like many medical applications, MR signal quantification faces a challenge in the **reproducibility** of results.

There are a large number of sources of variability for the results of a quantification run. Among these sources, **signal quality** has a critical impact: the noisier the signal ( $S$ ), the harder it will be for the algorithm to produce a reliable estimate ( $\hat{S}$ ) and the more likely metabolite concentrations ( $\alpha_m$ ) will vary between two estimation attempts.

Based on such indicators as signal quality, the aim of this internship is to identify which type(s) of MR signal(s) allow(s) a "reproducible" quantification of various metabolites.

In the [ReproVIP](#) project framework, we are not interested in the quantification algorithm itself (here considered as a black box). We seek to establish how far the *same calculation*, when repeated on the *same data*, will reproduce the *same results*. This is seldom the case for complex models (such as quantification), which compromises the ability of other people to reuse these models in different settings. Thus, the reproducibility of computations is a major scientific challenge for the next decade; on the one hand, to transfer models to the real world; and on the other hand, to spread them in the academic world according to Open Science standards.

## Internship Subject

**Objective** – Identify the link between the *quality* of a magnetic resonance signal and the *variability* of the results of a quantification algorithm.

**Main Tasks** – Some metrics have already been proposed to measure the variability of results. Similarly, the quality of a signal can be measured by indicators such as the signal-to-noise ratio or the peak width in the Fourier transform. As a first step, the trainee will propose algorithms to compute these indicators (or others that will seem useful) from an MR signal. In a second step, he or she will propose relevant criteria to define a set of signals that lead to a "reproducible" quantification of metabolites. These criteria will be implemented in a decision algorithm (e.g., classifier) capable of identifying new signals in an unknown database.

This work will be based on an existing quantification algorithm (cQUEST) and a collection of MR signals acquired during a study on small animals. Simulation (e.g., signal synthesis) can be used to characterize a phenomenon in a controlled way (for example, by removing biological sources of variation).

**Optional extension** – If there is time left, we will investigate the variations of the results between two different quantification algorithms (cQUEST & LCMoDel), then between two computing environments (change of machine or operating system).

## Required skills

- ❖ Knowledge of signal processing, statistics and/or machine learning;
- ❖ Scientific programming skills (ideally Python / R);
- ❖ Knowledge or interest in medical imaging.

## Working Environment

The trainee will be hosted at CREATIS lab (<https://www.creatis.insa-lyon.fr/site7/fr>), within the VIP platform. VIP (the Virtual Imaging Platform) is a web portal for the simulation and processing of massive data in medical imaging (<https://vip.creatis.insa-lyon.fr>). The platform is free and open-source; it currently has more than 1400 users and offers about twenty applications to the academic world. It uses computing and storage resources throughout Europe to provide users of the platform with high-performance calculations.

The intern will interact with VIP engineers and other laboratory members. S/he will work in a friendly environment, on an ambitious project aiming to lead medical imaging research towards Open Science standards.

Stage Recherche – Master 1 (Version Française)

## Vers des Calculs Reproductibles en Spectroscopie par Résonance Magnétique

### Durée

8 à 14 semaines

### Contacts

Gaël VILA : [gael.vila@creatis.insa-lyon.fr](mailto:gael.vila@creatis.insa-lyon.fr)

Hélène Ratiney : [helene.ratiney@creatis.insa-lyon.fr](mailto:helene.ratiney@creatis.insa-lyon.fr)

Sorina POP : [sorina.pop@creatis.insa-lyon.fr](mailto:sorina.pop@creatis.insa-lyon.fr)

## Contexte Scientifique

### Quantification d'un Signal de Résonance Magnétique

La spectroscopie par résonance magnétique (RM) est une technique d'imagerie médicale qui cherche à déterminer la composition chimique d'un tissu biologique en exploitant ses propriétés magnétiques. Cette « biopsie virtuelle » fournit une mesure *in vivo* de la concentration de certains biomarqueurs dans l'organisme, par exemple pour le suivi de maladies neurodégénératives comme la sclérose en plaques.

On s'intéresse au spectre (transformée de Fourier) du signal RM, dont les pics reflètent la concentration de certains métabolites (neurotransmetteurs, acides aminés) dans les tissus observés. Pour calculer ces concentrations, la **quantification** d'un signal  $S$  consiste à décomposer son spectre en éléments simples, comme dans le modèle simplifié suivant :

$$\hat{S} = \sum_m \alpha_m * S_m,$$

où  $\hat{S}$  est une estimation du signal mesuré ;  $S_m$  est le spectre de résonance du métabolite  $m$  ; et  $\alpha_m$  reflète la concentration de ce métabolite dans le tissu observé. Ce processus de quantification est illustré dans la Figure 1 ci-dessous.

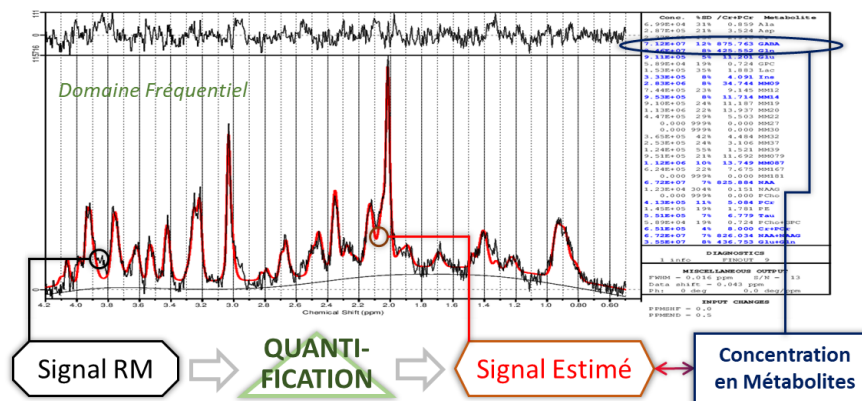


Figure 1. Quantification d'un signal en spectroscopie RM.

## Qualité du Signal et Reproductibilité des Calculs

Il n'existe pas de solution unique pour l'estimée  $\hat{S}$ , car la quantification du signal RM est un problème non-linéaire qui implique de nombreux paramètres. Par conséquent, la quantité d'un métabolite ( $\alpha_m$ ) obtenue pour un signal donné ( $S$ ) *peut varier d'un calcul à l'autre*, lorsqu'on cherche à l'estimer plusieurs fois. Comme beaucoup d'applications médicales, la quantification du signal RM fait face à un problème de **reproductibilité** des résultats.

Il existe un grand nombre de sources de variabilité pour ces résultats. Parmi ces sources, la **qualité** du signal a une influence décisive : plus le signal ( $S$ ) sera bruité, plus il sera difficile pour l'algorithme d'aboutir à une estimée fiable ( $\hat{S}$ ) et plus les concentrations ( $\alpha_m$ ) seront susceptibles de varier entre deux tentatives d'estimation.

L'objectif de ce stage est de déterminer ce qu'est un signal de « bonne qualité » ; c'est-à-dire permettant une estimation « reproductible » de la concentration d'un métabolite.

Dans le cadre du [projet ReproVIP](#), on ne s'intéresse pas à l'algorithme de quantification lui-même (ici considéré comme une boîte noire). On cherche à déterminer dans quelle mesure un *même calcul*, lorsqu'il est répété sur les *mêmes données*, reproduira toujours les *mêmes résultats*. C'est rarement le cas pour les calculs complexes (comme la quantification), ce qui compromet l'utilisation des modèles par d'autres personnes dans d'autres contextes. Ainsi, la reproductibilité des calculs s'impose comme un enjeu scientifique majeur de la prochaine décennie ; d'une part, pour transférer les modèles dans le monde réel ; et d'autre part, pour les diffuser dans le monde académique selon les standards de la Science Ouverte.

## Sujet de Stage

**Objectif** – Caractériser le lien entre la *qualité* d'un signal de résonance magnétique et la *variabilité* des résultats d'un algorithme de quantification.

**Travail à réaliser** – Certaines métriques ont déjà été proposées pour mesurer la variabilité des résultats. De même, la qualité d'un signal peut se mesurer à partir d'indicateurs comme le rapport signal sur bruit ou la largeur à mi-hauteur des pics dans la transformée de Fourier.

Dans un premier temps, le ou la stagiaire proposera des algorithmes qui calculeront ces indicateurs (ou d'autres qui lui sembleront utiles) à partir d'un signal RM. Dans un second temps, il ou elle proposera des critères pertinents pour définir un ensemble de signaux qui aboutissent à une quantification « reproductible » des métabolites.

Ce travail s'appuiera sur un algorithme de quantification existant (cQUEST) et sur un ensemble de signaux RM acquis dans le cadre d'une étude sur le petit animal. On pourra recourir à la simulation (synthèse de signaux) pour caractériser un phénomène de façon contrôlée (en enlevant par exemple les sources de variation biologiques).

**Extension possible** – S’il reste du temps, on s’intéressera aux variations des résultats entre deux algorithmes de quantification différents (cQUEST & LCMoel), puis entre deux environnements de calcul (changement de machine ou de système d’exploitation).

## Compétences requises

- ❖ Connaissances en traitement du signal et en statistiques ;
- ❖ 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 ou La stagiaire sera accueilli.e au laboratoire CREATIS (<https://www.creatis.insa-lyon.fr/site7/fr>), au sein de la plate-forme VIP. VIP (*the 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 des ressources de calcul et de stockage partout en Europe pour proposer aux utilisateurs de la plate-forme des calculs de haute performance.

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