



HAL
open science

L'intégration des signaux (incluant des résonances) dans les spectres unidimensionnels obtenus par spectroscopie de résonance magnétique quantitative in situ

Hervé This, Vo Kientza, Marc Bria, Laura Febvay

► To cite this version:

Hervé This, Vo Kientza, Marc Bria, Laura Febvay. L'intégration des signaux (incluant des résonances) dans les spectres unidimensionnels obtenus par spectroscopie de résonance magnétique quantitative in situ. Le Cahier des Techniques de l'INRA, 2021. hal-03156390v2

HAL Id: hal-03156390

<https://hal.inrae.fr/hal-03156390v2>

Submitted on 21 Dec 2021

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

L'intégration des signaux (incluant des résonances) dans les spectres unidimensionnels notamment obtenus par spectroscopie de résonance magnétique quantitative *in situ*

Hervé This ^{1*}, Marc Bria ², Laura Febvay ³

Résumé. Le dosage de composés présents dans les ingrédients alimentaires ou dans les aliments, par spectroscopie de résonance magnétique nucléaire (RMN), nécessite l'intégration de résonances parfois déformées dans des spectres dont la ligne de base peut être imparfaite, surtout pour les analyses *in situ*, sans préparation des échantillons. Or les logiciels commerciaux de traitement des spectres RMN donnent des résultats différents, même pour des résonances isolées des spectres. Ici on présente des feuilles de calcul (données en Matériels supplémentaires en *Maple* et en *pdf*) qui permettent des intégrations des signaux sans imposer de choix arbitraires des limites d'intégration, et, surtout, en permettant de suivre clairement les opérations effectuées.

Mots-clés :

spectroscopie de résonance magnétique nucléaire, RMN, NMR, integration, *is q* NMR, traitement du signal, intégration, dosage

Résumé.

Using nuclear magnetic resonance spectroscopy in order to determine quantitatively the compounds present in food ingredients or in food calls for the integration of resonances that are sometimes distorted, in spectra whose baseline can be imperfect, especially with *in situ* analyses, for which there is no preparation of samples. Commercial softwares for NMR spectra treatments give very different results even for resonances that are isolated. Here we present spreadsheets (given in Supplementary Materials) that integrate automatically signals without imposing any arbitrary choice of integration limits and explaining the details of calculations being made.

Keywords :

nuclear magnetic resonance spectroscopy, NMR, integration, *is q* NMR, signal treatment, dosage

Introduction

La spectroscopie de résonance magnétique (RMN, en anglais NMR) permet l'identification et le dosage de composés organiques dans des solutions ou des gels complexes (Gil *et al.*, 2000). Les dosages par RMN sont souvent moins sensibles que par les méthodes chromatographiques (Belton *et al.*, 1998 ; Noteborn *et al.*, 2000 ; Ward *et al.*, 2003), mais la RMN quantitative du proton (^1H NMR), même avec des sondes classiques (par

¹ Université Paris-Saclay, INRAE, AgroParisTech, UMR 0782 SayFood, 75005, Paris, France. Tel : 01 44 28 72 90.

Auteur pour correspondance : herve.this@inrae.fr

² Université Lille, CNRS, INRAE, Centrale Lille, ENSCL Université Artois, FR 2638-Institut Michel-Eugène Chevreul (IMEC), F-5900 Lille

³ Aerial, Strasbourg, France.

opposition aux cryosondes), permet de détecter, rapidement (en quelques minutes) et simultanément, un grand nombre de métabolites présents dans un milieu complexe, quand leur concentration est supérieure à environ 0,1 mg par g d'échantillon (pour une centaine d'acquisitions, avec un champ statique de 9.4 T) (Fan, 1996). Les spectres enregistrés peuvent ensuite être traités en vue d'études métabolomiques (Yue, 2017).

Les analyses par $q\ ^1\text{H}$ NMR se sont imposées pour des nombreuses applications scientifiques, technologiques ou techniques (Hatzakis, 2019), notamment pour l'exploration des ingrédients alimentaires (authentification, recherche de contaminants, etc.), pour l'étude de la production d'aliments (notamment gastronomie moléculaire ou suivi des effets des procédés industriels) (This, 2021) ou pour le contrôle des aliments (composés néoformés, par exemple). Pour la RMN dans le domaine fréquentiel, où les spectres résultent d'une transformation de Fourier appliquée au signal temporel de relaxation, le choix des spectromètres, caractérisés par des champs constants de différentes intensités, permet la détection soit de petites molécules en solution, soit de composés plus gros (oligomères, polymères) dans une matrice solide. Toutefois la préparation des échantillons, pour des analyses classiques, ralentit parfois notablement les travaux (quelques jours pour une extraction, suivie d'une double lyophilisation dans un solvant deutéré, par exemple) et peut conduire à des erreurs de dosage, quand les effets de matrice gênent les extractions. Par exemple, pour le dosage des mono- et oligo-saccharides dans les tissus végétaux, une méthode optimisée de détermination a été mise au point par Davis *et al.* (2007) sur la base des travaux de O'Donoghue (O'Donoghue *et al.*, 2004) et d'autres (Viola et Davies, 1992). Dans cette « méthode O'Donoghue modifiée » (MOD), les tissus végétaux sont d'abord lyophilisés, puis chauffés à reflux dans un mélange méthanol/eau, (62.5:37.5, w:w) pendant 15 min à 55 °C ; après filtration, évaporation du solvant et lyophilisation, le produit obtenu est analysé par diverses techniques analytiques, notamment la $q\ ^1\text{H}$ NMR. Dans cette dernière technique, le solide obtenu était naguère dissous dans du D_2O , puis relyophilisé, avant d'être redissous (Cazor *et al.*, 2006 ; Tardieu *et al.*, 2009). Par cette méthode, les analyses étaient longues, détruisaient les échantillons et imposaient des manipulations de solvants toxiques. En outre, les résultats obtenus par ces analyses diffèrent statistiquement (et l'on détecte moins de saccharides) de ceux qui sont obtenus par RMN quantitative *in situ* (is q NMR), où les tissus végétaux sont directement placés dans le tube d'analyse, et explorés par RMN liquide haute résolution dans le domaine fréquentiel (Weberskirch *et al.*, 2011).

Ces méthodes *in situ* évitent ainsi bien des inconvénients des dosages classiques. Elles sont fondées sur le fait que l'eau est abondante dans les tissus végétaux. Par exemple, pour des racines de carotte (*Daucus carota* L.), sa proportion atteint est environ 88 % de la masse fraîche (Bartolome et Hoff, 1972) : dans ces tissus, comme dans ceux des autres végétaux, le cytoplasme des cellules peut être considéré comme un gel (Noteborn *et al.*, 2000), les métabolites et les ions étant en solution dans un environnement liquide décrit par la formule DSF $\text{D}_0(\text{W})/\text{D}_3(\text{S})$ (This, 2012) ; quant aux vaisseaux (xylème, phloème), ils renferment respectivement des ions et des métabolites en solution dans la sève brute et élaborée (Campbell, 1995).

C'est parce que nous avons observé que les tissus végétaux ou animaux étaient formellement des gels (IUPAC, 2007 ; This, 2012) que nous avons eu l'idée d'étudier ces tissus directement par RMN du liquide, sans aucune autre préparation qu'une découpe permettant de les placer dans les tubes RMN. Initialement nous avons testé la méthode avec des échantillons de carotte, mais les tests ont été étendus à de nombreux autres systèmes (vin, gels laitiers à base blanche de type yaourt, gels de gélatine, tissus animaux, café, etc.). Des études *in situ* avaient été mises en œuvre par le passé (Monteiro-Marques *et al.*, 1991a ; 1991b ; 1992), mais il s'agissait d'étudier l'état d'hydratation des échantillons, plutôt que d'y identifier des métabolites à doser, ou des solutés pour le cas de gels. Au début de nos études *in situ*, les échantillons étaient placés dans les tubes RMN, immergés dans une solution d'un composé de référence dans du D_2O (Weberskirch *et al.*, 2012) ; ce dernier solvant était utilisé pour le calage en champ (« lock »), et le composé de référence était le sel sodique de l'acide (2,2,3,3-d4-triméthylsilyl)propionique (TSP). Des résultats équivalents sont obtenus quand cette solution de TSP dans du D_2O est dans un tube capillaire scellé, introduit à côté de l'échantillon, dans le tube RMN. Pour ce

dernier cas, le TSP du capillaire est préalablement étalonné à l'aide d'une solution précisément connue de phtalate de potassium, et, surtout, c'est le même capillaire qui sert à établir les courbes d'étalonnage des composés que l'on veut doser et à analyser les échantillons contenant des concentrations inconnues de ces composés (les acquisitions RMN étant alors faites dans les mêmes conditions, notamment pour ce qui concerne la relaxation). Parfois un séchage préalable des échantillons améliore la qualité des spectres obtenus, mais l'acquisition dans l'aimant peut durer moins de 10 minutes (pour un petit nombre d'acquisitions), et, en tout cas, on n'utilise plus ni solvant organique ni D₂O, ce dernier (quelques microlitres) étant placé une fois pour toutes dans le tube capillaire scellé ; l'échantillon est alors soit dans de l'eau, soit sans liquide environnant.

Tout cela étant posé, il reste la question essentielle du traitement des spectres obtenus, dont la ligne de base n'est pas toujours horizontale, ni confondue avec l'axe des abscisses (déplacements chimiques), en raison des fortes perturbations du champ magnétique dans le tube d'analyse, perturbations dues notamment (mais pas seulement) à la présence simultanée du tube capillaire et de l'échantillon. On n'observe pas les massifs élargis que l'on pouvait redouter (quand des particules sont en dispersion liquide), mais une distorsion de la ligne de base impose un traitement soigneux des signaux quand on cherche à utiliser les spectres pour des dosages.

D'ailleurs on peut également observer que, de même, l'intégration des signaux ne peut pas se faire facilement en routine, quand on dose de très petites quantités de produits (éventuellement à la limite de quantification) et que, de ce fait, les signaux deviennent du même ordre de grandeur que le bruit. Évidemment des acquisitions avec des nombres de balayages supérieurs permettent d'augmenter le rapport signal sur bruit, mais il n'en demeure pas moins que pousser la méthode à ses limites nécessite des intégrations soigneuses des signaux (que l'on espère être des résonances) correspondant aux composés que l'on veut doser, notamment pour les composés présents dans des mélanges complexes, tels que sont les aliments (Wishart, 2008). Parfois le recouvrement partiel de résonances complique l'intégration, et des décompositions sont indispensables (Bingol et Bruschiweiler, 2011), sans compter que les couplages doivent être pris en compte (surtout pour les complexes de résonance tels qu'en engendrent les saccharides), ainsi que les décalages des signaux en raison des variations de pH d'échantillon en échantillon (Fan, 1996 ; Aisala *et al.*, 2016) ou de la présence d'espèces paramagnétiques. Au total, l'analyse des spectres RMN peut être difficile et longue (Liu *et al.*, 2017).

Couramment les traitements de spectres sont effectués à l'aide de logiciels tels que *Xwin-NMR*, *TopSpin*, *MestReNova*, *NMRnotebook*, et d'autres. Partant des signaux de relaxation libre (*free induction decay*, FID), ces logiciels demandent des interventions des utilisateurs, qui peuvent être à l'origine de biais (Meyer *et al.*, 1988), même avec les programmes les plus récents (Mestrelab Research, 2017). Certes les nouveaux logiciels effectuent automatiquement ou semi-automatiquement des transformations de Fourier, pour produire les spectres dans le domaine fréquentiel, des décompositions de signaux, des réglages de la phase et de la ligne de base, des déconvolutions ou des intégrations, mais les auteurs de ces logiciels admettent eux-mêmes les limites de leurs systèmes. Par exemple, sous le titre *fully automatic processing*, on peut lire : « *Once the FID is opened Mnova will automatically and optimally process it using the information it discovered in Part 1, so you do not need to bother with processing details* » (traduction personnelle : une fois que le FID est ouvert, Mnova le traite automatiquement et de façon optimale en utilisant l'information trouvée en Partie 1, de sorte que vous n'avez pas à vous préoccuper des détails du traitement)... mais les mêmes auteurs enchaînent : « *The two steps outlined above will successfully process most NMR data. However, if you are not completely satisfied with the result you can, at anytime, reprocess using your own choice of processing parameters, changing or adjusting the window function, the FT, the phasing and baseline correction to meet your requirements* » (traduction personnelle : Les deux étapes décrites ci-dessus parviennent à traiter la plupart des données RMN avec succès. Cependant, si vous n'êtes pas complètement satisfait du résultat que vous obtenez, à tout moment, vous pourrez re-traiter avec les paramètres de votre choix, en changeant ou en ajustant les fonction fenêtre, la transformation de Fourier, le phasage ou la ligne de base, selon vos besoins) : cette seconde phrase est une réfutation de la première,

d'autant que les mêmes auteurs reconnaissent que « *every peak in an 1H-NMR spectrum is basically an envelope of a large number of transitions and its shape is dominated by the coupling pattern of the spin system. Even in molecules of modest size the number of distinct peaks is tens to thousands times smaller than that of quantum transitions.* » (traduction personnelle : chaque pic dans un spectre 1H-NMR est fondamentalement une enveloppe d'un grand nombre de transitions, et sa forme est déterminée par la structure de couplage du système de spin. Même dans les molécules de petite taille, le nombre de pics distincts est des dizaines à des milliers de fois plus petit que celui des transitions quantiques).

La question est ainsi loin d'être résolue, et, si les décompositions spectrales, partielles ou complètes, possiblement par ajustements spectraux, sont couramment effectuées pour l'analyse des spectres d'is q NMR ou de RMN *in vivo* (De Graaf et Bovee, 1990 ; Provencher, 1993 ; Vanhamme *et al.*, 1997), ces techniques s'appliquent mal à l'analyse des complexes de résonance dans les spectres RMN haute résolution ; elle sont notamment sensibles à de petits décalages des signaux, tels qu'en engendrent des variations de pH ou la présence d'ions paramagnétiques (Provencher, 1993). Ces difficultés sont notamment à l'origine de la mise au point d'algorithmes de réduction des données (*data reduction algorithms*) dans les études métabolomiques (Tsang *et al.*, 2005 ; Salek *et al.*, 2007 ; de Meyer *et al.*, 2008 ; McLoughlin *et al.*, 2009 ; de Graaf *et al.*, 2011) : après l'acquisition des spectres, ceux-ci sont divisés en intervalles nommés *bins* ou *buckets*, de largeur constante ou variable selon les algorithmes, mais plusieurs fois supérieure à la largeur des lignes spectrales. De la sorte, on évite des décalages spectraux et des changements de structures des multiplets. Puis on compare statistiquement les spectres divisés afin d'identifier des structures spectrales qui distinguent des classes. Pour de nombreuses études métabolomiques de spectres RMN ¹H, cette méthode est retenue, bien qu'il ait été montré que des classifications statistiques fondées sur des concentrations absolues conduisent à de meilleures discriminations (de Graaf *et al.*, 2011).

Toutefois de telles études imposent de disposer de nombreux spectres apparentés (Puig-Castellvi *et al.*, 2017 ; Le Gresley et Peron, 2019), alors que, parfois, les dosages ne peuvent reposer que sur un petit nombre de spectres (voire un seul) d'un même échantillon. Pour les analyses des aliments par des techniques *in situ*, et parce que l'on a souvent besoin de déterminer la quantité absolue des composés d'intérêts dans les échantillons, nous avons notamment mis au point des algorithmes d'intégration explicites pour lesquels les calculs sont exécutés par l'utilisateur, qui sait exactement quels traitements sont appliqués, et qui en voit le résultat. Pour les dosages, les valeurs d'intégration ainsi déterminées sont ensuite confrontées à celles qui sont produites pour des signaux homologues, dans les spectres produits à partir de solutions étalons de composés identifiés dans les spectres (courbes d'étalonnage). Enfin on observe que la méthode d'intégration proposée dans les feuilles de calcul nommées « integ » (voir les Matériels supplémentaires) s'applique à tous les spectres, et pas seulement ceux de RMN, quel que soit le mode de préparation des échantillons (moyennant des changements simples de quelques équations, pour tenir compte de la forme des signaux, généralement gaussiens plutôt que lorentziens pour la RMN).

De telles feuilles de calcul s'imposent d'autant plus que, comme dit précédemment, des comparaisons d'intégrations de signaux par des logiciels propriétaires couramment utilisés donnent des résultats notablement différents (voir la partie Discussion). Les feuilles de calcul que nous donnons en « Matériels supplémentaires » sont écrites avec le programme *Maple*, code compatible avec le langage C++, mais on en donne aussi une version .pdf, au cas où des utilisateurs voudraient faire des transcriptions dans d'autres langages). Elles ont été utilisées par des étudiants de niveau licence ou master, sans compétence initiale en RMN. Ces feuilles de calcul guident l'utilisateur pas à pas, explicitant les calculs pour bien montrer les opérations qui sont effectuées. Surtout ces feuilles de calcul (nommées « integ0 », « integ1 », ..., « integ10 ») correspondent à des conditions d'intégrations différentes (explicitées dans le sommaire du fichier « integ0 »), que l'on préconise d'appliquer après que les signaux du spectre ont été ordonnés par ordres de grandeur successifs (le programme classe les signaux

en classes successives, la première contenant tous les signaux dont le pic est supérieur au dixième du pic maximal, puis la deuxième contenant les signaux dont les pics sont dix à cent fois plus petits, et ainsi de suite). Cet ordonnancement, opéré automatiquement par le programme « integ1 », est utile, lors de l'exploration de milieux complexes, parce que se retrouvent dans une même classe (même ordre de grandeur) des signaux qui correspondent à des protons d'un même composé, ce qui facilite les identifications et, de là, les validations. Il permet également un affichage structuré des résultats d'intégration, ainsi qu'une estimation des incertitudes résultant de la superposition de résonances (par principe, une résonance au n -ième ordre est plus de 10 fois inférieure à une résonance à l'ordre $n - 1$).

On observera que, dans les spectres obtenus par $is\ q$ NMR, le signal de référence (celui du TSP, dans notre exemple) est particulièrement important, puisque la qualité de son intégration détermine celle de l'intégration des autres signaux. Cela fait l'objet de la feuille de calcul « integ1 », qui doit être mise en œuvre avant les feuilles suivantes.

Matériels et méthodes

Pour présenter les feuilles de calcul « integ » et pour comparer les résultats de ces feuilles de calcul et des programmes commerciaux, nous utilisons un même spectre S, dont la production est décrite ci-dessous. Pour la préparation des échantillons, nous avons pesé tous les produits et matériels trois fois, avec une balance Mettler Toledo XP 205 (précision 0.01 mg).

Préparation des échantillons

L'échantillon utilisé pour la production du spectre S a été obtenu à partir d'environ 2.0000 g de grains de café vert (*Coffea arabica* L.), qui ont été cryobroyés pendant 20 s, à l'aide d'un moulin Ika A 11 Basic (Wei *et al.*, 2010). Des échantillons de poudre (400 mg) ont été placés dans des tubes en verre clos avec 2.0000 g d'eau Milli-Q. Les extractions en phase aqueuse ont ainsi été effectuées pendant 1 h à la température de 21 °C, sous agitation. Puis les extraits ont été centrifugés à 15 557 g pendant 1 h à la température de 18 °C. Le surnageant a été récupéré et filtré en seringue (membranes Chromafil® Xtra PET 45/13).

Spectroscopie RMN

Pour les analyses RMN, on a placé 540 μ L d'extrait dans des tubes de quartz RMN Wilmad (diamètre 5 mm) (Bruker, 2017) ; puis on a ajouté un tube capillaire scellé contenant du TSP en solution dans du D_2O , de sorte que la hauteur finale des échantillons atteigne 4,0 cm, ce qui correspond aux recommandations Bruker pour le spectroscope utilisé.

Des spectres RMN unidimensionnels ont été enregistrés à la température de $295.0 \pm 0,1$ K, par un spectromètre Bruker (Bruker BioSpin, Wissembourg, France) composé d'un aimant Ascend (9,4 T) opérant à 400,19 MHz et équipé d'une sonde BBFO et d'une console Avance III HD, pilotée par le logiciel *TopSpin* 3.2.

Les réglages classiques ont été réalisés avant chaque acquisition : la sonde était réglée sur le signal 1H , le signal du D_2O a été utilisé pour le lock, et l'homogénéité du champ magnétique constant a été améliorée par un *shimming* automatique (« topshim »). La longueur d'impulsion a été calibrée à 90° du signal du TSP, et 84 134 points de mesure ont été enregistrés pour les signaux de relaxation (FID), soit un temps d'acquisition de 7 s. Après les impulsions, un temps de 13 s permettait une relaxation complète des principaux signaux d'intérêt. Le gain était réglé automatiquement, et, après deux impulsions factices (*dummy scans*), 256 FID ont été additionnés, afin d'augmenter le rapport *signal/bruit*.

Une multiplication exponentielle avec un coefficient d'élargissement de raie de 0,3 Hz a été appliquée aux FID obtenus (l'utilisation d'une telle fonction augmente le rapport *signal/bruit* en minimisant l'influence des derniers points du FID, le plus souvent associés à du bruit ; la constante de décroissance de cette fonction exponentielle détermine la largeur de la ligne des spectres.). Des zéro ont été ajoutés avant la transformation de Fourier, pour produire un spectre réel de 128 K points. Une correction manuelle de phase a été effectuée, et une calibration en fréquence a calé le signal du composé de référence (TSP) à 0 ppm. Puis on a comparé des corrections de la ligne de base par deux méthodes différentes : (i) à l'aide de la commande automatique de correction « absn » du logiciel *Topspin* et (ii) manuellement, en définissant des points et en effectuant la commande « sab » de *Topspin*.

Le spectre S est donné en Figure 1.

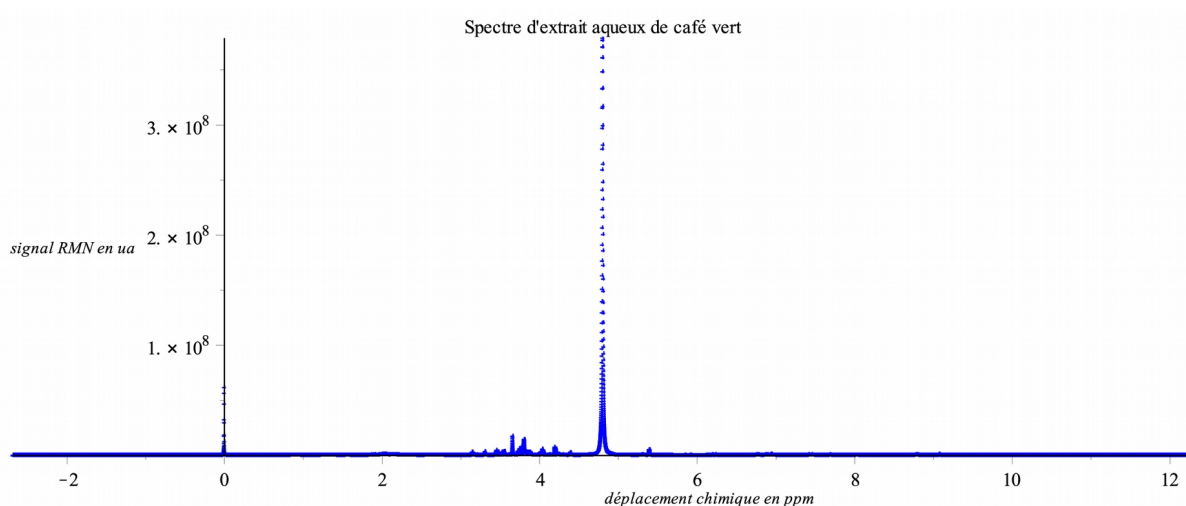


Figure 1. Le spectre RMN du proton (spectre S) qui a servi à toutes les comparaisons est affiché ici par valeurs croissantes de déplacement chimique, mais les feuilles de calcul permettent de le retourner facilement, quand l'utilisateur préfère cette option.

Calculs

Tous les calculs ont été réalisés à l'aide du programme *Maple 18.02* (Waterloo Maple Inc, Ontario). Les résultats ont été comparés à ceux qui ont été obtenus par des programmes d'interprétation des spectres RMN tels que *TopSpin 4.07* (Bruker Biospin, Wissembourg, France), *NMRnotebook* (NMRTEC SAS, Illkirch-Graffenstaden, France), *ACD* (ACD/Labs, Toronto, Canada), *Mestrec* (Mestrelab Research, Santiago de Compostela, Spain).

Résultats

Exploration des spectres par ordres de grandeur successifs

Une bonne pratique, pour l'analyse des spectres, notamment en vue de dosages de composés identifiés par leurs résonances, consiste à examiner d'abord la qualité des acquisitions (This, 2010), et cela est particulièrement important pour les spectres obtenus par *is q* NMR, parce que la présence de solides dans les tubes d'analyse, tels les tubes capillaires contenant le D₂O et le TSP, ou des particules végétales, peuvent perturber le champ magnétique (Bakhmutov, 2015). Pour de « bons » spectres, les résonances doivent être lorentziennes (Brown, 2016), et, notamment, les résonances telles que celles du TSP ou de tout autre composé utilisé comme référence, doivent être symétriques, avec des satellites symétriques (Rutledge, 1996). Pour déterminer quantitativement la « qualité » des spectres, on peut utiliser la largeur à mi-hauteur du signal de référence, ainsi qu'un coefficient d'asymétrie pour ce signal (Dyson, 1990). Pour des signaux gaussiens, le coefficient d'asymétrie est classiquement le rapport du troisième moment par le deuxième moment à la puissance 3/2. Toutefois, comme le troisième moment d'une lorentzienne n'existe pas, en toute rigueur (la somme de $-\infty$ à $+\infty$ du produit de la lorentzienne par x^3 diverge), les feuilles de calcul « integ » le déterminent en limitant l'intégration à un voisinage du pic, et, plus exactement, au voisinage où le signal n'est pas descendu dans la « bande morte », sous le bruit (on peut justifier cette pratique par le fait que l'essentiel du signal se trouve ainsi pris en compte, la décroissance du signal étant rapide autour de l'abscisse du pic).

Observons toutefois que même des spectres de mauvaise qualité, avec des résonances élargies ou très asymétriques, voire plurimodales, peuvent être utilisés pour des dosages, tant que les résonances ne se recouvrent pas trop et que l'on peut les distinguer par le calcul. Même si l'homogénéité du champ magnétique n'est pas bonne ou si des dérives surviennent, l'aire totale des résonances reste proportionnelle au nombre de protons (Sauriol, 2017), comme le confirment les utilisations de droites d'étalonnage réalisées pour des composés particuliers.

Une fois la qualité des spectres déterminée, c'est une bonne pratique que de considérer les spectres par ordres de grandeur décroissants de signaux (Raiman, 1991 ; Mavrovouniotis et Stephanopoulos, 1988). De la sorte, on analyse d'abord les signaux les plus importants, et l'on peut, pour des nombres restreints de signaux que l'on suppose être des résonances, chercher des rapports entiers ou fractionnaires d'aire entre les signaux de chaque ordre de grandeur, afin de mieux identifier les composés présents, et, aussi, de valider les intégrations. Dans les aliments analysés par *is q* NMR, le signal au premier ordre est généralement la résonance de l'eau, et l'on trouve souvent moins d'une dizaine de signaux au deuxième ordre de grandeur (tels que des mono- ou oligo-saccharides), et moins d'une vingtaine de signaux au troisième ordre de grandeur (tels qu'acides aminés, acides organiques, etc.) (Cazor *et al.*, 2006).

Théoriquement le classement des signaux par ordres de grandeurs décroissants devrait se faire sur l'aire des signaux, mais il est plus rapide de classer les signaux par ordres de grandeur de la hauteur de leur pic. A cette fin, une méthode de tri ordinal est mise en œuvre dans la feuille de calcul « integ0 » : la feuille de calcul commence par établir une liste de tous les pics, en cherchant des abscisses pour lesquelles les ordonnées correspondantes sont supérieures aux valeurs des ordonnées pour les valeurs immédiatement inférieure ou supérieure (pour les épaulements, des traitements séparés sont nécessaires). Puis on fait subir des permutations à la liste des pics, et l'on obtient des groupes en divisant successivement par 10 la valeur maximale de l'ordonnée.

L'intégration d'un signal de référence

Pour analyser quantitativement les spectres en minimisant les incertitudes, un signal de résonance doit être intégré aussi précisément que possible, sans quoi les incertitudes sur l'intégration des signaux d'intérêt sont entachées d'incertitudes supplémentaires, qui surviennent lors de la mise à 1 de l'intégrale du signal de référence

(valeur constante, de spectre en spectre, notamment lors des étalonnages). Puis ces incertitudes se répercutent sur les dosages, lors de l'utilisation des droites d'étalonnage.

Pour les intégrations, on est en droit de s'interroger sur la qualité des calculs, avec différentes méthodes d'intégrations : rectangles, trapèzes, etc. Dans le fichier « integ1 », des comparaisons de ces méthodes sont effectuées, et l'on verra, à titre d'exemple pour le spectre S, que les différences sont de l'ordre de 1/20 000.

Il est de toute première importance que les intégrations se fassent sur des spectres dont la ligne de base est aussi horizontale et confondue avec l'axe des abscisses que possible. Pour améliorer le spectre dans le voisinage du signal de référence (moins de deux fois la distance entre les satellites du signal de TSP, quand celui-ci est utilisé comme référence), les feuilles de calculs « integ » analysent la pente, en considérant les deux points de mesure extrêmes, puis deux points juste intérieurs, et ainsi de suite : la pente retenue est la moyenne des pentes calculées pour ces paires de points, étant entendu que la moyenne est calculée sur un nombre de points suffisants pour que la pente moyenne ne varie plus (lissage). Une fois cette pente déterminée, elle est soustraite au spectre, dans la région d'intérêt. Les figures 2 et 3 montrent les résultats de cette opération.

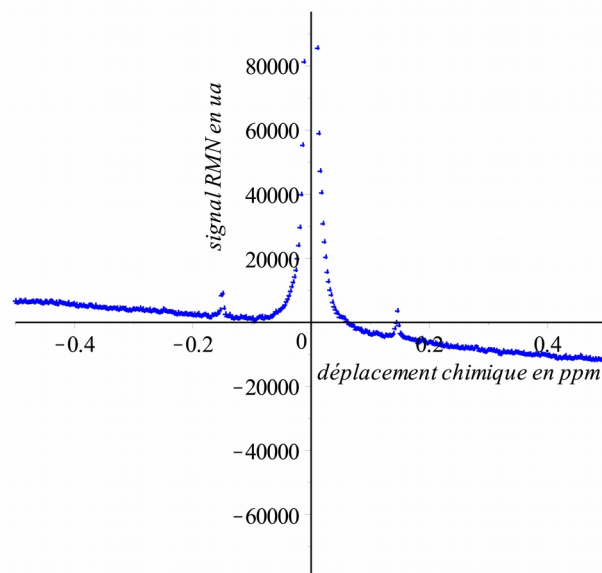


Figure 2: Un agrandissement de la base du signal de référence montre que sa ligne de base corrigée automatiquement par les logiciels commerciaux est loin d'être satisfaisante : cette imperfection nuit à la qualité des intégrations.

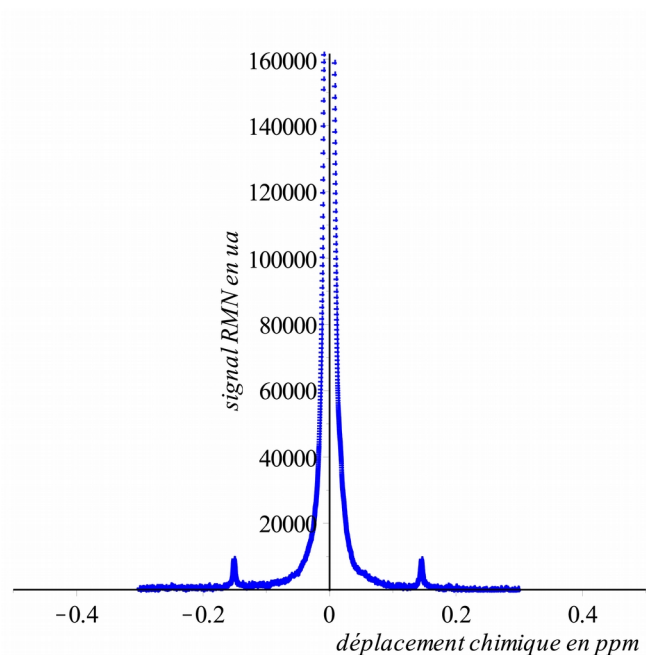


Figure 3: Rectification effectuée par la feuille de calcul « integ1 » pour le signal de référence de la figure 2 (ou tout autre signal).

La ligne de base étant rectifiée localement, le signal d'intérêt peut être intégré. Dans le cas particulier du spectre S, la différence relative entre l'intégration avec la ligne de base corrigée et l'intégration sans correction est environ 0.002 %.

Puis le choix des limites d'intégration est important. Pour les comparaisons précédentes, la largeur d'intégration était choisie arbitrairement au double de la distance entre les satellites du signal de référence, mais des tests ont montré que réduire cette largeur de 30 % ne change que de 2 % la valeur de l'intégrale. Reste que de tels choix arbitraires ne relèvent pas d'une bonne pratique ; aussi les feuilles de calcul « integ » calculent-elles des intégrales avec des bornes tirées au hasard un nombre de fois aussi grand que possible (selon le nombre de points de données du spectre), en vue de déterminer non pas une valeur d'intégration, mais une moyenne de telles valeurs, assortie d'un écart-type. Par exemple, pour la résonance à 9.1 ppm du spectre S, l'écart-type était d'environ 1 %.

Afin de valider cette première détermination, nous avons ajouté à l'algorithme mis en œuvre celui qui avait été proposé par This (Tardieu *et al.*, 2011) et qui ne rectifie pas la ligne de base *a priori*. Cet autre algorithme ouvre une fenêtre d'intégration centrée sur le pic d'intérêt, et intègre répétitivement, en élargissant progressivement la fenêtre, jusqu'à stabilisation de l'intégration (quand la ligne de base est horizontale et proche de l'axe des abscisses) ou jusqu'à son évolution asymptotique linéaire, quand une dérive de la ligne de base est observée. Là encore, un lissage s'impose.

Pour tous ces calculs, la perte de signal dans la « bande morte » (quand le signal est de l'ordre de grandeur du bruit) peut être notable (Dyson, 1990). En théorie, si le bruit était de moyenne rigoureusement nulle, l'intégrale du signal sur un intervalle extérieur à la partie centrale du signal (émergeant du bruit) serait l'intégrale d'un signal lorentzien.

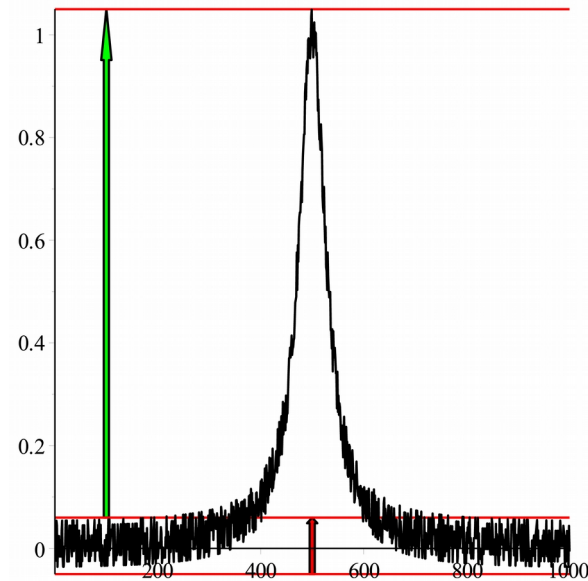


Figure 4: Un signal et le bruit qui l'entoure. La partie de signal entre les lignes rouges (la « bande morte ») est inaccessible.

Pour une lorentzienne d'équation $A/(B+x^2)$ et un bruit de hauteur h (Figure 4), la proportion de signal perdue dans le bruit (pp) est égale à :

$$pp = \frac{\left(\frac{A\pi}{\sqrt{B}} - \frac{2A \cdot \arctan\left(\frac{\sqrt{h(A-Bh)}}{h\sqrt{B}}\right)}{\sqrt{B}} \right) \sqrt{B}}{A\pi}$$

Quand cette proportion est analysée en fonction de h , on observe que la perte augmente rapidement (asymptote verticale sur la figure 5).

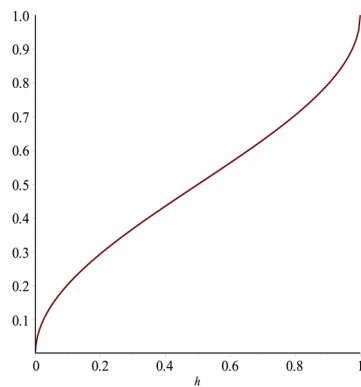


Figure 5: Proportion de signal perdu en fonction du bruit.

Dans le cas particulier du spectre S, l'épaisseur de la bande morte peut être trouvée par un algorithme qui cherche, s'écartant du pic d'une résonance isolée dans un domaine où n'apparaît pas d'autre résonance, avec une ligne de base corrigée, le premier minimum local, puis continue de chercher des maxima locaux ; on retient alors le plus grand de ces maxima. La bande morte étant ainsi identifiée, le programme part du pic et détermine la première valeur du spectre dans la bande morte. On intègre alors le signal au-dessus de la bande morte, et on le compare aux valeurs d'intégrations obtenues précédemment. Pour les singulets du spectre S, la perte dans la bande morte est de l'ordre de 5 % de l'aire totale du signal.

Toutefois cette valeur a été obtenue pour un signal qui a été intégré sans tenir compte du fait que les lorentziennes conservent une valeur notable loin du pic (par rapport à des signaux gaussiens). Aussi avons-nous également testé l'intégration de la fonction obtenue par un ajustement par une lorentzienne ayant le même pic, et une largeur à mi-hauteur telle que déterminée sur le spectre. De la sorte, l'ajustement est bon au-dessus de la bande morte, mais la lorentzienne ajustée diminue plus rapidement que le signal réel. Pour des signaux seulement dix fois supérieurs au bruit, la perte dans la bande morte est alors d'environ 3 %.

Le repérage des résonances

Dans les spectres, les résonances apparaissent dans des environnements différents, et nous proposons plusieurs algorithmes différents pour calculer au mieux les intégrations, dans les divers cas.

Le plus simple est celui de signaux (singulets ou multiplets) isolés (feuille de calcul « integ4 »). Puis un cas un peu plus complexe est celui de signaux « séparés », c'est-à-dire quand le signal revient à la ligne de base (en réalité, à la bande morte) entre elles (« integ5 »). Le cas de signaux « proches » d'intensités voisines, pour lesquelles le recouvrement est visible, impose un traitement particulier (« integ6 »). Plus complexe encore, le cas d'un petit signal dans le flanc d'un signal plus grand (« integ8 »), avant, enfin, le cas des complexes de signaux (« integ9 »).

Avant d'examiner les algorithmes proposés pour ces divers cas, on peut se demander quand un signal correspond à une ou plusieurs résonances, car on se souvient que les signaux peuvent être des enveloppes de plusieurs résonances (Mestrelab Research, 2017). A titre d'exemple de la difficulté à traiter, la figure 6 présente la somme de 4 signaux : une lorentzienne centrée en $x = 1$, une lorentzienne centrée en $x = 3$, une lorentzienne de même intensité que cette dernière, mais centrée en $x = 3.1$, et une gaussienne centrée en $x = 5$. La comparaison des lorentziennes pures avec la gaussienne confirme ce qui est connu, à savoir que, pour une lorentzienne et une gaussienne de même pic et avec les mêmes points d'intersection avec la bande morte, la lorentzienne perd plus de signal que la gaussienne dans la bande morte. La comparaison des deux premiers signaux de la figure montre combien il est parfois difficile de reconnaître quand un signal correspond à une ou plusieurs résonances.

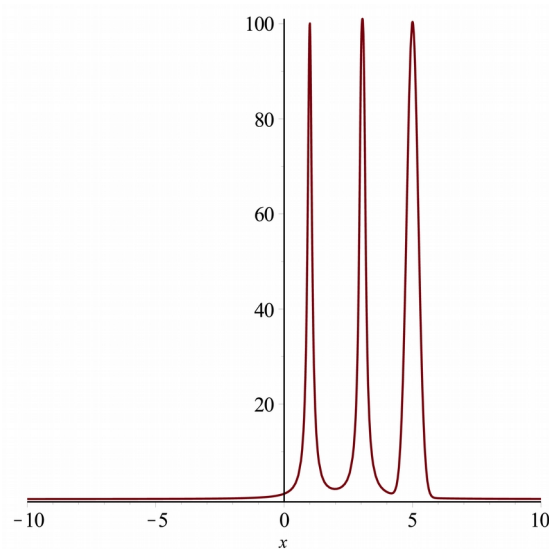


Figure 6: Une lorentzienne centrée en $x = 1$, une lorentzienne centrée en $x = 3$, une lorentzienne de même intensité que cette dernière, mais centrée en $x = 3.1$, et une gaussienne centrée en $x = 5$.

Avant de discuter l'algorithme que l'on peut mettre en œuvre pour une telle reconnaissance, on doit observer que la superposition de deux signaux correspondant à des résonances différentes conduit à des erreurs de quantification qui peuvent atteindre 100 %. Cela justifie, notamment, que l'intégration d'une résonance pour un composé soit validée par l'intégration d'un ou de plusieurs autres signaux du même composé, en d'autres endroits du spectre ; évidemment, on choisit des signaux de sorte qu'ils soient séparés de signaux attribuables à d'autres composés. En revanche, des variations du champ magnétique, durant l'acquisition de nombreux scans, peuvent conduire à des signaux non lorentziens sans que l'intégration soit invalidée, car l'intégrale d'un signal qui serait la somme de plusieurs résonances reste proportionnelle à la quantité de protons responsables de ces

résonances, quelle que soit la position de leur pic. Il en est de même quand l'homogénéité du champ magnétique est difficile à obtenir, notamment quand des tissus ou le tube capillaire sont ensemble dans le tube d'analyse : tant que les résonances ne recouvrent pas d'autres signaux, l'intégrale reste proportionnelle au nombre de protons. Cela étant posé, après que les conditions d'acquisition ont été optimisées, en vue d'éviter des recouvrements de résonances, l'algorithme proposé pour reconnaître si un signal particulier correspond à une ou plusieurs résonances consiste à décomposer le signal par des ajustements successifs à l'aide de 1, 2, 3, etc. lorentziennes, et à calculer, pour chaque cas, un « facteur de qualité » qui est la somme des carrés des résidus. En pratique, toutefois, un algorithme qui utiliserait des lorentziennes entièrement générales d'équation

$$\frac{A_i}{B_i + (x - x_i)^2}$$

où A_i , B_i et x_i sont à déterminer ne convient pas, car le nombre de points de données, dans la région d'intérêt, est souvent trop faible. Souvent l'ajustement est impossible, sauf à ajouter des 0 autour de l'intervalle d'intérêt.

L'intégration de signaux isolés

Pour la détermination de l'intégrale de signaux isolés, le réglage de la ligne de base et le choix des limites d'intégrations sont essentiels, comme on l'a vu pour l'intégration du signal de référence (TSP, par exemple). Afin de mieux estimer l'importance du choix des limites d'intégration, on a comparé une intégration manuelle « naïve » avec une moyenne de trois intégrations associées respectivement à des limites déterminées au hasard. La valeur arbitraire s'est trouvée au centre de l'intervalle défini par la moyenne et l'écart-type calculés par la seconde méthode, à moins d'un écart-type de la moyenne. Le coefficient de variation était de 0.01.

Pour deux signaux isolés, la question essentielle est de savoir quand ils sont véritablement séparés, auquel cas on peut utiliser, pour chacun, l'algorithme d'intégration correspondant aux signaux isolés. En théorie, deux résonances ne sont jamais séparées, parce que les lorentziennes ne s'annulent jamais, mais en pratique, le signal tombe rapidement dans la bande morte, et l'on peut donc considérer une « séparation pratique ».

Observons que la chimie analytique utilise les formules suivantes pour le calcul de R_s d'une paire de signaux gaussiens centrés en x_1 et x_2 (Dyson, 1990) :

$$R_s = \frac{2(x_2 - x_1)}{w_{b1} + w_{b2}}$$

$$R_s = \frac{1.18(x_2 - x_1)}{w_{h1} + w_{h2}}$$

Dans ces équations, w_{b1} et w_{b2} sont respectivement la largeur à la base, déterminée par les tangentes de chaque côté du signal, et w_{h1} et w_{h2} sont respectivement les largeurs du signal à mi-hauteur. Toutefois, en pratique, on peut considérer que les signaux sont « séparés » si au moins un point de signal, entre les deux maxima, est dans la bande morte. Comme indiqué, on peut alors utiliser l'algorithme de calcul des aires qui avait été proposé pour le composé de référence (TSP).

Deux signaux proches et d'intensités voisines

Quand deux signaux sont proches et que leurs intégrales sont du même ordre de grandeur, la hauteur des pics diffère de la hauteur du pic pour chaque signal qui serait considéré séparément. Ce cas est représenté sur la figure 7. Bien sûr, quand les deux moitiés de chaque signal sont symétriques, la valeur intégrale de chaque signal est égale à la moitié de l'intégrale totale, mais un tel cas est évidemment exceptionnel, et une solution générale s'impose.

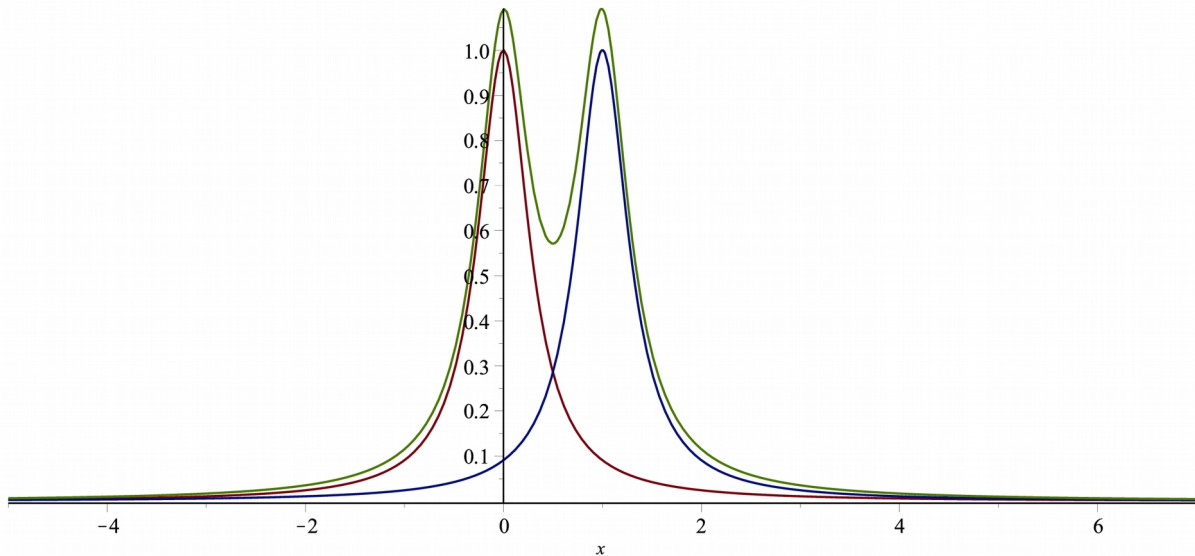


Figure 7: Décomposition d'un petit massif constitué de deux résonances.

Une possibilité consiste à ajuster chaque signal avec une équation lorentzienne, puis à intégrer chaque signal séparément. Comme, en pratique, le nombre de points expérimentaux est souvent insuffisant pour les ajustements, on gagne à réduire le nombre de paramètres inconnus en fixant, par exemple, l'abscisse des maxima. Se pose alors la question de savoir si l'abscisse des maxima est la même que celle des maxima des signaux qui seraient séparés.

Sur le spectre S, on peut observer que le décalage est de l'ordre de 0.8 %, mais les ajustements sont sensibles à de tels décalages (Jenson et Jeffreys, 1997). Les calculs, pour une paire de lorentziennes ou pour une paire de gaussiennes séparées par une valeur d entre les abscisses des maxima montre que, quand la distance d est inférieure à une certaine proportion de la largeur des signaux, un seul maximum apparaît pour la somme des deux signaux, mais quand la distance augmente, on observe deux maxima et un minimum local, avec un décalage des maxima qui varie en raison inverse de la distance d . En revanche, on peut observer que l'ordonnée des pics de deux maxima discernables -pour les lorentziennes comme pour les gaussiennes- est plus modifiée quand les pics sont proches.

Une autre possibilité consiste à modéliser les signaux en utilisant des lorentziennes définies par leur hauteur et leur largeur à mi-hauteur. Par cette méthode, la différence d'intégrales entre un signal réel et un signal ajusté atteint 10 %, ce qui justifie que d'autres solutions doivent être mises en œuvre.

Quand les intensités des deux signaux diffèrent, on peut ainsi considérer que le signal le plus important est peu modifié par le signal plus petit (d'environ 0.01 % quand les deux pics sont séparés par une largeur à mi-hauteur des deux), de sorte que l'on peut chercher à modéliser le plus gros signal en n'utilisant que sa moitié externe, pour laquelle la perturbation par le petit signal est faible ; on double alors la valeur de l'intégrale de cette moitié externe. Puis on détermine l'intégrale du petit signal en soustrayant l'intégrale ainsi calculée de l'intégrale totale pour les deux signaux. En pratique, des différences de 1 % sont calculées par rapport à des signaux purs qui seraient ajoutés.

Les épaulements

Pour ce qui concerne les épaulements, on doit commencer par observer que l'écrémage tangentiel (*tangent skimming*) est une technique médiocre : elle consiste à ne considérer que l'aire comprise entre le signal réel et une droite tangente en deux points au signal, sous l'épaulement (Bicking, 2006), mais elle conduit à des sous-estimations du signal d'intérêt dont nous avons parfois constaté qu'elles atteignent 75 % (on peut démontrer qu'ils surviennent quand le petit signal se trouve au maximum de courbure du signal principal qui porte l'épaulement) !

Pour déterminer l'intégrale d'un petit signal qui épaupe un signal plus grand, par exemple pour un signal dans le flanc du signal de l'eau en *is q* NMR, on peut observer que le grand signal est peu modifié par le petit signal qui forme l'épaulement, de sorte que l'on peut modéliser le grand signal par une lorentzienne à partir de la hauteur du pic et de la largeur à mi-hauteur, et -pour plus de précision- en ne considérant que la moitié du grand signal opposée à l'épaulement que l'on complète par symétrie par rapport à une droite passant par le pic. L'intégration de ce modèle permet encore, par soustraction, de déterminer l'intégrale du petit signal (Figure 8).

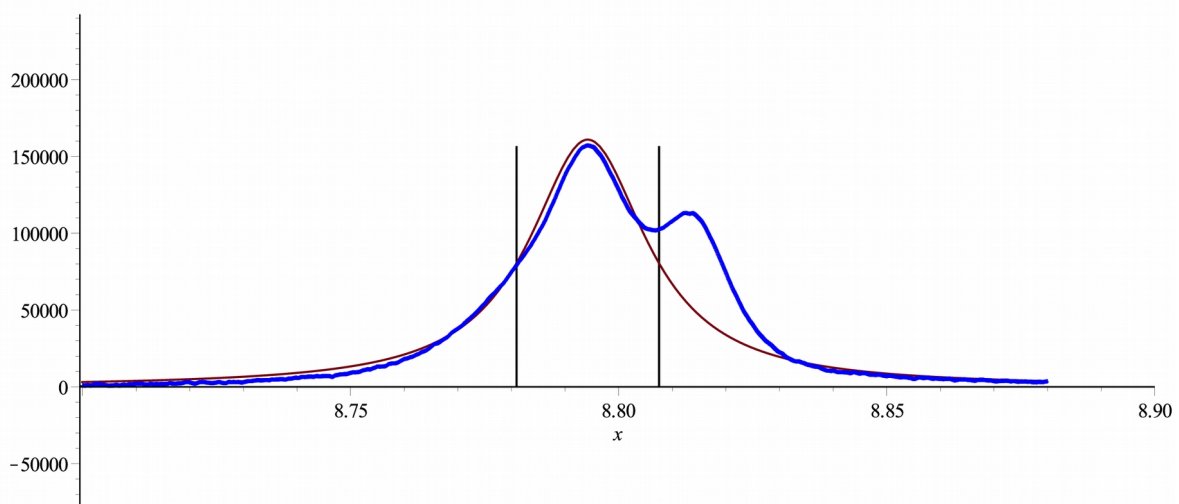


Figure 8: Traitement d'un épaulement.

Discussion

Pour mieux montrer l'intérêt des algorithmes décrits précédemment, mis en œuvre avec des commentaires détaillés dans les feuilles de calcul « integ » données en « Matériels supplémentaires », nous avons comparé les résultats des analyses effectuées par ces feuilles de calcul avec des intégrations du spectre S effectuées par des

logiciels commerciaux, en commençant par observer que ces logiciels commerciaux n'explicitent pas la nature des opérations effectuées, ni parfois même les méthodes mises en œuvre.

Pour le signal de référence (TSP), toutes les intégrations ont été faites avec les mêmes limites (- 0,1 et + 0,1 ppm), après un recalage en fréquence du pic de TSP à 0,00 ppm ; l'aire du signal de TSP a alors été fixée à 1.

Puis le signal isolé à 8,8 ppm a été intégré entre les bornes 8,60 et 8,90 ppm, ce qui a donné une aire de 50,83 avec le logiciel *Topspin* (correction automatique de la ligne de base), ou de 54,11, toujours avec *Topspin*, mais avec une ligne de base ajustée manuellement. Avec le logiciel *NMRnotebook*, l'intégrale du même signal a été trouvée égale à 47,8 ; avec *ACD* et *Mestrec*, on a trouvé respectivement des valeurs de 46,04 et de 43,94. Pour le signal entre 9 et 9,2 ppm, les intégrales, pour les divers cas précédents, ont été trouvées égale à 33,26, 28,83, 26,4, 28,24 et 40,28. Pour ces diverses valeurs, le coefficient de variation, avec trois paires de bornes fixées arbitrairement, est compris entre 4 et 10 %. Avec l'usage de la feuille de calcul « integ », la valeur trouvée est 26,59, dont s'approche le plus le résultat de *NMRnotebook*.

Comme de telles différences pouvaient dépendre du spectre, nous avons exploré d'autres spectres et, notamment, nous avons étudié le spectre ^1H NMR d'un extrait en phase aqueuse de café (*Arabica*) torréfié, avec un nombre de points de mesure réduit cette fois à 32768 et un nombre de scans égal à 1024 (zg, SW = 16 ppm, AQ = 3 s, RG = 1.6, P1 = 9.875 s), toujours avec le même tube capillaire scellé contenant 0.50 mL d'une solution de TSP (1 mmol/L) dans le D_2O . Pour ce spectre, les résultats de *TopSpin* 4.07 ont été comparés à ceux de *MestReNova* 11.0 et de *NMRnotebook* 2.80. Pour tous les traitements, la ligne de base était corrigée avec *TopSpin* 4.07, et une intégration manuelle était faite. Pour *MestReNova* 11.0, la méthode de correction de la ligne de base « Bernstein polynomial fit (order 3) » était utilisée avant la correction à points multiples et la correction en splines cubiques. Les résultats sont indiqués dans le Tableau 1. A nouveau, on observe des différences notables entre les divers logiciels commerciaux, mais, surtout, leur utilisation ne permet pas de savoir lequel donne les résultats les plus proches de la valeur réelle, contrairement aux feuilles de calcul, où l'utilisateur suit pas à pas, explicitement, les calculs.

Tableau 1. Comparaison de différents résultats d'intégration d'un signal d'un même composé, dans un spectre simple : les différences entre les résultats donnés par les logiciels commerciaux sont parfois supérieures à 30 %, ce qui est évidemment inacceptable.

Composé	Déplacement chimique (ppm)	Intégrations TopSpin 3.5	Intégrations MestReNova 11.0	Intégrations NMRnotebook 2.80	Intégration Maple 2016
TSP	0	1	1	1	1
Caféine	3.38-3.27	2.2004	2.31	2.3	2.35
	3.40-3.29	2.1949	2.29	2.0	
	3.40-3.27	2.2477	2.29	2.2	
Caféine	3.55-3.45	2.1012	2.08	1.9	2.20
	3.55-3.49	1.9273	2.32	1.9	
	3.50-3.45	2.0169	2.05	2.1	
Caféine	3.96-3.89	2.2384	2.76	2.0	2.46

	3.96-3.92	1.9854	2.71	2.0	
	3.92-3.89	2.2215	2.68	2.1	
Caféine	7.91-7.80	0.7768	0.74	0.8	0.78
	7.91-7.84	0.7618	0.73	0.8	
	7.87-7.80	0.7443	0.73	0.7	

Meyer *et al.* (1988) avaient déjà signalé des sous-estimations des intégrales des signaux RMN sur des spectres du ^{31}P . Ils observaient alors que l'intégration manuelle des signaux de ces spectres était l'objet de biais notables, conduisant à la sous-estimation de l'aire des petits signaux et des signaux larges. De tels biais peuvent être à l'origine d'erreurs importantes, au cours de l'interprétation des résultats d'expériences de spectroscopie, même quand les spectres ont des rapports *signal/bruit* importants, des pics bien résolus et des lignes de base de bonne qualité, comme cela est le cas, dans leur étude, pour les spectres du ^{31}P de muscles squelettiques. Meyer *et al.* concluaient que les intégrations seraient encore plus douteuses pour des spectres plus complexes. Évidemment l'automatisation des traitements n'est pas une solution à ce problème : des calculs erronés donnent toujours des résultats erronés, même si ces résultats sont reproductibles. En revanche, les feuilles de calcul « integ » montrent bien les calculs qui sont effectivement réalisés, et elles permettent à l'utilisateur de prendre conscience des difficultés d'intégrations éventuelles, ou, au pire, d'être en mesure de douter salutairement des résultats.

Conclusion et perspectives

A ce jour, les feuilles de calcul « integ », données en « Matériels supplémentaires », sont écrites à l'aide du logiciel *Maple*, et exécutables à l'aide de ce logiciel. Nous fournissons donc les versions .mw, exécutables par le logiciel *Maple*, mais aussi une version .pdf de calculs effectués sur le spectre S.

Les feuilles de calcul sont utilisables très facilement : l'utilisateur est en permanence guidé, et toutes les opérations sont non seulement explicitées, mais expliquées. Cela allonge les feuilles de calcul, mais permet aux utilisateurs de bien contrôler leurs calculs, pas à pas.

Références bibliographiques

Aisala H, Sinkkonen J, Kalpio M, Sandell M, This H, Hopia A (2016) In situ quantitative ^1H NMR spectroscopy discriminates between raw and steam cooked potato strips based on their metabolites. *Talanta* **161** : 245-252.

Antti H, Bollard ME, Ebbels T, Keun H, Lindon JC, Nicholson JK, Holmes E (2002) Batch statistical processing of ^1H NMR-derived urinary spectral data. *J Chemometr* **16** :461-468.

Bakhmutov VI (2015) *NMR Spectroscopy in Liquids and Solids*, CRC Press, Boca Raton.

Bartolome LG, Hoff JE (1972) Finning of potatoes: Biochemical effects of preheating. *J Agr Food Chem* **20** : 266-270.

Belton PS, Colquhoun IJ, Kemsley EK, Delgadillo I, Roma P, Dennis MJ, Sharman M, Holmes E, Nicholson JK, Spraul M (1998) Application of chemometrics to the ^1H NMR spectra of apple juices: discrimination between apple varieties. *Food Chem* **61** : 207-213.

Bingol K, Bruschiweiler R (2011) Deconvolution of Chemical Mixtures with High Complexity by NMR Consensus Trace Clustering. *Anal Chem* **83** : 7412-17.

Boyd Alan SF (1995), NMR sample preparation. <http://nmr.chem.umn.edu/samprep.html>, last access 10 July 2017.

Lindon JC, Holmes E, Nicholson JK (2001) Pattern recognition methods and applications in biomedical magnetic resonance. *Prog Nucl Mag Res Sp* **39** :1-40.

Brown KC (2016) Essential Mathematics for NMR and MRI Spectroscopists, Royal Society of Chemistry, London.

Bruker (2017) <https://www.bruker.com/service/information-communication/user-manuals/nmr.html>, last access 2017-04-30.

Campbell N (1995) Biologie, De Boeck-Wesmael, Bruxelles.

Cazor A, Deborde C, Moing A, Rolin D, This H (2006) Sucrose, glucose and fructose extraction in aqueous carrot root extracts prepared at different temperatures by means of direct NMR measurements. *J Agric Food Chem* **54** :4681-4686.

Colquhoun IJ (1998) High resolution NMR spectroscopy in food analysis and authentication. *Spectroscopy Europe* **10** : 8-18.

Davis FL, Terry LA, Chope GA, Faul CFJ (2007) Effect of Extraction Procedure on Measured Sugar Concentrations in Onion (*Allium cepa* L.) Bulbs. *J Agric Food Chem* **55** :4299–4306.

De Graaf AA, Bovee WM (1990) Improved quantification of in vivo ^1H NMR spectra by optimization of signal acquisition and processing and by incorporation of prior knowledge into the spectral fitting. *Magn Reson Med* **15** : 305–319.

De Graaf RA (2007) In Vivo NMR Spectroscopy. Principles and Techniques, 2 ed., John Wiley, Chichester.

De Graaf RA, Brown PB, McIntyre S, Nixon TW, Behar KL, Rothman DL (2006) High magnetic field water and metabolite proton T1 and T2 relaxation in rat brain in vivo. *Magn Reson Med* **56** : 386–394 .

De Graaf RA, Chowdhury GMI, Behar KL (2011) Quantification of High-Resolution ^1H NMR Spectra from Rat Brain Extracts. *Anal Chem* **83** : 216–224.

De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC, Van Crielinge W (2009) NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Anal Chem* **80** : 3783–3790 .

Dolan JW (2014) Estimating Resolution for Marginally Separated Peaks. LC-GC Europe 9 : 480-488 ; <http://www.chromatographyonline.com/estimating-resolution-marginally-separated-peaks-0?id=&pageID=1&sk=&date=>

Duarte IF, Barros A, Belton PS, Righelato R, Spraul M, Humpfer E, Gil AM (2002) High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. *J Agric Food Chem* **50** :2475-2481.

Ducauze C (2014) Chimie analytique, analyse chimique et chimométrie, Lavoisier, Paris.

Dyson N (1990) Chromatographic Integration Methods, Royal Society of Chemistry.

Fan TWM (1996) Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Prog Nucl Mag Res Sp* **28** :161-219.

Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* **48** : 155-171.

Gershon ND, Porter KR, Trus BL (1985) The cytoplasmic matrix: its volume and surface area and the diffusion of molecules through it. *P Natl Acad Sci USA* 5030–5034.

Gil AM, Duarte IF, Delgadillo I, Colquhoun IJ, Casascelli F, Humpfer E, Spraul M (2000). Study of the compositional changes of mango during ripening by use of nuclear magnetic resonance spectroscopy. *J Agr Food Chem* **48** : 1524-1536.

Harris DC (2007) Quantitative chemical analysis, 7th ed., W. H. Freeman and Company, New York.

Hatzakis E (2019) Nuclear magnetic resonance (NMR) spectroscopy in food science: a comprehensive review, *Comp Rev Food Sci Safety*, 18(1), 189-220.

IUPAC (2007) Definitions of terms relating to the structure and processing of sols, gels, networks, and inorganic-organic hybrid materials (IUPAC Recommendations 2007). *Pure and Applied Chemistry*, 79, 1801–1806.

Jenson VG, Jeffreys GV (1997) Mathematical methods in chemical Engineering, 2nd ed., Academic Press, San Diego.

Le Gall G, Colquhoun IJ, Davies, AL, Collins GJ, Verhoeyen ME (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ¹H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agr Food Chem* **51** :2447-2456.

Le Gall G, Puaud M, Colquhoun IJ (2001) Discrimination between orange juice and pulp wash by (¹H) Nuclear Magnetic Resonance spectroscopy: identification of marker compounds. *J Agric Food Chem* **49** : 580-588.

Le Gresley A, Peron JMR (2019) A semi-automatic approach to the characterisation of dark chocolate by Nuclear Magnetic Resonance and multivariate analysis, *Food Chem* **275** : 385–389

Liu Y, Cheng J, Liu H, Deng Y, Wang J, Xu F (2017) NMRSpec: An integrated software package for processing and analyzing one dimensional nuclear magnetic resonance spectra. *Chemometr Intelligent Laboratory Systems* **162** : 142–148.

Mavrovouniotis ML, Stephanopoulos G (1988) Formal Order-of-Magnitude Reasoning in Process Engineering. *Comput Chem Engng* **12** : 867-880.

McLoughlin GA, Ma D, Tsang DTM, Jones DNC, Cilia J, Hill MD, Robbins MJ, Benzel IM, Maycox PR, Holmes E, Bahn S (2009) Analyzing the effects of psychotropic drugs on metabolite profiles in rat brain using ¹H NMR spectroscopy. *J Proteome Res* **8**(4): 1943–1952.

Mestrelab Research. 2017. MestReNova Manual, 31 Dec 2017.

Meyer RA, Fisher MJ, Nelson SJ, Brown TR (1998) Evaluation of manual methods for integration of in vivo phosphorus NMR spectra. *NMR Biomed* **1**(3) : 131-135.

Mlynarik V, Cudalbu C, Xin L, Gruetter R (2008) Quantitative metabolomics using NMR. *J Magn Reson* **194** :163–168.

Monteiro-Marques JP, Rutledge DN, Ducauze CJ (1991) Low resolution pulse nuclear magnetic resonance study of water equilibration in dried carrots, *Int J Food Sci Tech* **17**–183.

Monteiro-Marques JP, Rutledge DN, Ducauze CJ (1991) Low resolution pulse NMR detection of the mobilisation point of solutes during the drying of carrots. *Lebensm Wiss Technol* **17** : 93–98.

Monteiro-Marques JP, Rutledge DN, Ducauze CJ (1992) A rapid low resolution pulse NMR method to detect particular water mobilities during the drying of carrots. *Sci Aliments* **6**–624.

Noteborn HP, Lommen A, Van der Jagt RC, Weseman JM (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J Biotechnol* **77** :103-114.

O'Donoghue EM, Somerfield SD, Bendall M, Hedderly D, Eason J, Sims I (2004) Evaluation of carbohydrates in Pukekohe Longkeeper and Grano cultivars of *Allium cepa*, *Journal of Agricultural and Food Chemistry* **52**: 5383-5390.

Ott KH, Aranibar N, Singh BJ, Stockton GW (2003) Metabolomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* **62** : 971-985.

Pfeuffer J, Tkac I, Provencher SW, Gruetter R (1999) Toward an in vivo neurochemical profile: quantification of 18 metabolites in short-echo-time (¹H) NMR spectra of the rat brain. *J Magn Reson* **141** : 104–120.

Provencher SW (1993) Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med* **30** : 672–679.

Puig-Castellví F, Alfonso I, Tauler R (2017) Untargeted assignment and automatic integration of ¹H NMR metabolomic datasets using a multivariate curve resolution approach, *Analytica Chimica Acta* **964** : 55-66.

Raiman O (1991) Order of magnitude reasoning. *Artif Intell* **51** : 11-38.

Rutledge DN (1996) *Signal Treatment and Signal Analysis in NMR*, Elsevier, Amsterdam.

Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, Nunez D, Sweatman BC, Haselden JN, Cox RD, Connor SC, Griffin JL (2007) A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat and human. *Physiol Genomics* **29** : 99–108.

Sauriol F (2017) The NMR facility at Queens University, <https://qshare.queensu.ca/Users01/sauriol/f/www/>, last access 2017-12-20.

Sindorf DW, Maciel GE (1983) Silicon-29 NMR study of dehydrated/rehydrated silica gel using cross polarization and magic-angle spinning. *J Am Chem Soc* 1487–1493.

Sobolev AP, Segre A, Lamanna R (2003) Proton high-field NMR study of tomato juice. *Magn Reson Chem* **41** :237-245.

Tardieu A, This H (2009) Comparison of Mono- and Disaccharides Release in Aqueous Solutions by Raw or Fried Dice of Onion (*Allium cepa* L.) Bulbs using Quantitative Nuclear Magnetic Resonance (qNMR). *J Food Sci* C319–325.

This H (2010) Modélisons au premier ordre, avant d'aller voir les détails. *Actual. Chimique* **342-343** : 109-111.

This H (2012) Solutions are solutions, and gels are almost solutions, *Pure Appl Chem* **5** : 257.

This H (2021) Gels. In Burke R, Kelly A, Lavelle C, This vo Kientza H (eds.) Handbook of molecular gastronomy, CRC Press, Boca Raton, Fl.

This H, Weberskirch L, Plassais M, Luna A, His A, Skoglund S (2010) La RMN du liquide voit le cœur des légumes et des viandes... puisque ce sont des gels. *Actual Chimique* **337** : 10-13.

Tsang TM, Griffin JL, Haselden J, Fish C, Holmes E (2005) Metabolic characterization of distinct neuroanatomical regions in rats by magic angle spinning 1H nuclear magnetic resonance spectroscopy. *Magn Reson Med* **53** : 1018–1024.

Vanhamme L, van den Boogaart LA, Van Huffel S (1997) Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J Magn Reson* **29** : 35–43.

Viola R, Davies HV (1992) A microplate reader assay for rapid enzymatic quantification of sugars in potato tubers, *Potato Res.*, 35:55-58.

Vogels JTWE, Terwel L, Tas AC, Van den Berg F, Dukel F, Van der Greef J (1996) Detection of adulteration in orange juices by a new screening method using proton NMR spectroscopy in combination with pattern recognition techniques. *J Agric Food Chem* **44** :175-180.

Ward JL, Harris C, Lewis J, Beale MH (2003), Assessment of 1H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **62** : 949-957.

Weberskirch L, Luna A, Skoglund S, This H (2011) Comparison of two liquid-state NMR methods for the determination of saccharides in carrot (*Daucus carota* L.) roots. *Anal Bioanal Chem* **399** : 483-487.

Wei K, Furihata K, Hu F, Miyakawa T, Tanokura M (2010) Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy. *Magn Reson Chem* **48** :857–865.

Wishart DS (2008) Quantitative metabolomics using NMR. *Trend Anal Chem* **27**(3) : 228-237.