

RÉGRESSION NON-LINÉAIRE POUR L'ANALYSE D'UN MÉDICAMENT ANTICANCÉREUX PAR DIFFUSION RAMAN EXALTÉE DE SURFACE

Tom Rohmer ¹, Laetitia Lê ^{1,2},
Antoine Dowek ^{2,3}, Eric Caudron ^{2,3} & Marc Lavielle ¹

¹ *Inria Saclay île-de-France, CMAP, école polytechnique Route de Saclay 91128
Palaiseau, France*

{tom.rohmer, marc.lavielle}@inria.fr

² *Service de Pharmacie, Hôpital européen Georges Pompidou, Assistance
Publique-Hôpitaux de Paris, 20 rue Leblanc, 75015 Paris*

laetitia.le@aphp.fr

³ *Lip(Sys)² Chimie Analytique Pharmaceutique, UFR de Pharmacie, 5 rue Jean
Baptiste Clément, 92290 Châtenay-Malabry, Université Paris-Saclay*

{eric.caudron, antoine.dowek}@u-psud.fr

Résumé. L'objectif de ces travaux est d'évaluer la faisabilité d'une technique en plein évolution, la spectroscopie Raman exaltée de surface (SERS) pour l'analyse de la concentration de médicaments anticancéreux. Cette technique utilisant des nanoparticules d'argent est appliquée à l'analyse quantitative du 5-fluorouracile, l'une des molécules les plus utilisées en cancérologie. Au vu des fortes variabilités spectrales entre les diverses répétitions de l'expérience et de l'interaction non-linéaire observée entre la concentration et l'intensité du signal, des méthodes de régressions non-linéaires permettant de prendre en compte ces variabilité ont été mis en places sur une base d'apprentissage puis comparés sur une base test.

Mots-clés. Spectroscopie Raman, Régression non-linéaire, Analyse de médicament

Abstract. The objective of this work is to evaluate the feasibility of an evolving technique, surface enhanced Raman spectroscopy (SERS) for the analysis of cytotoxic drug concentration. This technique using silver nanoparticles is applied for quantitative analysis of 5-fluorouracil, one of the most widely used molecules in oncology. In view of the high spectral variability observed between the various repetitions of the experiment, and the observed non-linear interaction between signal concentration and intensity, non-linear regression methods to take these variabilities into account have been developed on a training dataset and compared on a test dataset.

Keywords. Raman spectroscopy, non-linear regression, drug analysis

1 Introduction

Selon l’OMS, le cancer constitue de nos jours la deuxième cause de décès dans le monde. En 2015, plus de 2,5 millions de séances de chimiothérapies ont été réalisées pour traiter des patients atteints de cancer. La chimiothérapie repose sur l’administration de médicaments cytotoxiques dont l’action est de bloquer la division cellulaire et donc la multiplication des cellules cancéreuses. Cependant ces médicaments ne ciblent pas seulement les cellules cancéreuses, ils agissent également sur les cellules saines de l’organisme et peuvent être à l’origine de toxicité importante pour le patient. Ces médicaments sont préparés extemporanément pour un patient donné au sein des pharmacies à usage intérieur des établissements de santé. Malgré une optimisation des pratiques, des erreurs de préparation persistent. Dans une démarche de sécurisation, le contrôle analytique de ces médicaments représente de nos jours une étape incontournable pour garantir le bon produit à la bonne dose avant administration. Différentes stratégies de contrôle ont été développées pour garantir leur qualité physico-chimique. Malgré leurs performances, ces techniques exploitant les propriétés ultra-violettes des molécules ne permettent pas un contrôle exhaustif des préparations.

L’objectif de ce travail était d’évaluer la faisabilité d’une technique complémentaire, la spectroscopie Raman exaltée de surface pour l’analyse des médicaments anticancéreux. Les essais ont été réalisés pour l’analyse du 5-fluorouracile (5FU), l’une des molécules les plus utilisées.

2 Matériel et méthodes

2.1 Principe de la diffusion Raman exaltée de surface

La spectroscopie Raman exaltée de surface (*Surface Enhanced Raman Scattering*, SERS) est une spectroscopie vibrationnelle de la matière. Elle associe l’utilisation de nanoparticules (Nps) métalliques à la spectroscopie Raman conventionnelle permettant ainsi d’exalter le signal des molécules étudiées et de mesurer leurs spectres Raman avec une grande sensibilité. Depuis les années 70, son application à la caractérisation de molécules en solution notamment aqueuse très diluée a beaucoup évolué. Avec l’avènement des nanotechnologies, de nombreux substrats ont été développés et étudiés. Les substrats peuvent être des surfaces métalliques spécifiques nanostructurées ou des colloïdes de nanoparticules. Selon les substrats, plusieurs types d’exaltation basés sur des phénomènes de nature électromagnétique et/ou chimique ont été décrits. Les Nps les plus décrites sont des Nps de métal noble, par exemple d’or ou d’argent, dispersées en solution (substrats colloïdaux), ou bien encore des substrats solides voire des substrats hybrides combinant plusieurs types de Nps.

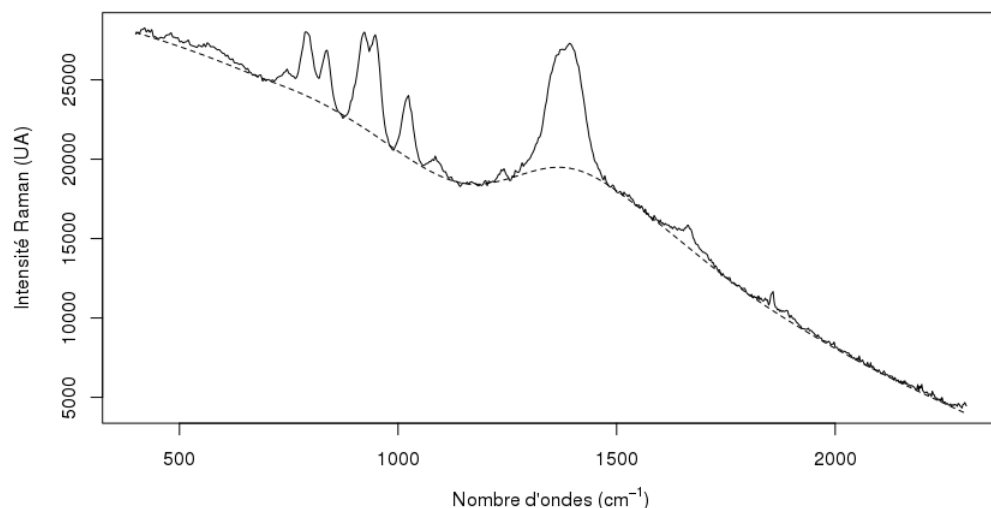


FIGURE 1 – Représentation d’un spectre brute (trait plein) et d’une estimation de sa ligne de base (ligne brisée)

2.2 Expérimentation

Au vu de leur stabilité et de leur reproductibilité de synthèse, l’utilisation des Nps d’argent a été privilégié. Les Nps d’argent de taille moyenne 50 nm ont été synthétisées selon la technique de Lee & Meisel (1982) par réduction chimique du nitrate d’argent par du citrate. Plusieurs solutions de concentrations croissantes (de $0,5$ à 12 mg/mL) correspondant aux concentrations thérapeutiques en 5-fluorouracile ont été préparées par dilution d’une solution concentrée dans une matrice aqueuse.

Les acquisitions ont été effectuées au sein de la plateforme de spectroscopie vibrationnelle de l’équipe Lip(Sys)² de l’Université Paris Sud sur un spectrophotomètre Raman portable équipé d’une diode laser à 785 nm (MIRA-1 Methrom) entre 400 à 2300 cm^{-1} . L’analyse SERS a permis de mettre en évidence douze bandes spectrales attribuables au 5FU avec une exaltation du signal Raman la solution à 1 mg/mL jusqu’à un facteur 10^4 pour la bande à 786 cm^{-1} caractéristique de la vibration du cycle pyrimidique du 5FU.

Pour notre analyse statistique, 63 échantillons ont été préparés (7 séries \times 9 concentrations), chaque série ayant été réalisée à des jours différents. Pour chaque échantillon, 3 analyses SERS ont été répétées avec 3 spectres acquis par répétition. Au total, 567 spectres ont été acquis.

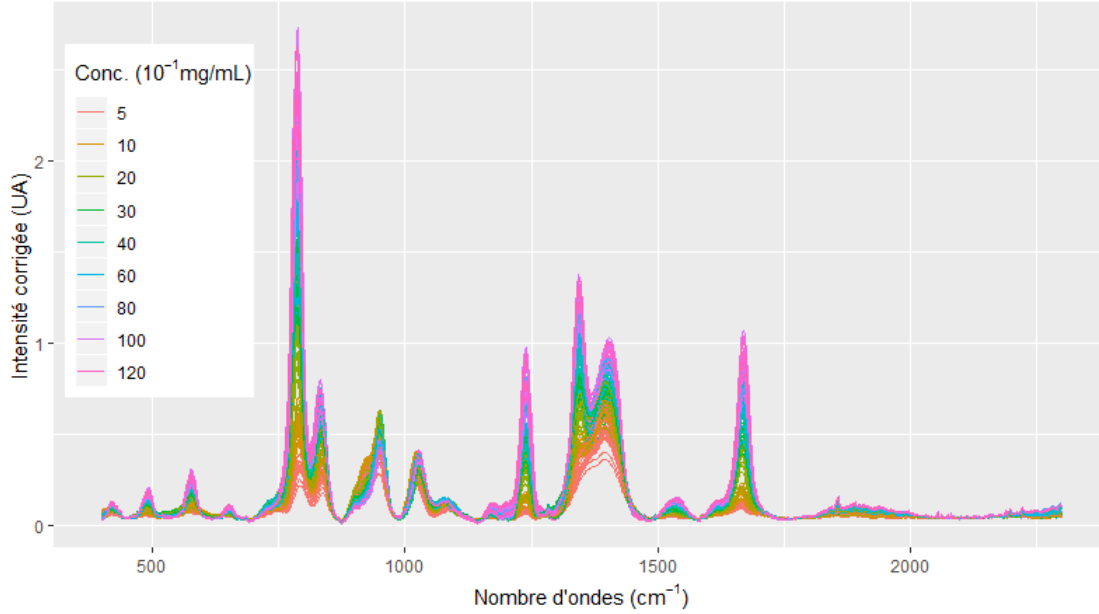


FIGURE 2 – Représentation de l’ensemble des 567 spectres corrigés en fonction des concentrations théoriques

3 Modèle de régression proposé

Au vu des fortes variabilités spectrales, les spectres moyens ont été calculés sur les 3 acquisitions pour réduire le bruit de fond puis corrigés individuellement (voir Figure 1) par soustraction d’une estimation de leur ligne de base (voir p. ex. Eilers & Boelens, (2005)). Une interaction non-linéaire ayant été observée entre la concentration et l’intensité du signal (qu’on pourra observer sur les Figures 2 et 3), des méthodes de régressions non-linéaires ont été mises en place.

Soient I le nombre de séries, J le nombre de concentrations, K de répétitions et L de longueurs d’onde. Pour $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$ et $\ell = 1, \dots, L$, nous considérons y_{ijkl} la variable réponse correspondant à l’intensité du signal Raman corrigé relative à la série i , la concentration c_j , la répétition k et la longueur d’onde λ_ℓ . Pour certaines longueurs d’onde λ_ℓ , $\ell \in \{1, \dots, L\}$, nous avons observé une dépendance (non-linéaire) forte entre la variable réponse y_{ijkl} et la concentration du 5FU (voir p.ex. Figure 3). Une modélisation de l’intensité du signal en fonction de la concentration du 5FU par une fonction sigmoïde nous apparait ici pertinente. Plus précisément nous avons considéré des modèles de régression de la forme

$$y_{ijkl} = f(c_j, \theta_{i\ell}) + e_{ijkl}, \quad (1)$$

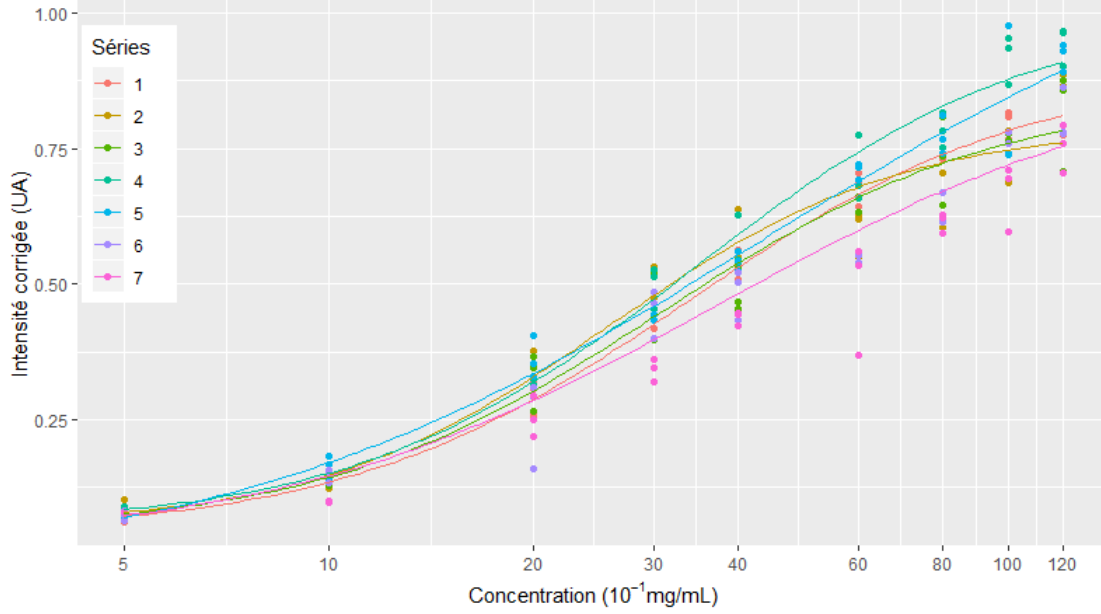


FIGURE 3 – Représentation du modèle de calibration pour l’un des nombres d’onde (1239 cm^{-1}) pour les 7 séries de mesure

où pour $\theta_{i\ell} = (S_{i\ell}, A_{i\ell}, \gamma_{i\ell}, \tau_{i\ell}) \in \mathbb{R}^4$, f est la sigmoïde définie par

$$f(c, \theta_{i\ell}) = S_{i\ell} + \frac{(A_{i\ell} - S_{i\ell})}{1 + \exp(-\gamma_{i\ell}(\log(c) - \tau_{i\ell}))}, \quad c \in \mathbb{R}^+, \quad i \in \{1, \dots, I\}, \quad \ell \in \{1, \dots, L\}.$$

Les résidus e_{ijkl} du modèle (1) sont supposés indépendants, de loi normale centrée et de variance $\text{var}(e_{ijkl}) = a_{i\ell}^2 + b_{i\ell}^2 f(c, \theta_{i\ell})^2$. Plusieurs choix ont été considérés et comparés pour le modèle structurel des erreurs : un modèle d’erreurs constant c.-à-d. $b_{i\ell}^2 = 0$, un modèle d’erreurs proportionnel c.-à-d. $a_{i\ell}^2 = 0$ et enfin une combinaison du modèle d’erreurs constant et proportionnel pour lequel $a_{i\ell}^2$ et $b_{i\ell}^2$ peuvent être différents de 0. Dans le modèle (1), les vecteurs de paramètres $(S_{i\ell}, A_{i\ell}, \gamma_{i\ell}, \tau_{i\ell}, a_{i\ell}^2, b_{i\ell}^2)$ étant inconnus pour $i \in \{1, \dots, I\}$ et $\ell \in \{1, \dots, L\}$, ils sont estimés de façon à maximiser la vraisemblance du modèle statistique. Dans la Figure 3, nous avons représenté pour un nombre d’onde choisi, la calibration du modèle réalisée sur chaque série de mesure constituant la base d’apprentissage.

Ces vecteurs de paramètres estimés sur la base d’apprentissage sont ensuite utilisés dans le problème inverse pour prédire la concentration de nouveaux échantillons de la base test. Cette approche repose sur l’analyse des ”plus proches voisins” par rapport aux $(I - 1)$ séries constituant la base d’apprentissage. La concentration pour un nouveau spectre et ses K répétitions est estimée par la valeur qui minimise la plus petite des erreurs quadratiques

moyennes (par rapport aux séries constituant la base d'apprentissage), calculées entre les intensités mesurées $(y_{k\ell}^{new})_{\ell=1,\dots,L}$ des K nouveaux spectres et les intensités prédites par le modèle (1), c.-à-d., dans le modèle d'erreurs combiné :

$$\hat{c} = \arg \min_{c \in \mathbb{R}^+} \min_{i \in \{1, \dots, I-1\}} \sum_{\ell=1}^L \sum_{k=1}^K \frac{(y_{k\ell}^{new} - f(c, \hat{\theta}_{i\ell}))^2}{\hat{a}_{i\ell}^2 + \hat{b}_{i\ell}^2 f(c, \hat{\theta}_{i\ell})^2}. \quad (2)$$

Les performances prédictives du modèle sont dès lors évaluées par validation croisée. Plus précisément, $(I - 1)$ séries sont successivement incluses dans le set d'apprentissage pour calibrer le modèle (1) qui est ensuite utilisé pour prédire les concentrations des données de la série restante constituant le set de test. Sur la base des concentrations prédites par l'équation (2), les erreurs relatives absolues par rapport aux valeurs nominales sont calculées. En vue de garantir la qualité des médicaments, une erreur relative maximale de 7,5% sur le set de test sera tolérée.

Bibliographie

Eilers, P.H.C. & Boelens, H.F.M., (2005) Baseline correction with asymmetric least squares smoothing, Leiden University Medical Centre Report. 1, 5.

Lee, P. C. & Meisel, D. (1982) *Adsorption and surface-enhanced Raman of dyes on silver and gold sols*. The Journal of Physical Chemistry 86, 3391–3395 .

Le, L.M.M., Kégl, B., Gramfort, A., & al. (2018) *Optimization of classification and regression analysis of four monoclonal antibodies from Raman spectra using collaborative machine learning approach*. Talanta.184, 260-265.