Approche bayésienne pour la décomposition conjointe d'une séquence de spectres de photoélectrons

Vincent MAZET[1], Sylvain FAISAN[1], Antoine MASSON[2], Marc-André GAVEAU[2], Lionel POISSON[2]

[1]LSIIT, Université de Strasbourg , CNRS–UMR 7005
Boulevard Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

[2]Laboratoire Francis Perrin, CNRS–URA 2453, CEA, IRAMIS
Service des Photons Atomes et Molécules, 91191 Gif-sur-Yvette Cedex, France.
{vincent.mazet,faisan}@unistra.fr, {prenom.nom}@cea.fr

Résumé – Ce travail traite de la décomposition d'une séquence temporelle de spectres de photoélectrons en une somme de raies dont on estime les positions, amplitudes et largeurs. Comme les raies évoluent lentement dans le temps, la décomposition est effectuée sur toute la séquence afin de prendre en compte cette information temporelle. À cette fin, nous avons développé un modèle bayésien où un champ de Markov gaussien favorise une évolution douce des raies. L'approche est non-supervisée et un échantillonneur de Gibbs couplé à un schéma de recuit simulé permet d'estimer le maximum a posteriori. Nous montrons la pertinence de cette approche par rapport à une méthode dans laquelle les spectres sont décomposés séparément et présentons une application sur données réelles de photoélectrons.

Abstract – This work deals with the decomposition of a temporal sequence of photoelectron spectra into a sum of peaks whose positions, amplitudes and widths are estimated. Since the peaks exhibit a slow evolution with time, the decomposition is performed on the whole sequence to take this temporal information into account. To this end, we have developed a Bayesian model where a Gaussian Markov random field favors a smooth evolution of peaks. The approach is unsupervised and a Gibbs sampler within a simulated annealing scheme enables to estimate the maximum a posteriori. We show the relevance of this approach compared with a method in which the spectra are decomposed separately and present an application on real photoelectron data.

1 Introduction

Un photoélectron est un électron émis d'un échantillon de matière suite à l'absorption d'une radiation électromagnétique. La distribution de ces électrons en fonction de leur énergie est appelée spectre de photoélectrons ; il peut être modélisé comme une somme de raies (qui informe sur la distribution d'énergie des électrons du système physique considéré) superposé sur un continuum [8]. Le système considéré dans la section 5 est un unique atome de barium (Ba) au sein d'un groupe de plusieurs centaines d'atomes d'argon (Ar). La distribution énergétique du système Ba-Ar$_n$ est mesurée en fonction du temps, résultant ainsi en une séquence temporelle de spectres de photoélectrons. L'objectif de cet article est de déterminer si l'énergie des raies évolue dans le temps, indiquant ainsi une solvatation progressive du barium dans le système.

Notre objectif est de décomposer chaque spectre de la séquence, c'est-à-dire d'estimer les centres, amplitudes et largeurs des raies, ainsi que le continuum. Ce n'est pas un problème de séparation de source ou de démelange spectral car les données ne peuvent pas être modélisées comme une somme de signaux sources ; c'est en revanche un problème de décomposition puisque chaque spectre est une somme de raies dont les positions et formes peuvent varier.

Plusieurs méthodes de décomposition *paramétrique* d'un spectre ont déjà été proposées [1, 5, 6, 12]. L'idée principale des trois premières méthodes, que nous reprenons dans notre travail, est d'estimer les paramètres des raies d'un spectre dans un cadre bayésien et à l'aide de méthodes de Monte Carlo par chaîne de Markov (MCMC). La dernière approche utilise une technique d'approximation parcimonieuse. Or, à notre connaissance, aucun travail n'a traité le problème de la décomposition d'une séquence de spectres. Nous avons récemment proposé [9] une telle méthode dans un cadre bayésien et montré qu'une approche séquentielle, dans laquelle les spectres sont décomposés indépendamment les uns des autres, n'est pas appropriée. En effet, la décomposition de deux spectres contigus (c'est-à-dire enregistrés à deux temps adjacents) peut aboutir à deux décompositions très différentes alors que les spectres sont très similaires. Nous avons également montré l'intérêt d'une décomposition conjointe qui prend en compte le fait que les raies évoluent dans le temps : en effet, la décomposition d'un spectre est guidée par les décompositions voisines grâce à l'a priori d'évolution douce des paramètres ; de plus les décompositions sont cohérentes car elles évoluent doucement ; enfin, une décomposition conjointe fournit une classification des raies, fournissant ainsi la possibilité de suivre les raies dans la séquence. Cependant, l'approche proposée dans [9] n'a été

validée que sur données simulées et plusieurs limitations sont apparues lors de l'application sur données réelles.

En effet, les données réelles contiennent plus de spectres que les données synthétiques utilisées dans [9] : cela conduit à une augmentation significative de la taille de l'espace de recherche et l'algorithme proposé dans [9] échoue ; une solution est présentée dans la section 3. En outre, le continuum n'a pas été modélisé dans [9], nous proposons ici de le modéliser par une fonction exponentielle (équation 1). Enfin, la méthode proposée dans [9] requiert le réglage d'hyperparamètres. La variabilité des données réelles rend ce réglage difficile et c'est pourquoi la méthode proposée dans cet article est non supervisée.

Le modèle et les a priori sont détaillés dans la section 2. L'estimateur du maximum a posteriori (MAP) est obtenu à l'aide d'un algorithme MCMC couplé à une approche de recuit simulé (section 3). Finalement, les performances de la méthode sont illustrées sur des données synthétiques (section 4) et réelles (section 5).

2 Modèle bayésien

Les données à traiter correspondent à S spectres composés chacun de N échantillons. Un spectre $\boldsymbol{y}_s$ ($s \in \{1,\dots,S\}$) est modélisé comme la somme bruitée de K raies gaussiennes et d'une fonction exponentielle [8] :

$$(\boldsymbol{y}_s)_n = \sum_{k=1}^{K} a_{s,k} \exp\left(-\frac{(n-c_{s,k})^2}{2w_{s,k}^2}\right) + \alpha_s \exp\left(-\frac{n}{\beta_s}\right) + (\boldsymbol{b}_s)_n, \tag{1}$$

où $(\boldsymbol{y}_s)_n$ est le n-ème élément ($n \in \{1,\dots,N\}$) du vecteur $\boldsymbol{y}_s$. Les paramètres de la k-ème raie sont son centre $c_{s,k}$, son amplitude $a_{s,k}$ et sa largeur $w_{s,k}$, les paramètres du continuum sont α_s et β_s et $\boldsymbol{b}_s$ est un buit additif modélisant les erreurs de modèle et de mesure. Le nombre K de raies est supposé connu et identique pour chaque spectre. Enfin, dans le but de pouvoir suivre l'évolution des raies au cours de la séquence et de régulariser leur évolution, il est nécessaire d'étiqueter chaque raie avec la variable $z_{s,k} \in \{1,\dots,K\}$.

Pour définir les distributions a priori des paramètres des raies et du continuum, il faut définir des vecteurs qui représentent leur parcours temporel. Ces derniers sont notés $\boldsymbol{c}^1,\dots,\boldsymbol{c}^K$, $\boldsymbol{a}^1,\dots,\boldsymbol{a}^K$, $\boldsymbol{w}^1,\dots,\boldsymbol{w}^K$, $\boldsymbol{\alpha}$, et $\boldsymbol{\beta}$ où l'exposant dénote les paramètres de même étiquette (ainsi par exemple : $\boldsymbol{c}^l = \{c_{s,k} | \; z_{s,k} = l, \forall s, \forall k\}$). Dans la suite, nous noterons $\boldsymbol{\theta}$ l'un de ces $3K+2$ vecteurs.

2.1 Lois a priori

Le bruit est supposé blanc gaussien de moyenne nulle et de variance $r_{\boldsymbol{b}}$:

$$\forall s, n, \quad (\boldsymbol{b}_s)_n | r_{\boldsymbol{b}} \sim \mathcal{N}(0, r_{\boldsymbol{b}}).$$

Les composantes de $\boldsymbol{\theta}$ évoluant lentement, un a priori de douceur permet de régulariser la solution : on propose donc de modéliser $\boldsymbol{\theta}$ à l'aide d'un champ markovien gaussien (GMRF : *Gaussian Markov random field*) [3] :

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto \frac{1}{r_{\boldsymbol{\theta}}^{S/2}} \exp\left(-\frac{1}{2r_{\boldsymbol{\theta}}}\|D\boldsymbol{\theta}\|^2\right) \mathbb{1}_{\Theta}(\boldsymbol{\theta}), \tag{2}$$

où $r_{\boldsymbol{\theta}}$ est un hyperparamètre qui caractérise l'évolution du paramètre $\boldsymbol{\theta}$, $\|\cdot\|$ représente la norme L^2, D définit une dérivée discrète (une dérivée d'ordre un pénalise toute évolution, une dérivée d'ordre deux favorise une évolution linéaire), et $\mathbb{1}_{\Theta}(\boldsymbol{\theta})$ est la fonction indicatrice ($\mathbb{1}_{\Theta}(\boldsymbol{\theta}) = 1$ si $\boldsymbol{\theta} \in \Theta$, 0 sinon) qui permet de définir le support du paramètre $\boldsymbol{\theta}$. L'espace Θ est $[1, N]$ pour les centres [1] et $\mathbb{R}^+$ pour les autres paramètres. Enfin, les paramètres $\boldsymbol{\theta}$ sont supposés mutuellement indépendants.

Les étiquettes sont supposées être distribuées uniformément sur l'ensemble $\mathcal{S}_K$ des permutations de $\{1,\dots,K\}$:

$$\forall s, \qquad \boldsymbol{z}_s \sim \mathcal{U}_{\mathcal{S}_K}.$$

La variance du bruit $r_{\boldsymbol{b}}$ suit une loi inverse gamma :

$$r_{\boldsymbol{b}} \sim \mathcal{IG}(\epsilon, \epsilon)$$

(ϵ est très petit afin que l'a priori ressemble à la distribution de Jeffreys traditionnellement utilisée mais malheureusement impropre).

Enfin, l'hyperparamètre $r_{\boldsymbol{\theta}}$ (équation (2)) fait référence à $r_{\boldsymbol{c}}$, $r_{\boldsymbol{a}}$, $r_{\boldsymbol{w}}$ (hyperparamètres relatifs à l'évolution de la position, de l'amplitude et de la largeur des raies au cours des séquences), $r_{\boldsymbol{\alpha}}$ ou $r_{\boldsymbol{\beta}}$ (hyperparamètres relatifs à l'évolution du continuum). Comme ces paramètres doivent être faibles pour permettre une évolution en douceur, nous choisissons un a priori similaire à celui utilisé pour la variance du bruit : $r_{\boldsymbol{\theta}} \sim \mathcal{IG}(\epsilon, \epsilon)$.

2.2 Lois a posteriori

La loi a posteriori globale est échantillonnée à l'aide d'un échantillonneur de Gibbs (section 3) qui consiste à simuler chaque variable selon sa loi a posteriori conditionnelle [11]. Celles-ci sont présentées brièvement ci-dessous (voir [9] pour une présentation détaillée).

Les centres, positions et β_s sont distribués selon

$$c_{s,k}| \dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{c}^l\|^2}{2r_{\boldsymbol{c}}}\right) \mathbb{1}_{[1,N]}(c_{s,k}),$$

$$w_{s,k}| \dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{w}^l\|^2}{2r_{\boldsymbol{w}}}\right) \mathbb{1}_{\mathbb{R}^+}(w_{s,k}),$$

$$\beta_s| \dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{\beta}\|^2}{2r_{\boldsymbol{\beta}}}\right) \mathbb{1}_{\mathbb{R}^+}(\beta_s)$$

avec $l = z_{s,k}$ et $\boldsymbol{e}_s$ est l'erreur de reconstruction.

Les amplitudes et α_s sont distribués suivant une loi normale à support positif :

$$a_{s,k}| \dots \sim \exp\left(-\frac{(a_{s,k}-\mu_{s,k})^2}{2\rho_{s,k}}\right) \mathbb{1}_{\mathbb{R}^+(a_{s,k})},$$

$$\alpha_s| \dots \sim \exp\left(-\frac{(\alpha_s-\lambda_s)^2}{2\nu_s}\right) \mathbb{1}_{\mathbb{R}^+}(\alpha_s)$$

1. L'espace $[1, N]$ est suffisamment grand par rapport aux faibles valeurs de $r_{\boldsymbol{c}}$ pour considérer la troncature négligeable et considérer que la densité de probabilité réelle $p(\boldsymbol{c}^l|\boldsymbol{z})$ peut être approximée par l'équation (2).

où $\mu_{s,k}$, $\rho_{s,k}$, $\lambda_{s,k}$ et $\nu_{s,k}$ dépendent de $\boldsymbol{e}_s$ et des autres variables [9].

Les étiquettes sont distribuées selon

$$\boldsymbol{z}|\cdots \sim \prod_s \exp\left[-\sum_l \left[\frac{\|D\boldsymbol{c}^l\|^2}{2r_{\boldsymbol{c}}}+\frac{\|D\boldsymbol{a}^l\|^2}{2r_{\boldsymbol{a}}}+\frac{\|D\boldsymbol{w}^l\|^2}{2r_{\boldsymbol{w}}}\right]\right]\mathbb{1}_{\mathcal{S}_K}(\boldsymbol{z}_s).$$

Enfin, les a posteriori des hyperparamètres sont :

$$\begin{aligned}
r_{\boldsymbol{b}}|\ldots &\sim \mathcal{IG}\left(NS/2+\epsilon, \textstyle\sum_l \|\boldsymbol{e}_s\|^2/2+\epsilon\right),\\
r_{\boldsymbol{c}}|\ldots &\sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l \|D\boldsymbol{c}^l\|^2/2+\epsilon\right),\\
r_{\boldsymbol{a}}|\ldots &\sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l \|D\boldsymbol{a}^l\|^2/2+\epsilon\right),\\
r_{\boldsymbol{w}}|\ldots &\sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l \|D\boldsymbol{w}^l\|^2/2+\epsilon\right),\\
r_{\boldsymbol{\alpha}}|\ldots &\sim \mathcal{IG}\left(S/2+\epsilon, \|D\alpha\|^2/2+\epsilon\right),\\
r_{\boldsymbol{\beta}}|\ldots &\sim \mathcal{IG}\left(S/2+\epsilon, \|D\beta\|^2/2+\epsilon\right).
\end{aligned}$$

3 Algorithme MCMC et recuit simulé

Un échantillonneur de Gibbs permet d'échantillonner la loi a posteriori car c'est un algorithme souvent utilisé avec des GMRF ainsi qu'en décomposition de spectres [1, 5, 6]. Il consiste à simuler chaque variable suivant sa loi a posteriori conditionnelle. Ainsi, les paramètres $c_{s,k}$, $w_{s,k}$ et β_s sont échantillonnés suivant un algorithme de Metropolis-Hastings à marche aléatoire [11]. Les paramètres $a_{s,k}$, α_s et les hyperparamètres sont échantillonnés directement. Enfin, les étiquettes sont échantillonnées à l'aide de l'algorithme symétrique de Metropolis-Hastings suivant : les étiquettes proposées sont identiques aux étiquettes courantes, sauf que les étiquettes k_1 et k_2 entre $s = s_1$ et $s = s_2$ sont permutées ; k_1, k_2, s_1 et s_2 étant uniformément choisis et 100 candidats sont proposés à chaque balayage de l'échantillonneur de Gibbs.

Or, en pratique, l'échantillonneur de Gibbs peut rester coincer dans un minimum local [11], c'est pourquoi il est inclus dans un schéma de recuit simulé [7, 2]. De cette manière, l'estimation du MAP est simplement l'échantillon de la chaîne de Markov qui maximise la loi a posteriori. Une descente de température géométrique est utilisée avec des températures initiales et finales respectives de 10 et $0,1$. Toutefois, le recuit simulé n'est pas implementé sur les variables $a_{s,k}$ et α_s ainsi que sur les hyperparamètres car il peut introduire des problèmes numériques et de divergence. Aussi, cette approche peut être vue comme une méthode de recuit simulé avec différentes températures [10].

4 Résultats sur données simulées

Pour quantifier les apports de la méthode proposée (DCNS : décomposition conjointe non-supervisée), nous la comparons aux méthodes de décomposition séquentielle (DS) et de décomposition conjointe supervisée (DCS) présentées dans [9]. La seule modification apportée à ces deux dernières méthodes est la prise en compte du continuum. Ainsi, les hyperparamètres sont fixés comme indiqués dans [9] ; la DS ne permet évidemment pas d'estimer les étiquettes des raies ni d'introduire une douceur sur leur évolution ; et la DCS simule les étiquettes avec l'algorithme de [9]. Chaque méthode effectue l'estimation grâce à un échantillonneur de Gibbs couplé à du recuit simulé sur $5\,000$ itérations afin d'estimer le MAP (utiliser plus d'itérations n'a pas conduit à une amélioration significative des résultats). Les trois méthodes ont été testées sur 30 séquences simulées avec $S = 10$, $K = 4$ et D est la dérivée seconde. Le code Matlab et les données sont disponibles librement (lsiit-miv.u-strasbg.fr/mazet/jointdec). Les performances ont été comparées en calculant l'erreur quadratique moyenne (EQM) entre les signaux observés et reconstruits.

Les DS et DCNS obtiennent des EQM respectives de $11,54 \cdot 10^{-2}$ et $3,78 \cdot 10^{-2}$. On peut expliquer ces résultats par le fait que la DS, qui ne peut pas tenir compte des estimations obtenues sur les spectres voisins, peut manquer des raies alors que l'a priori de douceur de la DCNS favorise la recherche de raies dans certaines zones de l'espace : en d'autres termes, l'estimation obtenue pour un spectre aide l'estimation des spectres voisins.

La DCNS surpasse également la DCS (dont l'EQM est $7,40 \cdot 10^{-2}$). Cela est dû au fait que le réglage des hyperparamètres n'est pas simple. En effet, les hyperparamètres contrôlent l'importance relative des termes de données et de régularisation et cela implique qu'ils doivent être réglés en fonction du rapport signal-sur-bruit et de l'évolution attendue des raies. Cela met clairement en évidence les apports de la méthode proposée.

5 Résultats sur spectres de photoélectrons

La séquence de spectres de photoélectrons (figure 2) est constituée de 44 spectres (couvrant une durée de $3,47$ ps), chacun étant constitué de $N = 182$ mesures (de $0,02$ eV à $2,52$ eV) avec des échantillonnages en temps (s) et en énergie (n) irréguliers. Sur la figure 2, l'axe horizontal correspond à s et l'axe vertical à n et on cherche à quantifier l'évolution des raies en fonction de s. L'approche proposée a été exécutée avec $K = 4$ raies, $I = 10^4$ itérations et en utilisant une dérivée d'ordre un (privilégiant donc aucune évolution, équation (2)). Les résultats sont représentés figures 1 et 2 : on y voit que l'évolution des paramètres est assez lisse et que les raies sont bien classées (il n'y a pas de permutation entre les raies). Cela valide l'algorithme utilisé pour l'échantillonnage des étiquettes alors que la méthode proposée dans [9] ne fournit pas de bons résultats lorsque S est grand. Il apparaît également que les raies évoluent plus doucement que si l'estimation avait été séquentielle, telle qu'elle a été implémentée dans [8] (résultat non présenté). On s'aperçoit par ailleurs que l'énergie d'une raie varie dans le temps : son énergie augmente légèrement de $1,18$ à $1,20$ eV entre 0 et $0,4$ ps puis elle diminue et atteint $1,05$ eV à $2,5$ ps. Après $2,5$ ps, la raie a une amplitude très faible et une largeur très importante, ce qui est typique d'un pic qui dis-

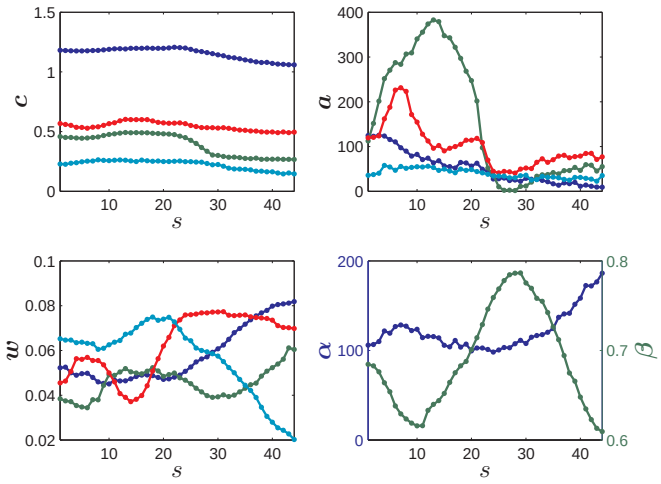

FIGURE 1 – Paramètres estimés : centres $\boldsymbol{c}$, amplitudes $\boldsymbol{a}$ et largeurs $\boldsymbol{w}$ des raies ; paramètres α et β du continuum. Chaque couleur correspond à une étiquette.

parait. C'est une limitation de l'approche dans la mesure où nous avons fait l'hypothèse, qui n'est pas toujours vérifiée, que le nombre de raies est constant dans le temps. Malgré tout, les résultats permettent de confirmer quantitativement les observations qualitatives exprimées dans [8], à savoir que l'énergie d'un niveau électronique du système Ba-Ar$_n$ varie dans le temps. Cela suggère que le niveau énergétique correspondant est très sensible à l'environnement des atomes d'argon, et que le barium est solvaté dans le groupe d'atomes d'argon.

6 Conclusion

Nous avons présenté une approche originale pour estimer les paramètres des raies dans une séquence de signaux spectroscopiques. L'idée principale est d'effectuer cette décomposition en traitant la séquence entière plutôt que chaque spectre séparément. Cela est fait en utilisant un modèle bayésien avec un GMRF pour adoucir l'évolution des paramètres estimés. Cette approche est complètement non-supervisée et la solution est obtenue à l'aide d'un algorithme MCMC combiné à un schéma de recuit simulé. Des simulations ont montré la pertinence de cette approche et les résultats obtenus sur des spectres de photoélectrons ont confirmé le point de vue des experts. Les futurs travaux se concentreront principalement sur l'estimation du nombre de raies à l'aide de l'algorithme RJMCMC (*reversible jump MCMC*) [4].

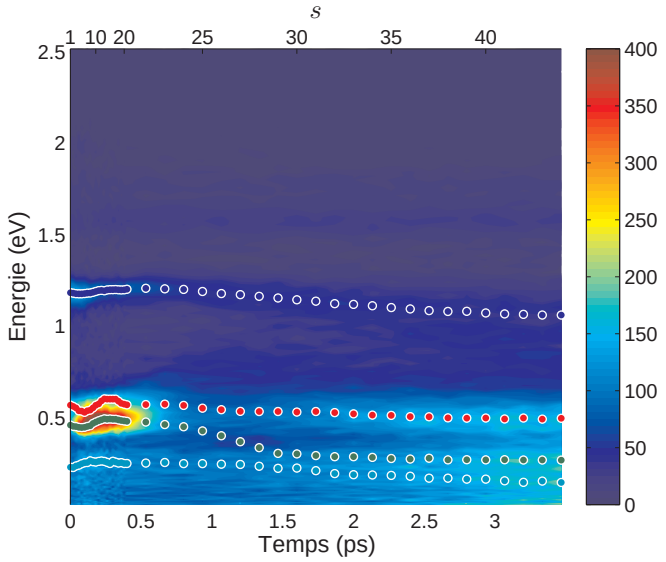

FIGURE 2 – Spectres de photoélectron : les données sont tracées en fonction de l'énergie des électrons (n, axe vertical) et du temps (s, axe horizontal) et les valeurs des spectres sont indiquées par l'échelle de couleur. Les points correspondent aux centres estimés des quatre raies et les couleurs de ces points représentent leur étiquette.

Références

[1] R. Fischer et V. Dose. « Analysis of mixtures in physical spectra ». In *Bayesian methods*, p. 145–154, 2001.

[2] S. Geman et D. Geman. « Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, p. 721–741, 1984.

[3] D. Geman et C. Yang. « Nonlinear image recovery with half-quadratic regularization ». *IEEE Transactions on Image Processing*, vol. 4, n°7, p. 932–946, 1995.

[4] P.J. Green. « Reversible jump Markov chain Monte Carlo computation and Bayesian model determination ». *Biometrika*, vol. 82, p. 711–732, 1995.

[5] S. Gulam Razul, W.J. Fitzgerald et C. Andrieu. « Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC ». *Nuclear Instruments and Methods in Physics Research A*, vol. 497, p. 492–510, 2003.

[6] N.M. Haan et S.J. Godsill. « Bayesian models for DNA sequencing ». *IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. IV–4020–IV–4023, 2002.

[7] S. Kirkpatrick, C.D. Gelatt Jr. et M.P. Vecchi. « Optimization by simulated annealing ». *Science*, vol. 220, p. 671–680, 1983.

[8] A. Masson, L. Poisson, M.-A. Gaveau, B. Soep, J.-M. Mestdagh, V. Mazet et F. Spiegelman. « Dynamics of highly excited barium atoms deposited on large argon clusters. I. General trends ». *The Journal of Chemical Physics*, vol. 133, n°5 (054307), 2010.

[9] V. Mazet. « Joint Bayesian decomposition of a spectroscopic signal sequence ». *IEEE Signal Processing Letters*, vol. 18, n°3, p. 191–184, 2011.

[10] B. Perret, V. Mazet, C. Collet et É. Slezak. « Hierarchical multispectral galaxy decomposition using a MCMC algorithm with multiple temperature simulated annealing ». *Pattern Recognition*, vol. 44, n°6, p. 1328–1342, 2011.

[11] C.P. Robert et G. Casella. *Monte Carlo statistical methods*. Springer, 2^e édition, 2004.

[12] S. Sahnoun, E.-H. Djermoune, C. Soussen, D. Brie. « Sparse multiresolution modal estimation ». *IEEE Statistical Signal Processing Workshop*, Nice, 2011.