

Une approche bayésienne non paramétrique à la reconstruction d’images spatiales 3D en TEP clinique

Mame Diarra FALL¹, Éric BARAT², Claude COMTAT³, Thomas DAUTREMER², Thierry MONTAGU², Simon STUTE³

¹Laboratoire de Mathématiques Appliquées à Paris 5, Université Paris Descartes and CNRS, Sorbonne Paris Cité.

²Laboratoire de Modélisation, Simulation et Systèmes (LM2S), CEA Saclay, 91191 Gif-sur-Yvette, France.

³ Service Hospitalier Frédéric Joliot, CEA/DSV/I2BM, 91401 Orsay, France.
diarra.fall@parisdescartes.fr, eric.barat@cea.fr

Résumé – Dans cette contribution, nous abordons le problème de la reconstruction d’images spatiales 3D en tomographie par émission de positons (TEP), à partir de données réelles. Nous formulons le problème de reconstruction dans un cadre bayésien non paramétrique. L’approche est *non paramétrique* au sens où l’image à reconstruire est vue comme une densité de probabilité dans \mathbf{R}^3 et inféré directement à partir des données, et *bayésienne* puisque nous mettons une loi *a priori* sur cette densité. Puis, nous approchons la loi *a posteriori* par un algorithme basé sur les méthodes de Monte-Carlo par chaînes de Markov (MCMC). La méthode proposée est évaluée sur des données cliniques et comparée avec celle fournie par le code constructeur. Cette dernière est basée sur la méthode du maximum de vraisemblance utilisant l’algorithme EM (ML-EM).

Abstract – In this contribution, we address the problem of 3D Positron Emission Tomography (PET) reconstruction, from real data. We cast the reconstruction problem in a Bayesian nonparametric context. The approach is *nonparametric* in the sense that the image to reconstruct is viewed as a probability density on \mathbf{R}^3 and directly inferred from the data, and it is *Bayesian* since we define a prior on this density. The posterior is approximated using a Markov Chain Monte-Carlo (MCMC) algorithm. The ability of the proposed model is assessed on data from clinical studies and we compare our results to those obtained with the constructor code. This latter is based on the maximum likelihood approach using EM algorithm (ML-EM).

1 Introduction

La tomographie par émission de positons (TEP) est une technique d’imagerie médicale fonctionnelle permettant de visualiser l’activité métabolique d’un organe. Elle est utilisée en oncologie, en neurologie, en cardiologie etc. L’une des contraintes majeures de la TEP réside dans l’administration au patient d’un traceur radioactif. Ainsi, pour favoriser l’émergence de cette technique à un plus grand public, il serait utile de développer des méthodes de reconstruction qui soient plus robustes dans un cadre de faibles doses. Dans nos travaux préliminaires [1], nous avons démontré la pertinence d’une approche bayésienne non paramétrique basée sur les mélanges par processus de Dirichlet pour la reconstruction d’images 3D, dans un contexte de dose relativement faible et sur des données simulées idéales. Dans cette contribution, nous abordons la reconstruction d’images cliniques en 3D. Cela nécessite toutefois une prise en compte de tous les phénomènes physiques biaisant les mesures, alourdissant ainsi la modélisation statistique. De plus, la spécificité de la géométrie du scanner (Biograph, Siemens HealthcareTM) doit être prise en compte.

Les résultats présentés dans la Section 4 montrent que, dans un contexte de dose réduite, la méthode de reconstruction proposée fournit des résultats compétitifs avec celle utilisée dans les plateformes cliniques (ML-EM) pour une dose normale. De plus,

l’erreur associée à notre estimée est quantifiée via l’écart-type.

2 Méthodologie

2.1 Formulation du problème

La TEP repose sur le principe physique de l’émission de positons. Le traceur administré au patient émet des positons qui s’annihilent avec des électrons du milieu. Cette annihilation produit deux photons qui partent simultanément en sens opposé et sont détectés à l’extérieur du patient. La détection simultanée de deux photons dans une fenêtre temporelle de quelques nanosecondes (dite détection en *coïncidence*) suppose qu’ils proviennent de la même annihilation et que cette dernière a eu lieu sur la ligne virtuelle joignant les centres des deux détecteurs (appelée ligne de réponse et notée LOR).

Néanmoins, ce modèle physique idéal sur lequel est basé la TEP est perturbé par des phénomènes sources de biais dans les mesures. En effet, la majeure partie des photons détectés sont des événements dits *simples*, l’autre photon d’annihilation pouvant être absorbé dans les tissus, non dirigé vers le système de détection, pas détecté etc. Lorsque deux événements simples sont détectés dans la même fenêtre temporelle, on parle de *coïncidences fortuites ou aléatoires*. De plus, lors de leur parcours

dans l'organisme, l'un des deux photons peut diffuser sur des électrons (diffusion Compton) : ce sont les *coïncidences diffusées*. Les photons peuvent aussi être absorbés, c'est le phénomène d'*atténuation*. Seules les coïncidences *vraies* constituent le signal permettant de reconstruire la distribution volumique du traceur. Les facteurs de dégradation des données doivent donc être corrigés directement sur les mesures ou inclus dans l'étape de reconstruction.

L'approche proposée consiste à considérer les coïncidences observées comme des échantillons d'une distribution de probabilité (elle même aléatoire) et nécessite l'accès à des données brutes non traitées. Ainsi, pour prendre en compte les phénomènes dits *parasites* (coïncidences diffusées et aléatoires), on les introduit dans les observations. Les données sont acquises sous le format *list-mode* qui enregistre pour chaque événement détecté, la position du couple de détecteurs. On note par y_i l'indice de la LOR correspondant au $i^{\text{ème}}$ événement détecté. Soit $\mathcal{B}(y)$ la distribution des événements parasites supposée connue. On formule le problème de reconstruction dans un cadre bayésien non paramétrique de la façon suivante :

$$\mathcal{F}(y) = p\mathcal{T}(y) + (1-p)\mathcal{B}(y) \quad (1)$$

où

$$\mathcal{T}(y) = \int_{\mathcal{X}} \mathcal{P}(\cdot|\mathbf{x}) G(d\mathbf{x}), \quad (2)$$

$p \in]0, 1]$ est une probabilité connue définissant le ratio entre les événements vrais (distribués suivant \mathcal{T}) et parasites (de loi \mathcal{B}). G est la distribution spatiale de l'activité du traceur que l'on cherche à estimer à partir des observations $\mathbf{y} = (y_1, y_2, \dots, y_n)$ distribuées suivant \mathcal{F} . Enfin, $\mathcal{P}(\cdot|\mathbf{x})$ est une loi de probabilité connue appelée *distribution de projection*. Elle donne la probabilité de détecter une paire de photons dans une LOR donnée sachant une émission en \mathbf{x} .

On souligne que l'on pourrait prendre en compte la distribution des événements diffusés en modélisant $\mathcal{P}(\cdot|\mathbf{x})$ comme un mélange de distributions de projections, avec un terme pour les événements vrais et un autre pour les diffusés. Mais un tel modèle suggère que les événements diffusés contribuent à la reconstruction de G , nécessitant de connaître leur distribution de projection. Cependant, il est d'usage en tomographie de ne pas utiliser les diffusés dans la reconstruction et de considérer que leur distribution de projection est non informative comparée à celle des vrais. Ainsi, cette distribution est non utilisée et sa contribution est considérée comme connue après une (ou plusieurs) reconstruction(s) préliminaire(s). C'est la raison pour laquelle les événements diffusés et aléatoires sont introduits dans les observations via l'équation (1).

2.2 Loi a priori sur la densité d'émission

Dans un cadre bayésien, on met une loi *a priori* sur les inconnues du modèle. Ici, cette inconnue est la densité d'émission du traceur que l'on cherche à reconstruire. On la notera f_G . Tout comme dans [1], la loi *a priori* sur la densité f_G est un mélange par processus de Dirichlet de Gaussiennes 3D et s'exprime

ainsi :

$$f_G(\mathbf{x}) = \int \mathcal{N}(\mathbf{x}|\boldsymbol{\theta})H(\boldsymbol{\theta})d\boldsymbol{\theta} = \sum_{k=1}^{+\infty} w_k f_{\mathcal{N}}(\mathbf{x}|\boldsymbol{\theta}_k^*), \quad (3)$$

où \mathcal{N} désigne la loi normale, $f_{\mathcal{N}}$ sa densité, H est générée suivant un processus de Dirichlet de paramètres $\alpha > 0$ et G_0 une mesure de probabilité. On note $H \sim DP(\alpha, G_0)$ et on a alors $H(\cdot) = \sum_{k=1}^{+\infty} w_k \delta_{\boldsymbol{\theta}_k^*}(\cdot)$, où $(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \dots)$ sont les valeurs uniques des $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)$ et sont i.i.d suivant G_0 . Cette dernière est choisie ici comme étant un modèle Normal-Inverse Wishart ($\mathcal{NIW}_{\rho, n_0, \mu_0, \Sigma_0}$), défini de la façon suivante :

$$\mathbf{m}|\Sigma \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma/\rho),$$

où ρ est le paramètre de précision, $\boldsymbol{\mu}_0$ la moyenne de la loi normale et

$$\Sigma^{-1} \sim \mathcal{W}(n_0, (n_0 \Sigma_0)^{-1}),$$

avec \mathcal{W} désignant la distribution de Wishart, n_0 le degré de liberté et Σ_0^{-1} la moyenne.

La séquence des poids $\mathbf{w} = (w_1, w_2, \dots)$ est construite de la façon suivante (stick-breaking) :

1. générer pour tout j , $V_j \sim \text{Beta}(1, \alpha)$,
2. calculer $w_1 = V_1$ et $\forall k \geq 2$, $w_k = V_k \prod_{j=1}^{k-1} (1 - V_j)$. Cette séquence de construction est appelée distribution GEM et est notée $\mathbf{w} \sim \text{GEM}(\alpha)$.

2.3 Modèle hiérarchique

Le modèle hiérarchique génératif des données TEP proposé dans [1] est étendu afin d'inclure les événements parasites. Ici, chaque événement doit être classé comme "vrai" ou "parasite". Pour cela, on définit pour $i = 1, 2, \dots, n$, une variable indicatrice B telle que $B_i = 0$ si la coïncidence est vraie (i.e $Y_i \sim \mathcal{T}$) et $B_i = 1$ sinon (i.e $Y_i \sim \mathcal{B}$). Tout comme dans [1], les lieux d'émissions $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ correspondant aux n observations sont introduits comme variables cachées manquantes. On propose le modèle génératif suivant pour les données cliniques en TEP 3D :

$$\begin{aligned} H|\alpha, G_0 &\sim DP(\alpha, G_0) \\ B_i|p &\overset{\text{ind}}{\sim} \text{Bernoulli}(p) \\ \boldsymbol{\theta}_i|H, B_i = 0 &\overset{\text{iid}}{\sim} H \\ \mathbf{x}_i|\boldsymbol{\theta}_i, B_i = 0 &\overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i|\boldsymbol{\theta}_i) \\ y_i|\mathbf{x}_i, B_i = 0 &\overset{\text{ind}}{\sim} \mathcal{P}(y_i|\mathbf{x}_i) \\ y_i|B_i = 1 &\overset{\text{ind}}{\sim} \mathcal{B}(y_i). \end{aligned} \quad (4)$$

2.4 Inférence

La structure du modèle hiérarchique (4) permet d'en dériver un algorithme MCMC pour inférer sur la distribution *a posteriori* d'activité $G(\cdot)|\mathbf{y}$. À chaque itération de l'algorithme MCMC et pour toutes les observations, on effectue un tirage

Bernouilli pour tester si l'événement est vrai ou non. La distribution conditionnelle *a posteriori* de la variable de Bernouilli B_i s'écrit :

$$B_i \sim \text{Bernouilli} \left(\frac{(1-p)\mathcal{B}(y)}{p\tilde{\mathcal{T}}(y) + (1-p)\mathcal{B}(y)} \right) \quad (5)$$

où

$$\tilde{\mathcal{T}}(y) = \int_{\mathcal{X}} \mathcal{P}(\cdot|\mathbf{x}) \tilde{G}(\mathrm{d}\mathbf{x}) \quad (6)$$

$$\text{avec } f_{\tilde{G}_i}(\mathbf{x}_i) \propto \sum_{k=1}^{K^*} \mathbf{1}(w_k > u_i) \max(w_k, \zeta) f_{\mathcal{N}}(\mathbf{x}_i|\boldsymbol{\theta}_k^*),$$

et où $f_{\tilde{G}}$ est la densité de \tilde{G} , $\mathbf{u} = (u_1, u_2, \dots, u_n)$ sont des variables auxiliaires uniformes permettant de générer un nombre fini K^* de composantes du mélange à chaque itération de l'algorithme (stratégie du "slice sampling") et ζ est un seuil que l'on propose afin d'améliorer les performances de l'algorithme MCMC [2, Chapitre 3].

L'intégrale (6) n'a pas de forme explicite et on l'approxime par échantillonnage d'importance, utilisant une loi d'échantillonnage G_i^* . Sa densité $f_{G_i^*}$ est obtenue en remplaçant dans l'intégrande de (6), la distribution de projection $\mathcal{P}(y_i|\mathbf{x}_i)$ par une densité Gaussienne $f_{\mathcal{N}}(\cdot|\mathbf{m}_{y_i}^*, \boldsymbol{\Sigma}_{y_i}^*)$ dont les paramètres $\mathbf{m}_{y_i}^*$ et $\boldsymbol{\Sigma}_{y_i}^*$ sont choisis de telle sorte à approximer $\mathcal{P}(y_i|\mathbf{x}_i)$. On obtient un mélange re-pondéré de lois normales, de densité proportionnelle à

$$\sum_{k=1}^{K^*} \mathbf{1}(w_k > u_i) \max(w_k, \zeta) \tilde{\eta}_{k,y_i} f_{\mathcal{N}}(\mathbf{x}_i^*|\tilde{\mathbf{m}}_{k,y_i}, \tilde{\boldsymbol{\Sigma}}_{k,y_i}),$$

où pour tout $k \leq K^*$ et $y \in \#\{LORs\}$,

$$\tilde{\eta}_{k,y} = f_{\mathcal{N}}(\mathbf{m}_y^*|\mathbf{m}_k, \boldsymbol{\Sigma}_y^* + \boldsymbol{\Sigma}_k),$$

$$\tilde{\mathbf{m}}_{k,y} = ((\boldsymbol{\Sigma}_y^*)^{-1} + \boldsymbol{\Sigma}_k^{-1})^{-1} ((\boldsymbol{\Sigma}_y^*)^{-1}\mathbf{m}_y^* + \boldsymbol{\Sigma}_k^{-1}\mathbf{m}_k),$$

$$\tilde{\boldsymbol{\Sigma}}_{k,y} = ((\boldsymbol{\Sigma}_y^*)^{-1} + \boldsymbol{\Sigma}_k^{-1})^{-1}.$$

On approxime l'intégrale de la façon suivante :

– pour un certain $L > 0$ et pour $l \in \{1, \dots, L\}$, échantillonner

$$\mathbf{x}_i^{(l)} \stackrel{\text{iid}}{\sim} G_i^* \text{ à l'itération } l,$$

– faire

$$\tilde{T}(y_i) \approx \frac{\mathcal{C}_i}{L} \sum_{l=1}^L \frac{\mathcal{P}(y_i|\mathbf{x}_i^{(l)})}{f_{\mathcal{N}}(\mathbf{x}_i^{(l)}|\mathbf{m}_{y_i}^*, \boldsymbol{\Sigma}_{y_i}^*)},$$

où \mathcal{C}_i est une constante de normalisation :

$$\mathcal{C}_i = \sum_{k=1}^{K^*} \mathbf{1}(w_k > u_i) \max(w_k, \zeta) \tilde{\eta}_{k,y_i}.$$

Toutes les distributions conditionnelles proposées dans [1] restent valables puisque l'on peut construire les événements vrais à partir de tous les événements et des variables Bernouilli. Ainsi, une fois que l'événement est classé comme étant "vrai", on déroule le même processus que dans [1, Section 3.2] à savoir :

1. Proposition de localisation de l'annihilation.
2. Affectation de l'annihilation à l'une des composantes du mélange Dirichlet.
3. Mise à jour des poids et des paramètres du mélange.

3 Matériel

La méthode de reconstruction proposée (BNP) a été testée sur des données cliniques acquises par le Biograph (Siemens Healthcare, Knoxville, TN, USA), pour une durée d'examen de 20 min. Le traceur utilisé est le PIB (Pittsburgh Compound-B), un analogue de la thioflavine T, utilisé en recherche clinique en TEP cérébrale pour l'étude de la maladie d'Alzheimer. Il est marqué au ^{11}C et noté $[^{11}\text{C}]\text{-PIB}$. Afin d'évaluer la technique proposée dans un contexte spécifique de faible dose, les données du fichier list-mode ont été aléatoirement décimées d'un facteur 5, conduisant à une dose artificiellement réduite de 5. La correction du phénomène d'atténuation a été effectuée en utilisant une mesure des coefficients d'atténuation obtenue grâce à un examen par tomодensitométrie pratiqué avant l'acquisition des données TEP. Quant à la correction des événements diffusés et aléatoires, les termes correctifs fournis par le constructeur ont été adaptés à notre méthode par une procédure de ré-échantillonnage.

4 Résultats

Nous avons effectué 7.000 itérations de l'algorithme MCMC, dont 1.000 écartées pour la période de burn-in. L'estimateur choisi pour la densité f_G est l'espérance *a posteriori* sur les tirages post burn-in. L'erreur associée à l'estimateur proposé est directement accessible et fournie ici via l'image d'écart-type. On présente aussi les reconstructions obtenues avec le code constructeur basé sur la méthode du maximum de vraisemblance [3] utilisant l'algorithme EM [4] (ML-EM), et pour une dose normale. Pour la régularisation de ML-EM, les images obtenues après convergence ont été post filtrées par un filtre Gaussien (LTMH=3 mm).

La figure 1 montre les résultats de reconstruction obtenus avec notre méthode BNP pour une dose divisée par 5, et ceux du code constructeur ML-EM pour une dose normale, suivant les vues axiale, coronale et sagittale.

5 Conclusion, discussion, perspectives

Nous avons proposé une modélisation bayésienne non paramétrique pour la reconstruction d'images quantitatives en TEP et incorporant la correction de tous les effets sources de biais dans les mesures. Les résultats obtenus démontrent la pertinence de cette approche puisque sous une dose divisée par 5, on obtient de bons résultats de reconstruction, compétitifs avec l'algorithme constructeur sous une dose normale. La comparaison de l'algorithme proposé avec celui du constructeur est, pour l'instant, limitée à l'aspect visuel même si les différences de reconstruction sont clairement visibles. Une évaluation plus poussée requiert une multitude de reconstructions et constitue un travail en cours.

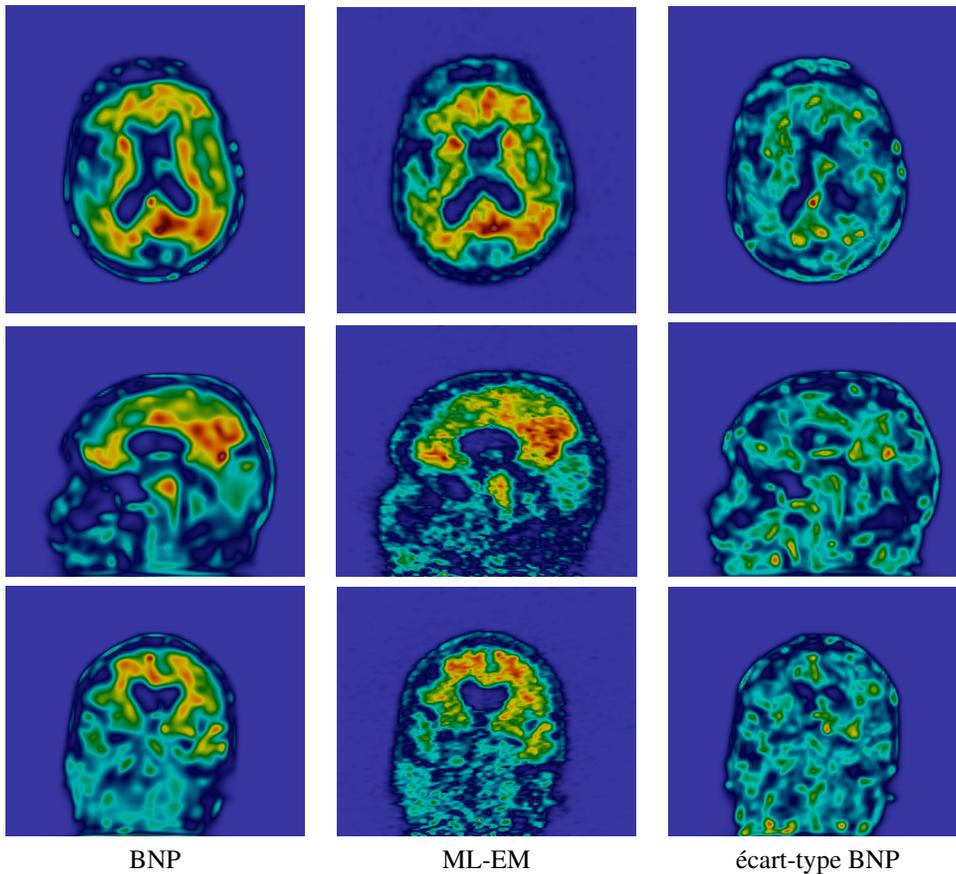


FIGURE 1 – Lignes : Vues axiale, coronale et sagittale.
Colonnes : BNP, dose $\div 5$; ML-EM + post-filtrage (LTMH=3mm), dose $\div 1$; écart-type BNP.

Références

- [1] M. D. Fall, É. Barat, C. Comtat, T. Dautremer, T. Montagu and A.M. Djafari, "Reconstruction bayésienne non paramétrique en tomographie par émission de positons", Actes du 23ème colloque GRETSI, Traitement du signal et des images, 2011.
- [2] M. D. Fall, "Modélisation stochastique de processus pharmaco-cinétiques, application à la reconstruction tomographique par émission de positrons (TEP) spatio-temporelle", Thèse de Doctorat, Université Paris Sud, Mars 2012.
- [3] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction in positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.
- [4] K. Lange and R. Carson, "EM reconstruction algorithm for emission and transmission tomography," *J. Comp. Assist. Tomo.*, vol. 8, pp. 306–316, 1984.