

# Reconstruction bayésienne non paramétrique en tomographie par émission de positons

Mame Diarra FALL<sup>1,2</sup>, Éric BARAT<sup>2</sup>, Claude COMTAT<sup>3</sup>, Thomas DAUTREMER<sup>2</sup>, Thierry MONTAGU<sup>2</sup> et Ali MOHAMMAD-DJAFARI<sup>1</sup>

<sup>1</sup>Laboratoire des Signaux et Systèmes (L2S), UMR 8506 (CNRS-SUPELEC-UNIV PARIS SUD),  
3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France.

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

<sup>3</sup> Service Hospitalier Frédéric Joliot, CEA/DSV/I2BM, 91401 Orsay, France.  
fall@lss.supelec.fr, eric.barat@cea.fr

**Résumé** – Dans cette contribution, nous traitons le problème de la reconstruction spatiale tridimensionnelle en tomographie par émission de positons (TEP) dans un cadre continu. L’objectif est de reconstruire la distribution spatiale d’activité directement à partir des données, sans aucune discrétisation préalable. Pour cela, nous formulons le problème dans un cadre bayésien non paramétrique basé sur les mélanges par processus de Dirichlet. La modélisation bayésienne permet de régulariser le problème inverse, l’approche non paramétrique quant à elle fournit un cadre robuste et flexible pour l’estimation. La difficulté majeure dans cette approche est de traiter le nombre infini de composantes pendant l’inférence. Nous proposons un schéma d’échantillonnage efficace basé sur les méthodes de Monte-Carlo par chaînes de Markov (MCMC) pour générer des échantillons suivant la distribution *a posteriori* d’activité. La méthode proposée est comparée avec celle du maximum *a posteriori* (MAP) et on présente des résultats sur données simulées à partir d’un fantôme réaliste.

**Abstract** – In this contribution, we address the problem of continuous Positron Emission Tomography (PET) spatial reconstruction in three dimensions. Our aim is to recover the spatial activity distribution directly from the data, without any ad hoc space discretization. We follow a Bayesian nonparametric approach using Dirichlet process mixture models. The Bayesian modeling allows to regularize the ill-posed inverse problem while being nonparametric offers a framework for robustness and flexibility to perform the estimation. The main challenge in this approach is to deal with the infinite number of components in the inference. We propose an efficient Markov Chain Monte-Carlo (MCMC) sampling scheme to generate samples from the posterior activity distribution. The proposed method is compared to the maximum *a posteriori* (MAP) approach and we present results on simulated data from a realistic phantom.

## 1 Introduction

En tomographie par émission de positons (TEP), on cherche à reconstruire une image volumique (3D) traduisant l’activité métabolique d’un organe à partir des données mesurées. Ces dernières proviennent de la détection en coïncidence de photons  $\gamma$  issus de la désintégration de la substance radioactive marquant la molécule d’intérêt. Deux approches sont alors possibles pour reconstruire la distribution spatiale du traceur radioactif. La première est dite analytique et repose sur la rétroprojection filtrée (FBP). Elle est basée sur une modélisation des données et de l’image dans un espace continu et utilise la transformée en rayons X [1]. Toutefois, la nature déterministe de cette approche conduit à des images présentant de forts artefacts de reconstruction. La seconde approche dite statistique permet de modéliser les processus aléatoires régissant l’émission et la détection. Les principales techniques sont celles maximisant la vraisemblance poissonnienne des données (ML) ([2], [3]), ou bien, pour limiter le bruit, les techniques de régularisation bayésienne maximisant la loi *a posteriori* [4]. Néanmoins dans ces méthodes statis-

tiques, la distribution spatiale de radioactivité est discrétisée et représentée sous forme vectorielle dans une base de fonctions prédéfinies et fixées. Le plus souvent, ces fonctions de base sont des volumes élémentaires appelés voxels. Se pose alors le problème de la détermination du nombre et des dimensions adéquats des voxels pour représenter une telle fonction. En effet, la taille du voxel agit comme un paramètre de régularisation puisque plus elle est grande, plus l’image est lissée et plus la résolution spatiale est mauvaise.

Dans cette contribution, nous présentons une méthode originale de reconstruction en TEP. Afin de s’affranchir de la base figée des voxels, nous adoptons une approche plus flexible et plus robuste dite non paramétrique, où le nombre de paramètres dans le modèle s’adapte automatiquement aux données et à leur structure. Pour régulariser le problème inverse mal posé, l’approche bayésienne est utilisée et cette méthode est alors appelée *bayésienne non paramétrique* [5]. De plus, cette approche permet de caractériser toute la loi de la distribution spatiale et l’on peut ainsi évaluer des paramètres importants pour l’imagerie quantitative tels que les intervalles de crédibilité.

## 2 Formulation bayésienne du problème non paramétrique

### 2.1 Formulation du problème

L'objet que l'on cherche à reconstruire sur  $\mathbb{R}^3$  est la distribution spatiale d'activité dans le champ de vue du tomographe. Soit  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ,  $n$  émissions localisées dans cet espace-objet. La densité à estimer est alors celle de la distribution spatiale  $G(\mathbf{x})$ , et on la notera  $f_G(\mathbf{x})$ .

En TEP comme dans tous les problèmes inverses, on n'observe pas directement ces émissions mais plutôt leurs projections dans l'espace des détecteurs. Une observation  $y_i$  désigne les coordonnées de la ligne virtuelle appelée ligne de réponse (LOR) qui joint les deux détecteurs ayant enregistré les photons en coïncidence issus de l'annihilation des positons ayant eu lieu en  $\mathbf{x}_i$ . Le problème inverse se formule alors ainsi,

$$F(\cdot) = \int_{\mathcal{X}} \mathcal{P}(\cdot|\mathbf{x}) G(d\mathbf{x}) \quad (1)$$

$$y_i \stackrel{\text{iid}}{\sim} F, \text{ pour } i = 1, \dots, n$$

où  $\mathcal{X} \subset \mathbb{R}^3$  désigne l'espace-objet;  $\mathbf{y} = \{y_1, \dots, y_n\}$  l'ensemble des observations distribuées suivant  $F$ ;  $G$  la distribution spatiale dont la densité est à estimer et enfin  $\mathcal{P}(\cdot|\mathbf{x})$  est une loi connue appelée distribution de projection. C'est la distribution donnant la probabilité de détecter une paire de photons dans une LOR  $l$  sachant une émission ayant eu lieu en  $\mathbf{x}$ . Dans les approches statistiques basées sur la discrétisation de l'espace (ML et MAP), cette loi de probabilité est discrétisée et représentée sous la forme d'une matrice appelée matrice-système.

### 2.2 Loi a priori sur la distribution d'activité

Dans la régularisation bayésienne, on munit les inconnues du modèle de lois *a priori*. Dans le contexte bayésien non paramétrique, cette loi *a priori* porte directement sur la loi  $G$  et s'exprime par  $G \sim \mathcal{G}$ , où  $\mathcal{G}$  est une distribution sur des distributions, *i.e.*, chaque tirage suivant  $\mathcal{G}$  est une mesure de probabilité sur  $\mathcal{X}$ . On dit alors que  $G$  est une mesure de probabilité aléatoire. La loi *a priori* que nous avons utilisée pour la densité  $f_G$  de  $G$  est un mélange par processus de Dirichlet (DPM) [6]. L'idée des DPM est de convoluer une mesure discrète générée par un processus de Dirichlet (DP) avec une fonction continue paramétrique afin d'obtenir un *a priori* sur une densité.

Plus précisément, soit  $H$  une mesure de probabilité générée suivant un processus de Dirichlet (DP). Ce dernier est paramétré par  $\alpha$  (un réel strictement positif) et  $G_0$  (une mesure de probabilité). On note  $H \sim DP(\alpha, G_0)$  et on a alors,

$$H(\cdot) = \sum_{k=1}^{\infty} w_k \delta_{\theta_k^*}(\cdot) \quad (2)$$

où  $\theta_k^* \sim G_0$  et la séquence infinie 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 des poids est appelée distribution GEM et est notée  $\mathbf{w} \sim \text{GEM}(\alpha)$ .

Pour obtenir un DPM, la fonction  $H$  est convoluée avec un noyau continu  $\phi(\cdot|\theta)$ . Dans notre cas, il s'agit d'une gaussienne 3D de paramètres  $\theta = (\mathbf{m}, \Sigma)$  où  $\mathbf{m}$  représente le vecteur moyen et  $\Sigma$  la matrice de covariance. Ceci conduit à la loi *a priori* suivante sur la densité de  $G$ ,

$$f_G(\mathbf{x}) = \int \phi(\mathbf{x}|\theta) H(\theta) d\theta = \sum_{k=1}^{\infty} w_k f_{\mathcal{N}}(\mathbf{x}|\theta_k^*). \quad (3)$$

### 2.3 Modèle hiérarchique pour les données TEP

Pour obtenir le modèle génératif des données TEP, on introduit les variables cachées non observées que sont les lieux d'émission  $\mathbf{x}_i$ . Le modèle hiérarchique s'écrit alors :

$$\begin{aligned} &\text{générer } H \sim DP(\alpha, G_0) \\ &\text{générer } \theta_i \sim H \text{ pour } i = 1, \dots, n \\ &\text{générer } \mathbf{x}_i \sim \mathcal{N}(\mathbf{x}_i|\theta_i) \text{ pour } i = 1, \dots, n \\ &\text{générer } y_i \sim \mathcal{P}(y_i|\mathbf{x}_i) \text{ pour } i = 1, \dots, n \end{aligned} \quad (4)$$

où  $\mathcal{P}$  est la distribution de la projection définie dans (1).

Notons que puisque  $H$  est discrète (cf. équation (2)), plusieurs  $\theta_i$  seront identiques; cela induit un regroupement *a priori* des données  $\mathbf{x}_i$  suivant les valeurs des paramètres des composantes, c'est l'effet de *renforcement statistique* de ce processus [5].

Le modèle (4) peut être ré-écrit de façon équivalente en introduisant des variables de classification  $c_i$  permettant d'identifier la gaussienne à partir de laquelle l'émission  $\mathbf{x}_i$  sera tirée, c'est-à-dire  $c_i = k$  ssi  $\theta_i = \theta_k^*$ . Cela s'exprime ainsi :

$$\begin{aligned} &\text{générer } \theta_k^* \stackrel{\text{iid}}{\sim} G_0 \text{ pour } k = 1, 2, \dots \\ &\text{générer } \mathbf{w} \sim \text{GEM}(\alpha) \text{ où } \mathbf{w} = (w_1, w_2, \dots) \\ &\text{générer } c_i \stackrel{\text{iid}}{\sim} \sum_{k=1}^{\infty} w_k \delta_k(\cdot) \text{ pour } i = 1, \dots, n \\ &\text{générer } \mathbf{x}_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i|\theta_{c_i}^*) \text{ pour } i = 1, \dots, n \\ &\text{générer } y_i \stackrel{\text{ind}}{\sim} \mathcal{P}(y_i|\mathbf{x}_i) \text{ pour } i = 1, \dots, n. \end{aligned} \quad (5)$$

Puisque  $G_0$  définit une distribution sur l'espace des paramètres des clusters, à savoir les  $\theta_k^* = (\mathbf{m}_k, \Sigma_k)$ , nous l'avons choisie suivant une loi Normal-Inverse Wishart ( $\mathcal{NIW}_{\rho, n_0, \mu_0, \Sigma_0}$ ), définie de la façon suivante :

$$\mathbf{m}|\Sigma \sim \mathcal{N}(\mu_0, \Sigma/\rho),$$

où  $\rho$  est le paramètre de précision,  $\mu_0$  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  $\Sigma_0^{-1}$  la moyenne.

## 3 Inférence

### 3.1 Inférence des DPM

L'inférence sur la loi *a posteriori* dans les modèles de mélange par processus de Dirichlet nécessite de faire appel à des techniques d'approximation dont la plupart sont basées sur les méthodes de Monte-Carlo par chaînes de Markov (MCMC). Il existe deux grandes classes d'algorithmes : les méthodes marginales et les méthodes conditionnelles. Les premières, comme leur nom l'indique, marginalisent la distribution aléatoire et génèrent seulement des échantillons suivant cette distribution [7]. Les méthodes conditionnelles quant à elles représentent explicitement la mesure. Pour traiter le nombre infini de composantes dans les algorithmes conditionnels, une approche consiste à tronquer la mesure [8]. Une autre approche permet d'éviter la troncature en introduisant une variable auxiliaire qui rend le modèle conditionnellement fini, c'est la stratégie dite de "slice sampling". Nous avons proposé un nouvel algorithme utilisant cette stratégie, modifié par l'introduction d'un seuil ainsi que la formulation de l'échantillonnage dans l'espace des classes d'équivalence des clusters [10]. Cela permet d'améliorer les propriétés de mélange de l'algorithme comparé aux autres méthodes conditionnelles. Cet algorithme a été utilisé dans notre reconstruction en TEP.

### 3.2 Algorithme MCMC pour la reconstruction en TEP

Se basant sur le modèle hiérarchique (5), nous proposons un algorithme MCMC pour inférer sur la distribution *a posteriori* d'activité. L'algorithme génère successivement des blocs de variables suivant les lois conditionnelles suivantes :

1. Proposition de localisation des annihilations :  $\mathbf{X}|\mathbf{w}, \Theta^*, \mathbf{y}$ .
2. Affectation des annihilations aux composantes du mélange :  $\mathbf{c}|\mathbf{w}, \Theta^*, \mathbf{X}$ .
3. Mise à jour des paramètres :  $\Theta^*|\mathbf{c}, \mathbf{X}$ .
4. Mise à jour des poids :  $\mathbf{w}|\mathbf{c}$ .

Grâce aux propriétés de conjugaison, toutes ces lois sont aisément simulables avec un échantillonneur de Gibbs sauf celle de  $(\mathbf{X}|\mathbf{w}, \Theta^*, \mathbf{y})$  pour laquelle nous avons recours à un algorithme Metropolis-Hastings (MH). La loi de proposition est choisie comme étant le produit entre le mélange Dirichlet et une loi normale dont les paramètres sont choisis de telle sorte à approximer la distribution de la projection  $\mathcal{P}$  définie dans l'équation (1). Le résultat est un mélange de gaussiennes dans la direction de la ligne de réponse considérée.

Après la convergence de l'algorithme, chaque itération complète de l'algorithme fournit des échantillons suivant la loi *a posteriori* jointe  $(\mathbf{X}, \mathbf{c}, \Theta^*, \mathbf{w}|\mathbf{y})$ . Les échantillons obtenus à l'itération  $t$ ,  $(\mathbf{X}^{(t)}, \mathbf{c}^{(t)}, \Theta^{*(t)}, \mathbf{w}^{(t)})$ , peuvent être utilisés pour calculer une estimée de la loi *a posteriori* d'activité  $G(\mathbf{x})|\mathbf{y}$ , de densité

$$f_{G^{(t)}}(\mathbf{x}) \approx \sum_{k=1}^{K^*} w_k^{(t)} f_{\mathcal{N}}(\mathbf{x}|\theta_k^{*(t)})$$

où  $K^*$  désigne le nombre de composantes retenues à chaque itération par l'algorithme du "slice sampler".

## 4 Application en TEP 3D

Afin de tester le modèle proposé, nous l'avons appliqué à la reconstruction TEP 3D et comparé avec une approche bayésienne basée sur la voxelisation de l'espace à savoir la méthode du maximum *a posteriori* (MAP).

On a d'abord généré  $10^7$  événements à partir d'un fantôme 3D. Les événements sont détectés par un scanner à 32 couronnes (FOV=155mm) et chaque couronne est composée de 576 détecteurs (rayon couronne=412mm). Pour des raisons de simplicité, nous n'avons pas généré des coïncidences aléatoires, diffuses et atténuées. Les probabilités de détection ont été supposées seulement liées à la géométrie du tomographe.

### 4.1 Approche bayésienne non paramétrique (BNP)

Pour la mise en œuvre de notre algorithme, les paramètres du DPM ont été choisis comme suit :  $\alpha = 500$ ,  $\Sigma_0 = 6.25 \times \mathbb{I}_3$ ,  $n_0 = 4$ . Pour l'algorithme MCMC, 15000 itérations ont été effectuées dont 5000 pour la période de chauffage. L'estimateur choisi comme estimée de la distribution d'activité est l'espérance conditionnelle de  $f_G$  sur les  $N = 10000$  tirages retenus,

$$\mathbb{E}(f_G|\mathbf{y}) \approx \frac{1}{N} \sum_{t=1}^N f_{G^{(t)}}.$$

### 4.2 Approche du maximum *a posteriori* (MAP)

Dans cette méthode, l'image à reconstruire est discrétisée et représentée sous forme vectorielle  $\mathbf{f} = (f_j, j = 1, \dots, J)$ ,  $f_j$  désignant l'intensité dans le voxel  $j$ . On cherche l'image  $\mathbf{f}$  qui maximise la loi *a posteriori*

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})\pi(\mathbf{f})}{p(\mathbf{y})}$$

avec  $p(\mathbf{y}|\mathbf{f})$  désignant la vraisemblance poissonnienne des données et  $\pi(\mathbf{f})$  la loi *a priori* sur l'image. Nous avons modélisé cette loi *a priori* par un champ de Gibbs dans le but de favoriser des intensités similaires dans les voxels voisins tout en préservant les contours. On a alors,

$$\pi(\mathbf{f}) \propto \exp\left(-\beta \sum_{r,s} w_{r,s} \psi\left(\frac{f_r - f_s}{\delta}\right)\right).$$

La fonction potentielle  $\psi$  est la fonction *log cosh* suggérée par [4]. Les paramètres  $\beta$  et  $\delta$  sont choisis de telle sorte à minimiser l'erreur moyenne quadratique par rapport au fantôme. Le voisinage d'un voxel est composé par l'ensemble des cinq voisins sur chacune de ses faces. Enfin, le poids  $w_{r,s}$  est donné par l'inverse de la distance quadratique entre les deux voxels  $r$  et  $s$ .

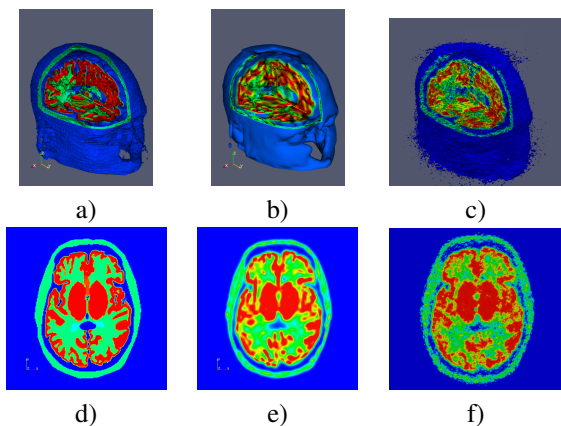


FIG. 1 – a) fantôme 3D, b) Estimée BNP, c) Estimée MAP, d) Vue axiale du fantôme, e) Vue axiale de l’estimée BNP, f) Vue axiale de l’estimée MAP.

L’image est reconstruite en maximisant la fonction de coût suivante,

$$\Phi(\mathbf{f}, \mathbf{y}) = L(\mathbf{y}|\mathbf{f}) - \beta \sum_{r,s} w_{rs} \psi \left( \frac{f_r - f_s}{\delta} \right)$$

où  $L$  désigne le logarithme de la vraisemblance.

### 4.3 Résultats

La figure FIG. 1 montre en 3D et sur une vue axiale le fantôme utilisé pour générer les données, notre estimée ainsi que celle obtenue par l’approche MAP. On peut y constater visuellement que notre reconstruction a lieu dans un espace continu tout en préservant les bords, et ce même dans les régions froides. L’approche MAP quant à elle fournit des images bruitées. Il est à signaler que dans notre approche, la discrétisation est seulement effectuée pour visualiser l’image et a été choisie *a posteriori* suivant la taille du fantôme, soit  $256 \times 256 \times 128$ . La même discrétisation a été utilisée pour l’algorithme MAP.

Ces résultats illustrent de la capacité de l’approche proposée à améliorer le signal-sur-bruit et la résolution de l’image lorsque l’on travaille sur un nombre relativement faible de données. En effet, le nombre de composantes allouées est adapté en fonction de l’information sur les données (nombre de composantes actives  $\approx 4000$ ). Le caractère non paramétrique se traduit par le fait que ce nombre peut augmenter avec les observations. De plus, la distribution *a posteriori* de l’incertitude est accessible pour l’analyse quantitative de n’importe quelle région d’intérêt.

La stratégie du “slice sampling” que nous avons adoptée permet d’éviter une troncature du modèle infini tout en imposant un nombre fini de composantes du mélange gaussien à chaque itération ( $\approx 10.000$ ).

## 5 Conclusion, discussions et perspectives

Nous avons proposé une modélisation flexible et numériquement accessible pour la reconstruction d’image TEP. Les résultats de

simulation présentés montrent la capacité de l’approche proposée à reconstruire de bonnes images dans un contexte de faibles doses injectées. L’introduction des variables cachées que sont les lieux d’émission permet d’effectuer la classification directement dans l’espace-objet et ainsi d’éviter toute discrétisation *a priori*. Cette absence de discrétisation implique qu’il n’y ait pas de matrice système à calculer. Mais la distribution de la projection  $\mathcal{P}(y_i|x_i)$ , qui dépend entre autres de la géométrie du tomographe, est explicitement incluse dans l’étape Metropolis-Hastings utilisée pour proposer les lieux des émissions.

Cependant, la méthode est assez coûteuse en temps de calcul comparée à l’approche EM (coût proportionnel à  $\#$  évènements  $\times$   $\#$  composantes). Ceci est en grande partie dû à l’échantillonnage MCMC. On pourrait envisager l’utilisation des méthodes bayésiennes variationnelles qui approximent analytiquement la distribution *a posteriori*.

Dans les résultats présentés, nous n’avons pas simulé des coïncidences aléatoires, diffuses et atténuées. La suite de ce travail consistera à prendre en compte ces effets au travers de la loi de projection  $\mathcal{P}(y|x)$ .

## Références

- [1] F. Natterer, “The Mathematics of Computerized Tomography”, *Society for Industrial and Applied Mathematics*, 1986.
- [2] L. A. Shepp and Y. Vardi, “Maximum likelihood reconstruction in Positron Emission Tomography”, *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.
- [3] K. Lange and R. Carson, “EM reconstruction algorithm for emission and transmission tomography”, *J. Comp. Assist. Tomo.*, vol. 8, pp. 306–316, 1984.
- [4] P. J. Green, “Bayesian reconstructions from emission tomography data using a modified EM algorithm”, *IEEE Trans. Med. Imaging*, vol. 9, no. 1, pp. 84–93, 1990.
- [5] N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker, “Bayesian Nonparametrics”, *Cambridge University Press*, April 2010.
- [6] C. E. Antoniak, “Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems”, *Ann. Statist.*, pp. 1152–1174.
- [7] R. M. Neal, “Markov Chain Sampling Methods for Dirichlet Process Mixture Models”, *Journal of Computational and Graphical Statistics*, pp. 249–265.
- [8] H. Ishwaran and L. F. James, “Gibbs sampling methods for stick-breaking priors”, *J. Am. Stat. Assoc.*, pp. 161–173.
- [9] S. G. Walker, “Sampling the Dirichlet mixture model with slices”, *Comm. Statist.*, vol. 36, pp. 45–54, 2007.
- [10] M. D. Fall and É. Barat, “Sampling the two-parameter Poisson Dirichlet distribution : state of the art and a new method”, Technical report, CEA, 2011.