Mame Diarra Fall¹

Bayesian Nonparametrics and Biostatistics: The Case of PET Imaging

¹ Laboratoire de mathématiques analyse probabilités modélisation d'Orléans, Orléans 45067, France, E-mail: diarra.fall@univ-orleans.fr

Abstract:

Biostatistic applications often require to collect and analyze a massive amount of data. Hence, it has become necessary to consider new statistical paradigms that perform well in characterizing complex data. Nonparametric Bayesian methods provide a widely used framework that offers the key advantages of a fully model-based probabilistic framework, while being highly flexible and adaptable. The goal of this paper is to provide a motivation of Bayesian nonparametrics (BNP) through a particular biomedical application, namely Positron Emission Tomography (PET) imaging reconstruction.

Keywords: Bayesian nonparametric, PET imaging, inverse problems, Markov Chain Monte Carlo

DOI: 10.1515/ijb-2017-0099

Received: December 1, 2017; **Revised:** October 7, 2019; **Accepted:** October 23, 2019

1 Introduction

Biostatistic applications often require to collect and analyze a massive amount of data. Hence, it has become necessary to consider new statistical paradigms that perform well in characterizing complex data. Nonparametric Bayesian methods provide a widely used framework that offers the key advantages of a fully model-based probabilistic framework, while being highly flexible and adaptable. The goal of this paper is to provide a motivation of Bayesian nonparametrics (BNP) through a particular biomedical application, namely Positron Emission Tomography (PET) imaging reconstruction. For a general review and motivations of BNP methods in various biostatistical applications, one can refer to [1] and [2]. Before going further, let us first recall what *Bayesian nonparametric* means.

1.1 Bayesian parametric vs nonparametric

Bayesian inference is often restricted to *parametric Bayesian* models, namely a family of models that can be indexed with a finite dimensional set of parameters. Inference builds on the posterior distribution of these parameters [3]. However, such an inference can be too restrictive and can lead to misleading decisions and inference. A more flexible inference is typically achieved by considering Bayesian models with infinitely many parameters known as *Bayesian nonparametric (BNP)* models [4]. The latter are used to avoid critical dependence on parametric assumptions, to robustify parametric models and to define model diagnostics and sensitivity analysis for parametric models, by embedding them in a larger encompassing nonparametric model.

The rest of this paper is organised as follows. In Section 2, we present the basis of PET. Section 3 is devoted to the proposed nonparametric Bayes approach. In Section 4, we perform reconstructions using our method and contrast it to the ML-EM algorithm. Conclusions are given in Section 5.

2 Introduction to Positron Emission Tomography

Positron Emission computed Tomography (PET) is an important diagnostic imaging modality used in nuclear medicine to visualize and measure *in vivo* biochemical and physiological processes [5]. To this aim, a molecule of interest labelled with a radionuclide, called *radiotracer* or *tracer*, is administered to the patient's bloodstream. The tracer is designed to provide information about a particular physiological function of interest, such as blood

Mame Diarra Fall is the corresponding author.

© 2019 Walter de Gruyter GmbH, Berlin/Boston.

This content is free.

flow, blood volume, a metabolic process, a transport step, a binding process, etc. The radionuclide undergoes nuclear decay by the emission of positrons (the electron's anti-particles) which at rest annihilate¹ with nearby electrons in the tissue. This results in a matter-antimatter annihilation from which two gamma photons are created. These gamma photons propagate in opposite directions (180deg apart in the center-of-mass system) and are detected in coincidence by detector crystals surrounding the patient. When two photons are detected at nearly the same time (within a few nanoseconds), a coincident *event* is recorded revealing that a positron emission occurred in the virtual straight line connecting the centres of two detectors, the so-called line of response (LOR). The acquisition process is illustrated in Figure 1.

Figure 1: A positron e^+ combines with an electron e^- and this produces two gamma photons moving in a straight line called LOR (red arrow). Photons are recorded by detectors around the patient (orange ring).

From the data collected by the detectors (the events), tomographic algorithms are used to reconstruct spatial (2D or 3D) or time-varying (2D+t or 3D+t) images showing the distribution of the tracer in the target organ. Therefore the PET image reconstruction problem emerges as an *inverse problem* in which we seek to obtain an estimate of the tracer density within the subject from the detector pair and timing information of photon pair coincidences. However, PET reconstruction is known as an *ill-posed inverse problem* and we are faced with a very challenging task [6]. The main shortcoming is the few number of detected events. Because a radioactive substance is administered to patients, the total amount of radiolabelled molecules is limited, resulting in relatively low number of detected events when compared to radiological X-ray computed tomography. This leads to a low signal-to-noise ratio (SNR) in the data. Therefore, the high level of noise in the acquired data, the limited sensitivity of the camera, the physical and geometrical complexity of scanners make tomographic reconstruction very challenging.

The reconstruction methods in PET are typically categorized into analytical methods (filtered back-projection [7]) and statistical iterative techniques, namely the ML-EM (Maximum Likelihood using the EM algorithm) [8, 9]). The former address the reconstruction problem (at least theoretically) in continuous data and image spaces. However, they are based on over-simplified models of the physical processes and are not designed to take into account the stochastic nature of positron emission and photon detection. This leads to reconstructed images with a lot of artefacts and significant noise. The statistical techniques have the advantage of modelling more realistically the data acquisition process (by incorporating and correcting for physics of acquisition, such as attenuation, scatter, patient body motion and breathing) and taking into account the random nature of emission and detection. Consequently, they usually provide images with a higher SNR than the analytical methods. Nevertheless, all these statistical methods rely on the decomposition of the object space using a finite set of fixed basis functions. Among others, the pixel-based basis function (or voxels in 3D) is mostly considered. The reason is that since PET scanner collects only a finite number of measurements, it has seemed inevitable that one must also represent the radiotracer distribution by a finite parameterization. Though usually taken for granted, this representation in a discrete basis is not the best since emission distribution is inherently continuous. The finiteness and rigidity of the basis may induce some bias in the estimated image.

These observations provide motivation for the study of alternative image representations. We have proposed a third category of image reconstruction algorithms for Emission tomographic data [10–12]. The approaches proposed fall in the class of statistical methods but are fundamentally different from the above. The tracer distribution is directly reconstructed from the data, without any *ad hoc* space or time discretisation. To this aim, we cast the reconstruction problem in the *Bayesian nonparametric framework*. The image to reconstruct is viewed as a spatial probability density on $\mathbf{R}^3$ (or a space-time density on $\mathbf{R}^3 \times \mathbf{R}^+$) and estimated from the data. Such an approach is *nonparametric* in the sense that the parameter of interest is an infinite-dimensional probability density, and it is *Bayesian* since we put a prior on the density and infer on its posterior distribution. The Bayesian context allows to regularize the ill-posed inverse problem. Furthermore, the entire distribution of the posterior activity distribution is directly accessible (not only a point estimate like in maximum likelihood approaches) and any

posterior uncertainty can be estimated (standard deviation, credible intervals, etc.). For quantitative images in “small samples” (referred to as “low doses” in PET), a rigorous handling of reconstruction uncertainty appears of interest in order to give credible intervals over profiles of any region of interest. This posterior exploration constitutes an argument for the Bayesian nonparametric choice. Preliminary investigations of this approach are described in our earlier conference articles, using simulated brain data or cardiac PET data.

In this paper, we pursue this investigation on the reconstruction of clinical brain data. However, reconstruction of real data requires to deal with all physical artefacts, the so-called *background events*, and then weighting the statistical modelling. Furthermore, the specific scanner geometry must be taken into account.

3 Theoretical framework

3.1 Problem formulation

We operate in *list-mode data acquisition* where each entry contains the photons’ arrival time and the detector coordinates for each individual coincident event. As we have already mentioned, an event detection in PET relies on the closeness of occurrence times of a pair of gamma photons produced in opposite directions by a positron-electron annihilation. This coincidence is detected through a timing window of few nanoseconds and a coincident event is recorded. This ideal physical system is however corrupted by the *background events* [13]. In fact, it is frequent that only one photon of the pair is detected, leading to a so-called *single event*. If two such single events occur during the timing window, we are faced with a *random event*. Furthermore, since annihilation photons propagate in the organism, one or both photons may be *scattered* due to Compton interactions. In addition, photons may be absorbed in the body leading to an *attenuation* phenomenon which has to be corrected through a weighting component, so as the PET scanner *normalization* correction. All these *background events* must be considered when dealing with real data. The spatial model proposed in [10] is completed by a background events contribution in the mixture. To be more precise, let y_i be the index of the LOR corresponding to the i -th observed event. Let $\mathcal{B}(y)$ (respectively $\mathcal{T}(y)$) denote the distribution of the background (respectively true) events. Details about the $\mathcal{B}$ distribution are provided in the results section. We formulate the spatial reconstruction in the Bayesian Nonparametric context as follows:

$$\begin{aligned} y_i | \mathcal{F} &\stackrel{\text{iid}}{\sim} \mathcal{F}, \text{ for } i = 1, \dots, n, \\ \mathcal{F}(y) &= p \mathcal{T}(y) + (1 - p) \mathcal{B}(y), \\ \mathcal{T}(y) &= \int_{\mathcal{X}} \mathcal{P}(y|\mathbf{x}) G(d\mathbf{x}), \\ G &\sim \mathcal{G}, \end{aligned} \tag{1}$$

where $\mathbf{y} = (y_1, \dots, y_n)$ is the set of observed data that are $\mathcal{F}$ -distributed. We model $\mathcal{F}$ as a mixture of true and background events, with $p \in [0, 1]$ a known probability defining the ratio between these two types of events. If p were not known, one could put a prior on it and treat it as an extra parameter. $\mathcal{X}$ is a bounded region of $\mathbf{R}^3$ representing the space over which the reconstruction is carried out. $\mathcal{P}(\cdot|\mathbf{x})$ is a known distribution called *projection distribution*. Based on the geometry of the tomographic apparatus and physics of the detection, $\mathcal{P}(y = l|\mathbf{x})$ is the probability for detecting an event in a LOR l assuming that it was emitted at location $\mathbf{x}$, where $\mathbf{x}$ is a three-dimensional vector describing the spatial coordinates. 2G stands for the spatial distribution of interest whose density is to be estimated from the observed data set. Being Bayesian, we put a prior on the investigated measure G and infer on its posterior. The prior $\mathcal{G}$ over G is a distribution over distributions, i. e. each draw G from $\mathcal{G}$ is a probability distribution. G is thus called a *random probability measure*.

Let us now focus on the prior for the spatial density probability.

3.2 Bayesian nonparametric prior for f_G

As mentioned in the last paragraph, we have to put a Bayesian nonparametric prior on G . To be more precise, the prior is put on f_G , the unknown density of G we want to recover from the data.

Dirichlet process mixtures (DPM) due to [14] and [15] play an important role in nonparametric density estimation and clustering problems. The idea of DPM is to convolve a sample from a Dirichlet process (DP) by a parametric kernel. We use a multivariate normal kernel, leading to a DPM of Gaussians as a prior for f_G . The

hierarchical DPM is:

$$H|\alpha, G_0 \sim DP(\alpha, G_0),$$

$$f_G(\mathbf{x}) = \int f_{\mathcal{N}}(\mathbf{x}|\boldsymbol{\theta})H(\boldsymbol{\theta})d\boldsymbol{\theta}$$

where $f_{\mathcal{N}}$ stands for the multivariate normal density parameterized by $\boldsymbol{\theta} = (\mathbf{m}, \boldsymbol{\Sigma})$, the mean and the covariance matrix; H is distributed according to a Dirichlet process with parameters $\alpha > 0$ and G_0 a fixed probability measure. H can be represented as an infinite sum,

$$H(\cdot) = \sum_{k=1}^{+\infty} w_k \delta_{\boldsymbol{\theta}_k^*}(\cdot),$$

where $\mathbf{w} = (w_1, w_2, \dots)$ is constructed as:

1. For all j , sample $V_j \stackrel{i.i.d.}{\sim} \text{Beta}(1, \alpha)$.
2. Set $w_1 = V_1$ and for all $k \geq 2, w_k = V_k \prod_{j=1}^{k-1} (1 - V_j)$.

The resulting law is called the GEM³ distribution. We write $\mathbf{w} \sim \text{GEM}(\alpha)$.

The parameters $(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \dots)$ are the unique values among $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots)$ where $\boldsymbol{\theta}_j = (\mathbf{m}_j, \boldsymbol{\Sigma}_j)$, and are i.i.d. G_0 . We choose for G_0 a Normal-Inverse Wishart ($\mathcal{NIW}$) model, with parameters ρ, n_0, μ_0 , and $\boldsymbol{\Sigma}_0$ such that

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

where ρ is a precision parameter, μ_0 the mean of the normal distribution,

$$\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(n_0, (n_0 \boldsymbol{\Sigma}_0)^{-1}),$$

with $\mathcal{W}$ the Wishart distribution, n_0 (resp. $\boldsymbol{\Sigma}_0^{-1}$) the degree of freedom (resp. the mean) of the distribution.

3.3 Hierarchical model

Suppose that during the PET scan, n disintegrations localized on the object at $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ were emitted. If all $\mathbf{x}_i$ were known, the estimation of the spatial tracer density would be straightforward since we would be faced with the estimation of a probability density function given samples from the distribution. Unfortunately, in the tomographic inverse problem, we do not observe the emission locations but rather the projections of these locations on detector pair units. Thus, as in ML-EM-family reconstruction, we use a data augmentation scheme where annihilation locations $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ are introduced as hidden variables. Furthermore, to deal with the background events, we introduce Bernoulli variables $B_i, i = 1, 2, \dots, n$, such that $B_i = 0$, if the i -th event is true (i.e. $Y_i \sim \mathcal{T}$); and $B_i = 1$ if it is background (i.e. $Y_i \sim \mathcal{B}$). The generative hierarchical model for clinical 3D PET data can be expressed as

$$\begin{aligned} H|\alpha, G_0 &\sim DP(\alpha, G_0) \\ B_i|p &\stackrel{\text{ind}}{\sim} \text{Bernoulli}(p) \\ y_i|B_i = 1 &\stackrel{\text{ind}}{\sim} \mathcal{B}(y_i) \\ \boldsymbol{\theta}_i|H, B_i = 0 &\stackrel{\text{iid}}{\sim} H \\ \mathbf{x}_i|\boldsymbol{\theta}_i, B_i = 0 &\stackrel{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i|\boldsymbol{\theta}_i) \\ y_i|\mathbf{x}_i, B_i = 0 &\stackrel{\text{ind}}{\sim} \mathcal{P}(y_i|\mathbf{x}_i) \end{aligned} \quad (2)$$

$\mathcal{N}(\mathbf{x}_i|\boldsymbol{\theta}_i)$ denotes the normal distribution parametrized by $\boldsymbol{\theta}_i = (\mathbf{m}_i, \boldsymbol{\Sigma}_i)$, the mean and the covariance matrix.

The hierarchical model (2) can be made more in adequacy with mixture models by introducing classification variables. We introduce allocation variables, $\mathbf{c} = (c_1, c_2, \dots, c_n)$ such that $c_i = k$ iff $\boldsymbol{\theta}_i = \boldsymbol{\theta}_k^*$, with $P(c_i = k) = w_k$.

3.4 Inference

Based on the generative hierarchical model (2), we propose a Markov Chain Monte-Carlo (MCMC) sampler to generate random posterior densities, denoted $f_G(\cdot|\mathbf{y})$. The sampler successively draws blocks of variables from the following full conditional distributions:

$$\begin{aligned} \text{Event classification, true vs background:} & \quad \mathbf{B}|\mathbf{y}, \mathbf{w}, \boldsymbol{\Theta}^*. \\ \text{Annihilation location proposal:} & \quad \mathbf{X}|\mathbf{y}, \mathbf{u}, \mathbf{w}, \boldsymbol{\Theta}^*, \mathbf{B}. \\ \text{Allocation of emission to atoms of } H: & \quad \mathbf{c}|\mathbf{w}, \mathbf{u}, \boldsymbol{\Theta}^*, \mathbf{X}. \\ \text{Components parameters:} & \quad \boldsymbol{\Theta}^*|\mathbf{c}, \mathbf{X}, \mathbf{B}. \\ \text{Weights and slice variables:} & \quad \mathbf{w}, \mathbf{u}|\mathbf{c}. \end{aligned} \quad (3)$$

The use of latent variables and priors conjugacy allows direct sampling of the last three conditional laws via a blocked Gibbs sampler. However, sampling the first two conditionals is more complicated.

In the first step of the sampler, we must perform a Bernoulli trial at each iteration of the MCMC algorithm to test if the event is true or not. The posterior conditional probability of success of the Bernoulli variable is :

$$\frac{(1-p)\mathcal{B}(y_i)}{p\tilde{\mathcal{T}}(y_i) + (1-p)\mathcal{B}(y_i)} \quad \text{where} \quad \tilde{\mathcal{T}}(y_i) = \int_{\mathbf{x}} \mathcal{P}(y_i|\mathbf{x}) \tilde{G}_i(d\mathbf{x}), \quad (4)$$

and $\tilde{G}_i$ is a distribution having density $f_{\tilde{G}_i}$,

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

where $\mathbf{u} = (u_1, \dots, u_n)$ are uniform auxiliary variables that allow to sample a finite number K^* of mixture components at each step. This is known as the *slice sampling strategy* [17]. The integral (4) does not exhibit any closed analytical expression. We use an importance sampling approximation [18], based on an important distribution G_i^* . Its density $f_{G_i^*}$ is a re-weighted mixture of normals, obtained by replacing in the integrand of (4), the projection distribution $\mathcal{P}(y_i|\mathbf{x}_i)$ with a Gaussian $\mathcal{N}(\mathbf{x}_i^*|\mathbf{m}_{y_i}^*, \boldsymbol{\Sigma}_{y_i}^*)$ whose parameters $\mathbf{m}_{y_i}^*$ and $\boldsymbol{\Sigma}_{y_i}^*$ are chosen to approximate $\mathcal{P}(y_i|\mathbf{x}_i)$. The result is a re-weighted mixture of normals, with density proportional to

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

where for all $k \leq K^*$ et $y \leq \text{number of LORs}$,

$$\begin{aligned} \tilde{\eta}_{k,y} &= f_{\mathcal{N}}(\mathbf{m}_y^*|\mathbf{m}_k, \boldsymbol{\Sigma}_y^* + \boldsymbol{\Sigma}_k), \\ \tilde{\mathbf{m}}_{k,y} &= ((\boldsymbol{\Sigma}_y^*)^{-1} + \boldsymbol{\Sigma}_k^{-1})^{-1} ((\boldsymbol{\Sigma}_y^*)^{-1} \mathbf{m}_y^* + \boldsymbol{\Sigma}_k^{-1} \mathbf{m}_k), \\ \tilde{\boldsymbol{\Sigma}}_{k,y} &= ((\boldsymbol{\Sigma}_y^*)^{-1} + \boldsymbol{\Sigma}_k^{-1})^{-1}. \end{aligned}$$

The integral (4) is approximated as follows:

– For some $L > 0$ and for all $l \in \{1, \dots, L\}$, sample

$$\mathbf{x}_i^{*(l)} \stackrel{\text{iid}}{\sim} G_i^* \text{ at iteration } l,$$

– Set

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

where $\mathcal{C}_i$ is a normalizing constant:

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

Note that this sampling step might be quite time consuming. In practice, it is likely that taking the approximation $\mathcal{P}(y_i | \mathbf{x}_i^{*(l)}) \approx \gamma_{y_i} f_{\mathcal{N}}(\mathbf{x}_i^{*(l)} | \mathbf{m}_{y_i}^*, \Sigma_{y_i}^*)$ for convenient γ_{y_i} , could give rise to good results while allowing closed expression for $\tilde{\mathcal{T}}(y_i)$. In this case, we obtain

$$\tilde{\mathcal{T}}(y_i) = \mathcal{C}_i \gamma_{y_i}.$$

The sampling of $(\mathbf{x}_i | y_i, u_i, \mathbf{w}, \Theta^*, B_i)$ is also complex since the distribution has no known form. We use the Metropolis-Hastings algorithm [18], with a proposal density being the product of the Dirichlet process mixture of Gaussians and a normal density extended in each observed LOR direction. The result is a re-weighted mixture of Gaussians in the LOR.

4 Application to real PET data

4.1 Materials

To illustrate the applicability of the proposed method, we have performed reconstructions of clinical PET data, acquired in list-mode format from a 90 mn brain study. The PET tracer is [^{11}C]-labeled Pittsburgh Compound ($[^{11}\text{C}] - \text{PIB}$). Since we aim to evaluate our reconstruction algorithm in the context of low injected dose, the full list-mode file was randomly decimated by a factor 30 ($\div 30$), leading to an artificially $\div 30$ reduced dose. Note that concerning the random events, the decimated list-mode is disadvantageous in comparison to a genuine low dose. Indeed, if the dose had been 30 times lower, the random events proportion would have been divided by a factor 30 whereas this proportion is not affected after decimation.

The scanner geometry has been modelled via $\mathcal{P}(y|\mathbf{x})$. We account for attenuation and random coincidence events in the BNP reconstruction as follows. Attenuation was taken into account by using a tissue attenuation map, which was measured by conducting a transmission scan prior to the dynamic PET scans. The tissue attenuation map was then projected to define a measure of attenuation for each LOR. The scanner normalization correction was derived from the one provided by the constructor. The distribution of random coincidences is estimated using a histogram of delayed coincidences for each detector. Estimating the distribution of scattered events needs to use the corrections terms established by the constructor. However, on the Biograph scanner, the scatters are known for sinogram bins while our method performs on list-mode data and requires correction terms for each individual LOR. We have to adapt corrective terms by a re-sampling procedure. This work is still in progress and the correction for scattered coincidences was not applied in the results presented thereafter.

Validation studies need to evaluate our proposed method performance against the constructor reconstruction method, namely the 3D maximum likelihood approach (ML-EM). Since the ML estimate is very noisy, in clinical platforms regularization is done by post-smoothing the reconstructed image.

4.2 Proposed BNP method

As mentioned earlier, our proposed reconstruction scheme has been evaluated under $\div 30$ reduced dose.

Parameters of the DPM were chosen as follows: $\alpha = 1000$, $n_0 = 4$, $\rho = 10^{-2}$, $\Sigma_0 = 6.25\text{mm}^2 \times \mathbb{I}_3$, where $\mathbb{I}_3$ is the identity matrix of $\mathbb{R}^3$. We computed 5,000 burn-in iterations of the sampler, followed by 10,000 which were retained for Monte-Carlo estimates.

At the convergence of the algorithm, we get draws from the joint posterior distribution of all model parameters. From each retained draw at iteration (s), we construct an estimate of the spatio-temporal density of recorded true events,

$$f_{G^{(s)}}(\mathbf{x}) \approx \sum_{k=1}^{K^*} w_k^{(s)} f_{\mathcal{N}}(\mathbf{x} | \theta_k^{*(s)}), \quad (5)$$

where K^* is the number of components retained by the slice sampler at the given iteration. The chosen estimator for the activity distribution is the conditional mean of f_G on the $N = 10,000$ retained draws after convergence,

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

The conditional mean is the estimator which minimizes the mean square error.

4.3 Results

Let us now give some details about the estimation of the background events. Attenuation was accounted for by using a Computerized Tomography (CT) transmission scan. The latter is performed prior to the PET scan and generates a tissue attenuation map which is projected to define a measure of attenuation for each LOR. The distribution of random coincidences is estimated using a histogram of delayed coincidences for each detector. Estimating the distribution of scattered events needs to use the corrections terms established by the constructor. However, on the Biograph scanner, reconstructions and corrective terms are based upon sub-sampled data. In fact, individual LORs are mixed after histogramming in parallel projection space using nearest neighbour approximation (each LOR is assigned to the nearest sinogram bin) and span (axial compression). Using list-mode data, we have to adapt corrective terms by a re-sampling procedure. We refer to [13] for more details on the various corrections of PET data.

Results obtained by the two competing algorithms are depicted in Figure 2. We compare our BNP results with $\div 30$ dose to $\div 1$ dose reconstruction based on filtered 3D ML-EM algorithm. The 3D ML-EM algorithm at convergence is followed by a post-smoothing Gaussian filter (FWHM= 3 mm). The ML-EM reconstruction has been performed using the full list-mode data (i.e., under “standard dose”).

Figure 2: Rows. top: sagittal view; middle: axial view; bottom: coronal view. Columns. left: dose $\div 30$ BNP; middle: dose $\div 1$ post. smooth. ML-EM (without scattering correction); right: dose $\div 1$ post. smooth. ML-EM (with scattering correction).

Since we do not make any scattered events correction in BNP yet, we provide two versions of ML-EM reconstruction, with and without scattered events correction. Lines in Figure 2 correspond to three 2D different views: sagittal, coronal and axial. Column 1 represents the proposed BNP reconstruction including the correction of attenuation and random events, column 2 is the post smoothed ML-EM reconstruction corrected only for attenuation and randoms, and column 3 is the post-smoothed ML-EM reconstruction corrected for all (attenuation, randoms, scattered). It is worth noting that in our results, the discretization of the spatial distribution is only for visualization purposes and was chosen a posteriori to $256 \times 256 \times 128$, the same as in the ML-EM. Results highlight the good performance of the new reconstruction scheme under reduced dose, in terms of both SNR

and spatial resolution. This is due to the ability of the Normal-Inverse Wishart model to estimate a covariance matrix for each component of the mixture. Due to this versatility, the potentially infinite mixture of normal components turns out to be a convenient regularization model for the brain spatial activity distribution.

Using a 4D extension of the model (a Dependent Dirichlet process mixture, not presented here because of lack of space and model complexity), we can also obtain an estimation of the so-called *functional volumes* which define, in this application, regions of the brain having the same temporal behaviour of radiotracer activity. Determining functional volumes is necessary to compute physiological kinetic parameters. Most of the time, the spatial extension of a functional volume is delineated from an anatomical image. Then, the mean activity in the functional volume (which is in fact an anatomical volume) is computed using the sequence of 3D images reconstructed for each time frame. Such an approach is based on the assumption that there is a one-to-one correspondence between anatomical and functional areas, which is not the case. We have sought to obtain this functional segmentation directly from the reconstruction process. Figure 3 shows the estimated functional volumes for the blood, the scalp, the grey matter and the white matter.

Figure 3: Estimated functional volumes. *Up row:* blood (left), scalp (right). *Bottom row:* grey matter (left), white matter (right).

Finally, the most encouraging result of the 4D extension BNP model is our ability to estimate the *input function*, representing the (cumulative) availability of the radiotracer in arterial plasma, which is of great interest in dynamic clinical PET. In fact, in order to estimate the physiological parameters of interest, arterial blood sampling is necessary. However, this is traditionally performed via the invasive and laborious procedure of arterial cannulation. In Figure 4, we depict the measured input function (arterial) over time (in minutes); and the estimated BNP input function, overlaid with its posterior 95 % credible interval. We underline that the lack of correction of scattered events turns out to be a severe drawback here, especially at early times. Indeed, during and just after tracer injection, most of the recorded events come from scattering. The estimation of the input function may suffer from this effect. We are aware that scattered events correction should improve estimation of this function. Nevertheless, one can notice that the estimated and true input functions are close after the early stages.

Figure 4: Measured (red) input function over time. The BNP estimated input function (blue) is overlaid with its credible interval (purple).

4.4 Discussion

This paper casts the problem of clinical PET image reconstruction in the framework of Nonparametric Bayesian methods. We have demonstrated the ability of Dirichlet process mixtures to extract the spatial density from clinical PET data. The nonparametric approach has several advantages. In particular, it is independent of any assumptions about the distribution of the radiotracer. In other words, one needs not pre-assume any particular functional form for the tracer spatial activity and the functional volumes. The results indicate that this approach is relevant when working with low injected doses. The extended 4D approach does not require arterial blood sampling and thus removes a serious impediment to routine application to human PET studies. This portends the application of the proposed method in clinical platforms.

Let us now consider the limitations of the proposed approach. The major drawback is its computational demands, although it was implemented on a parallel computer ($\approx$ one day on a high performance computer). The code is written in the C++ language with a Python interface. Since there is no voxelization in this approach, the computational cost is not given by the system matrix size (as in ML-EM approaches) but rather by the loops on emissions proposals and allocations in the MCMC sampling. Since our long term goal is the routine application, increasing the the speed of the algorithm is a challenge that we must take up. We believe that implementation on Graphical Processor Units hardware is a promising way to go.

5 Conclusion

A novel approach to emission tomographic reconstruction is explored in this paper. The method allows the lowering of the dose and shows significant improvement of image quality under reduced dose, over ML-EM algorithm under normal dose.

The Bayesian nonparametric approach allows for continuous (non-voxel) spatial representation with a conditional density function being derived at each spatial location. Although our method is presented here in the context of PET brain study, it has general applicability. We refer to [12] for an application to cardiac PET data.

Bayesian Nonparametrics is a new and exciting field, with however many important problems remaining to be solved. One clear issue is the computational time and complexity. A number of fast approximations to the posterior have been proposed for Dirichlet process mixtures, relying on variational approximations and other strategies [19].

Notes

- 1 Electron-positron annihilation occurs when an electron and a positron collide.
- 2 The interested reader can find all the details about the computation of this probability density in [16].
- 3 After Griffiths, Engen and McCloskey.

References

- [1] Dunson DB. Nonparametric Bayes applications to biostatistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2010.
- [2] Müller P, Mitra R. Bayesian nonparametric inference—why and how. *Bayesian Anal.* 2013;8:269–302.
- [3] Robert CP. The Bayesian choice. From decision-theoretic foundations to computational implementation (Springer texts in statistics). New York: Springer-Verlag, 2001.
- [4] Hjort NL, Holmes C, Müller P, Walker SG. Bayesian nonparametrics (Cambridge series in statistical and probabilistic mathematics). Cambridge: Cambridge University Press, 2010.
- [5] Valk P, Bailey D. Positron Emission Tomography: basic science and clinical practice. New York: Springer–Verlag, 2003.
- [6] Ollinger J, Fessler J. Positron-emission tomography. *Signal Process Magazine, IEEE*, 1997;14:43–55.
- [7] Natterer F. The mathematics of computerized tomography. Society for Industrial and Applied Mathematics, 1986.
- [8] Shepp L, Vardi Y. Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans Med Imag.* 1982;1:113–22.
- [9] Lange K, Carson R. EM reconstruction algorithm for emission and transmission tomography. *J Comp Assist Tomo.* 1984;8:306–16.
- [10] Fall MD, Barat É, Comtat C, Dautremer T, Montagu T, Mohammad-Djafari A. A discrete-continuous Bayesian model for Emission Tomography. In: *Proceedings of the IEEE International Conference on Image Processing*, 2011a:1373–6.
- [11] Fall MD, Barat É, Comtat C, Dautremer T, Montagu T, Stute S. Continuous space-time reconstruction in 4D PET. In: *Proceedings of IEEE Medical Imaging Conference NSS/MIC*, 2011b:2581–6.
- [12] Fall MD, Barat É, Comtat C, Dautremer T, Montagu T, Stute S. Dynamic and clinical PETdata reconstruction: a nonparametric Bayesian approach. In: *Proceedings of the IEEE International Conference on Image Processing*, 2013:345–9.
- [13] Townsend D. The theory and practice of 3D PET, Vol. 32. Boston: Kluwer Academic Publishers, 1998.
- [14] Ferguson T. Bayesian density estimation by mixtures of Normal distributions. *Recent advances in Statistics: papers in honor of Herman Chernoff on his sixtieth birthday*, 1983:287–302.
- [15] Lo AY. On a class of Bayesian Nonparametric estimates: I. density estimates. *Annal Stat.* 1984;12:351–57.
- [16] Fall MD, Barat É, Comtat C, Dautremer T, Montagu T, Mohammad-Djafari A. A Bayesian nonparametric model for 3D Positron Emission Tomography. *Bayesian Anal.* 2012, Submitted.
- [17] Fall MD, Barat É. Sampling the two-parameter Poisson Dirichlet distribution: state of the art and a new method. Technical report, CEA, 2011.
- [18] Robert CP, Casella G. Monte Carlo statistical methods (Springer texts in statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc, 2005.
- [19] Blei DM, Jordan M. Variational inference for Dirichlet process mixtures. *Bayesian Anal.* 2006;1:121–44.