



**HAL**  
open science

# Nonparametric posterior learning for emission tomography

Fedor Goncharov, Eric Barat, Thomas Dautremer

► **To cite this version:**

Fedor Goncharov, Eric Barat, Thomas Dautremer. Nonparametric posterior learning for emission tomography. *SIAM/ASA Journal on Uncertainty Quantification*, 2023, 11 (2), pp.452-479. 10.1137/21M1463367 . cea-04123345v5

**HAL Id: cea-04123345**

**<https://cea.hal.science/cea-04123345v5>**

Submitted on 9 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1 Nonparametric posterior learning for emission tomography with multimodal data\*

2 Fedor Goncharov<sup>†</sup>, Éric Barat<sup>†</sup>, and Thomas Dautremer<sup>†</sup>

---

4 **Abstract.** We continue studies of the uncertainty quantification problem in emission tomographies such as PET  
5 or SPECT when additional multimodal data (anatomical MRI images) are available. To solve the  
6 aforementioned problem we adapt the recently proposed nonparametric posterior learning technique  
7 to the context of Poisson-type data in emission tomography. Using this approach we derive sampling  
8 algorithms which are trivially parallelizable, scalable and very easy to implement. In addition, we  
9 prove conditional consistency and tightness for the distribution of produced samples in the small  
10 noise limit (i.e., when the acquisition time tends to infinity) and derive new geometrical and nec-  
11 essary condition on how MRI images must be used. This condition arises naturally in the context  
12 of identifiability problem for misspecified generalized Poisson models with wrong design. We also  
13 contrast our approach with Bayesian Markov Chain Monte Carlo sampling based on one data aug-  
14 mentation scheme which is very popular in the context of Expectation-Maximization algorithms for  
15 PET or SPECT. We show theoretically and also numerically that such data augmentation signifi-  
16 cantly increases mixing times for the Markov chain. In view of this, our algorithms seem to give a  
17 reasonable trade-off between design complexity, scalability, numerical load and assessment for the  
18 uncertainty.

19 **Key words.** tomography, inverse problems, MCMC, Bayesian inference, bootstrap

20 **AMS subject classifications.** 62-04, 62F15, 62C10

21 **1. Introduction.** Emission tomographies (further referred as ET) such as Positron Emis-  
22 sion Tomography (PET) or Single Photon Emission Computed Tomography (SPECT) are  
23 functional imaging modalities of nuclear medicine which are used to image activity processes  
24 and, in particular, metabolism in soft tissues. The level of metabolism and uptake of specific  
25 biomarkers provide crucial information for diagnostics and treatment of cancers; see e.g., [53],  
26 [36] and references therein. Therefore, quality of images in ET and their respective resolution  
27 are critical for the diagnostics-treatment pipeline. In this work we continue studies on the two  
28 following problems:

29 *Problem 1.* Quantify the uncertainty of reconstructions in ET.

30 *Problem 2.* Regularize the inverse problem using the multimodal data (e.g., images from  
31 CT or MRI).

32 *Problem 1* is not new and several approaches have been established already which in turn  
33 can be grouped according to the statistical view of the problem – frequentist ([12], [1], [30]),  
34 Bayesian ([23], [54], [10], [46], [3], [14]) and bootstrap ([20], [8], [28], [15]). Note that given  
35 list is far from being complete and it should include references therein.

36 *Problem 2* can be splitted further depending on which type of exterior data are used -  
37 CT or MRI. More generally, main reasons to use multimodal data in ET are the ill-posedness  
38 of corresponding inverse problems (in PET/SPECT forward operators are ill-conditioned; see

---

\*Submitted to the editors October 26, 2022.

**Funding:** This work is partly supported by the 'MMIPROB' project funded by ITMO Cancer (France).

<sup>†</sup>Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France ([fedor.goncharov@cea.fr](mailto:fedor.goncharov@cea.fr)).

39 e.g., [24]) and very low signal-to-noise ratio in the raw measured data. All this together  
 40 results in loss of resolution in reconstructed images and consequently in oversmoothing, e.g.,  
 41 when applying standard methods such as spatially invariant filters for post-smoothing. The  
 42 common way of using CT and MRI images consists in extracting boundaries of anatomical  
 43 features and embedding them into regularization schemes via special penalties and/or non-  
 44 invariant filters; see e.g., [13], [6], [22], [7], [52]. The foundation of the above approaches is that  
 45 there are correlations between PET and MRI signals starting from simple anatomical up to  
 46 biological ones (e.g., PET-MRI investigation on tumor imaging in [5]). Therefore, potentially  
 47 MRI data can be used to regularize accurately the inverse problem, however, it still requires  
 48 construction of fine models to describe such correlations. Finally, from very practical point  
 49 of view **Problem 2** with additional MRI data is of interest due to availability of commercially  
 50 available models of PET-MRI scanners [34], [26] which allow simultaneous registrations of  
 51 both signals. In this work for multimodal data we use series of presegmented anatomical MRI  
 52 images which are used differently than it was explained before. In **section 2** we explain in  
 53 detail how we use the MRI data and compare it with existing approaches.

54 Already the definition of uncertainty in **Problem 1** is not obvious: for exposure period  
 55  $[0, t]$  raw data  $Y^t$  (sinogram) is generated by unknown (binned) point process  $PP^t$  (typically  
 56 it is assumed to be Poisson with unknown intensity parameter  $\lambda_* \in \mathbb{R}_+^p$  and known design  $A \in$   
 57  $\text{Mat}(d, p)$ , i.e.,  $PP^t = PP_{A\lambda_*}^t = \text{Po}(t \cdot A\lambda_*)$ ). Therefore, for any estimator  $\hat{\lambda}^t$  the uncertainty  
 58 propagates directly from  $Y^t$ . This is known as aleatoric uncertainty which corresponds to  
 59 frequentist approach, and for ET it often leads to estimation of confidence bounds for the  
 60 maximum likelihood estimator (MLE) or the penalized maximum log-likelihood estimator  
 61 (pMLE or MAP; both are  $M$ -estimators [49]); see e.g., [12]. Frequentist approach has an  
 62 advantage of being relatively robust to model misspecification (i.e., when  $PP^t \neq PP_{A\lambda}^t$ ). In  
 63 this case, for large  $t$  consistent estimator  $\hat{\lambda}^t$  will tend a.s. to a projection of  $PP^t$  onto  $PP_{A\lambda}^t$   
 64 with respect to some chosen distance between probability distributions (e.g., for Kullback-  
 65 Liebler divergence). Under additional assumptions on  $PP^t$  even in misspecified case it is still  
 66 possible to establish asymptotic distribution of  $\hat{\lambda}^t$  such as asymptotic normality, from which  
 67 the confidence intervals can be retrieved. However, practical use of asymptotic estimates for  
 68 ET seems doubtful since very little data are available in a single scan.

69 Epistemic uncertainty is another type of uncertainty which corresponds to Bayesian or  
 70 bootstrap approaches in statistics. For the Bayesian case the initial uncertainty on the pa-  
 71 rameter of interest is encoded in some prior measure (using anatomical information from side  
 72 images, assumptions on support and smoothness) which is updated using model  $PP_{A\lambda}^t$  and  
 73 conditioning on  $Y^t$  to define the posterior distribution via the well-known Bayes' formula.  
 74 Sampling from such complicated posteriors is usually done via Markov Chain Monte Carlo  
 75 (MCMC) techniques [54], [23], [10], [14]. However, there are common bottlenecks: compli-  
 76 cated design of samplers and their implementations, high numerical load per iteration, lack of  
 77 scalability and most importantly – poor mixing in constructed chains; see e.g., [14], [50], [41].  
 78 Additional methodological issue is the misspecification of the model (e.g., incorrect design)  
 79 which cannot be included in the classical Bayesian framework and for robust inference it leads  
 80 to the recently proposed general Bayesian updating and bootstrap-type sampling; see [42],  
 81 Section 1.

82 As noted before bootstrap is another attractive technique to assess the uncertainty which  
83 can be also seen as some probabilistic sensitivity analysis or as approximate/exact sampling  
84 from nonparametric Bayesian posteriors; see e.g., [38], [35], [16]. Nontrivial questions for  
85 bootstrapping ET are the following ones: (1) how to define the procedure for Poisson-type  
86 raw data in ET and also include side information (2) provide guarantees (theoretical and  
87 numerical) on the coverage of the true signal by new credible intervals. A common approach to  
88 answer question (1) is to use resampling; see e.g., [20], [8]. For ET this one targets to resample  
89 photon counts and then propagate the uncertainty by using any reconstruction algorithm  
90 (e.g., FBP (Filtered backprojection), MLE or MAP (maximum a posteriori)). Question (2)  
91 is resolved theoretically often by demonstrating asymptotic equivalence between bootstrap,  
92 Bayesian and frequentist approaches via Bernstein von-Mises type theorems (see e.g., [49], [35],  
93 [39] or equivalence of Edgeworth's expansions for higher orders (see [42]) and numerically via  
94 calibration (e.g., using Q-Q plots).

95 In view of the above discussion, we note that for practice it seems that it is not of great  
96 importance which kind of uncertainty model is used – frequentist, Bayesian or bootstrap.  
97 The most important is to make usable the resulting framework and algorithms by practi-  
98 tioners, hence, they should be simple to implement, desirably with tractable parameters and  
99 numerically efficient (scalability is crucial for high-dimensional models in ET).

100 Being inspired with nonparametric posterior learning (further referred as NPL) originating  
101 from [35], [16], we propose sampling algorithms for ET of bootstrap type with and without  
102 MRI data at hand. Therefore, our main contribution is that we extend the NPL originally  
103 proposed for regular statistical models and i.i.d data to the non-regular generalized Poisson  
104 model of ET (see [3]), where the raw data are not i.i.d but a sample from a point process.  
105 The initial motivation for this work was the problem of poor mixing for the Gibbs-type  
106 sampler in [14] which was designed for posterior sampling in the PET-MRI context. Below  
107 we give a detailed analysis of this phenomenon and conclude with a few generic advices on  
108 design of MCMC-samplers for ill-posed inverse problems such as PET or SPECT. Our new  
109 algorithms solve the above problem since sampled images are automatically i.i.d, moreover,  
110 the scheme is trivially parallelizable, scalable and very easy to implement because it relies  
111 on the well-known EM-type reconstruction methods from [44], [11]. Our samplers are tested  
112 numerically on a synthetic dataset by demonstrating the regularization effect of MRI as well  
113 on calibration of the posterior. We also conduct a theoretical study for when large dataset  
114 is available (for ET this is equivalent to  $t \rightarrow +\infty$ ) and establish consistency and tightness  
115 of the posterior for almost any trajectory  $Y^t$ ,  $t \in [0, +\infty)$ . As a byproduct of our study, for  
116 the misspecified scenario with incorrect design matrix (which is always true in practice) we  
117 discover an intuitive sufficient condition for identifiability to persist. The latter can be of  
118 interest for further theoretical studies of ET model under misspecification.

119 This paper is organized as follows. In [section 2](#) we give notations and all necessary  
120 preliminaries on the statistical model of ET and on use of multimodal data. In [section 3](#)  
121 we give a very informative example for the problem of poor mixing for MCMC. In [section 4](#)  
122 we adapt the NPL for ET context and derive our sampling algorithms. In [section 5](#) we  
123 present results of the numerical experiment on a synthetic dataset. In [section 6](#) we study  
124 theoretically the asymptotic properties of our algorithms. In [section 7](#) we discuss our results  
125 and possibilities for future work.

## 2. Preliminaries.

**2.1. Notations.** By  $\mathbb{N}_0$  we denote the set of non-negative integers,  $\mathbb{R}_+^n$  denotes the positive cone of  $\mathbb{R}^n$ , by  $x \succeq y$ ,  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^n$ , we denote the property that  $x_j \geq y_j$  for all  $j = 1, \dots, n$ ,  $x \succ y$  denotes the same but with strict inequalities, both  $\langle x, y \rangle$  or  $x^T y$  stand for the scalar product,  $R_+(A)$  denotes the image of  $\mathbb{R}_+^p$  under action of operator  $A \in \text{Mat}(d, p)$ , by  $X \sim F$  we denote the property that r.v.  $X$  has distribution  $F$ ,  $\text{Po}(\lambda)$  denotes the Poisson distribution with intensity  $\lambda$ ,  $\lambda \geq 0$ ,  $\Gamma(\alpha, \beta)$  denotes the gamma distribution with shape  $\alpha$ , and scale  $\beta$  ( $\xi \sim \Gamma(\alpha, \beta)$ ,  $\mathbb{E}\xi = \alpha\beta$ ,  $\text{var}(\xi) = \alpha\beta^2$ ). Let  $A \in \text{Mat}(d, p)$ , then  $\text{cond}(A)$  denotes the condition number of  $A$ ,  $A_I$ ,  $I \subset \{1, \dots, d\}$  denotes the submatrix of  $A$  with rows indexed by elements in  $I$ ,  $\text{Span}(A^T)$  denotes the span of the rows of  $A$  being considered as vectors in  $\mathbb{R}^p$ . Let  $Z$  be a complete separable metric space equipped with metric  $\rho_Z(\cdot, \cdot)$  and boundedly finite non-negative measure  $dz$ ,  $B(Z)$  denotes the sigma algebra of borel sets in  $Z$ . By  $\mathcal{PP}$  we denote a (spatio-temporal) point process on  $Z \times \mathbb{R}_+$  and  $\mathcal{PP}_\Lambda$  denotes the Poisson point process on  $Z \times \mathbb{R}_+$  with intensity  $\Lambda(z) dz dt$ , where  $\Lambda$  is the nonnegative function  $\Lambda = \Lambda(z)$ ,  $z \in Z$ ,  $\Lambda$  is integrable w.r.t  $dz$ . Weighted gamma process on  $Z$  is denoted by  $GP(\alpha, \beta) = G_{\alpha, \beta}$ , where  $\alpha$  is the shape measure on  $Z$  and  $\beta$  is the scale which is a non-negative function  $Z$  and also  $\alpha$ -integrable; see, e.g., [33] for construction. Finally, by  $\mathcal{KL}(P, Q)$  we denote the standard Kullback-Leibler divergence between probability distributions  $P, Q$ .

**2.2. Mathematical model for ET.** Raw data in ET are described by the so-called sinogram  $Y^t = (Y_1^t, \dots, Y_d^t) \in (\mathbb{N}_0)^d$  which stands for the photon counts recorded during exposure period  $[0, t]$  along  $d$  lines of response (LORs). It is assumed that

$$(2.1) \quad \begin{aligned} Y_i^t &\sim \text{Po}(t\Lambda_i), \Lambda_i = a_i^T \lambda, \\ Y_i^t &\text{ are mutually independent for } i \in \{1, \dots, d\}, \end{aligned}$$

where  $\lambda \in \mathbb{R}_+^p$  is the parameter of interest on which we aim to perform inference. In practice, vector  $\lambda$  denotes the spatial emission concentration of the isotope measured in  $[\text{Bq}/\text{mm}^3]$ , that is  $\lambda_j$  is the concentration at pixel  $j \in \{1, \dots, p\}$ . Vector  $\Lambda = (\Lambda_1, \dots, \Lambda_d)$  denotes the observed photon intensities along LORs  $\{1, \dots, d\}$ , respectively. To separate the LORs with strictly positive intensities from those ones with zeros we introduce following notations:

$$(2.2) \quad I_0(\Lambda) = \{i : \Lambda_i = 0\}, I_1(\Lambda) = \{i : \Lambda_i > 0\}, I_0 \sqcup I_1 = \{1, \dots, d\}.$$

Collection of  $a_i \in \mathbb{R}^p$  in (2.1) constitute matrix  $A = [a_1^T, \dots, a_d^T]^T$ ,  $A \in \text{Mat}(d, p)$  which is called by projector or system matrix in applied literature on ET and by design (or design matrix in statistical literature). Each element  $a_{ij}$  in  $A$  denotes the probability to observe a pair of photons along LOR  $i \in \{1, \dots, d\}$  if both they were emitted from pixel  $j \in \{1, \dots, p\}$ . In view of such interpretation, for design  $A$  we assume the following:

$$(2.3) \quad a_{ij} \geq 0 \text{ for all pairs } (i, j),$$

$$(2.4) \quad A_j = \sum_{i=1}^d a_{ij}, 0 < A_j \leq 1 \text{ for all } j \in \{1, \dots, p\},$$

$$(2.5) \quad \sum_{j=1}^p a_{ij} > 0 \text{ for all } i \in \{1, \dots, d\}.$$

165 If any of formulas (2.4), (2.5) would not be satisfied, then, in practice it would mean that  
 166 either some pixel is not detectable at all (hence it can be completely removed from the model)  
 167 or some detector pair is broken and cannot detect any of incoming photons. These scenarios  
 168 are outside of our scope.

169 It is well-known that the inverse problems for PET and SPECT are mildly ill-posed (see  
 170 e.g., [24], [37]), which in practice means that

$$171 \quad (2.6) \quad \ker A \neq \{0\}.$$

172 *Remark 2.1.* Matrix  $A$  represents a discretization of weighted Radon transform operator  
 173  $R_a$  for ET with complete angle data on the plane (see [37], Chapter 2). Since  $A$  approximates  
 174  $R_a$  in strong operator norm we know that

$$175 \quad (2.7) \quad \sigma_k \asymp k^{-1/2}, \quad k = 1, \dots, p,$$

176 where  $\sigma_k$  are the singular values of  $A$ . In particular, even if  $A$  is injective for  $p$  large enough,  
 177 due to (2.7), it may happen that  $\text{cond}(A) > \varepsilon_F^{-1}$ , where  $\varepsilon_F$  is the floating-point precision. In  
 178 the latter case, due to the cancelling effect singular values of  $A$  numerically will be equivalent  
 179 to machine zeros which means then exactly the existence of a nontrivial kernel for  $A$ .

180 Likelihood and negative log-likelihood functions for model in (2.1) are given by the for-  
 181 mulas:

$$182 \quad (2.8) \quad PP_{A,\lambda}^t(Y^t) = p(Y^t | A, \lambda, t) = \prod_{i=1}^d \frac{(ta_i^T \lambda)^{Y_i^t}}{Y_i^t!} e^{-ta_i^T \lambda}, \quad \lambda \in \mathbb{R}_+^p, \quad t \geq 0,$$

$$183 \quad (2.9) \quad L(\lambda | Y^t, A, t) = \sum_{i=1}^d -Y_i^t \log(t\Lambda_i) + t\Lambda_i, \quad \Lambda_i = a_i^T \lambda.$$

185 Note that for  $A$  satisfying (2.6) and for any  $Y^t$  function  $L(\lambda | Y^t, A, t)$  is not strictly  
 186 convex even at the point of the global minima since  $L(\lambda + u | Y^t, A, t) = L(\lambda | Y^t, A, t)$   
 187 for any  $\lambda \in \mathbb{R}_+^p$  and  $u \in \ker A$ . To avoid numerical instabilities due to this phenomenon a  
 188 convex penalty  $\varphi(\lambda)$  is added to  $L(\lambda | Y^t, A, t)$ , so we also consider the penalized negative  
 189 log-likelihood:

$$190 \quad (2.10) \quad L_p(\lambda | Y^t, A, t, \beta^t) = L(\lambda | Y^t, A, t) + \beta^t \varphi(\lambda), \quad \lambda \in \mathbb{R}_+^p,$$

191 where  $\beta^t \geq 0$  is the regularization coefficient. Parameter  $\beta^t$  may increase with  $t$  at a certain  
 192 rate which is important for practice in order to increase the signal-to-noise ratio in recon-  
 193 structed images.

194 **2.3. Regularization penalty.** The role of penalty  $\varphi(\lambda)$  in (2.10) is to decrease the numer-  
 195 ical instability in the underlying inverse problem and to make function  $L_p(\lambda | Y^t, A, t, \beta^t)$   
 196 more convex, especially in directions close to  $\ker A$ .

197 In view of this we assume that

$$198 \quad (2.11) \quad \varphi \text{ is continuous and convex on } \mathbb{R}^p,$$

$$199 \quad (2.12) \quad g_u(w) = \varphi(u + w) \text{ is strictly convex in } w \in \ker A \text{ for any } u \in \text{Span}(A^T).$$

201 For numerical tests in [section 5](#) we choose  $\varphi$  to be the sum of two pairwise-difference  
 202 functions for neighboring pixels: first is of log-cosh type which is standard for ET (see [\[3\]](#),  
 203 [\[54\]](#)), and second is the pure  $\ell_2$ -squared norm to add more smoothness to sparse images  
 204 reconstructed with log-cosh type regularization.

205 Since  $A$  is not injective, even for infinite amount of data ( $Y^t \sim PP_{A,\lambda_*}^t$ ,  $t \rightarrow +\infty$ ), one is  
 206 able to find  $\lambda_*$  at most up to its projection  $\ker A$  (modulo extra information due to constraint  
 207  $\lambda_* \in \mathbb{R}_+^p$ ). With regularization the projection of  $\lambda_*$  onto  $\ker A$  will be defined uniquely by  $\varphi$   
 208 and positivity constraints. To describe this effect we define the following function:

$$209 \quad (2.13) \quad w_{A,\lambda}(u) = \arg \min_{\substack{\lambda+u+w \geq 0 \\ w \in \ker(A)}} \varphi(\lambda + u + w), \quad u \in \text{Span}(A^T), \lambda \geq 0.$$

211 Then, intuitively (this is made rigorous in [section 6](#)), the best one can hope to reconstruct  
 212 using MAP-estimator in [\(2.10\)](#) (or, equivalently, the penalized KL-projection) when  $t \rightarrow +\infty$   
 213 and  $\beta^t/t \rightarrow 0$ , will be

$$214 \quad (2.14) \quad \lambda_{*opt} = \lambda_* + w_{A,\lambda_*}(0) = \lim_{\beta \rightarrow +0} \arg \min_{\lambda \geq 0} L_p(\lambda | A\lambda_*, A, 1, \beta)$$

216 Thus, in what follows, the numerical quality of reconstructions, calibration etc., is tested  
 217 against  $\lambda_{*opt}$  rather than  $\lambda_*$  which is inaccessible no matter the amount of data.

218 **2.4. Multimodal data for ET.** In order to increase the SNR in reconstructed images and  
 219 not to loose a lot in resolution one can regularize the inverse problem using multimodal data  
 220 – scans from CT or MRI. We choose MRI since it provides anatomical information with high  
 221 contrast in soft tissues in comparison to CT (see [Figure 1](#) (a), (b)).

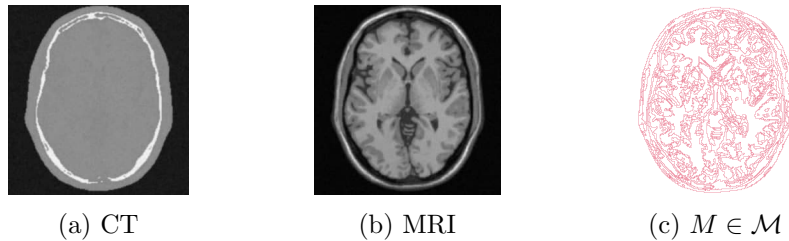


Figure 1: (a), (b) Multimodal data for ET of the brain; (c) segmented MRI-image in (b)

222 MRI-guided reconstructions in PET is an active topic of research (see the discussion  
 223 in [\[15\]](#) and references therein), however, still a lot of work is needed to describe precisely  
 224 correlations between ET and MRI signals (especially from biological point of view). Because  
 225 of the latter current use of MRI data is essentially image-based: spatially regularizing penalties  
 226 are constructed using MRI data in [\[4\]](#), [\[5\]](#), [\[52\]](#) (PET signals are penalized stronger when being  
 227 constant across edges in MRI images), models built upon MRI-segmented data for locally-  
 228 constant tracer distribution are used in [\[14\]](#) and also in our work.

229 In this work we assume that our side data consists of  $r$  presegmented MRI images  $\mathcal{M} =$   
 230  $\{M_1, \dots, M_r\}$  (see [Figure 1](#) (c); segmentations of MRI images are precomputed using the dd-  
 231 CRP algorithm from [\[17\]](#)), where segments are being disjoint and connected subsets of pixels.  
 232 First, using  $\mathcal{M}$  we construct a lower-dimensional model  $Y^t \sim \text{Po}(t\Lambda_{\mathcal{M}})$ ,  $\Lambda_{\mathcal{M}} = A_{\mathcal{M}}\lambda_{\mathcal{M}}$ ,  
 233  $A \in \text{Mat}(d, p_{\mathcal{M}})$ ,  $\lambda_{\mathcal{M}} \in \mathbb{R}_+^{p_{\mathcal{M}}}$  ( $p_{\mathcal{M}} \ll p$ ); see also [\(2.1\)](#). Second, randomized pseudo-  
 234 observations(-sinograms) from this model are mixed with observed  $Y^t$  into new sinograms.  
 235 Subsequent reconstructions from the latter constitute our samples being regularized by  $\mathcal{M}$ .

236 Now we explain the construction of  $A_{\mathcal{M}}$  and  $\lambda_{\mathcal{M}}$  and the actual sampling will be given  
 237 further in [subsection 4.4](#). Let  $p_k$  be the number of segments in  $M_k \in \mathcal{M}$ ,  $S(M_k)$  be their  
 238 collection. For each  $M_k$  we define new projector by the formulas:

$$239 \quad (2.15) \quad A_k = (a_{ij}^k) \in \text{Mat}(d, p_k),$$

$$240 \quad (2.16) \quad a_{is}^k = \sum_{j=1}^p a_{ij} \mathbb{1}\{\text{pixel } j \text{ belongs to segment } s \in S(M_k)\}, \quad k \in \{1, \dots, r\},$$

242 where  $A = (a_{ij})$  is the projector for the full model from [subsection 2.2](#). Finally, we stack all  
 243 segments and projectors into one model:

$$244 \quad (2.17) \quad A_{\mathcal{M}} = (A_1, \dots, A_r), \quad p_{\mathcal{M}} = \sum_{k=1}^r p_k,$$

$$245 \quad (2.18) \quad \lambda_{\mathcal{M}} = (\lambda_1^1, \dots, \lambda_{p_1}^1, \dots, \lambda_1^r, \dots, \lambda_{p_r}^r), \quad \Lambda_{\mathcal{M}} = A_{\mathcal{M}}\lambda_{\mathcal{M}}, \quad \Lambda_{\mathcal{M}} = (\Lambda_{\mathcal{M},1}, \dots, \Lambda_{\mathcal{M},d}).$$

247 Therefore,  $\lambda_{\mathcal{M}}$  is a positive linear combination of all segments from all images in  $\mathcal{M}$  with  
 248 constant signal in each segment, and  $A_{\mathcal{M}}$  being respective projector derived from  $A$ . For  $A_{\mathcal{M}}$   
 249 we assume that it is injective and well-conditioned, that is

$$250 \quad (2.19) \quad \ker A_{\mathcal{M}} = \{0\}, \quad \text{cond}(A_{\mathcal{M}}) < c_{\mathcal{M}},$$

252 where  $c_{\mathcal{M}}$  is some moderate constant. The latter assumption reflects the idea that images in  
 253  $\mathcal{M}$  consist of low number of large segments.

254 **3. A motivating example for NPL in ET.** Recently a Gibbs-type sampler was proposed  
 255 in [\[14\]](#) for Bayesian inference for PET-MRI. Despite a number of positive practical features  
 256 (spatial regularization, use of multimodal data) the problem of slow mixing for the corre-  
 257 sponding Markov chain was observed. Below we consider a simplified version which shares  
 258 the same mixing problem and explain the phenomenon numerically and theoretically.

259 In algorithms for ETs it is common to augment data  $Y^t$  by  $n^t = \{n_{ij}^t\}$ , where  $n_{ij}^t$  is the  
 260 number of photons being emitted from pixel  $j$  and detected in LOR  $i$ ,  $n_{ij}^t \sim \text{Po}(ta_{ij}\lambda_j)$ ,  $n_{ij}^t$   
 261 are mutually independent for all  $(i, j)$ ; see e.g., [\[44\]](#). In view of this physical interpretation,  
 262 for pair  $(n^t, Y^t)$  the following coherence condition must be satisfied:

$$263 \quad (3.1) \quad \sum_{j=1}^p n_{ij}^t = Y_i^t \text{ for all } i \in \{1, \dots, d\}.$$

264 By [\(3.1\)](#) one sees  $Y^t$  is a function of  $n^t$ , so  $(Y^t, n^t)$  is indeed a data augmentation of  $Y^t$ . Note  
 265 that  $n^t$  are not observed in a real experiment but they greatly simplify design of samplers (see



266 e.g., [25], [14]), because conditional distributions  $p(n^t | Y^t, A, \lambda, t)$ ,  $p(\lambda | n^t, A, t)$  admit simple  
 267 analytical forms even for nontrivial priors involving multimodal data. For our example below  
 268 we use only a simple pixel-wise positivity gamma-prior:

$$269 \quad (3.2) \quad \pi(\lambda) = \prod_{j=1}^p \pi_j(\lambda_j), \quad \pi_j = \Gamma(\alpha, \beta^{-1}), \quad \alpha > 0, \quad \beta > 0,$$

270 where  $\alpha, \beta$  are some fixed constants. For the prior in (3.2) and model (2.1) distributions  
 271  $p(n^t | Y^t, A, \lambda, t)$ ,  $p(\lambda | n^t, A, t)$  are as follows:

$$272 \quad (3.3) \quad p(n_{ij}^t | Y^t, A, \lambda, t) = \text{Multinomial}(Y_i^t, p_{i1}(\lambda), \dots, p_{ip}(\lambda)),$$

$$p_{ij}(\lambda) = \frac{a_{ij}\lambda_j}{\sum_k a_{ik}\lambda_k}, \quad i \in \{1, \dots, d\},$$

$$273 \quad (3.4) \quad p(\lambda_j^t | n^t, Y^t, A, t) = \Gamma\left(\sum_{i=1}^d n_{ij}^t + \alpha, (tA_j + \beta)^{-1}\right), \quad j \in \{1, \dots, p\},$$

274 where  $A_j$  is defined in (2.4).

275 Using (3.3), (3.4) the construction a Gibbs sampler for Bayesian posterior sampling from  
 276  $p(\lambda | Y^t, A, t)$  is straightforward.

278

---

**Algorithm 1** Gibbs sampler for  $p(\lambda | Y^t, A, t)$

---

1: **data** :  $Y^t$

2: **input** :  $\lambda_0 \in \mathbb{R}_+^p$ ,  $\pi(\lambda_j) = \Gamma(\alpha, \beta^{-1})$ ,  
 $B$  – number of samples

3: **for**  $k \leftarrow 1$  to  $B$  **do**

4:  $n_k^t \sim p(n^t | Y^t, A, \lambda_{k-1}, t)$

5:  $\lambda_k^t \sim p(\lambda | n_k^t, Y^t, A, t)$

6: **end for**

7: **return**  $\{\lambda_k^t\}_{k=1}^B$ ,

**Folklore:** empirical distribution of  $\{\lambda_k^t\}_{k=1}^B$  approximates posterior  $p(\lambda | Y^t, A, t)$

---

279 *Remark 3.1.* One may argue that prior in (3.2) is a very bad choice from practical point  
 280 of view, especially in view of ill-posedness of the inverse problem since it does not bring  
 281 any spatial regularization. However, the mixing rate for the Markov chain in [Algorithm 1](#)  
 282 asymptotically (i.e., when  $t \rightarrow +\infty$ ) will not depend on the choice of  $\pi(\lambda)$  in the small noise  
 283 limit due to Bernstein von-Mises phenomenon (see e.g., [3] and formulas (3.6), (3.5)). At the  
 284 same time, below we show that mixing is affected primarily by the choice of augmentation  
 285 scheme and the decision to sample  $n^t$ .

286 We consider the correlations between values of  $h(\lambda) = h^T \lambda$ ,  $h \in \mathbb{R}^p$ , for subsequent samples  
 287 from the Markov chain in [Algorithm 1](#):

$$288 \quad (3.5) \quad \gamma^t(h) = \text{corr}(h(\lambda_{k+1}^t), h(\lambda_k^t) | Y^t, t).$$

289 In formula (3.5) we assumed that the chain is in stationary state, i.e.  $k$  can be any.

290 Markov chain for the sampler in Algorithm 1 coincides with data augmentation schemes  
 291 from [31], [32], where the latter are exactly Gibbs samplers with only one layer of latent  
 292 variables. In Bayesian framework  $\gamma^t(h)$  is also known as fraction of missing information; see  
 293 [31]. In particular, in [31] authors gave an exact formula for  $\gamma^t(h)$  which can be written for  
 294 our example as follows:

$$295 \quad (3.6) \quad \gamma^t(h) = 1 - \frac{\mathbb{E}[\text{var}(h(\lambda) \mid n^t, Y^t, t) \mid Y^t, t]}{\text{var}(h(\lambda) \mid Y^t, t)}.$$

297 Exact formula for (3.6) for arbitrary  $t$  seem difficult (if possible) to obtain, however, in the  
 298 asymptotic regime  $t \rightarrow +\infty$  one can apply the Bernstein von-Mises type theorem from [3] and  
 299 arrive to the following simple expression:

$$300 \quad (3.7) \quad \gamma(h) = \lim_{t \rightarrow +\infty} \gamma^t(h) = 1 - \frac{h^T F_{aug}^{-1}(\lambda_*) h}{h^T F_{obs}^{-1}(\lambda_*) h}, \quad h \in \mathbb{R}^p, \quad \text{a.s. } Y^t, t \in (0, +\infty).$$

302 where

$$303 \quad (3.8) \quad \lambda_* \text{ is the true parameter, } \lambda_* \succ 0,$$

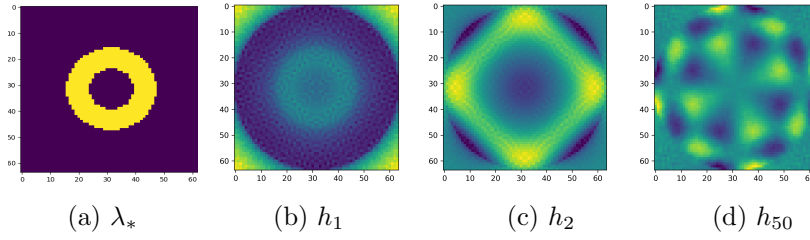
$$304 \quad (3.9) \quad F_{obs}(\lambda_*) = \sum_{i=1}^d \frac{a_i a_i^T}{\Lambda_i^*} = A^T D_{\Lambda^*}^{-1} A, \quad D_{\Lambda^*} = \text{diag}(\dots, \Lambda_i^*, \dots), \quad \Lambda_i^* = a_i^T \lambda_*,$$

$$305 \quad (3.10) \quad F_{aug}(\lambda_*) = \text{diag}(\dots, c_j, \dots), \quad c_j = A_j / \lambda_{*j}.$$

307 Note that from (2.5) and (3.8) it follows that  $\Lambda_i^* > 0$  for all  $i$ , therefore division by  $\Lambda_i^*$  in  
 308 (3.9) is well-defined. Matrices  $F_{obs}(\lambda_*)$ ,  $F_{aug}(\lambda_*)$  are the Fisher information matrices at  $\lambda_*$  for  
 309 Poisson models with observables  $Y^t$  and  $n^t$ , respectively. Note also that  $F_{obs}$  is not invertible  
 310 in the usual sense, so in (3.7) its pseudo-inversion in the sense of Moore-Penrose is considered.

311 *Remark 3.2.* Strict positivity assumption in (3.8) is not practical and a precise analytic  
 312 formula which extends (3.7) for  $\lambda_* \succeq 0$  can be established using the results from [3]. The  
 313 point is that model (2.1) is non-regular since the parameter of interest belongs to a domain  
 314 with a boundary, so a separate result for Bernstein von-Mises phenomenon is needed in this  
 315 case. For our toy example it is sufficient to consider the case in (3.8) as if we were interested  
 316 in mixing times of the chain in areas with positive tracer concentration.

317 Let  $h_1, \dots, h_p$  be the orthonormal basis of eigenvectors of  $F_{obs}(\lambda_*)$  with corresponding  
 318 eigenvalues  $s_1 \geq s_2 \geq \dots \geq s_p \geq 0$ . Intuitively, in  $\{h_m\}_{m=1}^p$  higher indices  $m$  correspond to  
 319 higher frequencies on images (see Figure 2 (a)-(d)).

Figure 2: eigenvectors  $h_m$  for  $F_{obs}(\lambda_*)$ 

320 From (3.7) it follows that

$$321 \quad (3.11) \quad \gamma(h_m) = 1 - s_m h_m^T F_{aug}^{-1} h_m.$$

323 Matrix  $F_{aug}$  is well-conditioned, continuously invertible and the quadratic term in (3.11)  
324 admits the following bound:

$$325 \quad (3.12) \quad F_{aug}^{-1}(\lambda_*) = \text{diag}(\dots, \frac{\lambda_{*j}}{A_j}, \dots) \Rightarrow h_m^T F_{aug}^{-1}(\lambda_*) h_m \leq \frac{\max_j(\lambda_{*j})}{\min_j(A_j)}.$$

326 Regular behavior of  $F_{aug}^{-1}$  is not surprising because this is the Fisher information matrix for  
327 latent variables  $n^t$  for which the inverse problem is not ill-posed at all. From (3.9) and the  
328 ill-conditioning nature of  $A$  it follows that  $F_{obs}(\lambda_*)$  is also ill-conditioned (see [18]), moreover,  
329  $s_m \asymp m^{-1}$  for large  $m$  (see Remark 2.1). From this and (3.11), (3.12) we conclude that

$$330 \quad (3.13) \quad \gamma(h_m) \approx 1 \text{ for large } m.$$

331 Formulas (3.5) and (3.13) constitute a clear evidence of poor mixing in the Markov chain  
332 in Algorithm 1. Though (3.7)–(3.13) were derived for  $t \rightarrow +\infty$ , they reflect well the be-  
333 havior of the chain for moderate  $t$  which is seen from the numerical experiment below (see  
334 Supplementary Materials, section SM5 for details).

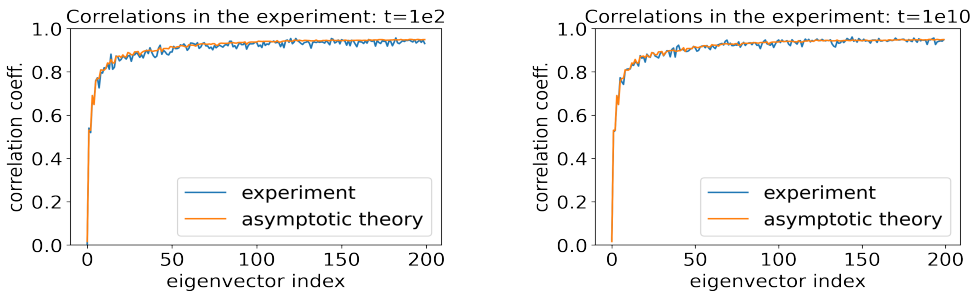


Figure 3:  $\text{corr}(h^T \lambda_k^t, h^T \lambda_{k+1}^t \mid Y^t)$  for  $t = 10^2, 10^{10}$  for  $h = h_m$ ; blue curve – empirical correlations computed from 2000 samples, orange curve – values for  $\gamma(h_m)$  for  $m = 1, \dots, 200$  by formula (3.7).

335 Here one concludes that mixing is much slower for high-frequency parts of images. There-  
336 fore, to estimate reliably, say mean  $h^T \lambda$  for some mask  $h \in \mathbb{R}^p$ , one needs almost infinite

337 number of samples if  $h$  contains a high-frequency component in terms of  $\{h_m\}_{m=1}^p$  (see Sup-  
 338plementary Material, [section SM4](#) for details). This also can be seen as a recommendation  
 339 for choosing  $h$  in practice:  $h$  should belong to  $\text{Span}(A^T)$  and  $|h^T h_m|$  should be as small as  
 340 possible for large  $m$ .

341 Note that such behavior of the sampler is not due to the choice of pixel-wise prior but  
 342 due to sampling of  $n_{ij}^t$ , which correspond to observations for the well-posed inverse problem.  
 343 A practical advice would be to avoid sampling of missing data in the Markov chain or to use  
 344 a strong smoothing prior/regularizer (for example by greatly increasing regularization coeffi-  
 345 cients so that asymptotic arguments in [\(3.7\)](#) will no longer hold but the posterior consistency  
 346 is still preserved). The latter approach will accelerate mixing at cost of oversmoothing in  
 347 sampled images.

348 By this negative but informative example we support the message in [\[50\]](#) saying that design  
 349 of a data augmentation scheme while preserving good mixing in the Markov chain is an ‘‘Art’’,  
 350 especially in the case of ill-posed inverse problems. In view of poor mixing, complexity of the  
 351 design and implementation, lack of scalability and high numerical load while using MCMC  
 352 ([\[54\]](#), [\[23\]](#), [\[10\]](#), [\[41\]](#), [\[14\]](#)) we turn to NPL as a practical relaxation of Bayesian sampling for  
 353 the problem of ETs.

354 **4. Nonparametric posterior learning for emission tomography.** To derive the NPL for  
 355 ET we prefer to start from the completely nonparametric setting as it was originally done  
 356 in [\[35\]](#). This allows us to concentrate on essential ideas behind and, moreover, all practical  
 357 algorithms can be directly obtained by binning nonparametric objects to finite dimensions. A  
 358 reader interested mainly in practical outcomes may skip this and go directly to [subsection 4.5](#).

359 **4.1. Nonparametric model for ET.** Nonparametric framework for ET can be seen as a  
 360 classical scanning scenario with a machine having infinite number of infinitely small detectors.  
 361 Let  $Z$  be the space of all detector positions in the acquisition geometry of a scanner (e.g.,  
 362 for one slice  $Z$  consists of all non-oriented straight lines in  $\mathbb{R}^2$ ). We also assume that  $Z$   
 363 is equipped with a boundedly-finite measure  $dz$  and with a metric  $\rho_Z$  (describing distances  
 364 between the lines). Then, for exposure period  $[0, t)$  the raw data are given by random measure  
 365  $Z^t$  generated by a point process:

$$366 \quad (4.1) \quad Z^t = \sum_{j=1}^{N^t} \delta_{(z_j, t_j)}, \quad (z_j, t_j) \in Z \times \mathbb{R}_+, \quad t_j < t_{j+1}, \quad t_j \leq t,$$

367 where

$$369 \quad (4.2) \quad N^t \text{ is total number of registered photons,}$$

$$370 \quad (4.3) \quad \{z_j\}_{j=1}^{N^t}, \{t_j\}_{j=1}^{N^t} \text{ are the LORs and arrival times of events, respectively.}$$

372 In practical literature on ET sample  $Z^t$  is known as list-mode data, whereas  $Y^t$  (sinogram) is  
 373 the version of  $Z^t$  integrated withing  $[0, t)$ . Under the assumption of temporal stationarity of  
 374 the generating process,  $Y^t$  contains the same amount of information as  $Z^t$  since the first one  
 375 is then a sufficient statistic.

376 For statistical model of  $Z^t$ , one takes the family of temporally stationary Poisson point  
 377 processes  $\mathcal{PP}_{A\lambda}$  on  $Z \times \mathbb{R}_+$ , where  $A, \lambda$  stand for nonparametric versions of the projector

378 and the tracer concentration from section 2. For intuition, in such model the intensity  
 379 parameter of the process in LOR  $z \in Z$  at time  $t$  is  $\Lambda(z) dz dt = [A\lambda](z) dz dt$ , therefore  
 380  $\Lambda(z) dt dz$  is the density function for the intensity measure of the Poisson process.

381 The negative log-likelihood for  $\mathcal{PP}_{A\lambda}$  with observation  $Z^t$  is defined via the following  
 382 formula (see, e.g., [24], Section 2; [9], Section 2.1):

$$\begin{aligned}
 383 \quad (4.4) \quad L(\lambda \mid Z^t, A, t) &= - \sum_{j=1}^{N^t} \log(\Lambda(z_j)) + \int_{Z \times [0, t]} \Lambda(z) dz dt \\
 384 \quad &= - \int_{Z \times [0, t]} \log(\Lambda(z)) Z^t(dz dt) + t \int_Z \Lambda(z) dz, \quad \Lambda(z) = A\lambda(z).
 \end{aligned}$$

385 **4.2. Misspecification and the KL-projection.** In reality our model assumption is always  
 386 incorrect and  $Z^t \sim \mathcal{PP}^t = \mathcal{PP}|_{Z \times [0, t]}$  (marginal for interval  $[0, t]$ ) for some point process  
 387  $\mathcal{PP}$  on  $Z \times \mathbb{R}_+$ , where  $\mathcal{PP}, \mathcal{PP} \neq \mathcal{PP}_{A\lambda}$  for any  $\lambda \succeq 0$ . Since the (penalized) maximum  
 388 log-likelihood estimates are the most popular in ET, we say that the best one can hope to  
 389 reconstruct using measurements from  $\mathcal{PP}_{A\lambda}$  on  $[0, t]$  is the projection of  $\mathcal{PP}^t$  onto  $\mathcal{PP}_{A\lambda}^t =$   
 390  $\mathcal{PP}_{A\lambda}|_{Z \times [0, t]}$  in the sense of Kullback-Leibler divergence:

$$391 \quad (4.5) \quad \lambda_*(\mathcal{PP}, [0, t]) = \arg \min_{\lambda \succeq 0} \mathcal{KL}(\mathcal{PP}^t, \mathcal{PP}_{A\lambda}^t).$$

392 Since  $A$  is ill-conditioned, in general,  $\lambda_*$  in (4.5) may not be defined uniquely. For this we  
 393 consider the penalized KL-projection defined by the formula:

$$394 \quad (4.6) \quad \lambda_*(\mathcal{PP}, [0, t], \beta^t) = \arg \min_{\lambda \succeq 0} [\mathcal{KL}(\mathcal{PP}^t, \mathcal{PP}_{A\lambda}^t) + \beta^t \varphi(\lambda)],$$

396 where  $\beta^t$  is the regularization coefficient and  $\varphi(\lambda)$  is a nonparametric version of penalty from  
 397 section 2. From (4.4) and the definition of Kullback-Leibler divergence it follows that (up to  
 398 terms independent of  $\lambda$ ):

$$399 \quad (4.7) \quad \mathcal{KL}(\mathcal{PP}^t, \mathcal{PP}_{A\lambda}^t) = - \int_{Z \times [0, t]} \log(\Lambda(z)) \mathbb{E}_{\mathcal{PP}^t}[Z^t(dz dt)] + t \int_Z \Lambda(z) dz,$$

401 where  $\mathbb{E}_{\mathcal{PP}^t}$  is the expectation on  $Z^t$  with respect to  $\mathcal{PP}^t$ . Putting together (4.6), (4.7), for  
 402 the penalized KL-projection we get the following formulas:

$$403 \quad (4.8) \quad \lambda_*(\mathcal{PP}, [0, t], \beta^t) = \arg \min_{\lambda \succeq 0} \mathbb{L}_p(\lambda \mid \mathcal{PP}, A, t, \beta^t),$$

$$\begin{aligned}
 404 \quad (4.9) \quad \mathbb{L}_p(\lambda \mid \mathcal{PP}^t, A, t, \beta^t) &= - \int_{Z \times [0, t]} \log(\Lambda(z)) \mathbb{E}_{\mathcal{PP}^t}[Z^t(dz dt)] + t \int_Z \Lambda(z) dz + \beta^t \varphi(\lambda), \\
 405 \quad &\Lambda(z) = A\lambda(z).
 \end{aligned}$$

406 **4.3. Propagation of uncertainty and the generic algorithm.** Following the idea from  
 407 [35], we say that uncertainty on  $\lambda$  propagates from the one on  $\mathcal{PP}$  via (4.8), (4.9). Let  $\pi_{\mathcal{M}}$  be  
 408 a prior in which we encode our beliefs over a set of possible  $\mathcal{PP}$ 's, that is  $\pi_{\mathcal{M}}$  is a nonparametric

409 prior on spatio-temporal point processes on  $Z \times \mathbb{R}_+$  and it is constructed using  $\mathcal{M}$ . Let data  
 410 be list-mode  $Z^t$  or sinogram  $Y^t$ , then our prior beliefs can be updated in form of posterior  
 411 distribution  $\pi_{\mathcal{M}}(\cdot | Z^t \vee Y^t, t)$ .

---

**Algorithm 2** Generic NPL for ET
 

---

```

1: data :  $Z^t$  or  $Y^t$ ,  $\mathcal{M}$ 
2: input :  $B$  – number of samples
3: for  $b \leftarrow 1$  to  $B$  do
4:    $\widetilde{\mathcal{P}\mathcal{P}} \sim \pi_{\mathcal{M}}(\cdot | Z^t \vee Y^t, t)$ 
5:    $\widetilde{\lambda}_b^t \leftarrow \arg \min_{\lambda \geq 0} \mathbb{L}_p(\lambda | \widetilde{\mathcal{P}\mathcal{P}}, A, t, \beta^t)$  for  $\mathbb{L}_p(\cdot)$  defined in (4.9)
6: end for
7: return  $\{\widetilde{\lambda}_b^t\}_{b=1}^B$ .
```

---

412 As it has already been outlined before and in [35], [16], the above scheme produces i.i.d  
 413 samples and is trivially parallelizable which is a strong numerical advantage in front of MCMC  
 414 sampling from pure Bayesian posteriors. In what follows ‘tilde’ will denote samples produced  
 415 by NPL in ET (either nonparametric or binned).

416 **4.4. Constructions of  $\pi_{\mathcal{M}}(\cdot)$  and  $\pi_{\mathcal{M}}(\cdot | Z^t \vee Y^t, t)$ . Binning.** In view of the physical  
 417 model of ET we assume that  $\mathcal{P}\mathcal{P}$  belongs to the family of temporally stationary Poisson  
 418 processes, that is

$$419 \quad (4.10) \quad \mathcal{P}\mathcal{P} = \mathcal{P}\mathcal{P}_\Lambda \text{ with some density } \Lambda(z) dz dt \text{ on } Z \times \mathbb{R}_+, \\ 420 \quad \Lambda(z) \geq 0 \text{ a.s. and integrable on } Z \text{ w.r.t. } dz.$$

422 Hence, to build  $\pi_{\mathcal{M}}$  we construct a prior on  $\Lambda$  using  $\mathcal{M}$ , and consequently, the posterior  
 423 will also defined on  $\Lambda$  while propagating the uncertainty via (4.10) on  $\mathcal{P}\mathcal{P}$ . For the sake  
 424 of accessibility, discussion of the above assumption (restrictivity and generalizations) with  
 425 detailed theoretical constructions of nonparametric  $\pi_{\mathcal{M}}(\cdot)$  and  $\pi_{\mathcal{M}}(\cdot | Z^t \vee Y^t, t)$  are put in  
 426 Supplementary Materials, [section SM6](#). Below we present finite-dimensional versions which  
 427 are also used in our numerical experiments.

428 In finite dimensions (after binning) process  $\mathcal{P}\mathcal{P}_\Lambda$  boils down to  $d$  independent stationary  
 429 Poisson processes on  $\mathbb{R}_+$  with intensities  $\Lambda_1, \dots, \Lambda_d$ . For the prior on  $\Lambda = (\Lambda_1, \dots, \Lambda_d)$  we  
 430 choose the mixture of independent gamma distributions (further denoted by MGP – mixture  
 431 of gamma processes (due to its nonparametric origin)):

$$432 \quad (4.11) \quad \Lambda_{\mathcal{M}} = (\Lambda_{\mathcal{M},1}, \dots, \Lambda_{\mathcal{M},d}) \sim P_{\mathcal{M}}(\cdot), \Lambda_i | \Lambda_{\mathcal{M},i} \sim \Gamma(\theta^t \Lambda_{\mathcal{M},i}, (\theta^t)^{-1}), i = 1, \dots, d,$$

434 where  $\Lambda_{\mathcal{M}}$  is the mixing parameter which also corresponds to the mean intensity in the MRI-  
 435 based model from [subsection 2.4](#),  $P_{\mathcal{M}}(\cdot)$  is the mixing distribution (hyperprior),  $\theta^t$  is a positive  
 436 scalar. The choice of such specific parametrization by  $\theta^t$  in (4.11) allows to center the gamma  
 437 distribution on  $\Lambda_{\mathcal{M}}$  ( $\mathbb{E}[\Lambda | \Lambda_{\mathcal{M}}] = \Lambda_{\mathcal{M},i}$ ), so  $\theta^t$  controls only the spread –  $\theta^t = 0$  corresponds to  
 438 improper uniform distribution on  $\mathbb{R}_+^d$ ,  $\theta^t = +\infty$  is equal to  $\Lambda = \Lambda_{\mathcal{M}} \sim P_{\mathcal{M}}$ . In short, for the  
 439 prior in (4.11) we will use the following notation

$$440 \quad (4.12) \quad \pi_{\mathcal{M}}(\cdot) = \text{MGP}(t, P_{\mathcal{M}}(\Lambda_{\mathcal{M}}), \theta^t \Lambda_{\mathcal{M}}, (\theta^t)^{-1}).$$

442 Conjugacy between Poisson distribution of  $Y^t$  and Gamma distributions of  $\Lambda | \Lambda_{\mathcal{M}}$  implies that

$$443 \quad (4.13) \quad \pi_{\mathcal{M}}(\cdot | Z^t \vee Y^t, t) = \text{MGP}(t, P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t), Y^t + \theta^t \tilde{\Lambda}_{\mathcal{M}}^t, (\theta^t + t)^{-1}),$$

444 where  $P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$  is the posterior for  $P_{\mathcal{M}}$  which we specify now. Distribution of  
445  $P_{\mathcal{M}}(\Lambda_{\mathcal{M}})$  is defined directly by sampling:

$$446 \quad (4.14) \quad \lambda_{\mathcal{M}} = (\lambda_1^1, \dots, \lambda_{p_1}^1, \dots, \lambda_1^r, \dots, \lambda_{p_r}^r) : \lambda_s^k \sim \Gamma(1, \infty), \Lambda_{\mathcal{M}} = A_{\mathcal{M}} \lambda_{\mathcal{M}}.$$

448 where  $\lambda_{\mathcal{M}}, A_{\mathcal{M}}$  are constructed in [subsection 2.4](#),  $\Gamma(1, \infty)$  is the uniform (improper) distri-  
449 bution on  $\mathbb{R}_+$ . Then, posterior  $P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$  is defined by the classical Bayes formula  
450 for model  $Y^t \sim \text{Po}(t\Lambda_{\mathcal{M}})$  and the prior in (4.14). In principle, due to moderate size of  $A_{\mathcal{M}}$   
451 and good conditioning it is possible to use MCMC-approach (e.g., a Gibbs sampler) to sample  
452 from  $P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$ , however, in order to keep the overall implementation as simple as  
453 possible we turn to WLB from [38] for approximate posterior sampling.

---

**Algorithm 3** Approximate sampling from  $P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$  via WLB

---

- 1: **data** :  $Y^t$
  - 2: **input** :  $A_{\mathcal{M}} \in \text{Mat}(d, p_{\mathcal{M}})$  from (2.15) and (2.17)
  - 3:  $\tilde{\Lambda}^t \leftarrow (\tilde{\Lambda}_1^t, \dots, \tilde{\Lambda}_d^t)$ , where independently  $\tilde{\Lambda}_i^t \sim \Gamma(Y_i^t, t^{-1})$
  - 4:  $\tilde{\lambda}_{\mathcal{M}}^t \leftarrow \arg \min_{\lambda_{\mathcal{M}} \succeq 0} L(\lambda_{\mathcal{M}} | \tilde{\Lambda}^t, A_{\mathcal{M}}, 1)$
  - 5:  $\tilde{\Lambda}_{\mathcal{M}}^t \leftarrow A_{\mathcal{M}} \tilde{\lambda}_{\mathcal{M}}^t$
  - 6: **return**  $\tilde{\Lambda}_{\mathcal{M}}^t$
- 

454 *Remark 4.1.* Since we assume that  $A_{\mathcal{M}}$  is well-conditioned, minimizer  $\tilde{\lambda}_{\mathcal{M}}^t$  in [Step 4](#) of  
455 [Algorithm 3](#) can be efficiently computed via the classical EM-algorithm from [44].

456 From (4.13) and construction of  $P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$  one can see that overall MGP  
457 posterior acts as (doubly randomized) linear combination of the raw sinogram  $Y^t$  and pseudo-  
458 sinogram  $t\tilde{\Lambda}_{\mathcal{M}}^t$  proposed by the MRI-based model; see also [Figure 4](#).

#### 459 **4.5. Final algorithm.**

---

**Algorithm 4** NPL for ET

---

- 1: **data** :  $Y^t$
  - 2: **input** :  $B$  – number of samples,  $\theta^t, A, \beta^t, \varphi(\lambda)$
  - 3: **for**  $b \leftarrow 1$  to  $B$  **do**
  - 4:  $\tilde{\Lambda}_{\mathcal{M}}^t \leftarrow (\tilde{\Lambda}_{\mathcal{M},1}^t, \dots, \tilde{\Lambda}_{\mathcal{M},d}^t) \sim P_{\mathcal{M}}(\tilde{\Lambda}_{\mathcal{M}}^t | Z^t \vee Y^t, t)$  via [Algorithm 3](#)
  - 5:  $\tilde{\Lambda}_b^t \leftarrow (\tilde{\Lambda}_{b,1}^t, \dots, \tilde{\Lambda}_{b,d}^t)$ , where independently  $\tilde{\Lambda}_{b,i}^t \sim \Gamma(Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t, (\theta^t + t)^{-1})$
  - 6:  $\tilde{\lambda}_b^t \leftarrow \arg \min_{\lambda \succeq 0} L_p(\lambda | \tilde{\Lambda}_b^t, A, t, \beta^t/t)$  for  $L_p(\cdot)$  defined in (2.10)
  - 7: **end for**
  - 8: **return**  $\{\tilde{\lambda}_b^t\}_{b=1}^B$
-

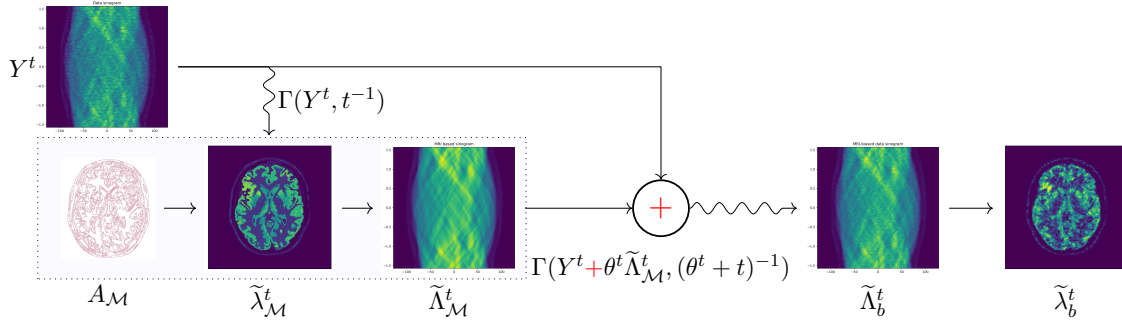


Figure 4: NPL-ET pipeline for one sample in Algorithm 4: wave-like arrows denote randomization of inputs, transparent blue region denotes steps within Algorithm 3.

460 *Remark 4.2.* In Step 6 of Algorithm 4 we have used the fact that binned version of  $\mathbb{L}_p(\cdot)$   
 461 from (4.9) coincides with  $L_p(\cdot)$  from (2.10). Moreover,

$$462 \quad (4.15) \quad L_p(\lambda \mid t\tilde{\Lambda}_b^t, A, t, \beta^t) = tL_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \beta^t/t) + R,$$

463 where  $R$  is independent of  $\lambda$ , hence, the minimization is directly applied to  $L_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \beta^t/t)$   
 464 instead of  $L_p(\lambda \mid t\tilde{\Lambda}_b^t, A, t, \beta^t)$ . If the numerical complexity of Step 4 is controlled by our choice  
 465 of  $P_{\mathcal{M}}(\cdot)$ , Step 6 is inevitable in the paradigm of NPL, hence, it must be numerically feasible  
 466 via some scalable optimization algorithm. This is the case for us in view of the well-known  
 467 the Generalized Expectation-Maximization(GEM)-type algorithm from [11] which is specially  
 468 designed for ET with Poisson-type log-likelihood  $L_p(\cdot)$ , where  $\varphi(\cdot)$  must be a  $C^2$ -smooth  
 469 convex pairwise difference penalty; see Supplementary Materials, section SM7 for details on  
 470 design of the algorithm.

471 *Remark 4.3.* Parameter  $\theta^t$  in Algorithm 4 admits the following interpretation: it is the  
 472 rate of creation of “pseudo-photons” in the model constructed from MRI data and being  
 473 conditioned with  $Y^t$ . By choosing  $\theta^t = \rho t$ ,  $\rho \geq 0$  in Step 5 we sum up sinograms  $Y^t$  and  $t\tilde{\Lambda}_{\mathcal{M}}^t$   
 474 in proportions  $1/(1 + \rho)$  and  $\rho/(1 + \rho)$ , respectively. For  $\theta^t = 0$  side information  $\mathcal{M}$  is not  
 475 used at all and we see Algorithm 4 as a version of WLB from [38] being adapted for the ET  
 476 context; see also [35], [16], [42] for connections between the WLB and NPL in the iid setting.

## 477 5. Numerical experiment. <sup>1</sup>

478 **5.1. Design.** We illustrate Algorithm 4 on synthetic PET data based on a realistic phan-  
 479 tom from the BrainWeb database [52]. Typical activity concentrations have been assigned to  
 480 annotated tissues (gray matter, white matter, skin, etc.) and we delineated a tumor lesion  
 481 area, not present in the initial phantom with an uptake of 50% compared to the gray mat-  
 482 ter activity; see Figure 5(a). We consider the worst case scenario for the prior, where the  
 483 anatomical MRI (T1) phantom (see Figure 5(b)) does not contain any information relative  
 484 to the lesion. Therefore, model  $Y^t \sim \text{Po}(tA_{\mathcal{M}}\lambda_{\mathcal{M}})$  in subsection 2.4 is strongly misspecified

<sup>1</sup>Source code in Python can be found at <https://gitlab.com/eric.barat/npl-pet>



485 (with increased bias) in the lesion area. For segmentation of MRI-images we used ddCRP [2]  
 486 with a concentration parameter equals  $10^{-5}$  leading to a few hundreds of random segments  
 487 for a 2D brain slice.

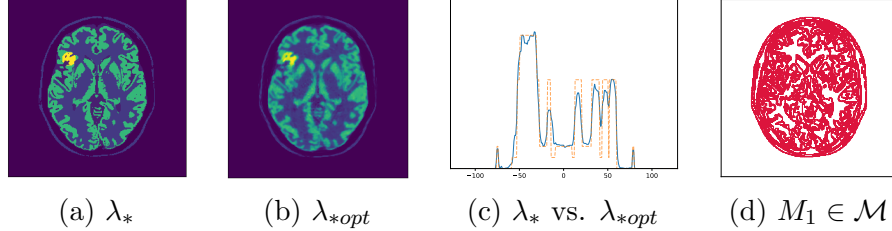


Figure 5: emission map with lesion hot spot at (a), optimal achievable reconstruction  $\lambda_{*opt}$  at (b), profile through lesion  $\lambda_*$  – orange dotted,  $\lambda_{*opt}$  – in blue at (c) segmented MRI at (d)

488 The reconstruction grid for images is of size  $256 \times 256$  ( $p = 2^{16}$ ) being identical to the  
 489 phantom's one. Acquisition geometry consists of LORs derived from a ring of 512 detectors  
 490 spaced uniformly on a circle. Design  $A$  was computed via classical Siddon's algorithm [45]  
 491 and  $A_{\mathcal{M}}$  was computed from  $A$  using formulas (2.15), (2.17). Intensity  $\lambda_*$  was set so that  
 492  $\sum_{j=1}^p \lambda_{*j} = 5 \cdot 10^5$  and for the experiment two sinograms were generated via formula (2.1) for  
 493  $t \in \{t_1, t_2\}$ ,  $t_1 = 1$ ,  $t_2 = 100$ . Case with  $t_1$  corresponds to realistic setting, whereas  $t_2 = 100$   
 494 is used to describe nearly asymptotic regime of the sampler. Below we present results for  
 495  $t = t_1$  (for  $t_2$  and additional experiments see Supplementary Materials, subsection SM7.1).  
 496 To compute  $\lambda_{*opt}$ , we have used (2.14) with  $\beta = \beta_{min} = 10^{-3}$ , where  $\beta_{min}$  was chosen  
 497 subjectively such that  $\lambda_{*opt}$  does not contain strong visible numerical artifacts related to the  
 498 implementation of projector  $A$  (see also Remark 4.2). For  $\varphi(\lambda)$  convex pairwise-difference  
 499 penalty from (SM8.1) with hyperparameters  $(\zeta, \nu)$ , where the latter were chosen to be always  
 500 fixed ( $\zeta = 0.05$ ,  $\nu = 0.15$ ) including  $\beta^t/t = 2 \times 10^{-3}$ .

501 For  $t_1 = 1$  we present results for  $\rho = \theta^t/t \in \{0, 0.5, 1.0, 2.0, 4.0\}$ ; see Remark 4.3. Using  
 502 Algorithm 4, for each combination of  $(t, \rho)$  we generated  $B = 1000$  bootstrap draws from  
 503 which further statistics (empirical mean, variance) as well calibration curves and plots were  
 504 computed. Main results are presented in Figure 6 and Table Table 1. First, we check visually  
 505 the effect of  $\rho$  on bias and variance (columns (a), (b); no need for  $\lambda_{*opt}$  to compute), and second,  
 506 calibration of the overall posterior (columns (c), (d), (e); requires  $\lambda_{*opt}$ ). For calibration  
 507 we employ the approach in [51], [21], which says that a model is well-calibrated if for any  
 508 level  $\alpha \in [0, 1]$  (target coverage), the corresponding posterior  $\alpha$ -level HPD-intervals (highest  
 509 probability density) computed pixel-wise will contain  $\lambda_{*opt}$  for  $\alpha \cdot 100\%$  of all pixels (achieved  
 510 coverage – fraction of  $j$ 's for which  $\lambda_{*opt,j} \in [\hat{q}_{j,\alpha}^L, \hat{q}_{j,\alpha}^U]$ , where  $[\hat{q}_{j,\alpha}^L, \hat{q}_{j,\alpha}^U]$  being the shortest  
 511 interval such that  $P(\tilde{\lambda}_{b,j}^t \in [\hat{q}_{j,\alpha}^L, \hat{q}_{j,\alpha}^U] | Y^t) = \alpha$ ) (column (c) – reliability curve). Thus, if  
 512 the achieved coverage is smaller than the target one, then the model is considered to be  
 513 overconfident and for vice versa – under-confident (or conservative). Note that for practice  
 514 it is preferable to have slightly conservative model than overconfident one, especially in such  
 515 domain as medical imaging; see the discussion in [21].

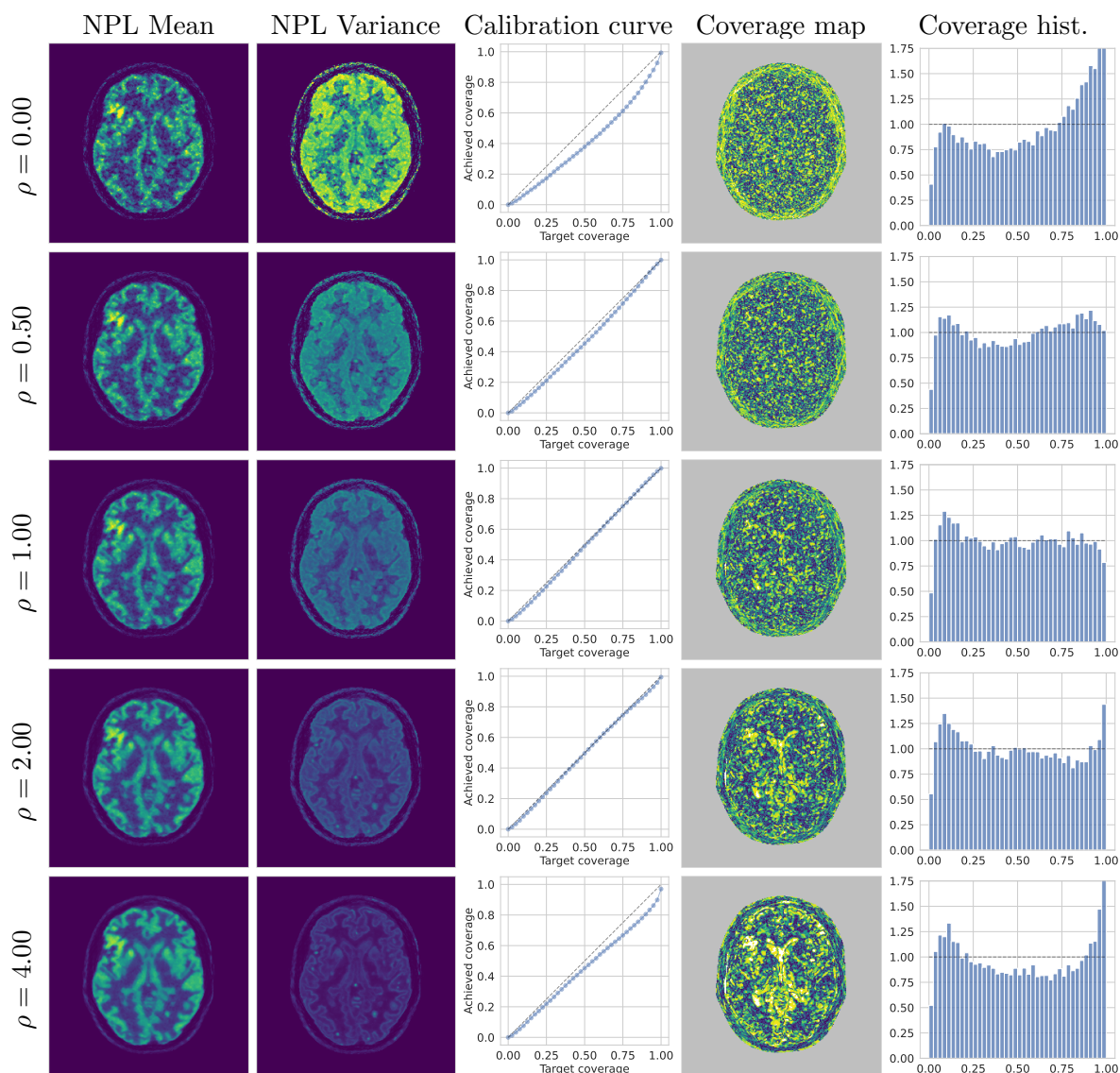


Figure 6: Columns : (a) NPL-mean, (b) NPL-variance (same color scale as mean), (c) calibration curve, (d) coverage probability map (mask in gray), (e) coverage histogram.

$\rho$	0.00	0.50	1.00	2.00	4.00
PSNR	21.42	24.15	25.29	<b>25.84</b>	25.66
MSWD	$8.74 \cdot 10^5$	1.16	<b>0.83</b>	1.04	1.78
ECE	$9.41 \cdot 10^{-2}$	$3.29 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	<b><math>1.03 \cdot 10^{-2}</math></b>	$5.29 \cdot 10^{-2}$
KLC	$5.52 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	<b><math>9.76 \cdot 10^{-3}</math></b>	$1.14 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$

Table 1: Performance metrics

516 Since the definition of calibration does not take into account correlations between pixels,  
 517 columns (d), (e) are used for diagnostic of the latter. Coverage map (d) shows for each pixel  
 518 the smallest probability so that the HPD-interval contains  $\lambda_{*opt}$  while the normalized coverage  
 519 histogram in (e) corresponds to the (empirical) probability density function for the coverage  
 520 curve in (c) being viewed as c.d.f. (note that for perfect calibration the c.d.f. in (c) and  
 521 p.d.f. in (e) correspond to the uniform distribution on  $[0, 1]$ ). In Table 1 we compute PSNR  
 522 for the NPL-mean, ECE – expected calibration error ( $\ell_1$ -norm between the calibration curve  
 523 in (c) and diagonal  $x = y$  on  $[0, 1]$ ), MSWD – mean-squared weighted deviation between  
 524  $\lambda_{*opt}$  and the NPL-mean, KLC – Kullback-Leibler divergence between uniform distribution  
 525 and coverage histogram in (e); see Supplementary Materials, subsection SM8.2 for precise  
 526 definitions and connections to other metrics.

527 **5.2. Interpretation.** The increase of  $\rho$  reduces the noise, but on the other hand, it in-  
 528 creases bias in the lesion area; see Figure 6(a) and PSNR in Table 1. The latter is due to  
 529 the aforementioned misspecification, therefore the high signal is being spread over the larger  
 530 segment in  $\mathcal{M}$  containing the lesion. Being subjective, for us the most visually appealing  
 531 results for the trade-off between noise and preservation of contours of the lesion were ob-  
 532 tained for  $\rho \in \{0.50, 1.00\}$ . Note also that pixel-wise variance in (b) decreases. However, for  
 533  $\rho \rightarrow +\infty$  the limit is not zero but the posterior variance in the MRI-based model<sup>2</sup> which is  
 534 much smaller, for example, than for  $\rho = 0$  (because  $Y^t$  contains much more information for  
 535 low-dimensional  $\lambda_{\mathcal{M}} \in \mathbb{R}^{p_{\mathcal{M}}}$  than for  $\lambda \in \mathbb{R}^p$ ). Spikes for variance in (b) (e.g., for  $\rho = 4.0$ )  
 536 correspond to smallest segments in  $\mathcal{M}$  where the signal is more sensitive to perturbations in  
 537  $\tilde{\Lambda}_b^t$  due to ill-conditioning nature of  $A$ . With calibration results in (c), (d), (e), and in Table 1  
 538 we can choose objectively one optimal  $\rho$  by arguing on guarantees of covering  $\lambda_{*opt}$  by the pos-  
 539 terior. First, note that for  $\rho = 0$  the posterior is essentially overconfident (columns (c), (e)) –  
 540 this is due to large amount of pixels in the slab between the cranium and soft tissues (exterior  
 541 yellow ring on images in (d)) where in fact the isotope concentration is zero. Coverage map  
 542 (d) and histogram (e) reveal that these pixels require very large credible levels to cover  $\lambda_{*opt}$   
 543 meaning that the posterior in this region is overcontracted. We explain the overcontraction  
 544 by the fact that for many LORs crossing such pixels and nearly tangential to the brain the  
 545 intensities  $\Lambda_i^*$  are so small (though positive) that for  $t = 1$  (mild regime) it happens that  
 546  $Y_i^t = 0$ . Then, in Step 5 one can see that  $\tilde{\Lambda}_{b,i}^t \sim \Gamma(0, t^{-1}) = \delta_0$  for  $\rho = 0$ , so  $\tilde{\Lambda}_{b,i}^t \equiv 0$  c.a.s.  
 547 and no uncertainty can propagate from such LOR in Step 6 which results in overcontrac-  
 548 tion. Moreover, in subsection 6.3 we show that for Poisson model the event  $Y_i^t = 0$  make the  
 549 posterior contract much stronger to zeros in pixels intersected by LOR  $i$  (effect of positivity  
 550 constraints in Step 6) which is another argument for overcontraction. Finally, overcontraction  
 551 was already reported for (non-Poisson) WLB in [40] with a proposal to fix it different from  
 552 NPL; see also Remark 4.3. An additional numerical experiment supporting our explanation is  
 553 given in the Supplementary Materials, subsection SM8.3. For  $\rho \in \{0.5, 1.00\}$ , since the afore-  
 554 mentioned empty slab is splitted into larger segments for which  $\tilde{\Lambda}_{\mathcal{M},i}^t > 0$ , the overcontraction  
 555 is corrected while improving the overall calibration and reaching the optimum for KLC and  
 556 MSWD at  $\rho = 1.0$  (see (c), (e) and ECE, KLC in Table 1). Further increase  $\rho \in \{2.00, 4.00\}$

---

<sup>2</sup> $\text{var}[\tilde{\Lambda}_b^t|Y^t] = (Y^t/t + \rho\mathbb{E}[\tilde{\Lambda}_{\mathcal{M}}^t|Y^t])/t^2(1+\rho)^2 + \rho^2\text{var}[\tilde{\Lambda}_{\mathcal{M}}^t|Y^t]/t^2(1+\rho)^2, \lim_{\rho \rightarrow +\infty} \text{var}[\tilde{\Lambda}_b^t|Y^t] = \text{var}[\tilde{\Lambda}_{\mathcal{M}}^t|Y^t]/t^2$

557 results in increased bias in the lesion area and since the posterior intervals are being more  
 558 contracted, the posterior again turns to be overconfident (see column (c) supported by sharp  
 559 increase for high confidence levels in (e) and also large yellow structures in (d) in the lesion  
 560 and central segments). In conclusion, calibration with  $\rho$  is simple and tractable, seemingly  
 561 with one optimum w.r.t bias (in the lesion) and (global-)variance trade-off.

562 **6. Asymptotic analysis of the algorithm.** Statistical model (2.1) is non-regular since  
 563 domain  $\mathbb{R}_+^p$  contains a boundary and, often it is the case that  $\lambda_* \in \partial\mathbb{R}_+^p$ . The results of [3] for  
 564 the classical Bayesian framework show that for the well-specified case and large class of priors  
 565 the posterior is consistent at  $\lambda_*$  and the asymptotic distribution is complex because it splits  
 566 in three modes due to the effect of positivity constraints (exponential, Gaussian and half-  
 567 Gaussian; two latter have the same standard contraction rates but the first one). Consistency  
 568 at  $\lambda_*$  and a very similar splitting are also present in NPL with the asymptotic distribution  
 569 being tight around strongly consistent estimator  $\hat{\lambda}_{sc}^t$  satisfying some contraction properties  
 570 in observation (sinogram) space. Interestingly, the aforementioned splitting depends not on  
 571  $\lambda_*$  (as it was in [3]) but again on  $\hat{\lambda}_{sc}^t$  because of which yet we fail to demonstrate fully the  
 572 asymptotic normality since it requires additional results on behavior of strongly consistent  
 573 estimators with constraints on the domain (detailed discussion is given in Supplementary  
 574 Materials, section SM9).

575 The problem of misspecification for the generalized Poisson model with wrong design  
 576 arises twice our setting: first, in Algorithm 3 when sampling  $\tilde{\Lambda}_{\mathcal{M}}^t$  (because we assume that  
 577  $Y^t \sim P_{A_{\mathcal{M}}, \lambda_{\mathcal{M}}}^t$  whereas  $Y^t \sim P_{A, \lambda_*}^t$ ) and, second, when we assume that model (2.1) is wrong, in  
 578 general. Surprisingly, in this simple case the identifiability of  $\lambda_*$  can be lost even for injective  
 579 designs which we show by an explicit example below. We propose an intuitive sufficient  
 580 condition on observed intensities along LORs and design  $A$  to retrieve it back.

581 **6.1. Convergence for conditional probabilities.** Let  $(\Omega, \mathcal{F}, P)$  be the common proba-  
 582 bility space on which process  $Y^t$ ,  $t \in [0, +\infty)$  and MGP prior in (4.12) are defined (see  
 583 Supplementary Materials, section SM1 for details). By  $U | Y^t$  we denote the distribution of  $U$   
 584 conditionally on the sigma algebra generated by  $Y^\tau$ ,  $\tau \in [0, t)$ .

585 **Definition 6.1.** *We say that  $U^t$  converges in conditional probability to  $U$  almost surely  $Y^t$*   
 586 *if for every  $\varepsilon > 0$  the following holds:*

$$587 \quad (6.1) \quad P(\|U^t - U\| > \varepsilon \mid Y^t) \rightarrow 0 \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

588 *This type of convergence will be denoted as follows:*

$$589 \quad (6.2) \quad U^t \xrightarrow{c.p.} U.$$

590 In our proofs for  $U^t \xrightarrow{c.p.} 0$  we also write

$$591 \quad (6.3) \quad U^t = o_{cp}(1).$$

592 **Definition 6.2.** *We say that  $U^t$  is conditionally tight almost surely  $Y^t$  if for any  $\varepsilon > 0$  and*  
 593 *almost any trajectory  $Y^t$ ,  $t \in [0, +\infty)$  there exists  $M = M(\varepsilon, \{Y^t\}_{t \in (0, +\infty)})$  such that*

$$594 \quad (6.4) \quad \sup_{t \in [0, +\infty)} P(\|U^t\| > M \mid Y^t) < \varepsilon.$$

595 In short, in the definitions above almost surely  $Y^t$  means that statements in (6.1), (6.4)  
596 hold for almost every trajectory  $Y^t$ ,  $t \in [0, +\infty)$ .

## 597 6.2. Consistency.

598 *Assumption 6.3.* Model (2.1) is well-specified, that is

$$599 \quad (6.5) \quad Y^t \sim PP_{A, \lambda_*}^t, \text{ for some } \lambda_* \in \mathbb{R}_+^p \text{ and all } t \in [0, +\infty),$$

600 where  $A$  satisfies (2.3)–(2.6),  $PP_{A, \lambda}^t$  is defined in (2.8).

601 **Theorem 6.4.** Let *Assumption 6.3* and conditions (2.11), (2.12) for  $\varphi$  be satisfied. Let also  
602  $\beta^t, \theta^t$  be such that

$$603 \quad (6.6) \quad \beta^t/t \rightarrow 0, \theta^t/t \rightarrow 0 \text{ when } t \rightarrow +\infty.$$

605 Then,

$$606 \quad (6.7) \quad \tilde{\lambda}_b^t \xrightarrow{c.p.} \lambda_{*opt},$$

607 where  $\tilde{\lambda}_b^t$  is sampled in *Algorithm 4*,  $\lambda_{*opt}$  is defined in (2.14).

608 The above result is merely a consequence a more general statement for any bootstrap-type  
609 procedure which is given below.

610 **Theorem 6.5.** Let conditions of *Theorem 6.4* be satisfied but *Assumption 6.3*. Assume also  
611 that

$$612 \quad (6.8) \quad \tilde{\Lambda}_b^t \xrightarrow{c.p.} \Lambda^* = A\lambda_* \text{ for some } \lambda_* \in \mathbb{R}_+^p.$$

614 Then, formula (6.7) remains valid.

615 Thus the conditional distribution of  $\tilde{\lambda}_b^t$  asymptotically concentrates at  $\lambda_*$  in the subspace  
616 where parameter  $\lambda$  is identifiable through design  $A$  and also regarding the positivity con-  
617 straints. Projection of  $\lambda_*$  onto  $\ker(A)$  which not “visible” by positivity constraints is not  
618 identifiable in model (2.1) and it is defined solely by  $w_{A, \lambda_*}(0)$ ; see formula (2.13).

## 619 6.3. Tightness.

620 *Assumption 6.6.*  $A_{\mathcal{M}} \in \text{Mat}(d, p_{\mathcal{M}})$  is injective.

621 *Assumption 6.7 (non-expansiveness condition).* Let  $\Lambda^* \in \mathbb{R}_+^d$ ,  $A_{\mathcal{M}} \in \text{Mat}(d, p_{\mathcal{M}})$ ,  $A_{\mathcal{M}}$  has  
622 only positive entries and analog of (2.4) for  $A_{\mathcal{M}}$  holds (i.e.,  $A_{\mathcal{M}, j} = \sum_{i=1}^d a_{\mathcal{M}, ij} > 0$ ). Define set

$$623 \quad (6.9) \quad \lambda_{\mathcal{M},*} = \arg \min_{\lambda_{\mathcal{M}} \geq 0} L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1),$$

625 where  $L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1)$  is given in (2.8). There exists at least one point in  $\lambda_{\mathcal{M},*}$  for which  
626 the following holds:

$$627 \quad (6.10) \quad I_0(\Lambda_{\mathcal{M}}^*) = I_0(\Lambda^*), \Lambda_{\mathcal{M}}^* = A_{\mathcal{M}}\lambda_{\mathcal{M},*},$$

628 where  $I_0(\cdot)$  is defined in (2.2).

629 The proposition below states that the non-expansiveness condition is always meaningful  
 630 and not very restrictive (for more details see Supplementary Materials, [section SM11](#)).

631 **Proposition 6.8.** *Let  $\Lambda^* \in \mathbb{R}_+^d$ ,  $A_{\mathcal{M}} \in \text{Mat}(d, p_{\mathcal{M}})$ ,  $A_{\mathcal{M}}$  has only positive entries and the  
 632 analog of (2.4) for  $A_{\mathcal{M}}$  holds (i.e.,  $A_{\mathcal{M},j} = \sum_{i=1}^d a_{\mathcal{M},ij} > 0$ ). Then, the set of minimizers in  
 633 (6.9) is non-empty and constitutes an affine subset of  $(p_{\mathcal{M}} - 1)$ -dimensional simplex  $\Delta_{A_{\mathcal{M}}}^p(\Lambda^*)$   
 634 defined by the formula:*

$$635 \quad (6.11) \quad \Delta_{A_{\mathcal{M}}}^{p_{\mathcal{M}}}(\Lambda^*) = \{\lambda_{\mathcal{M}} \in \mathbb{R}_+^{p_{\mathcal{M}}} \mid \sum_{j=1}^{p_{\mathcal{M}}} A_{\mathcal{M},j} \lambda_{\mathcal{M},j} = \sum_{i=1}^d \Lambda_i^* \geq 0\}.$$

636 Moreover, it always holds that

$$637 \quad (6.12) \quad I_1(\Lambda^*) \subset I_1(\Lambda_{\mathcal{M}}^*) \text{ or equivalently } I_0(\Lambda_{\mathcal{M}}^*) \subset I_0(\Lambda^*), \text{ where } \Lambda_{\mathcal{M}}^* = A_{\mathcal{M}} \lambda_{\mathcal{M},*}.$$

638 The aim of the non-expansiveness condition is to have a unique and stable KL-minimizer  
 639  $\lambda_{\mathcal{M},*}$  so that the the prior effect of  $\mathcal{M}$  on  $\tilde{\lambda}_b^t$  via  $\tilde{\Lambda}_{\mathcal{M}}^t$  (which concentrates near  $\Lambda_{\mathcal{M},*} =$   
 640  $A_{\mathcal{M}} \lambda_{\mathcal{M},*}$ ) is not spread ambiguously among different (but equivalent in terms of observations)  
 641 combinations of signals in segments of  $\mathcal{M}$ . This is provided by the theorem below.

642 **Theorem 6.9 (identifiability in the prior model).** *Let [Assumptions 6.6](#) and [6.7](#) be satisfied.  
 643 Then,  $\lambda_{\mathcal{M},*}$  defined in (6.9) has only one point and the following approximation holds:*

$$644 \quad L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) = \mu_{\mathcal{M},*}^T \lambda_{\mathcal{M}} + \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \Lambda_i^* \frac{(\Lambda_{\mathcal{M},i} - \Lambda_{\mathcal{M},i}^*)^2}{(\Lambda_{\mathcal{M},i}^*)^2}$$

$$645 \quad (6.13) \quad + o(\|\Pi_{A_{\mathcal{M}, I_1(\Lambda^*)}^T}(\lambda_{\mathcal{M}} - \lambda_{\mathcal{M},*})\|^2),$$

647 where  $\Pi_{A_{\mathcal{M}, I_1(\Lambda^*)}^T}$  denotes the orthogonal projector onto  $\text{Span}(A_{\mathcal{M}, I_1(\Lambda^*)}^T)$ ,

$$648 \quad (6.14) \quad \mu_{\mathcal{M},*} = \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \frac{a_{\mathcal{M},i}}{\Lambda_{\mathcal{M},i}^*} + \sum_{i=1}^d a_{\mathcal{M},i},$$

$$649 \quad \mu_{\mathcal{M},*} \succeq 0, \mu_{\mathcal{M},*,j} \lambda_{\mathcal{M},*,j} = 0 \text{ for all } j \in \{1, \dots, p_{\mathcal{M}}\}.$$

650 In particular,  $L(\lambda_{\mathcal{M}} \mid \Lambda^*, A, 1)$  is strongly convex at  $\lambda_{\mathcal{M},*}$ , so, there exists an open ball  $B_* =$   
 651  $B(\lambda_{\mathcal{M},*}, \delta_*)$ ,  $\delta_* = \delta_*(A_{\mathcal{M}}, \Lambda_*) > 0$  and constant  $C_* = C_*(A_{\mathcal{M}}, \Lambda_*) > 0$  such that

$$652 \quad (6.15) \quad L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) \geq C_* \|\lambda_{\mathcal{M}} - \lambda_{\mathcal{M},*}\|^2, \lambda \in B_* \cap \mathbb{R}_+^{p_{\mathcal{M}}}.$$

654 Result of [Theorem 6.9](#) is also a positive answer to the general identification problem when  
 655 model (2.1) is misspecified in the sense of wrong design. In [subsection 6.4](#) we show that the  
 656 non-expansiveness condition is essential and counterexamples are possible if it is removed.

657 Now we can turn to our main result on the tightness of the NPL-posterior.

658 Let  $\{e_j\}_{j=1}^p$  be the standard basis in  $\mathbb{R}^p$  and define the following spaces:

$$659 \quad (6.16) \quad \mathcal{V} = \text{Span}\{e_j \mid \exists i \in I_0(\Lambda^*) \text{ s.t. } a_{ij} > 0\},$$

$$660 \quad (6.17) \quad \mathcal{U} = \mathcal{V}^\perp \cap \text{Span}\{A_{I_1(\Lambda^*)}^T\},$$

$$661 \quad (6.18) \quad \mathcal{W} = (\mathcal{V} \oplus \mathcal{U})^\perp \cap \ker A.$$

663 Let also

$$664 \quad (6.19) \quad \Pi_{\mathcal{V}}, \Pi_{\mathcal{V}}, \Pi_{\mathcal{W}} \text{ be the orthogonal projectors on } \mathcal{V}, \mathcal{V}, \mathcal{W}, \text{ respectively.}$$

666 **Theorem 6.10.** *Let Assumptions 6.3 and 6.7 be satisfied and assume also that*

$$667 \quad (6.20) \quad \varphi \text{ satisfies (2.11), (2.12) and it is locally Lipschitz continuous.}$$

669 Let  $\tilde{\lambda}_b^t$  be defined as in Algorithm 4,  $\theta^t = o(\sqrt{t/\log \log t})$ ,  $\beta^t = o(\sqrt{t})$  and assume that there  
670 exists a strongly consistent estimator  $\hat{\lambda}_{sc}^t$  of  $\lambda_*$  on  $\mathcal{V} \oplus \mathcal{U}$  (i.e.,  $\Pi_{\mathcal{U} \oplus \mathcal{V}} \hat{\lambda}_{sc}^t \xrightarrow{a.s.} \Pi_{\mathcal{U} \oplus \mathcal{V}} \lambda_*$ ) such  
671 that

$$672 \quad (6.21) \quad \hat{\lambda}_{sc}^t \succeq 0,$$

$$673 \quad (6.22) \quad \limsup_{t \rightarrow +\infty} \left| \sum_{i \in I_1(\Lambda^*)} \sqrt{t} \frac{Y_i^t/t - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i \right| < +\infty \text{ a.s. } Y^t,$$

$$674 \quad (6.23) \quad t \hat{\Lambda}_{sc,i}^t \xrightarrow{a.s.} 0 \text{ for } i \in I_0(\Lambda^*),$$

676 where  $\hat{\Lambda}_{sc}^t = A \hat{\lambda}_{sc}^t$ . Then,  
(i)

$$(6.24) \quad t \Pi_{\mathcal{V}}(\tilde{\lambda}_b^t - \hat{\lambda}_{sc}^t) \xrightarrow{c.p.} 0.$$

677

(ii)

$$(6.25) \quad \sqrt{t} \Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \hat{\lambda}_{sc}^t) \text{ is conditionally tight a.s. } Y^t.$$

678 Statement in (i) claims that for pixels which are interested by LORs with  $\Lambda_i^* = 0$ , the  
679 posterior distribution contracts to zero with faster rate than for the ones intersected by LORs  
680 with positive intensities. Indeed, pixels in subspace  $\mathcal{V}$  are strongly forced to be zeros by the  
681 positivity constraints (i.e., if  $\Lambda_i^* = 0$  and  $\lambda_*$ ,  $a_i \in \mathbb{R}_+^p$ , then necessarily  $\lambda_{*,j} = 0$  where  $a_{ij} > 0$ ).  
682 Statement in (ii) claims that, in general, the posterior concentrates around  $\hat{\lambda}_{sc}^t$  in subspace  
683  $\mathcal{U}$  with standard scaling rate  $\sqrt{t}$ . This is not surprising since  $\mathcal{U}$  is orthogonal to  $\mathcal{V}$ , so the  
684 positivity constraints do not give extra information to achieve the faster contraction rate.  
685 Finally, requiring the non-expansiveness condition for the prior (Assumption 6.7) may seem  
686 surprising at first sight. The intuition behind is that it forbids our sampler to create “too  
687 many” pseudo-photons in LORs where intensity is zero a.s. ( $\Lambda_i^* = 0$  implies  $Y_i^t \equiv 0$ ) and  
688 significantly simplifies the theoretical analysis.

689 For  $\hat{\lambda}_{sc}^t$  we propose to take the MAP-estimate which is defined by the formula:

$$(6.26) \quad \widehat{\lambda}_{pMLE}^t = \arg \min_{\lambda \geq 0} L_p(\lambda \mid Y^t, A, t, \beta^t),$$

where  $L_p(\cdot)$  is defined in (2.10).

**Conjecture 6.11.** Let assumptions of Theorem 6.10 be satisfied and  $\widehat{\lambda}_{sc}^t = \widehat{\lambda}_{pMLE}^t$ , where the latter is defined by (6.26). Then,  $\widehat{\lambda}_{sc}^t$  is a strongly consistent estimator of  $\lambda_*$  on  $\mathcal{V} \oplus \mathcal{V}$  and formulas (6.21)–(6.23) hold.

The requirement for existence of a strongly consistent estimator is not new and already appears for WLB in [39]. However, in that case the sampling is performed via unconstrained optimization of quadratic functionals with  $\ell_1$ -penalties for which existence of such estimators is trivial by taking the standard OLS estimator or LASSO estimator; see the discussion after Theorem 3.3 in [39]. In our case, according to Kolmogorov’s 0-1 Law the statements in (6.22) and (6.23) either hold with probability one (i.e., almost surely  $Y^t$ ,  $t \in [0, +\infty)$ ) or zero, and the case of zero probability would mean a very exotic and unexpected behavior of the constrained MLE estimate for such model because they are trivially satisfied, for example, if  $A$  is diagonal. Another plausible argument in favour of existence of required  $\widehat{\lambda}_{sc}^t$  comes from [3] where the asymptotic posterior mean is strongly consistent and satisfies (6.21)–(6.23) (for details see Supplementary Materials, section SM9).

Finally, establishing tightness of the posterior is the first step towards the proof of asymptotic normality (see Bernstein von-Mises type theorems in [49], [39], [42]) which, in particular, implies that for large dataset the posterior distribution, in general (but not always if misspecified; see e.g., [27]; an interesting case of posterior inconsistency was found in [19]), is correctly calibrated against frequentist distribution of some strongly consistent estimator.

**6.4. Misspecification in design and identifiability.** Assumption 6.3 in subsection 6.2 reflects our belief that model (2.1) is correct. At the same time, for any practitioner in ET it is known that such model is by far approximate: the tracer inside the human body surely does not respect locally constant behavior, design  $A$  is known only approximately (with non-negligible errors, since it contains patient’s attenuation map which is reconstructed via a separate MRI or CT scan; see e.g., [48]), non-stationarity of the process due to kinetics of the tracer, scattered photons, errors from multiple events etc.; see e.g., [29], [43].

Assume that exposure period is  $[0, t)$  and  $PP^t$  is the unknown (binned) process that generates  $Y^t$ :

$$(6.27) \quad \begin{aligned} Y^t &\sim PP^t, Y^t \in (\mathbb{N}_0)^d, \\ \mathbb{E}_{PP^t}[Y^t] &= \text{var}_{PP^t}[Y^t] = \Lambda^*(t) \text{ for some } \Lambda^*(t) = (\Lambda_1^*(t), \dots, \Lambda_d^*(t)) \in \mathbb{R}_+^d. \end{aligned}$$

Formulas in (6.27) reflect our belief that  $Y^t$  has Poisson-type behavior (e.g., non-stationary Poisson process) at least for its two first moments which is not far from truth in practice [47]. Most importantly, we do not assume that  $\Lambda^*(t) \in R_+(A)$ . The main question now is the identifiability of  $\lambda$  which translated via (2.9) and (6.27) to the question of uniqueness in the following minimization problem:

$$(6.28) \quad \lambda_*(PP, [0, t)) = \arg \min_{\lambda \geq 0} \mathcal{KL}(PP^t, PP_{A\lambda}^t) = \arg \min_{\lambda \geq 0} L(\lambda \mid \Lambda^*(t)/t, A, 1),$$



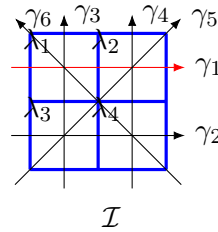
730 where  $PP_{A\lambda}^t$  is defined in (2.8). It appears that, in general, the answer is negative even for  
731 very meaningful choices of  $A$  and  $\Lambda^*(t)$ .

732 **Theorem 6.12.** *Let  $t = 1$ . There exist  $\Lambda^* = (\Lambda_1^*, \dots, \Lambda_d^*) \in \mathbb{R}_+^d$ ,  $\Lambda^* \neq 0$ ,  $A \in \text{Mat}(d, p)$   
733 which has only nonnegative entries, it is stochastic column-wise and injective such that solu-  
734 tions of the optimization problem (6.28) constitute a non-empty polytope of positive dimension  
735 of the  $(p - 1)$ -simplex  $\Delta_p(\Lambda^*) = \left\{ \lambda \in \mathbb{R}_+^p : \sum_{j=1}^p \lambda_j = \sum_{i=1}^d \Lambda_i^* \right\}$ .*

736 *Proof.* We construct  $\Lambda^*$  and  $A$  for  $p = 4$ ,  $d = 6$ . Let  $\mathcal{I}$  be the square of four pixels each  
737 with side length 1 as shown below, i.e.,  $\lambda = (\lambda_1, \dots, \lambda_4) \in \mathbb{R}_+^4$ , and  $\Gamma = \{\gamma_1, \dots, \gamma_6\}$  be the  
738 set of rays. Let  $A'$  be the classical Radon transform on  $\mathcal{I}$  for geometry  $\Gamma$  (i.e.,  $a'_{ij}$  being the  
739 length of intersection of ray  $\gamma_i$  with pixel  $j$ ):

$$A' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & \sqrt{2} & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 0 & \sqrt{2} \end{pmatrix}$$

$$\det(A'^T A') = 128 \neq 0.$$



741 Let  $A$  be a column-wise normalization of  $A'$ , i.e.,  $a_{ij} = a'_{ij} / (\sum_i a'_{ij})$  (this obviously does  
742 not break the injectivity of  $A'$ ). Let  $\Lambda^* = (1, 0, 0, 0, 0, 0)$ . Then, for (6.28) we get

$$743 \quad (6.29) \quad \lambda_* = \arg \min_{\lambda \geq 0} -\log \left( \frac{\lambda_1 + \lambda_2}{2 + \sqrt{2}} \right) + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4.$$

744 Note that in (6.29) we have used the fact that  $\sum_i a_{ij} = 1$  for all  $j \in \{1, \dots, 4\}$ . It is obvious  
745 that the set of minimizers in (6.29) is an affine set of the following form:

$$746 \quad (6.30) \quad \lambda_{*3} = \lambda_{*4} = 0, \lambda_{*1} + \lambda_{*2} = 1$$

747 which gives the desired non-uniqueness. Theorem is proved. ■

748 Finally, note that **Theorem 6.9** provides identifiability under the non-expansiveness condi-  
749 tion and injectivity of  $A$ .

750 **7. Discussion.** **Algorithm 4** solves **Problems 1** and **2** simultaneously and efficiently: gen-  
751 erated samples are automatically iid, algorithm is scalable because the crucial **Step 6** is per-  
752 formed via the classical GEM-type algorithm and, finally, our main calibration parameter  $\rho$   
753 ( $\theta^t = t\rho$ ,  $\rho \geq 0$ ; see **Remark 4.3**) can be interpreted as amount of pseudo-data (pseudo-  
754 photons) generated from the MRI-based posterior. Due to the latter the numerical calibration  
755 of the posterior is tractable. Moreover, in our experiment on the synthetic dataset for the  
756 worst case scenario (when MRI has no information on the lesion) we have observed that mod-  
757 erate values of  $\rho$ , indeed, improve calibration error as well PSNR and MSWD. Our principal  
758 theoretical results (posterior consistency and tightness) are complicated by the non-standard  
759 form of ET but show a great number of connections to existing works ([35], [16], [3]). The  
760 new non-expansiveness condition (**Assumption 6.7**) is of independent geometric interest and  
761 is a key to extend all previous results to the fully misspecified case. Among possible exten-  
762 sions, one most interesting for us is to relax the independence of increments of the Gamma

763 process in the prior and consider ones with correlations (for example, scaled Polyà-tree priors  
764 for  $\Lambda_{\mathcal{M}}$ ). These correlations can be used to smooth out sinogram  $Y^t$  by projecting it (non-  
765 linearly) on the stable part of  $\text{Span}(A^T)$  using an MRI-based model and, in addition, remove  
766 completely the need for regularizer  $\varphi$  (high frequencies are still regularized by  $\varphi$  whereas  $\mathcal{M}$  is  
767 used for low-frequencies). Our preliminary numerical results show that it improves resolution  
768 while retaining the interpretability of calibration parameters as before. Another improvement  
769 could be to replace the (random) segmentations of MRI-images via ddCRP with other ma-  
770 chine learning techniques (such as DNNs) that on input will take MRI-scans with sinograms  
771 and output possible low-dimensional models  $A_{\mathcal{M}}, \lambda_{\mathcal{M}}$  (possibly corrected by medical experts).  
772 This has a chance to reduce bias in the lesion while non-increasing the calibration error and  
773 variance. Finally, an experiment on real PET-MRI data is of great importance and will be  
774 given elsewhere.

775 **Supplementary materials.** Supplementary materials include discussion of the assumption  
776 in (4.10) and remarks on nonparametric constructions in subsection 4.4, all details of numeri-  
777 cal experiments in sections 3 and 5 (with additional numerical experiments for large  $t$ ), proofs  
778 of all theoretical results in section 6, a separate discussion of results on ET from [3] with con-  
779 nections to Conjecture 6.11, a remark on the geometric intuition behind the non-expansiveness  
780 condition (Assumption 6.7) and a remark on the choice of centering term in Theorem 6.10.

781 **Acknowledgments.** We are grateful to Zacharie Naulet from Université d’Orsay for many  
782 valuable comments on statistical side of the paper, to our colleagues from Service Hospitalier  
783 Frédéric Joliot (SHFJ) – Marina Filipović, Claude Comtat and Simon Stute for many practical  
784 insights on the topic of PET-MRI reconstructions and also to anonymous referees for remarks  
785 that have helped to improve greatly the presentation of this work, especially in the numerical  
786 part.

## 787 REFERENCES

- 788 [1] H. H. BARRETT, D. W. WILSON, AND B. M. TSUI, *Noise properties of the em-algorithm. i. theory*, Phys.  
789 Med. Biol., 39 (1994), p. 833.
- 790 [2] D. BLEI AND P. FRAZIER, *Distance dependent chinese restaurant processes*, Journal of Machine Learning  
791 Research, 12 (2011).
- 792 [3] N. A. BOCHKINA AND P. J. GREEN, *The bernstein–von mises theorem and nonregular models*, The Annals  
793 of Statistics, 42 (2014), pp. 1850–1878.
- 794 [4] J. BOWSHER, V. JOHNSON, T. TURKINGTON, R. JASZCZAK, C. FLOYD, AND R. COLEMAN, *Bayesian re-  
795 construction and use of anatomical a priori information for emission tomography*, IEEE Transactions  
796 on Medical Imaging, 15 (1996), pp. 673–686.
- 797 [5] J. BOWSHER, H. YUAN, L. HEDLUND, T. TURKINGTON, G. AKABANI, A. BADEA, W. KURYLO,  
798 C. WHEELER, G. COFER, M. DEWHIRST, AND G. JOHNSON, *Utilizing mri information to estimate  
799 f18-fdg distributions in rat flank tumors*, in IEEE Symposium Conference Record Nuclear Science,  
800 vol. 4, IEEE, 2004.
- 801 [6] S. Y. CHUN, J. A. FESSLER, AND Y. K. DEWARAJA, *Post-reconstruction non-local means filtering methods  
802 using ct side information for quantitative spect*, Physics in Medicine & Biology, 58 (2013), p. 6225.
- 803 [7] C. COMTAT, P. E. KINAHAN, J. A. FESSLER, T. BEYER, D. W. TOWNSEND, M. DEFRISE, AND  
804 C. MICHEL, *Clinically feasible reconstruction of 3d whole-body pet/ct data using blurred anatomical  
805 labels*, Physics in Medicine & Biology, 47 (2001), p. 1.

- 806 [8] M. DAHLBOM, *Estimation of image noise in pet using the bootstrap method*, in IEEE Nuclear Science  
807 Symposium Conference Record, vol. 4, IEEE, 2001.
- 808 [9] D. J. DALEY AND D. VERE-JONES, *An introduction to the theory of point processes: volume I: elementary*  
809 *theory and methods*, Springer Science & Business Media, 2005.
- 810 [10] A. R. FERREIRA AND K. H. LEE, *Single photon emission computed tomography example*, in Multiscale  
811 Modeling, Springer Series in Statistics, 2007.
- 812 [11] J. FESSLER AND A. HERO, *Penalized maximum-likelihood image reconstruction using space-alternating*  
813 *generalized EM algorithms*, IEEE Transactions on Image Processing, 4 (1995), pp. 1417–1429.
- 814 [12] J. A. FESSLER, *Mean and variance of implicitly defined biased estimators (such as penalized maximum*  
815 *likelihood): Applications to tomography*, IEEE Transactions on Image Processing, 5 (1996), pp. 493–  
816 506.
- 817 [13] J. A. FESSLER, N. H. CLINTHORNE, AND W. L. ROGERS, *Regularized emission image reconstruction*  
818 *using imperfect side information*, IEEE Transactions on Nuclear Science, 39 (1992), pp. 1464–1471.
- 819 [14] M. FILIPOVIĆ, E. BARAT, T. DAUTREMER, C. COMTAT, AND S. STUTE, *Pet reconstruction of the pos-*  
820 *terior image probability, including multimodal images.*, IEEE transactions on medical imaging, 38  
821 (2018), pp. 1643–1654.
- 822 [15] M. FILIPOVIĆ, T. DAUTREMER, C. COMTAT, S. STUTE, AND E. BARAT, *Reconstruction, analysis and*  
823 *interpretation of posterior probability distributions of pet images, using the posterior bootstrap*, Physics  
824 in Medicine & Biology, 66 (2021), p. 125018.
- 825 [16] E. FONG, S. LYDDON, AND C. HOLMES, *Scalable nonparametric sampling from multimodal posteriors with*  
826 *the posterior bootstrap*, in Proceedings of the 36th International Conference on Machine Learning,  
827 vol. 97, PMLR, 09–15 Jun 2019, pp. 1952–1962.
- 828 [17] S. GHOSH, A. B. UNGUREANU, E. B. SUDDERTH, AND D. M. BLEI, *Spatial distance dependent chinese*  
829 *restaurant processes for image segmentation*, in Advances in Neural Information Processing Systems  
830 24, Curran Associates, Inc., 2011, pp. 1476–1484.
- 831 [18] P. J. GREEN, *Bayesian reconstructions from emission tomography data using a modified em algorithm*,  
832 IEEE Trans. Med. Imag., 9 (1990), pp. 84–93.
- 833 [19] P. GRÜNWARD AND T. VAN OMMEN, *Inconsistency of Bayesian Inference for Misspecified Linear Models,*  
834 *and a Proposal for Repairing It*, Bayesian Analysis, 12 (2017), pp. 1069 – 1103.
- 835 [20] D. R. HAYNOR AND S. D. WOODS, *Resampling estimates of precision in emission tomography*, IEEE  
836 Transactions on Medical Imaging, 8 (1989), pp. 337–343.
- 837 [21] J. HERMANS, A. DELAUNOY, F. ROZET, A. WEHENKEL, AND G. LOUPPE, *Averting a crisis in simulation-*  
838 *based inference*, arXiv preprint arXiv:2110.06581v2, (2021).
- 839 [22] A. O. HERO, R. PIRAMUTHU, J. A. FESSLER, AND S. R. TITUS, *Minimax emission computed tomog-*  
840 *raphy using high-resolution anatomical side information and b-spline models*, IEEE Transactions on  
841 Information Theory, 45 (1999), pp. 920–938.
- 842 [23] D. HIGDON, J. BOWSER, V. JOHNSON, T. TURKINGTON, D. GILLAND, AND R. JASZCZAK, *Fully bayesian*  
843 *estimation of gibbs hyperparameters for emission computed tomography data*, IEEE Transactions on  
844 Medical Imaging, 16 (1997), p. 516.
- 845 [24] T. HOHAGE AND F. WERNER, *Inverse problems with poisson data: statistical regularization theory, ap-*  
846 *plications and algorithms*, Inverse Problems, 32 (2016), p. 093001.
- 847 [25] L. F. JAMES, *Bayesian calculus for gamma processes with applications to semiparametric intensity models*,  
848 Sankhyā: The Indian Journal of Statistics, (2003), pp. 179–206.
- 849 [26] M. JUDENHOFER, H. WEHRL, D. NEWPORT, C. CATANA, S. SIEGEL, M. BECKER, A. THIELSCHER,  
850 M. KNEILLING, M. LICHY, M. EICHNER, K. KLINGEL, G. REISCHL, S. WIDMAIER, M. RÖCKEN,  
851 R. NUTT, H. MACHULLA, K. ULUDA, S. CHERRY, C. CLAUSSEN, AND B. PICHLER, *Simultaneous*  
852 *pet-mri: a new approach for functional and morphological imaging*, Nature medicine, 14 (2008),  
853 pp. 459–465.
- 854 [27] B. J. K. KLEIJN AND A. W. VAN DER VAART, *The bernstein-von-mises theorem under misspecification*,  
855 Electronic Journal of Statistics, 6 (2012), pp. 354–381.
- 856 [28] C. LARTIZIEN, J.-B. AUBIN, AND I. BUVAT, *Comparison of bootstrap resampling methods for 3-d pet*  
857 *imaging*, IEEE Transactions on Medical Imaging, 29 (2010), pp. 1442–1454.
- 858 [29] C. S. LEVIN, M. DAHLBOM, AND E. J. HOFFMAN, *A monte carlo correction for the effect of compton*  
859 *scattering in 3-d pet brain imaging*, IEEE Transactions on Nuclear Science, 42 (1995), pp. 1181–1185.

- 860 [30] Y. LI, *Noise propagation for iterative penalized-likelihood image reconstruction based of fisher information*,  
861 Phys. Med. Biol., 56 (2011), p. 1083.
- 862 [31] J. S. LIU, *The fraction of missing information and convergence rate for data augmentation*, Computing  
863 Science and Statistics, (1994), pp. 490–497.
- 864 [32] J. S. LIU, W. H. WONG, AND A. KONG, *Covariance structure of the gibbs sampler with applications to  
865 the comparisons of estimators and augmentation schemes*, Biometrika, 81 (1994), pp. 27–40.
- 866 [33] A. Y. LO, *Bayesian nonparametric statistical inference for poisson point processes*, Zeitschrift fur  
867 Wahrscheinlichkeitstheorie und verwandte Gebiete, 59 (1982), pp. 55–66.
- 868 [34] A. LUNA, J. C. VILANOVA, L. C. HYGINO DA CRUZ JR, AND S. E. ROSSI, *Functional imaging in oncology:  
869 biophysical basis and technical approaches - Vol. 1*, Springer Science & Business Media, 2013.
- 870 [35] S. LYDDON, S. WALKER, AND C. HOLMES, *Nonparametric learning from bayesian models with randomized  
871 objective functions*, Advances in Neural Information Processing Systems, (2018).
- 872 [36] L. G. MARCU, L. MOGHADDASI, AND E. BEZAK, *Imaging of tumor characteristics and molecular pathways  
873 with pet: developments over the last decade toward personalized cancer therapy*, International Journal  
874 of Radiation Oncology Biology Physics, 102 (2018), pp. 1165–1182.
- 875 [37] F. NATTERER, *The mathematics of computerized tomography*, Society for Industrial and Applied Mathe-  
876 matics, 2001.
- 877 [38] M. A. NEWTON AND A. E. RAFTERY, *Approximate bayesian inference with the weighted likelihood boot-  
878 strap*, Journal of the Royal Statistical Society: Series B (Methodology), 56 (1994), pp. 3–26.
- 879 [39] T. L. NG AND M. A. NEWTON, *Random weighting in LASSO regression*, Electronic Journal of Statistics,  
880 16 (2022), pp. 3430 – 3481.
- 881 [40] L. NIE AND V. ROČKOVÁ, *Bayesian bootstrap spike-and-slab lasso*, Journal of the American Statistical  
882 Association, (2022), pp. 1–16.
- 883 [41] O. PAPASPILIOPOULOS, G. O. ROBERTS, AND M. SKÖLD, *A General Framework for the Parametrization  
884 of Hierarchical Models*, Statistical Science, 22 (2007), pp. 59 – 73.
- 885 [42] E. POMPE, *Introducing prior information in weighted likelihood bootstrap with applications to model mis-  
886 specification*, arXiv preprint arXiv:2103.14445, (2021).
- 887 [43] A. RAHMIM, J. TANG, AND H. ZAIDI, *Four-dimensional (4d) image reconstruction strategies in dynamic  
888 pet: Beyond conventional independent frame reconstruction*, Medical Physics, 36 (2009), pp. 3654–  
889 3670.
- 890 [44] L. A. SHEPP AND Y. VARDI, *Maximum likelihood reconstruction for emission tomography*, IEEE trans-  
891 actions on medical imaging, 1 (1982), pp. 113–122.
- 892 [45] R. SIDDON, *Fast calculation of the exact radiological path for a three-dimensional ct array.*, Medical  
893 physics, 12 2 (1985), pp. 252–5.
- 894 [46] A. SITEK, *Data analysis in emission tomography using emission count posteriors*, Physics in Medicine &  
895 Biology, 52 (2012), p. 6779.
- 896 [47] A. SITEK AND M. A. CELLER, *Limitations of poisson statistics in describing radioactive decay*, Physica  
897 Medica, 31 (2015), pp. 1105–1107.
- 898 [48] S. STUTE AND C. COMTAT, *Practical considerations for image-based psf and blobs reconstruction in pet*,  
899 Physics in Medicine & Biology, 58 (2013), p. 3849.
- 900 [49] A. W. VAN DER VAART, *Asymptotic statistics*, vol. 3, Cambridge university press, 2000.
- 901 [50] D. A. VAN DYK AND X.-L. MENG, *The art of data augmentation*, Journal of Computational and Graphical  
902 Statistics, 10 (2001), pp. 1–50.
- 903 [51] F. VASCONCELOS, B. HE, N. SINGH, AND Y. W. TEH, *Uncertainr: Uncertainty quantification of end-  
904 to-end implicit neural representations for computed tomography*, arXiv preprint arXiv:2202.10847v2,  
905 (2022).
- 906 [52] K. VUNCKX, A. ATRE, K. BAETE, A. REILHAC, C. M. DEROOSE, K. VAN LAERE, AND J. NUYTS,  
907 *Evaluation of three mri-based anatomical priors for quantitative pet brain imaging*, IEEE transactions  
908 on medical imaging, 31 (2011), pp. 599–612.
- 909 [53] W. A. WEBER, *Use of pet for monitoring cancer therapy and for predicting outcome.*, Journal of Nuclear  
910 Medicine, 46 (2005), pp. 983–995.
- 911 [54] I. S. WEIR, *Fully bayesian reconstructions from single-photon emission computed tomography data*, Jour-  
912 nal of the American Statistical Association, 92 (1997), pp. 49–60.