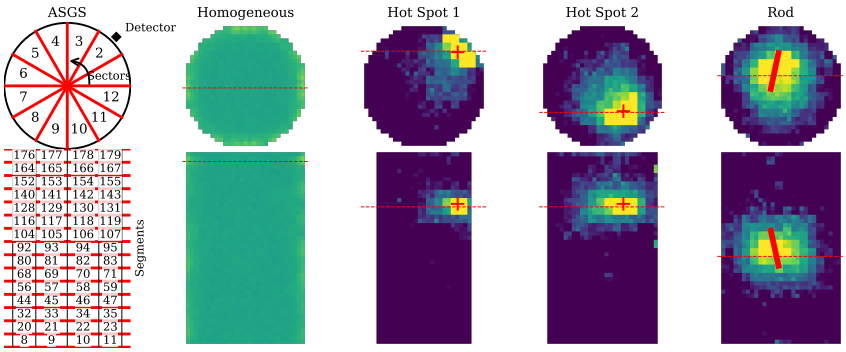
Graphical Abstract

3D Activity reconstruction from Angular Gamma Scanning via variational Bayes: a proof of concept

Victor J. Casas-Molina, Eric Laloy, Bart Rogiers, Tom Dhaene, Ivo Couckuyt



Highlights

3D Activity reconstruction from Angular Gamma Scanning via variational Bayes: a proof of concept

Victor J. Casas-Molina, Eric Laloy, Bart Rogiers, Tom Dhaene, Ivo Couckuyt

- 3D reconstruction of the spatial activity distribution from AGS measurements.
- High-dimensional Bayesian inference using SVI.
- Multi-resolution prior yields realistic activity fields with preserved coherence.
- Accurate results for isotopes and isotopic vectors with multiple gamma emission lines.

3D Activity reconstruction from Angular Gamma Scanning via variational Bayes: a proof of concept

Victor J. Casas-Molina[a,b], Eric Laloy[a], Bart Rogiers[a], Tom Dhaene[b], Ivo Couckuyt[b]

[a] *Sustainable Waste & Decommissioning, Belgian Nuclear Research Centre (SCK CEN), Boeretang 200, Mol, 2400, Belgium*
[b] *Department of Information Technology, Ghent University-imec, Technologiepark-Zwijnaarde 126, Ghent, 9052, Belgium*

Abstract

This study serves as a proof of concept for a Bayesian variational framework enabling high-resolution 3D activity reconstruction in 220 liter waste drums using Angular Segmented Gamma Scanning (ASGS) data and transmission-derived attenuation maps. Our proposed inference and uncertainty quantification approach is demonstrated using virtual experiments that simulate typical waste characterization scenarios. Computations are made tractable by using stochastic variational inference (SVI) together with a multi-resolution spatial prior to infer the spatial activity distribution. Results show that the approach can recover the spatial activity distribution within the considered drum, while also providing more accurate total activity estimates than conventional methods, thereby enhancing the accuracy of radiological waste characterization.

Keywords: Segmented Gamma Scanning, Bayesian Inference, Variational Inference, Uncertainty, Legacy waste, non-destructive testing

Email address: victor.casas.molina@sckcen.be (Victor J. Casas-Molina)

1. Introduction

Non-destructive waste characterization methods can be used to assess the radioactive inventory of waste drums in a cost-effective way, while reducing operator exposure. One of these methods is Segmented Gamma Scanning (SGS). SGS involves scanning the package in horizontal segments, employing a detector,normally a HPGe detector, equipped with a collimator to reduce crosstalk and ensure proper spatial separation of segments. In combination with transmission data, SGS can handle heterogeneous matrices, and therefore has the potential to provide more accurate inventory estimates.

The SGS variant, known as Angular Segmented Gamma Scanning (ASGS) allows the horizontal segments to be further divided into equi-angular cylindrical sectors by rotating the waste package around its vertical axis. This technique provides a coarse estimate of the spatial activity distribution within the considered waste package.

Traditional estimation methods are usually based on a prescribed source distribution (either point or homogeneous) and on the assumption of a homogeneous drum matrix in terms of density and chemical composition. For the case of highly inhomogeneous waste or source distributions, which is frequent for legacy, non-conditioned waste, this results in a biased activity quantification with unreliable uncertainty estimates (Laloy et al., 2021, 2024).

In recent years, different Bayesian approaches have been developed to improve the inference of activity distributions (Bücherl et al., 2021, Laloy et al., 2021, Carasco, 2021). The approach by Bücherl et al. (2021) combines ASGS measurements with information related to density obtained from transmission measurements. Transmission measurements, e.g., (Pham et al., 2013), provide estimates of the sectorial Linear Attenuation Coefficients (LACs) within the drum.

However, the progress made so far using Bayesian approaches together with ASGS has been limited to 2D inferences, where activity was estimated on a grid of approximately 5,000 possible source locations (Bücherl et al., 2021).

In some cases, the measured count rates were also used to calibrate the prior distributions. While this helps the inference to be more stable, it is known to potentially lead to uncertainty underestimation in the inferred quantities, since the same data are used twice: to shape the prior and to calculate the likelihood (Gelman et al., 2013).

Extending this Bayesian inference to 3D is not straightforward due to the considerable computational complexity of the problem. Nevertheless, such an extension would provide a more general solution to the probabilistic interpretation of ASGS data and could significantly enhance inventory estimates. This improvement would be particularly beneficial for tasks requiring precise source determination, such as waste batch repacking (International Atomic Energy Agency, 2024). To address this challenge, we have developed a methodology capable of performing inference of high-resolution 3D activity distributions within 220 liter waste drums and quantify the uncertainty associated with the derived spatial activity distribution. Our approach is rooted within the Bayesian paradigm. In this work, we provide a proof-of-concept under the assumption that suitable high-resolution LAC data are available from transmission measurements.

Using synthetic data simulated with Monte Carlo (MC) and representative of real, archetypal cases, we aim to improve the methodology presented by Bücherl et al. (2021) and to extend it from the 2D to the 3D case. Amongst others this means that the number of unknowns goes from about 5,000 in 2D to about 32,000 in our 3D setup. To address the computational challenges of full Bayesian inference, in such a high-dimensional space, we approximate the posterior activity distribution parametrically, thereby turning the problem into a high-dimensional optimization task. For this purpose, our approach employs Stochastic Variational Inference (SVI) (Hoffman et al., 2013). We consider a 3D activity field with a voxel size of 1.92 cm in our inferences, while the observed data are the observed counts per sector. In addition, our approach accounts for the attenuation effects of the drum matrix and the spatial and spectral response of the detector.

Combining this method with techniques for measuring or estimating matrix attenuation, such as Gamma Tomography or Computed Tomography (CT), can significantly enhance the characterization process for heterogeneous waste drums for which the content and/or origin are unclear, such as in the case of orphan waste (International Atomic Energy Agency, 2024). This approach also allows for uncertainty quantification, providing a more comprehensive and reliable assessment than conventional methods based on averaging or best-estimates.

The remainder of this paper is organized as follows. Section 2 presents the considered physical and probabilistic models and their implementation, the computational aspects of the method, and the design and suitability of the synthetic drums used as case studies. Section 3 presents our results, which is followed by a discussion in Section 4. Lastly, Section 5 concludes the paper and outlines some future directions of work. At the end of the document, in the Data Availability section, there are indications for the reader to access the code and data used in this work.

2. Methods

The methodology used in this is presented in the diagram shown in Fig. 1.

As shown in the diagram, four distinct workflows can be identified and are described in the following subsections. The forward model generates detector counts from the ASGS process given an unknown activity field and a set of problem-specific variables. The probabilistic model combines this forward model with prior information on the source distribution to infer the true activity for the synthetic cases used to illustrate the method. Synthetic drums are generated through simulations based on predefined test case specifications, including filler density, source location, and detector coordinates. The quantities of interest are the posterior mean of the activity field and the total recovered activity within a drum, computed as the sum of the posterior mean over all voxels.

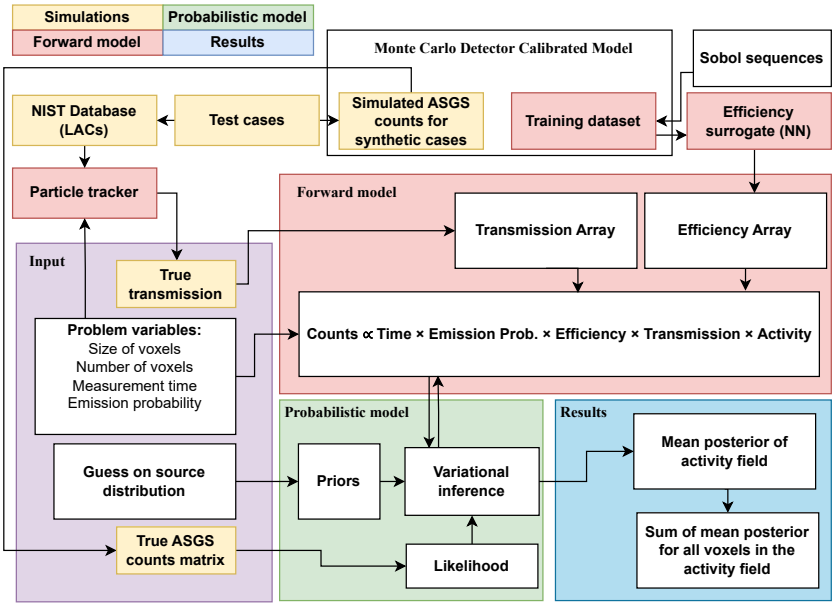


Figure 1: Diagram showing the methodology followed in this work.

2.1. Forward model

To infer the spatial activity distribution within a waste drum, it is required to devise a functional form of the model that governs the registered counts in the detector positions given an activity distribution. Based on the model developed by Bücherl et al. (2021) the following expression is used for obtaining the counts for a generic RI in the forward calculation:

$$s_{n,m,e} = t_{n,m} \cdot \eta_e \cdot \sum_{i,j,k} [(w_{i,j,k})_{n,m,e} \cdot a_{i,j,k}] \tag{1}$$

The counts, $s_{n,m,e}$ for a given energy line e measured at vertical segment n and angular sector m during ASGS is equal to the product of measuring time at that position $t_{n,m}$, the emission probability of the energy line from the identified RI η_e, and a matrix term which accounts for the attenuation and detection efficiency of the system for that given energy line, $(w_{i,j,k})_{n,m,e}$, coming from an activity $a_{i,j,k}$ positioned in the 3D Cartesian space xyz, which in

this case, is discretized in finite equal-sized elements or voxels with coordinates i, j, k. This model is only limited by the geometrical approximation done by the voxelization approach of the drum, i.e., certain small and thin details such as small elements and the drum wall would be coarsened with all the limitations that this implies. Nonetheless, reducing this geometrical approximation error would imply to dramatically increase the number of voxels to be inferred. It is worth recalling that about 32,000 voxels are needed, using a 1.92 cm size voxel, to cover the space within the drum.

The parameters, namely $t_{n,m}$, η_e, and $(w_{i,j,k})_{n,m,e}$ in Eq. 1, are to be determined prior to the inversion. The measuring time is considered to be known and in this application, set to one since we are working with MC simulations. Similarly, probability of emission for different energies and different RIs are well known and tabulated. In this application, all the emission probabilities have been retrieved from Bé et al. (2006).

The weighting term $(w_{i,j,k})_{n,m,e}$ is computed via element-wise multiplication of two arrays:

$$(w_{i,j,k})_{n,m,e} = (\mathcal{E}_{i,j,k})_{n,m,e} \cdot (\mathcal{T}_{i,j,k})_{n,m,e} \tag{2}$$

where $(\mathcal{E}_{i,j,k})_{n,m,e}$, is the detector response at a given voxel, detector position and energy, and $(\mathcal{T}_{i,j,k})_{n,m,e}$, is the transmittance value for the path between the voxel (i, j, k) to the center of the detector window for detector position (n, m) and for the energy e.

2.1.1. Detector efficiency

The detector efficiency must account for both its intrinsic efficiency and its spatial and spectral behavior. Even though there are many ways to model or approximate a detector response, whether numerical, analytical or using commercial software such as ISOCS (Venkataraman et al., 1999), our goal is to set up a methodology for flexible and accurate modeling of the efficiency for complex detectors used in ASGS.

Therefore, we built a MC model of an existing collimated HPGe detector for ASGS using the PHITS code (Sato et al., 2018). Once the model was constructed we calibrated its effective parameters using Bayesian optimization and actual efficiency measurements. For a comprehensive explanation of this MC model construction and evaluation, we refer the reader to Casas-Molina et al. (2024).

Unfortunately, the MC model is too computationally expensive to be integrated within the proposed inversion method. Instead we thus surrogated the MC model and replaced it with its very efficient surrogate based on a Feedforward Neural Network (NN).

The data used for training the NN was generated by the calibrated MC model. A total of 20,000 training points were created in two steps. A first batch of 5,000 points was sampled using a space-filling Sobol sequence (Sobol, 1967) from a reduced space with bounds described in Table 1.

Table 1: Input bounds for the first batch of the NN training dataset.

Dimension	**Units**	**Lower**	**Upper**
x(lateral, drum width)	cm	-30	30
y (vertical, drum height)	cm	-50	50
z (depth, towards drum)	cm	15	100
e (photopeak energy)	keV	40	2,000

The spatial boundaries were chosen based on preliminary simulations that helped us assess at what point counts were becoming practically zero even for a considerable activity value ($\mathcal{E} \sim 10^{-12}$). The spectral boundaries were based on the detector specifications.

A second batch of 15,000 points, that were extracted from the central, uncollimated area of the detector with the objective of properly capturing the spatial region where nonlinearities are most pronounced. This second batch is described in Table 2.

Table 2: Input bounds for the second batch of the NN training dataset.

Dimension	**Units**	**Lower**	**Upper**
x (lateral, drum width)	cm	-15	15
y (vertical, drum height)	cm	-30	30
z (depth, towards drum)	cm	15	100

In this case, a second Sobol sequence was used to generate 1,500 values for x, y and z and then, this sequence was repeated for 10 different energies of typical RIs, which are 59.54 (^{241}Am), 121.78 (^{152}Eu), 244.70 (^{152}Eu), 344.27 (^{152}Eu), 661.66 (^{137}Cs), 779.36 (^{152}Eu), 964.45 (^{152}Eu), 1112.39 (^{152}Eu), 1332.49 (^{60}Co) and 1408.45 keV (^{152}Eu).

These isotopes, commonly used for detector calibration, are especially helpful for evaluating the efficiency curve in regions where it deviates most from linearity.

Once the design of the dataset was determined, the efficiency of a point source in the center of the voxel was evaluated in these 20,000 points (x, y, z, e) through MC simulations. To evaluate the detector's response across these 20,000 space and energy configurations (x, y, z, e), 20,000 input files for PHITS were generated using the calibrated detector model. For each file, the point source was positioned at the spatial coordinates defined by the experimental design and assigned the corresponding photon energy. Once the simulations were completed, the tally of electron interactions in the detector crystal, resulting from photons with energies corresponding to the experimental energy lines, was processed. The background was then subtracted in every independently generated peak to obtain the Full Energy Peak Efficiency (FEPE) of each simulation. The input and output data for training was then independently scaled from 0 to 1.

The NN surrogate model was implemented in Keras (Chollet et al., 2015) and its hyperparameters were found through neural architecture search via Optuna (Akiba et al., 2019). The final model contains three hidden layers. The first hidden layer has 8 neurons, the second 16, and the third 64, all using the

Exponential Linear Unit (ELU) activation function. The output layer consists of a single neuron with a linear activation, suitable for regression tasks. The model was trained with the Adam optimizer (Kingma and Ba, 2014) and Huber loss function, and trained for 300 epochs with a batch size of 4. The trained network was then used to prepare precomputed arrays of $(\mathcal{E}_{i,j,k})_{n,m,e}$. The performance of the surrogate was validated by k-fold cross-validation using 10 folds via quantitative evaluation based on the Root Mean Squared Error (RMSE). In Fig 2 a schematic of the neural network architecture is shown.

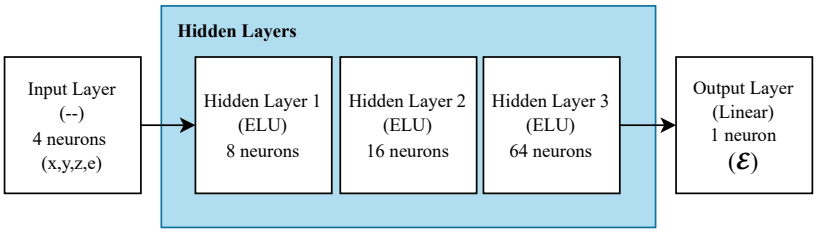


Figure 2: Schematic view of the neural network architecture used in this work to surrogate the detector efficiency. Activation functions (ELU and Linear) are between brackets indicating the connection with the preceding layer. Input is 4-dimensional (x,y,z,e) and output is single-dimensional ($\mathcal{E}$)

2.1.2. Transmittance matrix

The term $(\mathcal{T}_{i,j,k})_{n,m,e}$ of Eq. 2 is the attenuation of a photon beam calculated by the Beer-Lambert law, that states in its general formulation that the transmittance for multiple materials on a path of a photon beam is calculated as:

$$\mathcal{T} = \exp\left(-\sum_{s} \mu_s l_s\right) \tag{3}$$

where μ_s is the LAC and l_s is the path length through the material s. As explained by Bücherl et al. (2021), this concept can be used for describing the probability that photons emitted within the drum escape it.

For a drum embedded within a regular grid, dividing its space into $i \times j \times k$

voxels, the contribution of each voxel (i, j, k) to the signal detected at position (n, m) is determined by accounting for the attenuation caused by all voxels along the path $\mathcal{P}$ from voxel (i, j, k) to detector position (n, m). Using Eq. 3, this contribution is expressed as:

$$(\mathcal{T}_{i,j,k})_{n,m,e} = \exp\left(-\sum_{\substack{p \\ (i,j,k)\to(n,m)}}^{\mathcal{P}} \mu_p^e \cdot l_p\right) \tag{4}$$

Where p represents every voxel in the path, $\mathcal{P}$, between voxel (i, j, k) and detector position (n, m), μ_p^e represents the LAC for an energy e of the material in the voxel p and l_p, represents the distance traversed by a photon in the voxel p that was emitted from a source located at the voxel (i, j, k). The LAC can be estimated from transmission or CT tomography techniques as they are common solutions available for the characterization of waste drums (Pham et al., 2013, Carasco, 2021, Estep et al., 1994). This approach has been proven to work in Bücherl et al. (2021). Nonetheless, as stated before, we assume here that the LACs are known values.

2.2. Probabilistic model

In Bayesian statistics, the probability density function of random variables is no longer representing only the frequency or repetition of the observed states as in the case of frequentist analysis, but a degree of belief in the actual value for that variable. Bayes' theorem is used to calculate the posterior probability of a hypothesis, or in this case, a parameter set, given prior evidence or belief and fit to available data encoded in a likelihood function.

$$p(A \mid S^{obs}, \vec{\theta}) = \frac{p(S^{obs} \mid A, \vec{\theta}) \cdot p(A \mid \vec{\theta})}{p(S^{obs} \mid \vec{\theta})} \tag{5}$$

Where $p(A \mid S^{obs}, \vec{\theta})$ is the posterior probability distribution which represents the activity field A given the observed counts array S^{obs} from the different detector positions during ASGS and a deterministic vector of parameters $\vec{\theta}$, ($t_{n,m}$, η_e, and $(w_{i,j,k})_{n,m,e}$ in our general model, Eq. 1) that describe both the

drum matrix and the measurement setup. According to Bayes' theorem, this posterior combines the likelihood, $p(S^{obs}|A,\vec{\theta})$, with prior information on the activity, $p(A \mid \vec{\theta})$. The denominator, $p(S^{obs} \mid \vec{\theta})$, is the normalization constant:

$$p(S^{obs} \mid \vec{\theta}) = \int p(S^{obs} \mid A, \vec{\theta}) \cdot p(A \mid \vec{\theta}) dA \tag{6}$$

Even though the normalization constant (Eq. 6) is difficult to estimate in practice, it is not indeed required for inference as long as the dimensionality of A remains fixed. Therefore we focus on drawing samples from the unnormalized posterior:

$$p(A \mid S^{obs}, \vec{\theta}) \propto p(S^{obs} \mid A, \vec{\theta}) \cdot p(A \mid \vec{\theta}) \tag{7}$$

In principle, jointly inferring $\vec{\theta}$ together with A could possibly help accounting for uncertainty in the characterization process. However, doing so would significantly increase the dimensionality of the problem and make inference more challenging. For this reason, all parameters other than the activity are treated as deterministic and fixed in this work.

2.2.1. Likelihood function

The Likelihood function represents the probability of observing a certain set of counts for a given activity present within the waste drum, $p(S^{obs}|A,\vec{\theta})$.

Radiation phenomena, such as decay and gamma emission is fundamentally described by a Poisson distribution (Gilmore, 2008). Therefore the counts registered by the detector in the different segments are assumed to follow a Poisson distribution and the likelihood of observing these counts is provided by:

$$S^{obs} \sim \text{Poisson}(s_{n,m,e}) \tag{8}$$

However, in a realistic setting, the likelihood may not be strictly Poisson-like. Due to measurement noise (here, MC noise) and complex model errors (e.g., approximations and epistemic uncertainty), the Poisson assumption may not hold—other noise sources are also present.

Using a pure Poisson likelihood leads to an overconfident, too narrow, posterior, leading to incorrect recovery of the total activity. To address this, a normal approximation to the Poisson is used, with heteroscedastic noise added via a noise scaling factor, σ_{noise}, to be inferred for each case:

$$S^{obs} \sim \mathcal{N}(s_{n,m,e},\, \sigma_{noise}\sqrt{s_{n,m,e}}) \tag{9}$$

In our model, the σ_{noise} factor is sampled from a Gamma distribution prior, which is flexible for modeling positive-valued uncertainties and is a common choice for scale parameters (Aksoy, 2000):

$$\sigma_{noise} \sim Gamma(\alpha = 2, \theta = 1) \tag{10}$$

2.2.2. The activity prior

The information on the problem parameters can be incorporated by using prior probability distributions. In this context, stronger priors could be used for the parameters that are better known for instance through calculations. While for other parameters, little or no information may be available, and we can fall back to using literature values or expert elicitation to come up with prior distributions, this is the case for the activity.

The activity values can be assumed to be non-negative and bounded by a maximum related to a worst-case scenario derived from expert criteria, available information or direct interpretation of S^{obs}, since different activity distributions will yield different detection signals.

In this work, we have experimented with different prior distributions for the activity. Weak, uninformative priors such as uniform, independent distributions have proven to be difficult or even impossible to work with. The reasoning behind this is two-fold:

Trying to infer completely independent voxel activities makes the problem harder than it actually is. In reality, activity is often spatially correlated: neighboring voxels are likely to contain similar activities, especially in small voxels like those considered here. This expected smoothness effectively reduces the

dimensionality of the problem and helps regularize the solution, avoiding compensating errors between voxels that could fit the data but produce unrealistic reconstructions.

In addition, using uninformative priors can allow individual voxels to take on extreme or unrealistic values. While the smoothness prior helps constrain the solution spatially, uninformative priors alone do not prevent scientifically meaningless activity reconstructions, especially in high-dimensional settings.

Other typically employed methods such as Markov Random Fields (MRFs), Conditional AutoRegressive (CARs) models and Gaussian Process (GPs) priors (Li, 2009),(Banerjee et al., 2014),(Rasmussen and Williams, 2006).

The first two models did not perform as expected since neither MRFs nor CARs where able to impose the expected smoothness and to locate the source, additionally, the true activity values were not properly recovered in the posterior distribution. On the other hand, GPs proved impractical to implement: maintaining a voxel resolution of 1.92 cm would require inverting a covariance matrix with at least 32,062 variables, which leads to out-of-memory issues and computational bottlenecks.

Instead of investigating more scalable approaches, we decided to adopt a simple multi-resolution, spatially structured model. The activity field is first defined on a coarse three-dimensional grid, which serves as a low-resolution backbone representing large-scale spatial patterns. This coarse field is then expanded to a finer resolution by interpolating its structure and introducing small, localized deviations around it. These finer details are not arbitrary, but constrained relative to the coarse structure, allowing for spatial variability without sacrificing coherence. Our prior is built as follows. For a single energy peak we have:

$$a_{i,j,k} = a_{i,j,k}^{normalized} \cdot a_{i,j,k}^{fine} \cdot \delta_{i,j,k} \times \lambda \tag{11}$$

where $a_{i,j,k}$ is the activity (Bq) in voxel i, j, k, $a_{i,j,k}^{normalized}$ is the inferred normalized activity variation field at block level (8 voxels group), $a_{i,j,k}^{fine}$ is the

inferred activity variation field at the voxel level, $\delta_{i,j,k}$ is a fixed masking array that allows inference only in voxels inside the drum, and λ, an inferred parameter that scales the activity fields product to the inferred activity scaling factor.

To produce $a_{i,j,k}^{\text{normalized}}$, we start by defining $a_{u,v,w}^{\text{coarse}}$, an intermediate variable representing a block-level activity field. Each coarse block groups 8 voxels ($2 \times 2 \times 2$), and the total number of coarse blocks is therefore 6,144. Specifically, the original voxel grid has dimensions $31 \times 48 \times 31$; after padding each dimension to be divisible by 2 (resulting in $32 \times 48 \times 32$) and downsampling by a factor of 2, the coarse grid dimensions become $16 \times 24 \times 16$. Here, u, v, w are the indices of the coarse blocks. Each of these blocks is modeled using independent Beta priors:

$$a_{u,v,w}^{coarse} \sim Beta(\alpha, \beta) \tag{12}$$

The Beta distribution has its support between 0 and 1 and its shape can be modeled via the α and β parameters. For this reason, the Beta distribution is frequently employed in the modeling of proportions. As illustrated in Fig. 3, one can play with the alpha and beta values to represent the a-priori knowledge we have about the spatial activity distribution in the drum.

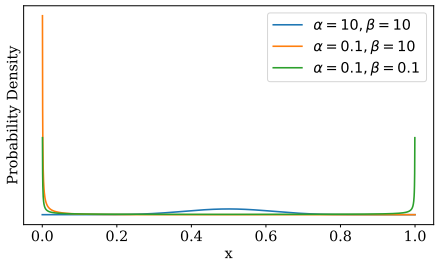


Figure 3: Probability density functions of generic Beta distributions with different set of parameters

For instance, if the detector signals are very similar across different positions, one could expect a fairly homogeneous drum and therefore try to force

homogeneity by choosing symmetric parameters such as $\alpha = \beta = 10$, which yields a distribution concentrated around the midpoint.

Conversely, if prior knowledge suggests that the drum is likely to be nearly empty, one might select parameters like $\alpha = \beta = 0.1$ to reflect this expected bimodal aspect. If only hot spots are expected, a more aggressive distribution can be placed on the proportion by using for example α =0.1, $\beta = 20$, encoding in this way that almost all the activity field is expected to be 0 but with potentially some larger outliers.

Thus, the flexibility of the Beta distribution allows for encoding a wide range of prior assumptions about the spatial distribution of activity across the considered drum blocks.

To sum up, the main task of this first level prior is to assess the proportion of the total activity in each coarse block. In this work, for the homogeneous case, we used $\alpha = \beta = 10$, while we set $\alpha = 0.1$, $\beta = 20$ for the other cases.

In order to broadcast the coarse field $a_{u,v,w}^{coarse}$ down to the voxel grid dimensions, the Kronecker product is used in such a way that:

$$a_{i,j,k}^{broadcasted} = a_{u,v,w}^{coarse} \otimes 1_{2\times2\times2} \tag{13}$$

The Beta distribution is used only to localize the activity source, therefore, $a_{i,j,k}^{broadcasted}$ is normalized to sum up to one:

$$a_{i,j,k}^{normalized} = \frac{a_{i,j,k}^{broadcasted}}{\sum a_{i,j,k}^{broadcasted}} \tag{14}$$

To perform fine tuning of the resulting activity field, a second prior for a finer activity field, $a_{i,j,k}^{fine}$, is now placed at voxel-level. This prior "accommodates" the coarse values inferred at the block-level before performing fine tuning of the full activity spatial distribution.

The easiest way to implement this is to select independent probability distribution functions that multiply each block-level value by a correction factor close to 1. In total, as one distribution per voxel is needed, 32,062 distributions are inferred with this strategy. After some preliminary testing, we used a Half

Normal distribution with standard deviation of 1 (Normal distribution centered on 0, restricted to the positive range):

$$a_{i,j,k}^{fine} \sim HalfNormal(\sigma = 1) \tag{15}$$

Half Normal distributions allow positive variations at the fine-field level while keeping values close to those defined at the coarse-field level, if necessary.

In order to avoid including voxels located outside the drum, the resulting activity field is multiplied by a deterministic masking array ($\delta_{i,j,k}$ in Eq. 11) which elements are 0 when the voxel is located outside the drum, and 1 if they are inside.

The values for the product of $a_{i,j,k}^{normalized}$ and $a_{i,j,k}^{fine}$ fields are expected to be very close to [0,1] interval given the support of Beta distribution and the low variability allowed by the Half Normal. Therefore a Log Uniform prior is set for a scalar (λ) that multiplies all the activity field and that accounts for the scaling factor of the activity:

$$\lambda \sim LogUniform(L = 4, U = 12) \tag{16}$$

This scaling factor represents the total activity within the drum. Its bounds [4,12] are set such that the resulting prior distribution over the total activity lies within the expected range for a given drum. In this case, we set lower and upper bounds to [4,12].

By sampling from the full probabilistic model using these bounds, the effective prior distribution over the total activity is obtained, as shown in Fig. 4. Note that this distribution is close to log-uniform due to the log-uniform prior placed on the activity scaling factor λ (Eq. 16), and it reflects the sum of $a_{i,j,k}$ across all voxels.

This prior is designed to allow the inference to recover the true total activity if it lies within approximately 5 kBq to 500 GBq. By sampling from the full probabilistic model, the resulting effective prior over the sum of all voxel activities is log-uniform (uniform in the logarithmic space of the total activ-

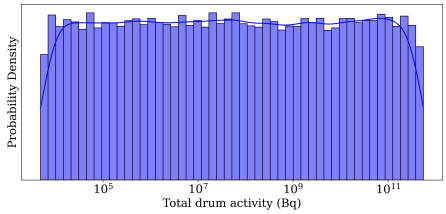


Figure 4: Actual prior distribution for the total activity in the drum (sum of $a_{i,j,k}$), obtained by prior sampling of the full probabilistic model. This effective prior distribution is close to log-uniform (note the base 10 logarithmic scale of the x-axis), given the log-uniform placed on λ (Eq. 16).

ity), meaning that all total activity values within this range are expected to occur with equal probability, without incorporating the data yet to the inference (only using prior knowledge). This procedure leads to the final expression for the activity field presented in Eq. 11.

Fig. 5 presents a schematic overview of the entire computational graph.

It can be seen that, the top-level inference includes priors for the scaling factor (λ) and the coarse activity field ($a^{coarse}_{u,v,w}$). Then dimensions are matched according to Eq. 13 with the one of the fine level ($a^{fine}_{i,j,k}$). After masking, the activity ($a_{i,j,k}$) is passed to the forward model. The resulting counts are compared to observations through the likelihood (within the dashed box).

2.2.3. SVI in inverse problems

In this work we are solving an inverse problem to find the activity: the observed counts are a function of an unknown ideal model, the inferred activity, and some fixed parameters in such a way that:

$$S = f(A, \vec{\theta}) \tag{17}$$

The problem is ill-posed since the number of available observations in S is typically much smaller than the number of elements of the three dimensional activity field A, i.e., 180 observations ($n \times m$) vs. 32,062 voxels ($i \times j \times k$)

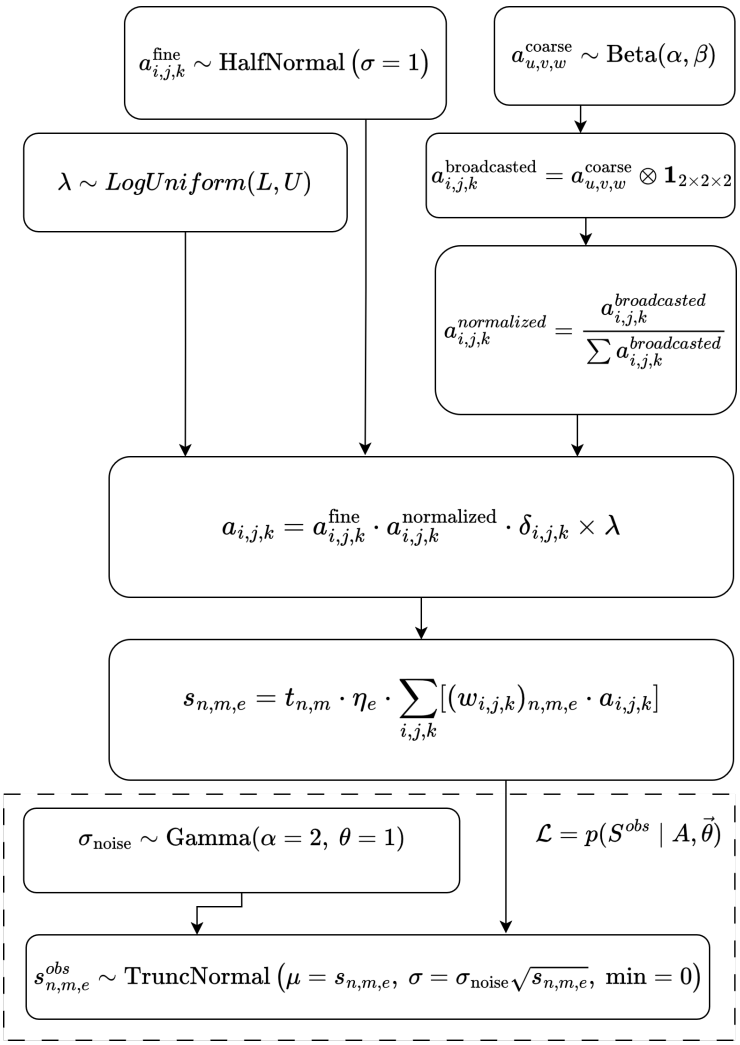


Figure 5: Sketch of the entire probabilistic model.

within the drum in our case. This ill-posedness is exacerbated as the discretization of the activity grid becomes finer. Additionally, there is measurement noise present in the observations and, more importantly, uncertainties arising from the model error as our forward model (see Eqs. 1 and 2) is based on different simplifications. First, the detector's efficiency is provided by a data-driven surrogate model trained beforehand and second, matrix attenuation is assumed to be known.

These factors render the problem analytically intractable and one has to resort to methods like Markov Chain Monte Carlo (MCMC) (Gilks et al., 1995) to approximate the exact posterior, or SVI to optimize a parametric approximation of the posterior distribution.

Unfortunately, given the high dimensionality of this problem, our trials of using state-of-the-art MCMC samplers, like the No U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) were unfruitful. Regardless of any prior structure, regularization or reparameterization, convergence was never reached for all the sampled variables.

As a consequence, we resorted to Stochastic Variational Inference (SVI) (Hoffman et al., 2013). Instead of drawing samples directly from the true posterior $p(\vec{\psi} \mid S^{obs})$, SVI defines a simpler family of distributions $q_{\vec{\phi}}(\vec{\psi})$, known as the *variational distribution*, *guide*, or *variational family*, and adjusts its parameters $\vec{\phi}$ to resemble the true posterior. This is achieved by maximizing the Evidence Lower Bound (ELBO) (Hoffman et al., 2013), defined as

$$\mathrm{ELBO}(\vec{\phi}) = \mathbb{E}_{q_{\vec{\phi}}(\vec{\psi})}\big[\log p(S^{obs} \mid \vec{\psi})\big] - \mathrm{KL}\big(q_{\vec{\phi}}(\vec{\psi}) \,\|\, p(\vec{\psi})\big) \tag{18}$$

where S^{obs} is the observed data, $\vec{\psi}$ are the model inferred parameters (for the priors and noise), and $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback–Leibler divergence. The first term encourages the variational distribution to explain the observed data well, while the second term penalizes deviations from the prior.

Using stochastic gradient descent, SVI updates the variational parameters $\vec{\phi}$ iteratively, making it well suited for high-dimensional problems like ours.

Intuitively, maximizing the ELBO corresponds to minimizing the KL divergence between $q_{\vec{\phi}}(\vec{\psi})$ and the true posterior $p(\vec{\psi} \mid S^{obs})$, ensuring that the variational distribution approximates the posterior as closely as possible.

2.2.4. Computational aspects

The probabilistic model was implemented in NumPyro (Phan et al., 2019), a probabilistic programming library (PPL) built on top of JAX (Bradbury et al., 2018). JAX is a Python library that enables automatic differentiation, GPU/TPU acceleration and Just-in-Time compilation of the models.

All the cases run on a NVIDIA RTX A6000 GPU with 48 GB of dedicated memory. Using half of the dedicated memory, the average speed for solving the single peak cases was about 400 iterations per second. As for the multi peak cases (7 energy peaks), the speed was about 100 iterations per second. This resulted in a total runtime of about 5 minutes for the single peak case and about 20 minutes for the multi peak case. There was no significant increase of speed using more memory than 24 GB.

After substantial testing, using a Normal variational family was deemed to yield the most stable solutions. The chosen ELBO implementation was *TraceELBO*, calculated with 10 particles. The chosen optimizer was Adam (Kingma and Ba, 2014) with a learning rate of 10^{-3}. All the cases were run for 120,000 iterations. All cases converged according to a visual assessment of the ELBO evolution over iterations.

2.3. Simulations

The process conceived for testing this approach is based on MC simulations of mock-up drums to obtain a realistic estimate of $s_{n,m,e}$ for a given drum matrix and source distribution, and on LAC values (μ_p^e) for the matrix obtained from the NIST database (Hubbell and Seltzer, 2004).

Since LAC values can be obtained through various measurement techniques, we chose to first address the uncoupled high-resolution distributed activity inference, keeping the full problem including derivation of LAC values from a

separate series of measurements, for a later stage. The MC simulation procedure is to be repeated for each given combination of vertical and angular position of the detector (n, m) and simulated photopeak (e) to assemble $s^{obs}_{n,m,e}$.

2.3.1. Monte Carlo model for ASGS

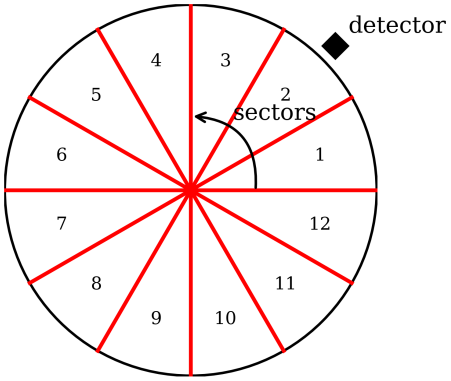
The simulations employed to test the methodology are cases simulated in the PHITS MC code (Sato et al., 2018). The same full ASGS process was simulated for all the synthetic drums using as a detector the model used for training the surrogate model of the efficiency. This model represents a collimated HPGe p-type detector, more information on its design can be found in Casas-Molina et al. (2024).

The detector is placed in such a way that its window is 45 cm away from the waste drum axis, and 180 positions are tracked in total, i.e., 15 vertical segments (n) and 12 angular sectors (m). The collimator presents a 60 mm aperture. The drum is 88.5 cm tall and 57.4 cm wide, therefore, the scanned segments are about 6 cm high and cover a sector of 30°. The drum casing is modeled as steel, with a thickness of 1 mm. For the different simulations, only the energy of the emitted photons is changed, all other simulation parameters remained unchanged. Fig. 6 depicts the simulated arrangement for ASGS in PHITS.

2.3.2. Simulated observations in ASGS

The observed data are the counts per sector, $s^{obs}_{n,m,e}$. In our synthetic approach, these counts are derived from the tally of a MC simulation, used to estimated the Net Peak Area (NPA) for a given energy line, e.

The region to be tracked is the active volume of the detector crystal. In this region, the tracked score is the deposited energy by the photons emitted within the drum, that causes electron-hole pairs to be generated within the active Ge crystal. The energy discretization of the tally was fixed to 150 equally-sized bins. These groups are analogous to the channels of a Multi Channel Analyzer (MCA) used in gamma spectrometry. The NPA is derived by subtracting the



	176	177	178	179	
	164	165	166	167	
	152	153	154	155	
	140	141	142	143	
	128	129	130	131	
	116	117	118	119	
	104	105	106	107	segments
	92	93	94	95	
	80	81	82	83	
	68	69	70	71	
	56	57	58	59	
	44	45	46	47	
	32	33	34	35	
	20	21	22	23	
	8	9	10	11	

Figure 6: Dihedric depiction of ASGS simulation, with 15 vertical segments and 12 angular sectors, the numbering for the resulting discretization, i.e., detector position indexes, can be observed in the frontal view.

background counts (B) from the gross counts (G), following the Covell method (Gilmore, 2008) across all the energy groups (n_G=150).

$$NPA = G - B = \sum_{i=L}^{U} C_i - \frac{n_G}{2m_G}\left(\sum_{i=L-m_G}^{U} C_i + \sum_{i=U+1}^{U+m_G} C_i\right) \tag{19}$$

The Covell method estimates the background using the counts (C) in groups of channels located at the upper and lower energy edges of the peak region.

Given that for the bins away from the center of the peak the count statistics are poorer due to scarce counts, the upstream and downstream values for calculating the peak are the average of the counts of the channels (m_G) located in the Upper edge (U) and the Lower edge (L) of the region of interest. For this case, using the four farthest channels ($m_G = 4$) for the average calculation was found to provide a good estimation of the background counts. Fig. 7 provides an illustrative description on the Covell method.

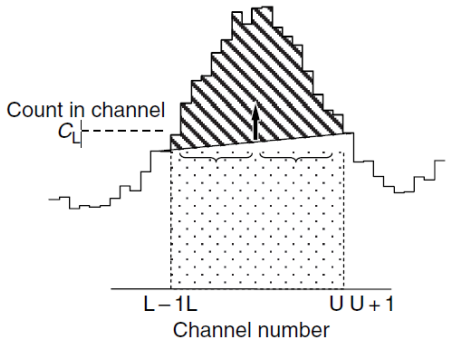


Figure 7: Calculation of peak area with the Covell method. Figure taken from Gilmore (2008).

As an example, a simulated ASGS count matrix for a hot spot of ^{137}Cs in a drum is shown in Fig. 8.

As observed, because the simulated drum matrix is homogeneous, the ASGS simulations can approximately locate the hotspot 20 cm above the center of the drum, at an angle of roughly 50° clockwise relative to the coordinate origin.

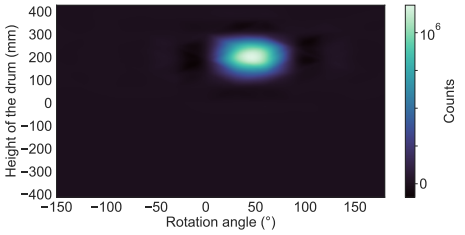


Figure 8: Example of the count matrix assembled from the simulation of ASGS performed on a drum with a hot spot of ^{137}Cs using a homogeneous filler.

2.3.3. Synthetic drums

The considered synthetic drums include three possible source distributions: (1) a homogeneously distributed activity within the drum, (2) all the activity lies in one voxel of size 1.92 cm, and (3) an activated, radioactive cylinder of 1 cm of radius and length of 19 cm embedded in the drum matrix.

We consider all three distributions to be representative of real life cases: the homogeneous source is representative of those drums filled with radioactive material that has little variation in activity across the drum, the all-in-one voxel scenario represents cases where the hot spots to be identified are very small and can be considered as rare events. The last scenario is representative of objects like pipes and tools that are activated or contaminated.

The sources were placed to avoid any kind of symmetry with respect to the detector positions of the drum, i.e., both the voxel and cylinder sources are placed out of the drum axis and the cylinder is rotated by 38° with respect to one of the horizontal axes (X) and by 10° with respect to the vertical axis (Y).

The assumed drum matrix is the same for the three cases: it represents a 220 liter drum filled with sand. We did not take into account the packing effect of the gravity in the sand, so the density was set to be constant all over the drum, i.e., 1.52 g/cm^3. The sand composition values for the MC simulations were obtained from Mcconn et al. (2011) for the reference "Sand", the drum wall was assumed to be 1 mm thick and fabricated in steel with a density of 7.82 g/cm^3, its composition was obtained from (Mcconn et al., 2011) for the

reference "Steel, Carbon".

The LAC values (μ_p^e) for both materials were derived based on their elemental composition by summing of the LAC values of the elements alone, weighted by their proportions in weight; according to the procedure described in the NIST database (Hubbell and Seltzer, 2004) for obtaining LAC from compounds and mixtures. The voxels were then initialized to the volumetric average value of the materials they contained. A representation of the drum matrix can be observed in Fig. 9

Regarding the simulated RIs, ^{137}Cs, and ^{152}Eu isotopes have been considered in each scenario so both single peak and multi peak isotopes can be tested in the approach. In total, eight simulations were conducted, i.e., four source positions with one of the two studied RIs at a time. A comprehensive view of the different source distributions is provided in Table 3:

Table 3: Source distributions for synthetic mockups.

Source Geometry	^{137}Cs **(GBq)**	^{152}Eu **(GBq)**
(S1) Homogeneously distributed	3	5
(S2) Hot spot at [153.6, 201.6, 153.6] mm	3	3
(S3) Hot spot at [96.0, 201.6, -115.2] mm	5	5
(S4) Rod with center at [-27.5, 0, 68.0] mm	3	9

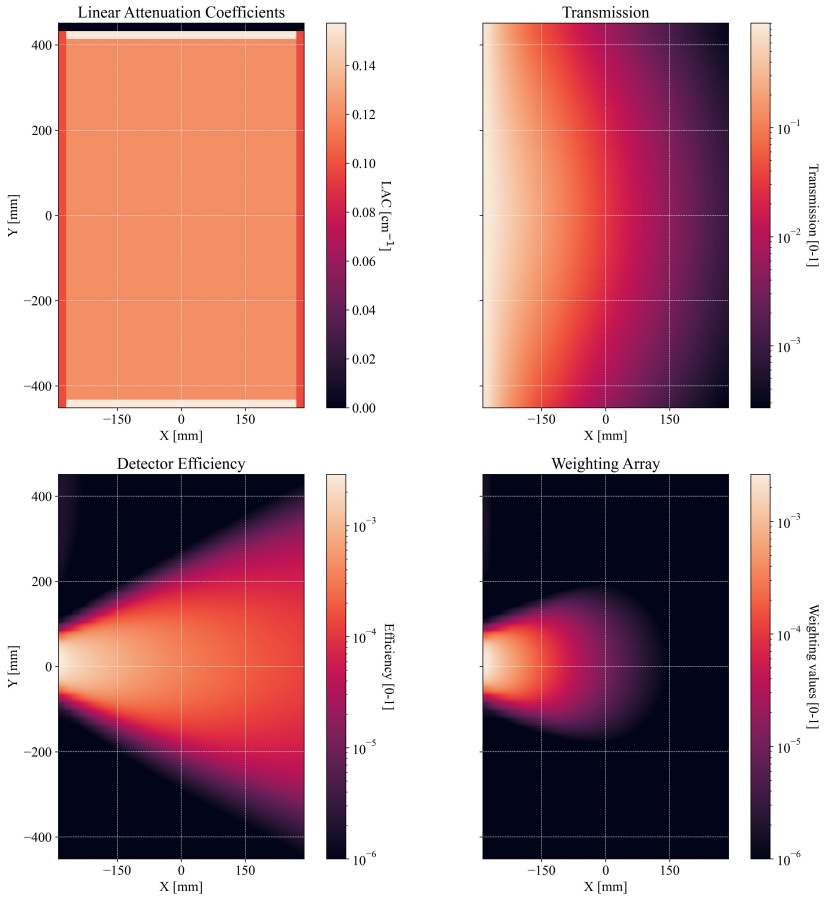


Figure 9: Cross section of the drum showing the LAC, Transmission, Detector efficiency, and Weighting array per voxel for the energy line of ^{137}Cs (661.7 keV) and for the detector located in the middle of the drum as an example. The Weighting array is the product of Detector efficiency array and the Transmission array. Note that because of averaging, the drum walls are presenting slightly smaller values due to the surrounding air while the bottom plate and drum lid cause the LAC values to increase.

3. Results

3.1. Single peak inference

3.1.1. Diagnostics

In Bayesian inference, the ability of the calibrated model to reproduce the observed data is commonly verified with posterior predictive checks, where posterior predictions are simulated from the likelihood and compared to the observations. Fig. 10 presents the posterior predictive distribution for the single peak inference.

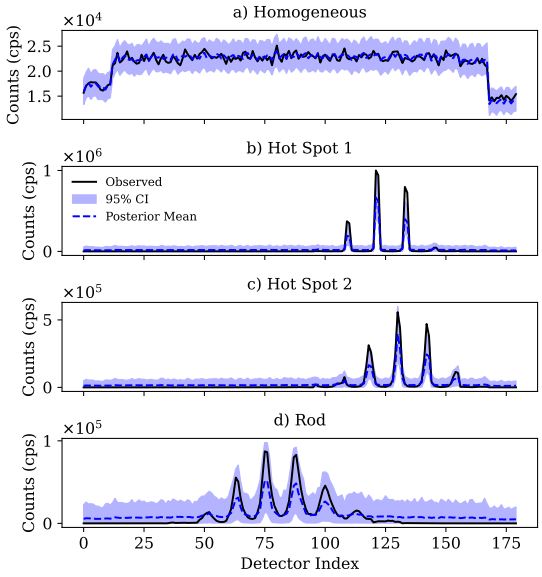


Figure 10: Posterior predictive check for ^{137}Cs cases. Detector indexes match with those of Fig. 6.

The mean posterior predictive counts from the homogeneous case (Fig. 10a.) are in close agreement with the observations, the 95% credible interval covers all the observed points. Therefore, it can be said that the model explains the data well, even though the posterior uncertainty seems to be slightly overestimated.

However, this is not the case for the Hot spots (Fig. 10b. and c.) where the large count rates are underestimated by the model. The high uncertainty

assigned to low-count voxels may indicate that the inference activates voxels that are actually inactive, exploiting additional fitting freedom to better match the observed counts. A similar underestimation of the large count rates occurs for the rod case (Fig. 10d.), but now with a much larger posterior uncertainty in the model predictions.

In general, posterior predictive checks for the hot spots and the rod indicate that the model is unsure about how to explain low signal regions given the high uncertainty. Additionally, an underfitting can be noticed for the high count peaks. This indicates that the model is unable to fit the signal properly.

3.1.2. Spatial distribution

Fig. 11 displays a series of cross sections corresponding to the mean of the inferred posterior activity samples by SVI. The top sections of Fig. 11 depict horizontal cross section of the drum, while the bottom part shows the vertical cross sections for the Homogeneous, Hot Spot 1, Hot Spot 2 and Rod cases.

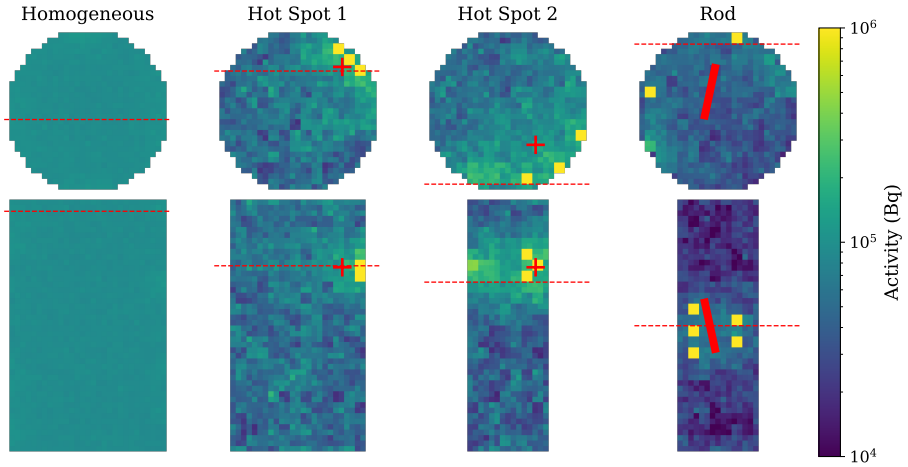


Figure 11: Cross sections for ^{137}Cs. Dashed red lines indicate the maximum values found within that cross-section and show the corresponding slice of the opposite cross-section. Red markers represent the ground truth, with crosses highlighting the hot spots, and a depiction of the rod position for the rod case.

As can be observed, for the single energy peak inference, the Homogeneous case is the only case where the approach performs reasonably well. For the other synthetic drums the inference seems to assign the activity to a handful of separated voxels. They all tend to be located near the edge of the drum, and many are in fact located just in between detector positions. It can also be noticed that the vertical position is more accurately retrieved than the "radial" or horizontal position within the drum. This is not surprising since the detector collimator favors this phenomenon, i.e., the angular aperture of the collimator allows the detector to see a wide but flat slice, which reduces horizontal and increases vertical effective resolution.

3.1.3. Total recovered activity

Fig. 12 presents the total inferred activity for single peak mock-ups. For comparison, we have included an 'Averaged' measurement of the activity. This Averaged activity is obtained in a way that mimics the conventional way of assessing waste drums, where the counts and the density of the matrix are averaged and then, used to yield an activity estimate. As discussed in the introduction, this can result in strong biases given that it does not take into account the spatial variability. Fig. 12 displays the results of the total recovered activity for the single peak mock-ups.

It can be seen that the truth is properly included in the posterior for the Homogeneous and Hot Spot 1 cases. In contrast, the posteriors for the Hot Spot 2 and Rod cases are very underestimated, which is consistent with Fig. 11: the model is putting activity in the outer region where it is more sensitive to counts, thereby needing less activity than the actual one to fit the observations.

3.2. Multi peak inference

3.2.1. Diagnostics

Fig. 13 displays the posterior predictive distributions associated with the multi-energy peak inferences. Here we can observe a different behavior with respect to Fig. 10.

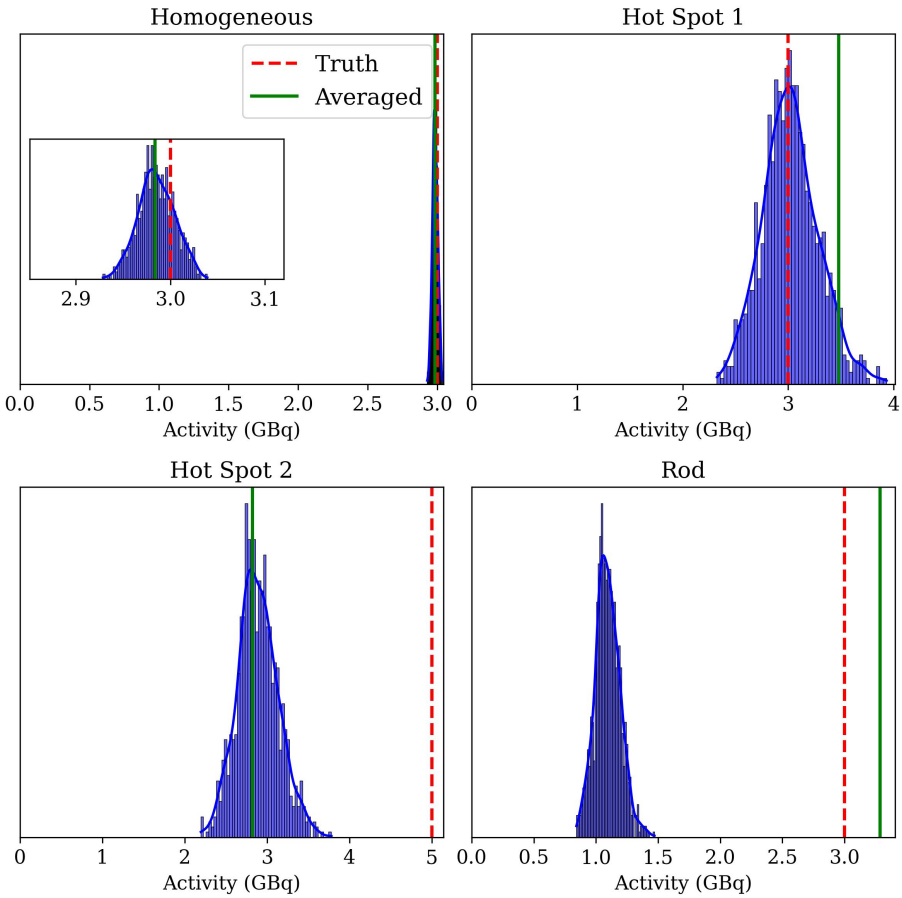


Figure 12: Recovered activity for the ^{137}Cs. The truth, shown in red (dashed), represents the actual activity content of the radionuclide. The averaged estimate, shown in green (solid), corresponds to the result obtained using the conventional method of averaging counts over the estimated detector efficiency and transmission measurements. The posterior distribution reflects the degree of belief produced by the inference regarding the likely activity level and its associated uncertainty.

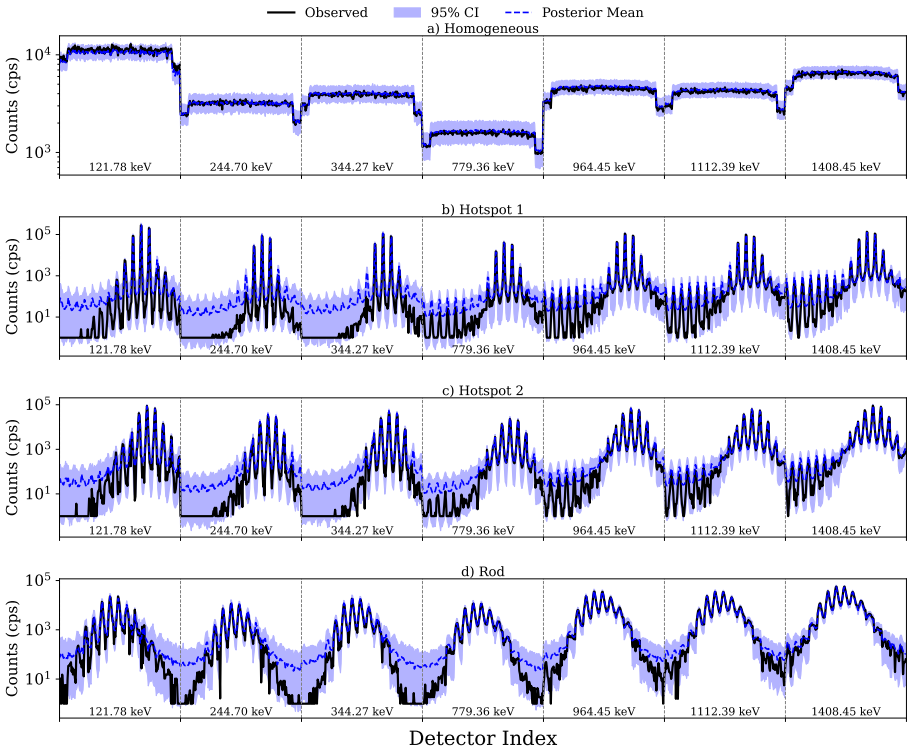


Figure 13: Posterior predictive check for the ^{152}Eu cases. Detector indexes span from 0 to 179 for each energy (same detector positions), but are omitted for visual convenience. Counts are in logarithmic scale.

The posterior predictive distribution indeed resembles the observed data much better with no excessive uncertainty (note that the y-axis scale is logarithmic when comparing uncertainties). This indicates that our probabilistic model is adequate to properly explain the observed data from the multiple peak cases.

3.2.2. Spatial distribution

Fig. 14 presents a collection of the cross-sections for the cases of ^{152}Eu.

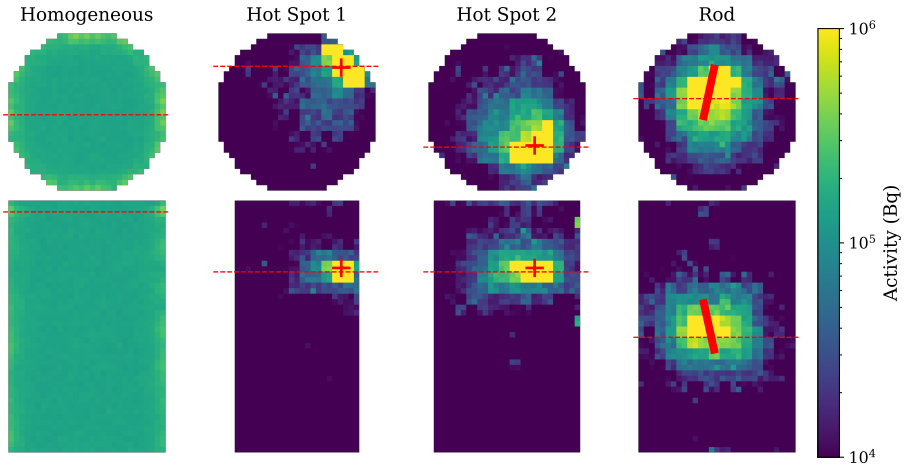


Figure 14: Cross sections for ^{152}Eu. Dashed red lines indicate the maximum values found within that cross-section and show the corresponding slice of the opposite cross-section. Red markers represent the ground truth, with crosses highlighting the hot spots, and a depiction of the rod position for the rod case.

For this multi-peak ^{152}Eu case a much better performance can be noted, as the inference is now able to accurately locate each of the activity spatial distributions for all the cases. It is noticeable that, in the hot spot cases where the source solely occupies a single voxel, many more voxels are deemed to contain activity by the inference. Consequently, the total inferred activity is distributed across more voxels than one as in the "true" drum. Note that Fig. 15 displays the posterior mean activity field and the latter could be a smoothed version of

the actual posterior field realizations. Yet individual posterior field realizations similarly tend to distribute activity over more than one voxel for the hot spot cases (not shown). However, this behavior is not surprising given the small voxel size and the facts that (1) the measured data contain Poisson noise and (2) our count simulation model (Eq. 1) is not error free but is affected by the error associated with the used surrogate detector efficiency model (to predict $(\mathcal{E}_{i,j,k})_{n,m,e}$ in Eq. 2).

3.2.3. Total recovered activity

Fig. 15 presents the results of the total recovered activity for the multiple peak mock-ups.

It is observed that, the truth is properly contained within the posterior distributions for all cases. In general, the inference however tends to underestimate the true values, which may be due to the fact that our sources only occupy very small fractions of the drum volume except for the homogeneous case, which yields a true value centered in a very narrow posterior. Additionally, it is worth noting that here the results from the conventional way of activity averaging are outside of the posterior distribution.

4. Discussion

4.1. On the model performance

Our results show that for the considered synthetic cases, our proposed approach is able to accurately retrieve the spatial activity distribution together with the associated uncertainty, as long as it is dealing with multiple energy peaks simultaneously. Unfortunately, our approach is found to fail when tackling only a single gamma energy peak at a time.

Considering multiple energy peaks simultaneously, with known emission probabilities, leads to a richer set of measured counts which contain enough information to appropriately constrain the inferred activity. More specifically, since peaks of different energies are attenuated in a different way, considering

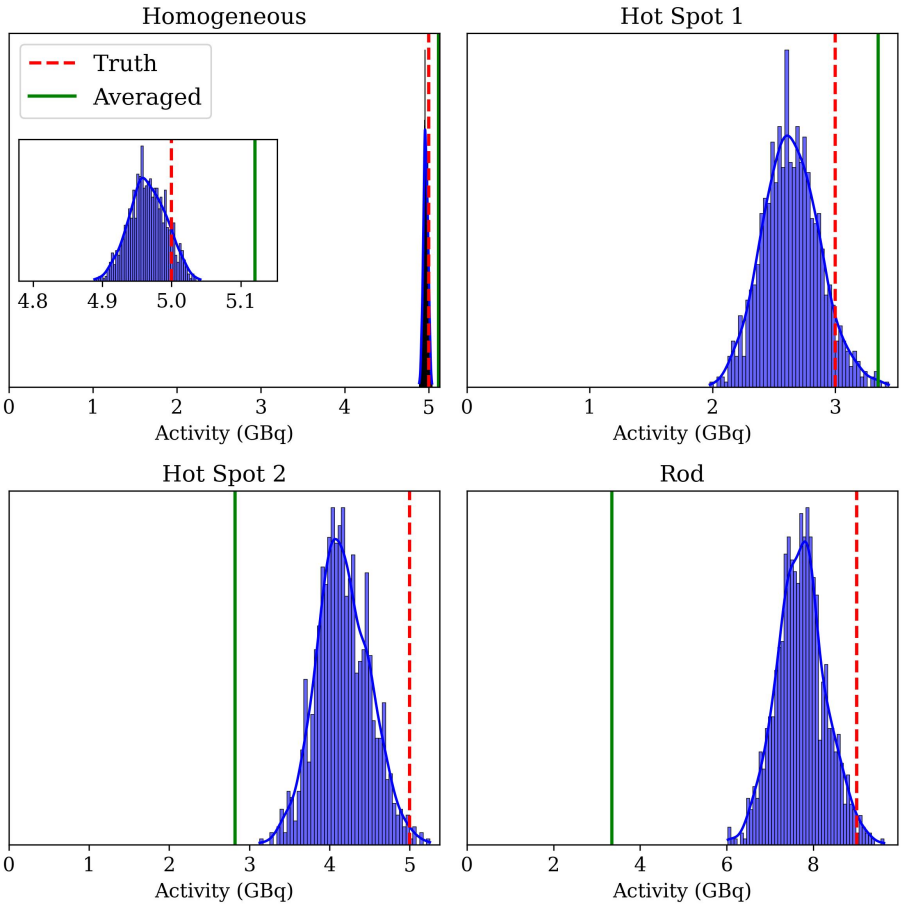


Figure 15: Recovered activity for the ^{152}Eu. The truth, shown in red (dashed), represents the actual activity content of the radionuclide. The averaged estimate, shown in green (solid), corresponds to the result obtained using the conventional method of averaging counts over the estimated detector efficiency and transmission measurements. The posterior distribution reflects the degree of belief produced by the inference regarding the likely activity level and its associated uncertainty.

them jointly, or their ratio, provides information on the scale of attenuation and detection efficiency, and hence helps constraining the position of the source(s) within the drum. For our ^{137}Cs cases, the fact that this information is not present, results in non-identifiability, as the signature of any activity deeper in the drum can be mimicked by lower activities closer to the drum walls. Even with the rather large amount of ASGS measurements (180 herein), such non-identifiability remains. Given that the effective prior distribution placed on the total activity of the drum is approximately log-uniform (see Fig. 4), this non-identifiability results in underestimation, placing source positions too close to the high-efficiency region near the drum wall.

As a verification measure, we conducted the following experiment for Hot Spot 1 of ^{137}Cs: a second mask was added to force the inference to place activity only in the true voxel. In this case, the results are perfect, showing a perfect posterior predictive check and retrieving an average measurement of 3 GBq (the ground truth).

The free, unconstrained configuration returned an ELBO of about 12,489, whereas the restricted configuration returned an ELBO of about 11,427. This suggests that the inference should recognize the restricted (ideal) configuration as a better solution.

To further compare the two configurations, we drew posterior samples $\vec{\psi}^{(s)} \sim q_{\vec{\phi}}(\vec{\psi} \mid S^{obs})$ from the variational guide learned through SVI, and then computed their log posterior density and log prior density values. Analyzing both configurations confirmed that the issue does not stem from the prior density, since both configurations are feasible a priori.

These observations (ELBO and log densities) suggest that here SVI may converge to a local optimum. Moreover, the information provided by the observations for a single peak is not sufficient to steer SVI towards the correct solution.

The fact that our proposed approach requires the detection of multiple energy peaks to work should not pose much of a challenge in real life applications, where actual waste drums usually contain a rich isotopic vector. If this vector

can be assumed to be constant for all the non-zero activity regions in the drum, then different peaks can be jointly fitted since they are coming from the same species. Furthermore, even if an isotopic vector would be uncertain, it could likely be jointly inferred along with the activity, (see Laloy et al. (2023a,b, 2024)), and still help in constraining the location of the source(s).

Another question concerns the slightly underestimated total activity of the drum (Fig. 15) in the multiple-peak case. This bias can be explained by the overspread of active voxels selected by the inference, which effectively changes the efficiency of the source. In fact, our prior probability of having all significant activity concentrated in a single voxel is very low, and the evidence in the data is insufficient to overcome this low prior probability. This can be verified by running a case in which activity is allowed only in the true voxel: for a ground-truth total activity of 3 GBq, the posterior peaks at 3.4 GBq, indicating that the posterior overestimates rather than underestimates the total activity in this constrained scenario.

Similarly, analyzing the Hot Spot 1 results for ^{152}Eu, we observe that the SVI samples with the highest activity values consistently correspond to those with more active voxels (i.e., containing a significant fraction of the maximum voxel activity). This suggests that the spatial extent of the inferred activity blob plays a key role in the bias observed in the posterior for total activity. Nonetheless, the true total activity for the multi-peak case is still properly captured within the posterior distribution.

Additionally, it should not be forgotten that to be computationally efficient, SVI makes the necessary assumption that the posterior distribution is Gaussian, which might not be the case in reality. If the true posterior differs significantly from the Gaussian distribution, then this deviation might create a significant bias in the SVI approximation.

Finally, we rely here on a priori knowledge to set the different parameters for the Beta distribution to discriminate between hot spots and homogeneous cases which can be problematic; nonetheless, peeking at the observed signal (Data-dependent scaling) may help the operator in selecting the proper parameter

configuration and that should be relatively straightforward given the differences in signal (please, compare observed signals in Fig. 10a. against b., c. and d.).

4.2. On error modeling

Accounting for measurement noise and epistemic uncertainty of the model is perhaps the most critical aspect to succeed in solving the inverse problem.

In this approach we resorted to inflating the likelihood by inferring a heteroscedastic noise scale that multiplies the standard deviation of the Normal approximation to the Poisson distribution (Eq. 9). We consider this to be a flexible approach that has been proven to work for the multiple peak case.

Attempts to propagate the uncertainty of the detector's efficiency by letting the inference adjust the $(\mathcal{E}_{i,j,k})_{n,m,e}$ term of the weighting term $(w_{i,j,k})_{n,m,e}$ have been unsuccessful. This is to be expected, because $(w_{i,j,k})_{n,m,e}$ multiplies the activity to be inferred (Eq. 1), which introduces a non-identifiability problem. Two strategies were explored: first, a single scalar multiplier was assigned to $(w_{i,j,k})_{n,m,e}$ under a positively truncated Normal prior, so that only one parameter needed to be inferred; second, MC dropout was applied to the neural network to obtain predictive uncertainty for each element of $(w_{i,j,k})_{n,m,e}$ to use them as standard deviations for positively truncated Normal and Multivariate Normal distributions to infer $(w_{i,j,k})_{n,m,e}$. Both approaches failed to properly account for the efficiency surrogate model error on the recovered posterior activity.

In general, we would like to point out that the error modeling is a crucial step in this inverse problem solution and that accounting for it in the likelihood is the only solution that appeared to work. Analytical propagation of uncertainty through the forward model assuming independency and normality in $(w_{i,j,k})_{n,m,e}$ elements was not satisfactory. We also remark that the forward model is very sensitive to the accuracy of the surrogate model that yields $(\mathcal{E}_{i,j,k})_{n,m,e}$, which is the only estimated element of $(w_{i,j,k})_{n,m,e}$ (Eq. 2). However, obtaining a (nearly) perfect surrogate model is not possible. All of these findings highlight the need of finding a way to appropriately model the forward

model error.

Finally, note that there are sources of geometric error that are not taken into account here as previously stated in Subsection 2.1. Those can cause some deviation from the truth as already pointed out in Bücherl et al. (2021), such as the rendering of the drum wall, the loss of information occurring during down-sampling from the LAC image to the voxel grid, the masked zone inside the drum, etc. However, we consider these errors to be minor in comparison with the weighting term $(w_{i,j,k})_{n,m,e}$, and therefore, we checked different masks, analytical addition of the attenuation of 1 mm of steel to avoid dealing with the drum wall, and checked that averaged values of the voxel grid were equivalent to the weighted average of what they were containing. After all these tests, we did not observe any significant difference on the results for the considered application.

5. Conclusion

In this work, we presented a new proof of concept for efficiently reconstructing high-resolution 3D activity distributions in radioactive waste drums containing multiple energy-emitting sources. The method combines conventional ASGS measurements with a Bayesian approach, demonstrating strong potential based on synthetic test cases. These results suggest that the framework can capture complex activity patterns while remaining computationally feasible.

At the same time, several challenges and limitations were identified. Future work will focus on extending the framework toward more realistic scenarios by using accurate linear attenuation coefficient (LAC) values and progressing from synthetic data to mock-up experiments and, ultimately, to real waste drums. This gradual validation will be key to assessing how robust the method is under realistic conditions. Another important direction is improving how the model parameters, particularly $(w_{i,j,k})_{n,m,e}$, are represented, to reduce biases and improve the reliability of the reconstructed activity. It may also be worth exploring approximate posterior recalibration (Cai et al., 2024) to help correct for differ-

ences between the model and real measurements.

While computational cost remains a limiting factor, addressing these aspects will be crucial for improving the method's overall performance. With continued development, the proposed approach could become a practical and reliable tool for non-destructive characterization of radioactive waste, supporting safer and more efficient waste management in the future.

Data availability

To ensure reproducibility, all source code and test data have been provided through the *BayRad3D* GitHub repository (Casas-Molina, 2025).

Acknowledgments

This work was funded by the Recovery and Resilience Facility (RRF) of the European Union under the NextGenerationEU program as part of the ANUBIS project [Advancing NUclear dismantling in Belgium through Improving Sustainability], by the Flemish Government under the 'Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen' and by the Research Foundation—Flanders (FWO) (grant number G0A2824N).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that influenced the work reported in this paper.

CRediT author statement

Victor J. Casas-Molina: Conceptualization, Methodology, Data Curation, Software, Validation, Visualization, Formal Analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing. **Eric Laloy:** Conceptualization, Supervision, Project administration, Methodology, Resources, Writing - Review & Editing. **Bart Rogiers:** Conceptualization, Supervision, Project administration, Methodology, Writing - Review & Editing. **Ivo Couckuyt:** Supervision, Project administration, Writing - Review & Editing. **Tom Dhaene:** Supervision, Project administration, Writing - Review & Editing.

References

T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

H. Aksoy. Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences*, 24, 01 2000.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2014.

J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL `http://github.com/jax-ml/jax`. Version 0.3.13.

M.-M. Bé, V. Chisté, C. Dulieu, E. Browne, C. Baglin, V. Chechev, N. Kuzmenko, R. L. Helmer, F. Kondev, and T. D. Macmahon. *Table of radionuclides (Vol. 3 - A = 3 to 244)*, volume 3. Bureau international des poids et mesures (BIPM), 2006. URL `https://cea.hal.science/cea-02476243`.

T. Bücherl, S. Rummel, and O. Kalthoff. A Bayesian method for the evaluation of segmented gamma scanning measurements – Description of the principle. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1019:165887, Dec. 2021. ISSN 0168-9002. doi: 10.1016/j.nima.2021.165887. URL https://www.sciencedirect.com/science/article/pii/S0168900221008676.

T. Cai, P. Greengard, B. Goodrich, and A. Gelman. Approximate posterior recalibration, 2024. URL https://sites.stat.columbia.edu/gelman/research/unpublished/Approximate_posterior_calibration.pdf. Accessed via Google Scholar.

C. Carasco. Coupling gamma ray spectrometry and tomography in a bayesian frame. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 990: 164985, 2021. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2020.164985. URL https://www.sciencedirect.com/science/article/pii/S0168900220313826.

V. J. Casas-Molina. BayRad3D: 3d activity reconstruction from angular gamma scanning via variational bayes. https://github.com/vcasasmo/BayRad3D, 2025. Accessed: 2025-09-30.

V. J. Casas-Molina, E. Laloy, B. Rogiers, T. Dhaene, and I. Couckuyt. Bayesian optimization of a collimated HPGe detector model for Segmented Gamma Scanning. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1067:169687, Oct. 2024. ISSN 0168-9002. doi: 10.1016/j.nima.2024.169687. URL https://www.sciencedirect.com/science/article/pii/S0168900224006132.

F. Chollet et al. Keras. https://keras.io, 2015. GitHub repository: https://github.com/keras-team/keras.

R. J. Estep, T. H. Prettyman, and G. A. Sheppard. Tomographic gamma scanning to assay heterogeneous radioactive waste. *Nuclear Science and Engineering*, 118(3):145–152, 1994. doi: 10.13182/NSE94-A19380.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 3rd edition, 2013. ISBN 9781439840955. See Chapter 5: Hierarchical Models, especially Section 5.1 on the parameterized prior distribution, Logic of combining information and empirical Bayes.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1995.

G. Gilmore. Statistics of Counting. In *Practical Gamma-Ray Spectrometry*, pages 101–129. John Wiley & Sons, Ltd, Apr. 2008. ISBN 978-0-470-86198-1. doi: 10.1002/9780470861981.ch5. URL https://doi.org/10.1002/9780470861981.ch5.

M. D. Hoffman and A. Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1593–1601, 2014.

M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

H. Hubbell, J. and M. Seltzer, S.˙ Tables of x-ray mass attenuation and mass energy-absorption coefficients, version 1.4. NIST Standard Reference Database 126, July 2004. Online. Available: https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients (accessed 2025-07-05).

International Atomic Energy Agency. Addressing challenges in managing radioactive waste from past activities. Technical Report IAEA-TECDOC-2039, International Atomic Energy Agency, Vienna, Austria, 2024. URL https://www-pub.iaea.org/MTCD/Publications/PDF/TE_2039web.pdf. Accessed: 2025-07-05.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

E. Laloy, B. Rogiers, A. Bielen, and S. Boden. Bayesian inference of 1D activity profiles from segmented gamma scanning of a heterogeneous radioactive waste drum. *Applied Radiation and Isotopes*, 175:109803, Sept. 2021. ISSN 0969-8043. doi: 10.1016/j.apradiso.2021.109803. URL `https://www.sciencedirect.com/science/article/pii/S0969804321002062`.

E. Laloy, B. Rogiers, A. Bielen, A. Borella, and S. Boden. Improving bayesian radiological profiling of waste drums using dirichlet priors, gaussian process priors, and hierarchical modeling. *Applied Radiation and Isotopes*, 194:110691, 2023a. doi: 10.1016/j.apradiso.2023.110691.

E. Laloy, B. Rogiers, A. Bielen, A. Borella, W. Kubiński, C. Carasco, C. Mathonat, and S. Boden. Probabilistic radwaste characterization: Findings of a multi-method multi-mockup exercise using interpolation-based surrogate efficiencies. *Annals of Nuclear Energy*, 194:110065, 2023b. doi: 10.1016/j.anucene.2023.110065.

E. Laloy, B. Rogiers, A. Bielen, A. Borella, G. Gandolfo, L. Lepore, G. Marzo, N. Cherubini, B. Perot, Q. Ducasse, C. Eleon, and S. Boden. First steps towards productionizing probabilistic radwaste characterization. *Nuclear Engineering and Design*, 424:113257, 2024. doi: 10.1016/j.nucengdes.2024.113257.

S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 2009.

R. Mcconn, C. Gesh, R. Pagh, R. A. Rucker, and R. Williams. Compendium of Material Composition Data for Radiation Transport Modeling. Technical report, Pacific Northwest National Laboratory, Jan. 2011.

H. P. Pham, X. H. Nguyen, D. H. Nguyen, X. T. Nguyen, Q. D. Ong, C. T. Phan, N. S. Pham, H. T. Ho, Q. D. Tran, and V. N. Ho. Research method for assay activity and activity distribution of gamma emitting radionuclides in the

model waste drum by tomography gamma scanning technique. *VINATOM Annual Report*, pages 232–238, 2013.

D. Phan, N. Pradhan, and M. Jankowiak. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554*, 2019.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

T. Sato, Y. Iwamoto, S. Hashimoto, T. Ogawa, T. Furuta, S.-i. Abe, T. Kai, P.-E. Tsai, N. Matsuda, H. Iwase, N. Shigyo, L. Sihver, and K. Niita. Features of Particle and Heavy Ion Transport code System (PHITS) version 3.02. *Journal of Nuclear Science and Technology*, 55(6):684–690, June 2018. ISSN 0022-3131. doi: 10.1080/00223131.2017.1419890. URL `https://doi.org/10.1080/00223131.2017.1419890`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00223131.2017.1419890.

I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. doi: 10.1016/0041-5553(67)90144-9.

R. Venkataraman, M. Villani, R. T. Kouzes, S. D. Strom, and C. S. Bryan. Validation of in situ object counting system (isocs) mathematical efficiency calibration software. *Nuclear Instruments and Methods in Physics Research Section A*, 422:450–454, 1999. doi: 10.1016/S0168-9002(98)01099-9.