I
MPROVING BAYESIAN RADIOLOGICAL PROFILING OF WASTE DRUMS USING DIRICHLET PRIORS, GAUSSIAN PROCESS PRIORS, AND HIERARCHICAL MODELING

Eric Laloy∗ Bart Rogiers An Bielen Alessandro Borella Sven Boden

Institute for Environment, Health and Safety, Belgian Nuclear research Centre (SCK CEN)

December 8, 2022

ABSTRACT

We present three methodological improvements of the “SCK CEN approach” for Bayesian inference of the radionuclide inventory in radioactive waste drums, from radiological measurements. First we resort to the Dirichlet distribution for the prior distribution of the isotopic vector. The Dirichlet distribution possesses the attractive property that the elements of its vector samples sum up to 1. Second, we demonstrate that such Dirichlet priors can be incorporated within an hierarchical modeling of the prior uncertainty in the isotopic vector, when prior information about isotopic composition is available. Our used Bayesian hierarchical modeling framework makes use of this available information but also acknowledges its uncertainty by letting to a controlled extent the information content of the indirect measurement data (i.e., gamma and neutron counts) shape the actual prior distribution of the isotopic vector. Third, we propose to regularize the Bayesian inversion by using Gaussian process (GP) prior modeling when inferring 1D spatially-distributed quantities. As of uncertainty in the efficiencies, we keep using the same stylized drum modeling approach as proposed in our previous work to account for the source distribution uncertainty across the vertical direction of the drum. A series of synthetic tests followed by application to a real waste drum show that combining hierarchical modeling of the prior isotopic composition uncertainty together with GP prior modeling of the vertical Pu profile across the drum works well. We also find that our GP prior can handles both cases with and without spatial correlation. Of course, our GP prior modeling framework only makes sense in the context of spatial inference. Furthermore, the computational times involved by our proposed approach are on the order of a few hours, say about 2, to provide uncertainty estimates for all variables of interest in the considered inverse problem. This warrants further investigations to speed up the inference.

1 Introduction

Different variants of Bayesian data inversion or inference (see, e.g., Gelman et al., 2014) have been recently introduced in the field of radiological characterization, independently by different groups (Carasco, 2021; Laloy et al., 2021; Buecherl et al., 2021; Clément et al., 2021). All these approaches resort to Markov chain Monte Carlo (MCMC) sampling to derive the posterior activity or, equivalently, mass distribution of one or more radionuclides contained in the considered object, from radiological measurement data (e.g., gamma measurements). The “SCK CEN approach” proposed by Laloy et al. (2021) has the following distinctive features: (i) it relies on an efficient MCMC algorithm called Hamiltonian Monte Carlo (HMC), (ii) accounts for two important sources of uncertainty, namely the measurement uncertainty and the uncertainty in the source distribution within the drum, and (iii) formulates an efficiency model that permits spatially-distributed inference along the vertical direction of the drum. In this work, we significantly improve upon this approach in the following four ways.

∗Corresponding Author, eric.laloy@sckcen.be
First, we consider prior distributions that allow to work with a total Pu mass and associated isotopic vector directly. This permits accounting for more prior knowledge and results into a more realistic inference than when considering the activities of the Pu isotopes to be independent from each other. Our joint inference of total Pu mass and associated isotopic vector is done by putting a Dirichlet prior distribution on the isotopic vector.

Second, we regularize the inferred spatial distribution of each considered (Pu-related) nuclide by using a Gaussian process (GP) prior for the spatially distributed total Pu mass (e.g., Rasmussen and Williams, 2006). The kernel parameters of the GP are then jointly inferred with the other variables. By regularizing the inverse solution, this GP prior largely improves computational tractability of the inference process.

Third, we use Bayesian hierarchical modeling (also called Bayesian multilevel modeling, see, Gelman et al., 2014) to account for the uncertainty in the parameters of the Dirichlet prior that is put on the isotopic vector. Doing so allows the isotopic vector to vary spatially (here between horizontal drum segments), while still pooling information across the considered spatial locations (drum segments).

Lastly, we combine two measurement techniques for a more accurate radiological characterization. Hence, we combine segmented gamma scanning (SGS) with passive neutron coincidence counting (PNCC, Borella et al., 2021). This permits inference of both gamma emitting and non-gamma emitting nuclides, albeit not spatially for the non-gamma emitting nuclides. This because the PNCC technique provides a single measurement for the whole drum, of which the spatial information content is too low for resolving the nuclides’ spatial distribution.

To the best of our knowledge, this is the first application of Dirichlet priors, Gaussian process priors and Bayesian hierarchical modeling, respectively, to Bayesian radiological characterization. Furthermore, the combination of gamma measurements with neutron coincidence counting (and other measurement techniques) within a Bayesian framework is a key task of the EU-funded CHANCE project and is new to this study and the companion CHANCE study of Carasco et al. (under review).

The remainder of this paper is organized as follows. Section 2 describes the considered waste drum and the measurement data, while sections 3 and 4 detail the various aspects of our MCMC-based Bayesian uncertainty quantification approach. Section 5 then presents and discusses the obtained spatially distributed radionuclide inventory for both a synthetic and a real case, before section 6 summarizes our main findings and outlines some future developments.

2 Measurements

2.1 Segmented gamma scanning

The segmented gamma scanning (SGS) system used in this work to quantify the gamma-emitters present in a radioactive waste drum is the same as the system used by Laloy et al. (2021). For brevity we refer the reader to that study for a full description.

2.2 ISOCS measurement

An open geometry ISOCS gamma measurement with fixed detector position was performed (Boden, 2021) in addition to the 3AX-SGS measurement. In this work, this open geometry gamma measurement only serves to derive preliminary estimates of the isotopic vector components for the whole drum, by application of the FRAM isotopic composition software (Sampson et al., 2003). This software provides estimates of the relative proportions of gamma-emitting Pu isotopes based of their spectrum. It also calculates the relative proportions of $^{241}$Pu, $^{237}$Np and $^{235}$U based on empirical relationships. We would like to stress that our approach does not at all strictly require this open geometry measurement. Here we simply use it to derive a prior estimate for the isotopic vector but using any other source of information (e.g., expert knowledge, measured data from a relatively similar drum, ...) to derive this estimate would be just fine too.

2.3 Passive neutron coincidence counting

Passive neutron coincidence measurements (PNCC) were also carried out on the considered drum. Generally, radioactive materials mainly emit neutrons according to three processes: ($\alpha$, n) reactions, spontaneous fission, and neutron-induced fission. While ($\alpha$, n) reactions always emit only one neutron per event, more than one neutron may be emitted in a fission event. This results in a time correlation between detected events that is used in PNCC to estimate the quantity of material undergoing spontaneous fission in the sample being measured (Dierckx and Hage, 1983). In this study, PNCC measurements were carried out with a transportable system consisting of two identical slab counters equipped with $^3$He detectors each and coupled to shift register electronics. The system is fully described in Borella et al. (2021) and for brevity, we refer the reader to this publication for more information.
2.4 Waste package and measurement setups

We focus on the real non-conditioned 200-liter radioactive waste drum considered in our previous publication on Bayesian radiological characterization (Laloy et al., 2021). This drum is fully described in Laloy et al. (2021) and here too we refer the reader to this publication for more details. Similarly as for our previous study, the drum dimensions, filling degree, matrix composition and matrix density distribution are assumed to be known and kept fixed. However, our approach accounts for the source distribution uncertainty. The source distribution uncertainty is a major source of uncertainty for this type of measurements.

As detailed in section 3.1, the drum is discretized into 20 horizontal segments (Figure 1). The bottom and top segments (S1 & S20) are 23.4 mm thick will the other 18 segments all have a thickness of 46.8 mm (S2 – S19). To each segment corresponds a measurement with the detector directly in front of it. The 3AX-SGS measurement setup details are as follows:

- The detector is surrounded by a rectangular lead slit collimator with a 1 mm copper coating (5° opening; 18 mm x 110 mm; 100 mm depth);
- The 200-liter drum is placed on a turning table and is continuously rotated during the measurements, with a rotation rate of 10 rotations per minute.
- The measured 20 segments result in 20 individual spectra (M1 up to M20) and measurement time is 300 s. For the first measurement, the bottom of the drum is placed in front of the middle of the detector (M1). For each following measurement the detector is raised by 46.8 mm (Figure 1).

Table 1 lists the identified nuclides in the drum (independently by both the 3AX-SGS and open geometry ISOCS measurements) and, for the gamma-emitting ones, their associated photon energy peaks.
Table 1: Studied nuclides and whenever applicable, associated photon energies, whether a FRAM-based prior estimate is available and the FRAM-based isotopic composition estimates together with their relative uncertainties (as provided by the FRAM software). Regarding the FRAM-derived isotopic composition, “spectrum” signifies that the nuclides’ spectrum is used for the estimation while “empirical” means application of an empirical relationship. Furthermore, NA stands for “not applicable” while COMP and RSD denote the FRAM-based estimate and its uncertainty in terms of relative standard deviation. The COMP values are percentages of the total Pu mass for the five Pu isotopes: $^{238}$Pu, $^{239}$Pu, $^{240}$Pu, $^{241}$Pu and $^{242}$Pu. As of $^{241}$Am, $^{237}$Np and $^{235}$U, the COMP values are the considered nuclide to total Pu mass ratios. The relatively low RSD values are associated with a very long measurement time of about 13 hours.

| Nuclide | Energy [keV] | FRAM   | COMP  | RSD [%] |
|---------|-------------|--------|-------|---------|
| $^{241}$Am | 125.3 | spectrum | 0.043 [-] | 0.3 |
|          | 662.4      |        |       |         |
|          | 722.01     |        |       |         |
| $^{244}$Cm | NA         | NA     | NA    | NA      |
| $^{237}$Np | NA         | empirical | 0.002 [-] | 0.5 |
| $^{238}$Pu | 152.72     | spectrum | 0.64 [%] | 0.5 |
|          | 766.39     |        |       |         |
|          | 129.30     |        | 77.69 [%] | 0.3 |
|          | 345.01     |        |       |         |
|          | 375.05     |        |       |         |
|          | 413.71     |        |       |         |
|          | 451.48     |        |       |         |
| $^{240}$Pu | NA         | empirical | 1.80 [%] | 1.1 |
| $^{241}$Pu | 148.57     | spectrum | 1.32 [%] | 0.3 |
| $^{242}$Pu | NA         | empirical | 1.80 [%] | 1.1 |
| $^{235}$U | NA         | empirical | 0.074 [-] | 1.9 |

For the PNCC measurement, we consider the same drum discretization and treatment of the source distribution uncertainty as for the SGS measurement (Figure 1).

3 Efficiency models

3.1 Efficiency model of the SGS measurement

This subsection is slightly modified from our previous study (Laloy et al., 2021). For our approach to account for the effect of the source distribution uncertainty within the drum on the detector efficiencies, it is required to model the whole measurement system for a limited set of prescribed configurations. To perform this modeling we used the complex cylinder model of the Geometry Composer V4.3 library of the ISOC/LabSOCS software by [ISOC/LabSOCS - Mirion Technologies (Canberra), Inc. (n.d.)]. The following two relatively extreme source distributions are considered (Figure 1):

- Optimistic efficiency hypothesis: each segment contains an homogeneous source distribution (hypothesis h, Figure 1a). This is our chosen upper limit in efficiency for the considered fixed matrix properties.
- Conservative efficiency hypothesis: each segment contains a point source (hypothesis p) that is located on the drum axes on the border of a segment (on the bottom of the segment for S1 up to S10 and on the top of the segment for S11 up to S20, see Figure 1b). This is the minimum efficiency for the considered fixed matrix properties.

The drum dimensions, filling degree, mean matrix composition and density remain constant and fixed in the ISOC/LabSOCS model. In contrast, the source distribution is varied. In case an homogeneous source distribution is modeled, the matrix shielding is minimal for the part of the source close to the detector and increases more towards the center of the drum. For the modeling of a point source located on the drum axes, the matrix shielding thickness is larger than the drum
radius. This because the detector points to the center of a sub-volume but in this case the point source is located at the bottom or top of the sub-volume (Figure 1b). The symmetry between measurements M1-M10 and measurements M11-M20 (Figure 1) allows for duplicating the efficiency calculations for the lower part of the drum to the upper part of the drum, thereby reducing the required amount of calculations by two. An individual model is thus designed for every individual measurement associated with the lower part of the drum (M1 up to M10), with the source located stepwise in segments S1 up to S20 for both the homogeneous (hypothesis h) and point source (hypothesis p) scenarios, respectively. This leads to a total of 400 models (10 detector locations × 20 source locations × 2 source distributions). For both the h and p assumptions, the detector’s efficiency is therefore computed for each configuration of source location (S, the source is located in a given segment while the other 19 segments do not contain any source), detector location (M), and photon energy (among the 12 energies listed in Table 1). These efficiency values are then used to create two \( n_{\text{source}} \times n_{\text{seg}} \) 3D arrays, \( \mathbf{E}_{\text{SGS}} \) and \( \mathbf{E}_{\text{SGS}} \), for the h and p assumptions, respectively. Here \( n_{\text{source}} = n_{\text{seg}} = 20 \) and \( n_{\text{peaks}} = 12 \). Note that an even more favorable or “positively” extreme distribution than hypothesis h consists of a circularly distributed source at a radius close to the drum’s radius. Also our h and p assumptions are relatively extreme source distributions for the considered (fixed) matrix properties only. Other matrices, such as one including one or more self-shielding sources such as fuel pellets for instance, could lead to other extreme situations (see also [Martens and Filss 1990] for other examples). It is worth noting that, in principle, our approach allows for including not just two (h and p) but a few “end-member” (that is, “extreme” case) efficiencies that are considered relevant. This will be explored in future work.

The efficiency \( \epsilon(e) \) corresponding to a given gamma detection at energy \( e \) can be balanced between the \( h \) and \( p \) assumptions for a given segment by using a \( \lambda \in [0, 1] \) coefficient: \( \epsilon(e) = \lambda \epsilon(h)(e) + (1 - \lambda) \epsilon(p)(e) \). This forms the basis of our count simulation model. For a given count rate, \( c_{\text{rate}} \), we have

\[
\epsilon_{i,k}^{\text{rate}} = \sum_{l=1}^{n_{\text{seg}}} a_l^j P_{\gamma,j,k} \left[ \lambda \epsilon_{i,l}^h(e_k) + (1 - \lambda) \epsilon_{i,l}^p(e_k) \right],
\]

where the subscript \( i \) indexes the detector location, \( i = 1, \ldots, 20 \), the subscript \( k \) indexes the considered energy peak, \( k = 1, \ldots, 12 \), the superscript \( j \) indexes the considered nuclide, \( j = 1, \ldots, 5 \), that is responsible for peak \( k \), \( a_l^j \) denotes the activity of nuclide \( j \) at segment location \( l \), and \( P_{\gamma,j,k} \) is the emission probability of nuclide \( j \) at energy \( k \).

The simulated gross count at detector location \( i \) for the energy \( k \), \( c_{i,k}^{\text{gross}} \), is then obtained from

\[
c_{i,k}^{\text{gross}} = c_{i,k}^{\text{rate}} t_m + b_{i,k},
\]

where \( t_m \) is the measurement time and \( b_{i,k} \) is the background continuum count for time \( t_m \) at detector location \( i \) for the energy \( k \).

### 3.2 Efficiency model of the PNCC measurement

As stated earlier, the PNCC setup and simulation model used herein are detailed in [Borella et al. 2021] and a full mathematical development is given in [Borella and Rossa 2022]. We only describe here what is needed to understand our specific inversion approach.

Similarly as for the 3AX-SGS setup, PNCC efficiencies for the different source configurations shown in Figures 1a-b were simulated, thereby leading the creation of two 20-dimensional vectors, \( \mathbf{e}_{\text{PNCC}} \) and \( \mathbf{e}_{\text{PNCC}} \). The spontaneous fission (SF) decay of \( ^{240}\text{Pu} \) was simulated in [MCNP6 2014] to determine the efficiencies, here defined as number of reals per spontaneous fission event. The modeled observable is the number of reals per second, \( R_s \), and in this work is expressed as

\[
R_s = F_{\text{SF}}^{240\text{Pu}} \sum_{l=1}^{n_{\text{seg}}} m_{240\text{Pu}_{\text{eq}},l} \left[ \lambda_l \epsilon_{\text{PNCC},1}^{\text{PNCC}} + (1 - \lambda_l) \epsilon_{\text{PNCC},1}^{\text{PNCC}} \right],
\]

\[
m_{240\text{Pu}_{\text{eq}},l} = M_l \sum_{i=1}^{n_i} \frac{Q_i}{Q_{240\text{Pu}}} m_{i,l},
\]

\[
Q_i = F_{\text{SF}}^{\nu_{\nu}(2),i},
\]

where \( m_{240\text{Pu}_{\text{eq}}} \) represents the mass of \( ^{240}\text{Pu} \) that results in the same \( R_s \) as the considered sample, \( F_{\text{SF}} \) is the spontaneous fission rate per mass unit of radionuclide \( i \) (negligibly small for odd Pu isotopes, \( ^{237}\text{Np} \) and \( ^{235}\text{U} \),
\(\epsilon_{\text{PNCC}}^h\) and \(\epsilon_{\text{PNCC}}^p\) are the pre-calculated efficiencies for the homogeneous and point source distribution assumptions, respectively, in segment \(l\), \(m_i^l\) is the inferred total mass of the considered radionuclides in segment \(l\), \(m_i^f\) is the inferred mass fraction of radionuclide \(i\) in segment \(l\), and \(\nu_{\alpha(2),i}\) is the second factorial moment of the neutron emission distribution by spontaneous fission associated to radionuclide \(i\).

### 4 Bayesian inference

#### 4.1 Bayesian paradigm

The forward problem is commonly represented as

\[
d = F(\theta) + \eta,  \tag{4}\]

where \(d = [d_1, \ldots, d_{N_d}] \in \mathbb{R}^{N_d}, N_d \geq 1\) is the measurement data, \(F(\theta)\) is a deterministic forward model with parameters \(\theta\) and the noise term \(\eta\) lumps all sources of errors.

In the Bayesian paradigm, parameters in \(\theta\) are viewed as random variables with a posterior probability density function (pdf), \(p(\theta|d)\), given by

\[
p(\theta|d) = \frac{p(\theta)p(d|\theta)}{p(d)} \propto p(\theta)p(d|\theta),  \tag{5}\]

where \(p(d|\theta)\) signifies the likelihood function of \(d\) given \(\theta\). The normalization factor \(p(d) = \int p(\theta)p(d|\theta)\,d\theta\) is not needed for parameter inference when the parameter dimensionality is kept fixed. In the remainder of this paper, we will thus focus on the unnormalized density \(p(\theta|d) \propto p(\theta)p(d|\theta)\).

#### 4.2 Inferred variables

Here the vector of inferred variables, \(\theta\), consists of the following

1. The isotopic vector (\%) of the following five Pu isotopes, \(^{238}\text{Pu},\,^{239}\text{Pu},\,^{240}\text{Pu},\,^{241}\text{Pu}\) and \(^{242}\text{Pu}\), in each of the 20 segments

\[
v_i = \left[ m_{238\text{Pu},i}^f, m_{239\text{Pu},i}^f, m_{240\text{Pu},i}^f, m_{241\text{Pu},i}^f, m_{242\text{Pu},i}^f \right]  \tag{6}\]

with \(i = 1, \ldots, 20\), the segment number.

For notational convenience, we encapsulate the 20 isotopic vectors in the 2D \(V\) array

\[
V = \begin{pmatrix} v_1 \\ \vdots \\ v_{20} \end{pmatrix} \tag{7}\]

2. The total Pu mass (g) in each of 20 segments, \(p = [p_1, \ldots, p_{20}]\).

3. The \(^{241}\text{Am}\) mass (g), expressed as the ratio of the \(^{241}\text{Am}\) mass to the total Pu mass, in each of the 20 segments, \(a = [a_1, \ldots, a_{20}]\).

4. The (unique) \(cr\) parameter that is used to derive the \(^{244}\text{Cm}\) mass (g), \(c\), in each of the 20 segments, \(c = [c_1, \ldots, c_{20}]\). As detailed in section 4.4.3, the \(cr\) parameter is equal to \(\log_{10}\left[ (m_{239\text{Pu},i} + m_{240\text{Pu},i}) / m_{244\text{Cm},i} \right]\) where \(m_{239\text{Pu},i}\) and \(m_{244\text{Cm},i}\) are the \(^{239}\text{Pu}\) and \(^{244}\text{Cm}\) masses in a given segment \(i\).

5. The \(^{237}\text{Np}\) mass (g), expressed as the ratio of the \(^{237}\text{Np}\) mass to the total Pu mass, in each of the 20 segments, \(n = [n_1, \ldots, n_{20}]\).

6. The \(^{235}\text{U}\) mass (g), expressed as the ratio of the \(^{235}\text{U}\) mass to the total Pu mass, in each of the 20 segments, \(u = [u_1, \ldots, u_{20}]\).

7. The \(\lambda = [\lambda_1, \ldots, \lambda_{20}]\) vector.

8. The 240 background continuum counts of the 3AX measurement, \(b = [b_1, \ldots, b_{240}]\) (12 energy peaks \(\times 20\) detector locations).
9. The five parameters, \( \alpha = [\alpha_1, \cdots, \alpha_5] \), of the Dirichlet prior distribution put on \( v_1, \cdots, v_{20} \), when Bayesian hierarchical modeling is used to account for the uncertainty in our preliminary estimate of the Pu isotopic vector. Here, \( \alpha_1, \alpha_2, \alpha_3, \alpha_4, \) and \( \alpha_5 \) therefore represent the prior mass percentages of \( ^{238}\text{Pu}, ^{239}\text{Pu}, ^{240}\text{Pu}, ^{241}\text{Pu} \) and \( ^{242}\text{Pu} \), respectively. Notice that although they share the same prior, the \( v_1, \cdots, v_{20} \) vectors are free to take different values if needed. In other words, the posterior isotopic vectors can be different between the 20 segments.

10. When a GP prior is put on \( \mathbf{p} \) (see later on), the mean, \( \mu_{\text{GP}} \), variance, \( \sigma_{\text{GP}}^2 \), and lengthscale, \( \tau_{\text{GP}} \), of the chosen GP kernel.

Overall, we therefore jointly infer between 441 and 449 parameters depending on the considered setup.

Our measurement vector, \( \mathbf{d} \), contains the 240 measured gross counts by 3AX-SGS, \( d_{\text{SGS}} \) and the number of reals per second measured by PNCC, \( d_{\text{PNCC}} \), \( \mathbf{d} = [d_{\text{SGS}}, d_{\text{PNCC}}] \). In addition, our derived posterior distribution is not only conditioned to \( \mathbf{d} \) but also to the predefined \( E_{\text{SGS}}^h, E_{\text{PNCC}}^h, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p \) arrays. This results in assessment of

\[
p\left( \theta | \mathbf{d}, E_{\text{SGS}}^h, E_{\text{PNCC}}^p, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p \right) \propto p\left( \mathbf{d}, E_{\text{SGS}}^h, E_{\text{PNCC}}^p, \theta \right) p\left( \theta \right).
\]

### 4.3 Likelihood function

Our likelihood function, \( p\left( \mathbf{d}, E_{\text{SGS}}^h, E_{\text{PNCC}}^p, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p | \theta \right) \) is the product of the likelihoods of the SGS and PNCC data, respectively, given \( \theta \)

\[
p\left( \mathbf{d}, E_{\text{SGS}}^h, E_{\text{PNCC}}^p, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p | \theta \right) = p\left( \mathbf{d}_{\text{SGS}}, E_{\text{SGS}}^h, E_{\text{SGS}}^p | \theta \right) \times p\left( \mathbf{d}_{\text{PNCC}}, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p | \theta \right).
\]

If we assume \( d_{\text{SGS}} \) to follow a Poisson process, which is the norm for count data, we have

\[
p\left( \mathbf{d}_{\text{SGS}}, E_{\text{SGS}}^h, E_{\text{SGS}}^p | \theta \right) = \prod_{i=1}^{N_{\text{SGS}}} \exp\left( -\tilde{d}_i \right) \frac{\tilde{d}_i^{d_i} d_i!}{d_i!},
\]

where \( \tilde{d}_{\text{SGS}} = [\tilde{d}_1, \ldots, \tilde{d}_{N_{\text{SGS}}}] \) contains the simulated gross counts for a given \( \theta \).

For \( d_{\text{PNCC}} \), we assume a normal distribution with a 5% relative error, \( \sigma_{\text{PNCC}} = 0.05 \times \hat{d}_{\text{PNCC}} \)

\[
p\left( \mathbf{d}_{\text{PNCC}}, e_{\text{PNCC}}^h, e_{\text{PNCC}}^p | \theta \right) = N\left( \hat{d}_{\text{PNCC}}, 0.05 \hat{d}_{\text{PNCC}} \right),
\]

where \( \hat{d}_{\text{PNCC}} \) is the simulated number of reals per second corresponding to a given \( \theta \).

### 4.4 Prior distributions

#### 4.4.1 Dirichlet prior and hierarchical modeling of the prior uncertainty in the vector

With respect to the Pu isotopic vector associated with each segment, that is, each row \( v_i \) of \( \mathbf{V} \) with \( i = 1, \cdots, 20 \) and \( v_i = [m_{238\text{Pu},i}^f, m_{239\text{Pu},i}^f, m_{240\text{Pu},i}^f, m_{241\text{Pu},i}^f, m_{242\text{Pu},i}^f] \), we use a Dirichlet prior distribution, \( \mathbf{v} \propto \text{Dir} \left( \alpha \right) \). This multidimensional distribution has the nice property that the values of the elements of any sampled vector \( \mathbf{v} \) sum up to one

\[
p\left( \mathbf{v} \right) = \frac{1}{B \left( \alpha \right)} \prod_{j=1}^{K} v_j^{\alpha_j - 1},
\]

where

\[
B \left( \alpha \right) = \frac{\prod_{j=1}^{K} \Gamma \left( \alpha_j \right)}{\Gamma \left( \sum_{j=1}^{K} \alpha_j \right)}, \quad \alpha = [\alpha_1, \cdots, \alpha_K],
\]

with \( \sum_{j=1}^{K} v_j = 1 \) and \( v_j \geq 0 \ \forall j \in \{1, \cdots, K\} \).
In case there is no additional prior information on the 20 $v_i$ vectors, we set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1$ in $\alpha$, which amounts to a flat Dirichlet prior where for each segment, each considered Pu isotope receives the same prior probability under the constraint that the Pu isotope fractions sum up to 1.

Besides a flat Dirichlet prior, we also consider the case where prior information about the Pu isotopic vector is available. As written earlier, here this information comes from isotopic composition analysis performed using the FRAM software (see section 2.2) and standard deviations set to half the associated means, $\sigma = \frac{\text{coefficient of variation of 50 %}}{2}$.

In this study, the FRAM-based prior isotopic composition analysis gives: $m_{238\text{Pu}} = 0.0064, m_{239\text{Pu}} = 0.7769, m_{240\text{Pu}} = 0.1855, m_{241\text{Pu}} = 0.0132$ and $m_{242\text{Pu}} = 0.0180$. The degree of enforcement of the prior isotopic vector is controlled by the width of the Dirichlet prior that is put on each $v_i$, which is itself defined by the actual $\alpha$ values. For $\alpha$ values $> 1$, the larger the value the more the probability mass is concentrates near the predefined proportions.

As of $p(\alpha)$, we put truncated normal priors on $\alpha_1, \cdots, \alpha_5$ with means equal to our preliminary estimates obtained from prior isotopic composition analysis, $\mu_{\alpha} = [\mu_{\alpha,1}, \cdots, \mu_{\alpha,5}] = [\mu_{\alpha,238\text{Pu}}, \mu_{\alpha,239\text{Pu}}, \mu_{\alpha,240\text{Pu}}, \mu_{\alpha,241\text{Pu}}, \mu_{\alpha,242\text{Pu}}]$, (see section 2.2) and standard deviations set to half the associated means, $\sigma_{\alpha,i} = 0.5\mu_{\alpha,i}$, which corresponds to a coefficient of variation of 50 %.

$$p(\alpha) = \begin{cases} N(\mu_{\alpha},C_{\alpha}), & \text{if all } \alpha_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$C_{\alpha} = \begin{pmatrix} \sigma_{\alpha,1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\alpha,5}^2 \end{pmatrix}$$ (14)

In this study, the FRAM-based prior isotopic composition analysis gives: $m_{238\text{Pu}} = 0.0064, m_{239\text{Pu}} = 0.7769, m_{240\text{Pu}} = 0.1855, m_{241\text{Pu}} = 0.0132$ and $m_{242\text{Pu}} = 0.0180$. The degree of enforcement of the prior isotopic vector is controlled by the width of the Dirichlet prior that is put on each $v_i$, which is itself defined by the actual $\alpha$ values. For $\alpha$ values $> 1$, the larger the value the more the probability mass is concentrates near the predefined proportions. In this work we set $\mu_{\alpha,238\text{Pu}} = 0.64, \mu_{\alpha,239\text{Pu}} = 77.69, \mu_{\alpha,240\text{Pu}} = 18.55, \mu_{\alpha,241\text{Pu}} = 1.32$ and $\mu_{\alpha,242\text{Pu}} = 1.80$. This is deemed to feed the inference with a good level of prior knowledge with still allowing for enough deviations if required to fit the data.

### 4.4.2 Gaussian process prior for the Pu mass profile

Gaussian processes (GPs) form flexible nonparametric priors that are widely used for various machine learning tasks such as regression and classification (see the book by [Rasmussen and Williams, 2006] for details). A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP, $f(x)$, is completely determined by its mean function, $\mu(x)$, and covariance function, $k(x, x')$

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$$

$$\mu(x) = \mathbb{E}[f(x)]$$

$$k(x, x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x'))]$$ (15)

Putting a GP prior on the vertical Pu mass profile, $p$, helps to enforce its smoothness. Here for computational tractability we assign a GP prior on $y = \log(p)$. Assuming a constant mean for $y, \mu_{\text{GP}}$, we thus have

$$y \sim \mathcal{GP}(\mu_{\text{GP}}, k(y, y'))$$ (16)

There are many possibilities for the choice of the covariance kernel $k(\cdot, \cdot)$. In this study we chose the squared exponential or radial basis function (RBF) kernel which is the de-facto default kernel for GPs. This kernel is infinitely differentiable, which means that the GP with this covariance function is very smooth ([Rasmussen and Williams, 2006])

$$k_{\text{RBF}}(y, y') = \sigma_{\text{GP}}^2 \exp \left(-\frac{(y - y')^2}{2\tau_{\text{GP}}^2}\right)$$ (17)
where \( \sigma_{GP}^2 \) and \( \tau_{GP} \) are the so-called kernel variance and lengthscale of the GP. Jointly inferring \( \mu_{GP}, \sigma_{GP}^2 \) and \( \tau_{GP} \) along with \( y \) and the other unknown variables will induce a degree of smoothness in the \( y \) profile that is consistent with the measurement data \( d \). In other words, the MCMC sampling will pick up \( y \) profiles that are as smooth as possible while still leading to an appropriate fit of \( F ( \theta ) \) to \( d \).

Under the GP prior model, the prior distribution of \( y, p ( y ) \), becomes

\[
p ( y ) = \int \int \int p ( y | \mu_{GP}, \sigma_{GP}^2, \tau_{GP} ) p ( \mu_{GP} ) d \mu_{GP} d \sigma_{GP}^2 d \tau_{GP}.
\]  

(18)

The \( p ( y | \mu_{GP}, \sigma_{GP}^2, \tau_{GP} ) \) density is known as the marginal likelihood of the GP model (see Rasmussen and Williams 2006 for details) and can be computed analytically. For our considered application we have

\[
\log p ( y | \mu_{GP}, \sigma_{GP}^2, \tau_{GP} ) = -\frac{1}{2} (y - \mu_{GP})^T K^{-1} (y - \mu_{GP}) - \frac{1}{2} \log |K| - \frac{n_y}{2} \log 2\pi,
\]  

(19)

where \( K \) is the \( n_y \times n_y \) matrix of two-point covariances given by equation (17). With respect to the kernel hyperparameters, \( \mu_{GP}, \sigma_{GP}^2 \), and \( \tau_{GP} \), we select the following priors.

- \( \log ( \mu_{GP} ) \sim N (0, 1) \). This means that for each of the 20 drum segments, the mean (total) Pu mass is assigned a lognormal prior distribution with mode, median and standard deviation of \( \exp ( -1 ) \) g, 1 g and 2.16 g, respectively.
- \( \log ( \sigma_{GP}^2 ) \sim N (0, 1) \), which is deemed wide enough to allow for the sampled \( y \) values to depart from \( \mu_{GP} \) by orders of magnitude if necessary for fitting the measurement data \( d \).
- \( p ( \tau_{GP} ) \sim U (0.5, 20) \). The \( \tau_{GP} \) parameter is the lengthscale of the GP prior, also loosely called correlation length, and controls how much two values, \( y \) and \( y' \) separated by a given distance will be correlated. Here the considered distance measure is the number of segments and by using a \( U (0.5, 20) \) prior distribution we thus allow \( \tau_{GP} \) to vary from half a segment to the entire profile length of 20 segments. For the considered Gaussian kernel in equation (17), \( \tau_{GP} = 0.5 \) means a negligibly small spatial correlation of about 0.14 between adjacent segments while \( \tau_{GP} = 20 \) translates into a spatial correlation of about 1 between adjacent segments and about 0.61 between the top (segment 20) and bottom (segment 1) of the drum.

4.4.3 Prior for the \( 244Cm \) profile

For inference of the 20 \( 244Cm \) masses, we devise a prior based on the NEA-SFCOMPO 2.0 dataset (available online, see Michel-Sendis et al. 2017). The NEA-SFCOMPO database contains \( 239Pu, 240Pu \) and \( 244Cm \) experimentally measured masses from fuel samples irradiated in power reactors over the past 50 years. It basically includes all reactor and fuel types with very low to high burnup. From this database, it is observed that the quantity \( \log_{10} \left( \frac{m_{239Pu} + m_{240Pu}}{m_{244Cm}} \right) \) approximately follows a gamma distribution with shape parameter 7.88 and rate parameter 2.51 (Figure 2). We therefore infer the \( cr \) coefficient with a gamma prior, \( p ( cr ) = \Gamma (7.88, 2.51) \) and calculate the corresponding \( 244Cm \) mass in a given segment, \( i \), as \( m_{244Cm,i} = \frac{m_{239Pu,i} + m_{240Pu,i}}{10^{cr}} \) where \( m_{239Pu,i} \) and \( m_{240Pu,i} \) are jointly inferred along \( cr \) and the other considered variables by the MCMC sampling (see section 4.2).
Figure 2: Distribution of the quantity $\log_{10}\left(\frac{(m_{239Pu} + m_{240Pu})}{m_{244Cm}}\right)$ from the NEA-SFCOMPO dataset and corresponding fitted gamma distribution.

### 4.4.4 Other priors

Alternatively to using a GP prior for $y = \log(p)$, we also test with $p(y) \sim N(0, 1)$ and $p(y) \sim U(\log(0.01), \log(100))$, respectively. Using $p(y) \sim N(0, 1)$ is practically the same as putting a standard normal, that is, $N(0, 1)$, prior on the GP mean (as done herein) and fixing $\sigma^2_{GP} = 1$ and $v_{GP} = 0$ in equation (17).

Furthermore, using $p(y) \sim U(\log(0.01), \log(100))$ basically signifies that the Pu mass in any given segment, $p$, is allowed to vary between 0.01 g and 100 g while lower masses are preferred over larger ones if consistent with $d$.

The prior distributions for the $^{241}$Am to Pu mass ratios, $a$, $^{237}$Np to Pu mass ratios, $n$, and $^{235}$U to Pu mass ratios, $u$, are all taken as bounded uniform distributions, of which the lower and upper bounds are chosen on the basis of the isotopic composition analysis applied to the ISOCOM gamma measurement described in section 2.2. We use $p(a) = U(0.01, 0.15)$, $p(n) = U(0.0001, 0.01)$ and $p(u) = U(0.01, 0.15)$.

Since for technical reasons we cannot assign Poisson priors to the (SGS) background continuum counts (SGS), $b$, we use instead an uncorrelated and independent normal prior distribution with mean and variance vectors both equal to the measured count values: $p(b) = N(b, C_b)$ with $C_b$ a diagonal matrix with the $b_1, \ldots, b_{240}$ values as diagonal elements. A $N(x, x)$ distribution provides an increasingly accurate approximation to $Pois(x)$ as $x$ increases (see, e.g., Barbour et al., 1992). The approximation is commonly deemed excellent for $x > 1000$ and reasonably accurate for $x > 10$ (SOCR/UCLA, 2007).

Lastly, the prior distribution for the 20-dimensional $\lambda$ vector is taken as a bounded uniform distribution: $U(0, 1)$.
4.5 Sampled posterior

We assume the prior distributions for $V, y, a, cr, n, u, \lambda$ and $b$ to be independent. Using Bayesian hierarchical modeling of $p(V)$ and a Gaussian process prior for $y = \log (p)$, the sampled posterior parameter distribution becomes

$$p \left( V, y, a, cr, n, u, \lambda, b, \mu_G, \sigma_G^2, \tau_G | d, E^h, E^h_{SGS}, e^b_{PNCC}, e^b_{PNCC} \right) \propto$$

$$p \left( d, E^h_{SGS}, E^h_{PNCC}, e^b_{PNCC}, e^b_{PNCC} | V, y, a, cr, n, u, \lambda, b \right) \times$$

$$p \left( V | \alpha \right) p \left( \alpha \right) p \left( y | \mu_G, \sigma_G^2, \tau_G \right) p \left( \mu_G \right) p \left( \sigma_G^2 \right) p \left( \tau_G \right) \times$$

$$p \left( a \right) p \left( cr \right) p \left( n \right) p \left( u \right) p \left( \lambda \right) p \left( b \right).$$

No analytical solution of the posterior distribution in equation (20) can be derived and we therefore sample from this distribution by MCMC simulation (see [Gelman et al., 2014], [Betancourt, 2018] for an extensive description of the HMC algorithm). Convergence of the MCMC to the posterior target is monitored by means of the potential scale reduction factor, $R$ ([Gelman and Rubin, 1992], [Gelman et al., 2014]), using eight interacting Markov chains evolved in parallel (see next section). The $R$ statistic compares for each parameter of interest the average within-chain variance to the variance of all the Markov chains mixed together. The closer the values of these two variances, the closer to one the value of the $R$ diagnostic. To declare convergence to a limiting distribution we require that the values of $R$ are jointly smaller than 1.1 for all sampled variables.

4.6 Software implementation

Similarly as in [Laloy et al., 2021], we used the open-source greta package ([Golding, 2019]) to perform the HMC-based MCMC sampling. The greta package provides a probabilistic programming language embedded in R ([R Core Team, 2020]) interfacing to some of the MCMC sampling algorithms implemented in the Tensorflow-probability package ([TFP, Dillon et al., 2018]), which itself relies on the Tensorflow (TF) machine learning platform ([Abadi et al., 2016]). The most useful MCMC sampler available through greta and used herein is HMC ([Neal, 2011], [Betancourt, 2018]). This TFP-based HMC implementation can evolve several Markov chains in parallel on both CPUs and GPUs, with the different chains exchanging information during warmup to speedup convergence. In this study, we evolved 8 interacting Markov chains in parallel over 6 CPUs. Additionally, most of the pre- and post-processing was performed with the tidyverse collection of packages ([Wickham et al., 2019]), together with a few other specific packages (mainly by [Plummer et al., 2006], [Kay, 2021], [Peterson, 2020]).

5 Results and discussion

We first investigate the behavior of our proposed inference method on a synthetic experiment for which the “truth” is thus known, before applying it to the same real waste drum as considered by [Laloy et al., 2021].

5.1 Synthetic experiment

The reference or “true” drum was built by as follows. The true total Pu mass (g) profile across the 20 segments, $p_{true}$, was taken as

$$p_{true} = 0.15 + [\sin (0.5x) + 1] \times 1.175$$

with $x = [1, 2, \ldots, 20].$ (21)

Furthermore, the true $\lambda$ value was set to 1 (homogeneous source) in every segment except for segment 16 where it was set to 0 (point source). True values for all other inferred variables were randomly sampled from their respective prior distributions.

The true SGS and PNCC measurements were then created by randomly perturbing the forward-simulated data corresponding to the true parameter values. More specifically, the $i = 1, \ldots, 240$ simulated net counts (associated with SGS), $c^{net}_{true,i}$, were used as mean parameters of 240 Poisson distributions from which the 240 observed net counts were drawn: $c^{net}_{obs,i} \sim Pois \left( c^{net}_{true,i} \right)$. The 240 observed background continuum counts, $b_{obs,i}$, were set to the measured ones for the real case (section 5.2). In addition, the observed number of reals per second (associated with PNCC), $r_{obs}$, was drawn from a normal distribution centered around the true simulated number of reals per second, $r_{true}$, and relative standard deviation of 5%. Note that due to the added errors in the observed net counts and reals per second, no perfect inverse solution exists which is required for Bayesian inference.
We first consider Bayesian hierarchical modeling of the prior uncertainty in the 20 Pu isotopic vectors (equation 13 in section 4.4.1) and putting a GP prior model on the total Pu mass across segments (section 4.4.2). This inversion strategy is referred to as strategy 1. The HMC was ran for 10,000 warmup iterations and 5000 consecutive sampling iterations, thereby providing 40,000 posterior samples.

Figures 3 and 4 depict the posterior distributions of the Pu isotopic vectors and total nuclide masses, respectively, across the considered 20 segments. It is seen that the posterior distribution is overall well resolved and always contains the true parameter value. With respect to the posterior \( \lambda \) distribution (source distribution uncertainty in each segment), we find that the posterior uncertainty is larger in case of lower total radioactive content in the segment (Figure 4f). This is likely due to the fact that a lower amount of Pu produces less net counts (for gamma-emitting isotopes) and reals per seconds (for the odd Pu isotopes) which reduces the available information about \( \lambda \).

Figure 5 shows the posterior distribution of the inferred GP parameters, \( \mu_{GP}, \sigma^2_{GP}, \) and \( \tau_{GP} \), together with the associated “estimated” true values obtained by fitting our GP model to \( \log(p_{true}) \) directly. The marginal posterior GP parameter distribution appears to be in good agreement with the estimated truth.

Figure 3: Prior and posterior Pu isotopic vector distributions in each of the 20 segments for the synthetic test case and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment. Segment numbering goes from bottom (1) to top (20).
Figure 4: Prior and posterior (a) Pu mass, (b) $^{241}$Am mass, (c) $^{244}$Cm, (d) $^{237}$Np, (e) $^{235}$U and (f) $\lambda$ distributions for the synthetic test and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment. The black dots denote the true values used to generate the SGS (gross count) and PNCC (reals per second) data. The outcome of equation (21) is the distribution of black dots in subplot (a). Note that the $x$-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding $x$-axes do not cover the full prior range (especially for subplots b, c and d). For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).

Figure 5: Prior and posterior (a) $\sigma^2_{GP}$, (b) $\tau_{GP}$ and (c) $\mu_{GP}$ distributions for the synthetic test case and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment.
We now turn attention to the case where no GP prior model is used for the total Pu mass but either a standard normal prior or a loguniform prior bounded between 0.1 g and 100 g in each segment. We refer to these two inversion strategies as strategies 2 and 3, respectively. As written above, using a standard normal prior (strategy 2) is equivalent to using the GP prior model described in equations (16) - (17) (strategy 1) but with fixing $\sigma^2_{GP} = 1$ and $\tau_{GP} = 0$. In case of the loguniform prior (strategy 3), the warmup period needed to be extended to 20,000 iterations to achieve $\hat{R}$-convergence. Figures 6 presents the posterior mass and $\lambda$ distributions in each segment corresponding to strategy 2 while Figure 7 displays the same distributions for strategy 3. These distributions are fairly (strategy 2) to strongly (strategy 3) wider than when a GP prior model is used for the total Pu mass (strategy 1, Figure 4). Furthermore, although narrower the posterior distributions associated with strategy 1 always include the true values. This nicely illustrates the benefits of GP-based regularization to reduce posterior uncertainty.

Figure 6: Prior and posterior (a) Pu mass, (b) $^{241}\text{Am}$ mass, (c) $^{244}\text{Cm}$, (d) $^{237}\text{Np}$, (e) $^{235}\text{U}$ and (f) $\lambda$ distributions for the synthetic test case and strategy 2. Strategy 2 relies on Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and puts a standard normal prior on the logarithm of the total Pu mass per segment. The black dots denote the true values used to generate the SGS (gross count) and PNCC (reals per second) data. The outcome of equation (21) is the distribution of black dots in subplot (a). Note that the $x$-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding $x$-axes do not cover the full prior range. For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).
Figure 7: Prior and posterior (a) Pu mass, (b) $^{241}$Am mass, (c) $^{244}$Cm, (d) $^{237}$Np, (e) $^{235}$U and (f) $\lambda$ distributions for the synthetic test case and strategy 3. Strategy 3 relies on Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and puts an uniform prior on the logarithm of the total Pu mass per segment. The black dots denote the true values used to generate the SGS (gross count) and PNCC (reals per second) data. The outcome of equation (21) is the distribution of black dots in subplot (a). Note that the $x$-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding $x$-axes do not cover the full prior range. For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).

To verify whether using a GP prior model remains useful when $\log(p_{true})$ does not at all obey a GP, we repeated the exercise with a GP prior put on $\log(p)$ (strategy 1) but for a case where $p_{true}$ is set to 0.015 g in every segment but segments 5 and 16 where it is set to 2.5 g. Hence, the $\log(p_{true})$ profile is now flat with two spikes (in segments 5 and 16). Such a “flat and spiky” profile is arguably far from Gaussian. Figure 8 presents the corresponding posterior mass and $\lambda$ distributions in each segment. It is observed that the inferred posterior mass profiles always include the true mass values, though at the sharp interfaces (upper and lower neighboring segments of segment 5 and 16), most of the probability mass is put at somewhat too large values (see posterior mass distributions in segments 4 and 6 and segments 15 and 17, respectively). This can be fully mitigated by using a log-uniform rather uniform prior for $\tau_{GP}$, together with a lower bound of 0.01 instead of 0.5 (not shown). Nevertheless, such setting of the prior $\tau_{GP}$ distribution assumes that some prior knowledge of the true $\log(p_{true})$ profile is available. Overall, we conclude from this last synthetic test that even if the true Pu mass profile does not at all show smooth transitions between segments, using a GP prior still allows for finding a regularized solution that is reasonably accurate.
Figure 8: Prior and posterior (a) Pu mass, (b) $^{241}$Am mass, (c) $^{244}$Cm, (d) $^{237}$Np, (e) $^{235}$U and (f) $\lambda$ distributions for the synthetic test case and strategy 1, when using a “flat and spiky” true Pu mass profile. The black dots denote the true values used to generate the SGS (gross count) and PNCC (reals per second) data. Note that the $x$-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding $x$-axes do not cover the full prior range. For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).

5.2 Real drum

For the real data inversion, our reference strategy is strategy 1 which therefore includes hierarchical modeling of the Pu isotopic vector uncertainty and putting a GP prior model on the total Pu mass.

Some of the associated inversion results are illustrated in Figures 9 (posterior distribution of the isotopic vectors), 10 (posterior mass and $\lambda$ distributions) and 11 (posterior GP parameter distributions). All sampled variables appear to be relatively well constrained, with marginal posterior distributions that are substantially narrower than their associated prior distributions. Consistently with our previous study (Laloy et al., 2021), the total Pu mass is lower at the top and, to a lesser extent, at the bottom of the drum (Figure 10a), while the assumption of an homogeneous source distribution is strongly favored for segments 2 - 18 and more uncertain at the top (segments 19 - 20) and bottom (segment 1). As detailed in Laloy et al. (2021), the smaller posterior Pu masses in the top and bottom segments is likely caused by an incomplete drum’s filling (top segments) and the fact that the bottom plate of the drum does not touch the floor, only the drum’s outer bottom ring does (bottom segment).

As of the GP model parameters, the posterior lengthscale distribution is concentrated around its lower bound of 0.5. This basically means that no spatial correlation is inferred from the measurement data. This does not prevent the derived log ($p$) profile to be somewhat smooth and well resolved (Figure 10b). In addition, for this application we find that strategy 2 results into a posterior distribution (not shown) that resembles closely the posterior distribution achieved by strategy 1. This is mainly due to the absence of detectable spatial correlation which causes the prior distribution for $y$ to become relatively similar between strategies 1 and 2. As detailed earlier, for sufficiently small $\tau_{GP}$ values (say $\leq 0.5 - 0.6$), the only practical difference between the two priors is that for strategy 1 the $\sigma_{GP}^2$ parameter is jointly inferred with the other variables with a lognormal prior that has a median of 1 while for strategy 2 the $\sigma_{GP}^2$ parameter is fixed to 1.
In contrast to our strategies 1 and 2, Figure 12 illustrates that if an uniform prior is put on \( \log(p) \) instead of a GP or standard normal prior, which as written already corresponds to inversion strategy 3, then posterior uncertainty becomes substantially larger while bimodality appears in the posterior mass distributions in some segments (see segments 11, 12 and 15). This bimodality is due to the large negative correlations between the posterior masses that appear for some adjacent segments. In addition, likewise for the synthetic case study strategy 2 incurs a larger computational cost to reach \( \hat{R} \)-convergence of the MCMC: 20,000 warmup iterations instead of 10,000 warmup iterations for strategies 1 and 2.

Figure 9: Prior and posterior Pu isotopic vector distributions in each of the 20 segments for the real test case and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment. Segment numbering goes from bottom (1) to top (20).
Figure 10: Prior and posterior (a) Pu mass, (b) $^{241}$Am mass, (c) $^{244}$Cm, (d) $^{237}$Np, (e) $^{235}$U and (f) $\lambda$ distributions for the real test case and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment. Note that the x-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding x-axes do not cover the full prior range. For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).

Figure 11: Prior and posterior (a) $\sigma_{GP}^2$, (b) $\tau_{GP}$ and (c) $\mu_{GP}$ distributions for the real test case and strategy 1. Strategy 1 uses Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and a GP prior model for the logarithm of the total Pu mass per segment.
Strategy 3 relies on Bayesian hierarchical modeling of the Pu isotopic vector uncertainty and puts an uniform prior on the logarithm of the total Pu mass per segment. Note that the x-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding x-axes do not cover the full prior range. For the λ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).
Figures 13 and 14 present the posterior Pu isotopic vectors’ distribution and posterior nuclide masses’ distribution, respectively, obtained when there is no prior information available on the 20 isotopic vectors and a flat Dirichlet prior is thus put on the \( i = 1, \cdots, 20 \) \( \nu_i \) vectors, while a GP prior model is used for \( \log(p) \). This variant is referred to as strategy 4. Herein too, the posterior mass uncertainty increases a lot for each nuclide compared to strategy 1 (Figure 10) and some bimodality is observed (see posterior masses in segments 5 and 15 of Figure 14). Furthermore, the computational effort required to achieve \( \hat{R} \)-convergence now becomes even larger than for the other two considered inference strategies as after 30,000 warmup iterations (followed by 5,000 sampling iterations) only 27 (12) out of the 444 sampled parameters associated with strategy 3 show a \( \hat{R} < 1.1 (1.2) \). The results depicted in Figures 13 and 14 are therefore not officially converged and should be considered with some caution, although we do not expect the main trends seen in Figures 13, 14 to change drastically with more warmup iterations, only the currently observed small bumps should be smoothed out.

Figure 13: Prior and posterior Pu isotopic vector distributions in each of the 20 segments for the real test case and strategy 4. Strategy 4 includes putting flat Dirichlet priors on the 20 Pu isotopic vectors and putting a GP prior model on the logarithm of the total Pu mass per segment. Segment numbering goes from bottom (1) to top (20).
Figure 14: Prior and posterior (a) Pu mass, (b) $^{241}$Am mass, (c) $^{244}$Cm, (d) $^{237}$Np, (e) $^{235}$U and (f) $\lambda$ distributions for the real test case and strategy 4. Strategy 4 includes putting flat Dirichlet priors on the 20 Pu isotopic vectors and putting a GP prior model on the logarithm of the total Pu mass per segment. Note that the x-axes of subplots (a - e) have a base 10 logarithmic scale. When plotted with this scale, all prior distributions in subplots (a - e) are symmetric but, for visual convenience, the corresponding x-axes do not cover the full prior range. For the $\lambda$ variable, we show a posterior histogram instead of a kernel density estimate. This is because when applied to a bounded data sample, kernel density smoothing tends to create artifacts near the bounds. Segment numbering goes from bottom (1) to top (20).

Figure 15 displays the posterior mass distributions over the drum for all considered nuclides and strategies 1, 3 and 4. We find that although strategies 3 and 4 lead to a much larger uncertainty at the segment level compared to strategy 1, strategies 3 and 4 provide similarly well resolved posterior mass distributions for the whole drum. With respect to strategy 3, the posterior mass distributions are basically the same as for strategy 1. A similar situation is observed for strategy 4 for all nuclides but $^{244}$Cm (dark blue curve) $^{240}$Pu (yellow curve) and $^{242}$Pu (pink curve). The relatively big shift towards higher $^{242}$Pu masses for strategy 4 is primarily caused by the use of a flat Dirichlet prior for the Pu isotopic vector together with the fact that $^{242}$Pu is not sensed by the (3AX-SGS) gamma measurement. The relatively big shift towards higher $^{244}$Pu masses for strategy 4 can also be explained by the use of a flat Dirichlet prior for the Pu isotopic vector, while an informative Dirichlet prior centered near a $^{240}$Pu mass fraction of 0.18 is used with strategies 1 and 3. With respect to $^{244}$Cm, the difference is likely caused by the difference in inferred $^{240}$Pu content as our inferred $^{244}$Cm content is co-derived from the inferred $^{239}$Pu and $^{240}$Pu masses on the one hand, and the inferred $c_T$ parameter on the other hand (see section 4.4.3 for details).
Figure 15: Posterior mass distributions over the whole drum for the considered radionuclides and MCMC inversion strategies 1, 3 and 4. For this case study, strategy 2 leads to quite similar results as strategy 1 (see main text for details).

Going back to our reference strategy 1, Figure 16 presents the prior and posterior background count distributions for some selected peaks. Very similar distributions are obtained for strategies 2, 3 and 4 (not shown) while our previous work with the same 3AX-SGS data but a different MCMC inversion setup also produced rather similar posterior count distributions (see Figure 6 in Laloy et al. (2021)). As stated by Laloy et al. (2021), it is worth noting that the prior $p(b)$ is based on a measured background continuum count that is considered to be a realization of a Poisson distribution of which the shape parameter is unknown. By using $p(b) = \text{Pois}(b)$ (or $p(b) = N(b, C_b) \approx \text{Pois}(b)$ as done herein) one implicitly postulates that the measured background continuum count is equal to the mean of its underlying distribution, which is obviously not necessarily the case. Some deviations of $p(b|d)$ from $p(b)$ are thus to be expected.

Lastly, Figure 17 shows a so-called posterior predictive check where the fit to the 3AX-SGS data provided by the derived posterior solutions is verified. A base 10 logarithmic scale is used to make it possible to visualize the discrepancies between (very) small measured and simulated counts while vertical bars denote the 95% posterior uncertainty intervals of the simulated counts. This plot is essentially similar to that obtained in our previous study (see Figure 7 in Laloy et al. (2021)). The measured gross counts are globally well approximated as the 95% uncertainty intervals are relatively tight and most often include the 1:1 line. Moreover, the most important relative deviations are for some pretty small gross counts, below $10^{-15}$. Overall, these fitting results are consistent with both our expectations and the Poisson count statistics. As of the PNCC measurement, the measured number of reals per second is also very well fitted. The maximum likelihood simulated value is practically equal to the measured value and its associated relative standard deviation is equal to assumed relative standard deviation of 5% in the PNCC likelihood (equation 11).
Figure 16: Prior and posterior distributions of the inferred background continuum counts. Each facet’s label mentions the considered combination of nuclide and energy peak (keV). For visual convenience, the counts have been standardized using the normal prior parameters and only 7 out of the considered 12 peaks are shown (the following energy peaks are left out of the plot: $^{241}$Am & 662.24 keV, $^{238}$Pu & 152.72 keV, $^{239}$Pu & 129.3 keV, $^{239}$Pu & 375.05 keV and $^{239}$Pu & 451.48 keV). Segment numbering goes from bottom (1) to top (20).
Figure 17: Measured gross counts against a posteriori simulated gross counts. The green points represent the maximum likelihood solution (solution with the largest Poisson likelihood) among the collected 40,000 posterior samples, and the vertical bars denote the 95% posterior uncertainty intervals. A base 10 logarithmic scale is used to make the deviations between the smallest observed and simulated gross counts visible.
5.3 Computational time

On the used 6-core workstation and for the used algorithmic parameters of greta/HMC, performing one MCMC iteration in each of the Markov chains takes about 0.5 s (this is an average over warmup and sampling iterations as greta warmup iterations incur a larger computational cost that sampling iterations). Therefore, performing 10,000 warmup iterations and 5000 sampling iterations, as done with our inversion strategies 1 and 2, roughly takes 2 hours. Such runtime per drum might be prohibitively long for routine analyses, let alone that inversion strategy 3 requires twice more warmup iterations that strategies 1 and 2 to reach \( R \)-convergence at the 1.1 level. This warrants further investigations on how to accelerate the inference.

6 Concluding remarks

We demonstrate in this study three methodological improvements for Bayesian inference of the radionuclide inventory in radioactive waste drums. First we rely on the Dirichlet distribution for the prior distribution of the isotopic vector as this distribution has the nice property that the elements of its vector samples sum up to 1. Second, we show that such Dirichlet priors can be incorporated within an hierarchical modeling of the prior uncertainty in the isotopic vector, when prior information about isotopic composition is available. Our used Bayesian hierarchical modeling framework makes use of this available information but also acknowledges its uncertainty by letting to a controlled extent the information content of the indirect measurement data (i.e., gamma and neutron counts) shape the actual prior distribution of the isotopic vector(s). Third, we propose to take advantage of GP prior modeling when inferring 1D spatially-distributed mass or, equivalently, activity distributions, to help regularize the MCMC inversion. Combining hierarchical modeling of the prior isotopic composition uncertainty together with GP prior modeling of the vertical Pu profile across the drum appears to work rather well. We find that our GP prior approach can handle both cases with and without spatial correlation, and additionally speeds up the inference, even when no correlation is present, compared to using a log-uniform prior for the Pu content in each considered drum’s segment.

Of course, our proposed GP prior modeling framework only makes sense in the context of spatial inference. Furthermore, the computational times involved by our proposed approach are on the order of a few hours, say about 2, to provide uncertainty estimates for all variables of interest. This might prevent our approach to be applied for routine analyses and future work will focus on speeding up the inference. With respect to uncertainty in the efficiencies, our study uses the same stylized drum modeling approach as proposed by Laloy et al. (2021) to account for the source distribution uncertainty across the vertical direction of the drum within the Bayesian inversion. Ongoing investigations on combining different measurement methods within a Bayesian framework that are also performed within the CHANCE project rely on the same approach to handle matrix-related uncertainties. These results will be presented in due course.

7 Acknowledgments

This work received funding by the EU project CHANCE “Characterization of Conditioned Nuclear Waste for its Safe Disposal in Europe”. An example code of the proposed approach is available at https://github.com/elaloy/UQ-RADWASTE.

References

Abadi, M., Agarwal, A., Barham, P., et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv e-prints, March 2016, https://www.tensorflow.org.

Bai, Y.F., Mauerhofer, E., Wang, D.Z., Odoj, R. 2009. An improved method for the non-destructive characterization of radioactive waste by gamma scanning. Appl. Radiat. Isot., 67,1897–1903. https://doi.org/10.1016/j.apradiso.2009.05.017.

Barbour, A.D., Holst, L., Janson, S. 1992. Poisson Approximation. Oxford University Press (1992), 277 pp., ISBN 0-19-852235-5.

Betancourt, M. 2018. A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

Boden, S. 2021. Open geometry HRGS measurements performed within the H2020 CHANCE project: ISOCS & FRAM. I-1715 SCK CEN / CHANCE report.

Borella, A., Boden, S., Bruggeman, C., Rogiers, B., Smets, S., Valcke, E. 2021. Neutron coincidence measurements and Monte Carlo modelling of waste drums containing reference nuclear material. EPJ Web of Conferences 253, 07001. https://doi.org/10.1051/epjconf/202125307001.
Borella, A., Rossa, C. 2022. A data analysis suite for neutron time correlation measurements. BLG-2940 SCK CEN report. [https://publications.sckcen.be/portal/files/7735110/BLG_2940.pdf]

Buecherl, T., Kaciniel, E., Lierse, C. H. 1998. Synopsis of gamma scanning systems - comparison of gamma determining systems and measuring procedures for radioactive waste packages. European Network of Testing Facilities for the Quality Checking of Radioactive Waste Packages, Report WG-A-01, September 1998. [http://www.en-trap.eu/doc/gammasySynopsis.pdf]

Buecherl, T., Rummel, S., Kalthoff, . 2021. A Bayesian method for the evaluation of segmented gamma scanning measurements – Description of the principle. Nucl. Instr. and Meth. in Phys. Res. A, 10, 165887. https://doi.org/10.1016/j.nima.2021.165887.

Carasco, C. 2021. Coupling gamma ray spectrometry and tomography in a Bayesian frame. Nuclear Inst. and Methods in Physics Research, A, 990, 164985. https://doi.org/10.1016/j.nima.2020.164985.

Carasco, C., Pérot, B., Guetos, O., Morales, F., Loridon, J., Evrard, S., Mathonat, S., Andrea, F., Narbot, J. C., Mettan, X., Laloy, E., Rogiers, B., Kubinski, W. R. Combination of neutron, gamma and calorimetry measurements within a Bayesian frame. Under review.

Clément, A., Saurel, N., Perrin, G., Gombert, N. 2021. Bayesian approach for multi gamma radionuclide quantification applied on weakly attenuating nuclear waste drums. IEEE Transactions On Nuclear Science, 68 (9). 2342–2349. https://doi.org/10.1109/TNS.2021.3100863.

Curie, L.A. 1968. Limits for qualitative detection and quantitative determination. Application to radiochemistry. Anal. Chem., 40(3), 586–593. https://doi.org/10.1021/ac60259a007.

Dierckx, R., Hage, W. 1983. Neutron signal multiplet analysis for the mass determination of spontaneous fission isotopes. Nuclear Science and Engineering, 85(4), 325–338. https://doi.org/10.13182/NSE83-3.

Dillon, J.V., Ian Langmore, Dustin Tran, Eugene Brevido, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matthew D. Hoffman, and Rif A. Saurous, 2017. Tensorflow distributions. CoRR, abs/1711.10604.

Frosio, T., Menaa, N., Duchemin, C., Riggaz, N., Theis, C. 2020. A new gamma spectroscopy methodology based on probabilistic uncertainty estimation and conservative approach. Appl. Radiat. Isot., 155, 108929. https://doi.org/10.1016/j.apradiso.2019.108929.

Gelman, A.G, and D. N Rubin (1992), Inference from iterative simulation using multiple sequences, Statistical Science,7, 457–472.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari A., and Rubin, D.B., 2014. Bayesian Data Analysis, Third Edition, Chapman & Hall/CRC.

Genie 2000 Spectroscopy Software. Customisation Tools Manual 9230847F 3/01. Mirion Technologies (Canberra), Inc. [https://www.mirion.com/products/genie-2000-gamma-analysis-software]

Golding, N., 2019. ggdist: Visualizations of Distributions and Uncertainty. R package version 2.4.0, [https://doi.org/10.5281/zenodo.3879620].

Kim, J.H., Hornibrook, C., Yim, M.-S. 2020. The impact of below detection limit samples in residual risk assessments for decommissioning nuclear power plant sites Journal of Environmental Radioactivity, 222, 106340. https://doi.org/10.1016/j.jenvrad.2020.106340.

Kirkpatrick, J. M., Venkataraman, R., Young, B. M. 2013. Minimum detectable activity, systematic uncertainties, and the ISO 11929 standard. J. Radioanal. Nucl. Chem., 296, 1005–1010. https://link.springer.com/article/10.1007/s10967-012-2083-5.

Knoll, G.F. 2010. Radiation detection and measurement, , 4th Edition. Wiley.

Krings, T., Mauerhofer, E. 2011. Reconstruction of the activity of point sources for the accurate characterization of nuclear waste drums by segmented gamma scanning. Appl. Radiat. Isot., 69(6), 880–889. https://doi.org/10.1016/j.apradiso.2011.02.009.

Laloy E., Rogiers B., Bielen A., Boden S. 2021. Bayesian inference of 1D activity profiles from segmented gamma scanning of a heterogeneous radioactive waste drum. Appl. Radiat. Isot., 0969-8043, 175. https://doi.org/10.1016/j.apradiso.2021.109803.
Marijuàn, M.J., Quiñones, J., Suàrez-Navarro, J.A. Garcia-Toraño, E. 2017. Efficiency calibration and validation of a system for the assay of radioactive waste drums containing homogeneous material. J. Radioanal. Nucl. Chem., 311, 513–518. https://doi.org/10.1007/s10967-016-4939-6.

Martens, B.R, Filss, P. 1990. Determination of the radionuclide inventory for waste packages using gamma-ray-detectors of low beam widths. Commission of the European Communities (CEC), Topical Meeting on Non-Destructive Assay of Radioactive Waste, EUR—12890. https://inis.iaea.org/search/search.aspx?orig_q=RN:22067014

MCNP6 Users Manual - Code Version 6.1.1 beta, LA-CP-14-00745, June 2014.

Michel-Sendis, F. et al. 2017. SFCOMPO-2.0: An OECD NEA database of spent nuclear fuel isotopic assays, reactor design specifications and operating data. Annals of Nuclear Energy 110, 779–788. https://www.sciencedirect.com/science/article/pii/S0306454917302104. https://www.oecd-nea.org/jcms/pl_21515/sfcompo-2-0-spent-fuel-isotopic-composition.

Multi-Group Analysis Software. https://www.mirion.com/products/mga-multi-group-analysis-software

Neal R (2011). MCMC Using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G.L. Jones, X.L. Meng (eds.), Handbook of Markov Chain Monte Carlo, pp. 116–162. Chapman & Hall/CRC.

Patra, S., Agarwal, C. 2019. Segmented gamma-ray assay of large volume radioactive waste drums containing plutonium lumps. Appl. Radiat. Isot., 153, 108827. https://doi.org/10.1016/j.apradiso.2019.108827.

Peterson, T.L. 2020. patchwork: The Composer of Plots, https://patchwork.data-imaginist.com/

Plummer, M., Best, N., Cowles, K., Vines, K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News, 6(1), 7–11. https://journal.r-project.org/archive/

Python Software Foundation. 2020. http://www.python.org

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

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

Sampson, T.E., Kelly, T.E., Vo, D.T. 2003. Application guide to gamma-ray isotopic analysis using the FRAM software, LA-14018, Los Alamos.

SOCR/UCLA. http://wiki.stat.ucla.edu/socr/index.php/AP_Statistics_Curriculum_2007_Limits_Norm2Poisson

Wickham et al. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686.