A reliable absolute and relative bayesian method for nuclear decommissioning: low-level radioactivity detection with gamma-ray spectrometry

Hanan Arahmane, Jonathan Dumazert, Eric Barat, Thomas Dautremer, Nicolas Dufour, Frederick Carrel, Maugan Michel, Frederic Laine

A Reliable Absolute and Relative Bayesian Method for Nuclear Decommissioning: Low Level Radioactivity Detection with Gamma-Ray Spectrometry

Hanan Arahmane, Jonathan Dumazert, Eric Barat, Thomas Dautremer, Nicolas Dufour, Frédérick Carrel, Maugan Michel, Frédéric Lainé

***Abstract*— In this paper, we aim at measuring a low-activity uranium contamination deposited on concrete surfaces of a nuclear facility and presenting varying enrichment levels. Considering this application field, CEA LIST has developed an original approach, combining gamma-ray spectrometry based on high-purity germanium (HPGe) measurements and specific Bayesian algorithms. This methodology gives access to an indirect surface activity estimation, assuming that the ratio between the number of alpha particles to be quantified and the number of detected gamma rays is known. The Bayesian approach characteristic property is able to inject a representation of the physical context in the form of a probabilistic *a priori*. It enables to improve the trade-off between the true detection rate (TDR) and the false alarm rate (FAR) at low count rates and takes benefit of a richer time-energy information structure than the algorithms used in conventional detection procedures. The performance evaluation and characterization of Bayesian statistical tests is performed using classical receiver operating characteristic (ROC) curves by comparison to frequentist hypothesis tests. Results indicate that the Bayesian approach has a superior detection performance for low-activity uranium contamination compared to the frequentist approach. The estimated gain is contained between $30\,\%$ and $50\,\%$, considering a variable or stable radiological background. The Bayesian approach offers the best trade-off between the TDR, FAR and the response time and is compatible with the user's requirements.**

***Index Terms*— Bayesian hypothesis test, decommissioning, gamma-ray spectrometry, surface activity of uranium in concrete, receiver operating characteristic (ROC) curves.**

I. Introduction

As any industrial facility, nuclear installations, have a finite lifetime. At the end of this activity period, decommissioning and dismantling operations come into play. Decommissioning, as a general concept, involves tasks such as dismantling of plant, decontamination of structures, nuclear waste management, etc. Amongst the various technical challenges of decommissioning is to carry out accurate radioactivity measurements on a large range of wastes. In this paper, we focus on the detection of low-activity uranium contamination on concrete surfaces, presenting varying enrichment levels. Our surface activity limit will be taken as $500\ \mathrm{Bq}/0.1\ \mathrm{m}^2/\Omega, \Omega = 2\pi \times sr$, where Ω is the solid angle of emission of the source. This choice is related to the maximal surface activity defined by the French Nuclear Safety Authority (ASN).

Alpha, beta and gamma emissions are three distinct signatures [3-5] for measuring a potential radiological contamination involving uranium isotopes; namely, uranium 234, 235 and 238 (U-234, U-235, U-238). Alpha emission has the advantage of being homogenous with the decommissioning criteria. However, the large linear energy transfer of alpha particles and concrete surface irregularities, leading to different source-detector distances, are severe limitations for accurately quantifying uranium activity. The beta and gamma measurement can be theoretically converted to alpha equivalent surface activity provided that we have prior knowledge of the isotopic uranium composition. However, for high U-235 enrichment ($\geq 5\,\%$), beta activity is reduced and thus detection limits are increased for the same alpha activity. Compared to alpha and beta particles, gamma rays appear less affected by surface concrete states and enrichment level of U-235.

Scintillators and semiconductors are two basic detectors used for gamma-ray spectrometry that enable the identification and/or quantification of radionuclides by analysis of the gamma-ray energy spectrum [6]. Detection limits increase with background signal and detector resolution. Gamma-ray spectrometry based on HPGe detector (HPGe) is the best available solution to detect low uranium activity because of its high energy resolution [7]. This figure of merit of HPGe, indeed, allows to define narrower regions of interest (ROI) around selected gamma-ray peaks in order to increase signal-to-noise ratio (SNR) for low activity measurement [2], [8, 9]. Nevertheless, high detection limits under low SNR (typically in the order of 1 over 1s), are still observed with traditional algorithmic approaches, which forms a major limitation in the application field of nuclear decommissioning.

To tackle the constraints of standard decision approaches, and the limitations encountered with low count rates in the context of radioactive detection, several methods have been proposed. Such methods included the use of Bayesian inference for nuclide ratio computation and source identification from high-resolution gamma-ray spectra [7]. The development of a sequential Bayesian approach for the detection and identification of radioactive materials allowed to reduce uncertainties at low count rate [10, 11]. Recently, specific and performing frequentist hypothesis tests were used with bilateral, unilateral, and either absolute or relative properties for the detection

Hanan Arahmane, Eric Barat, Thomas Dautremer, Nicolas Dufour, Frédérick Carrel, Maugan Michel and Frédéric Lainé are with Université Paris-Saclay, CEA, List, F-91120 Palaiseau, France (e-mail: hanan.arahmane@cea.fr, eric.barat@cea.fr, thomas.dautremer@cea.fr, nicolas.dufour3@cea.fr, frederick.carrel@cea.fr, maugan.michel@cea.fr, frederic.lainé@cea.fr).

Jonathan Dumazert is with CEA-DAM, DIF, 91297 Arpajon, France (e-mail: jonathan.dumazert@cea.fr).

of low-activity uranium contamination on nuclear facilities' concrete surfaces [2]. These tests allowed to perform a computation of the measurement statistical power, and thus to define the confidence interval of the true detection rate (TDR) and false alarm rate (FAR). A basic bilateral solution consists in introducing the relative test of Kolmogorov-Smirnov (KS) sensitive to any variation in the spectrum shape. We can, however, expect that the trade-off between TDR and FAR for the KS test will be degraded compared to those obtained for a unilateral test, the latter being only sensitive to an increase in the counting statistics. For example, the KS test yields a TDR $= 55\,\%$ for an enrichment level of $1\,\%$wt U-235 and an integration time $t = 500\,s$ at the above cited surface activity. For that reason, a second solution was proposed, using some prior knowledge of the expected signal. This enables to use the ROI of the signal using to define an absolute and unilateral test, based on the cumulative function of a Negative Binomial law (CNB). From these features, we could expect that the CNB test would give access to a better trade-off between TDR and FAR than the KS test. Considering the same enrichment level as KS ($1\,\%$wt U-235) and an integration time $t = 500\,s$, a TDR= 94.6 % was indeed obtained. However, the absolute CNB test has the weakness of being impacted by any significant variation ($\geq 5\,\%$) associated with the background signal intensity. Hence, it was necessary to introduce an alternative version of the test, taking into account the dispersion of the counting intensity [2]. To this aim, a relative and unilateral test, based on the cumulative function of a Beta Binomial law (CBB), was developed. The CBB test uses the same ROI as the CNB test, but also another ROI where no signal from the contamination is expected, called "control region." Against a non-stationary background, the CBB test yielded a better trade-off between TDR and FAR, by 10% compared to the CNB test. Considering the same enrichment level as KS and CNB for an integration time $t = 3000\,s$, the CBB test yielded a TDR $= 99.9\,\%$. It thus appears that the CBB test gives access to an acceptable trade-off between sensitivity to the uranium signal, and vulnerability to the characteristics of the radiological background. Nevertheless, in the presence of a very low SNR (in order of 10^{-1} over $1s$), or of a background noise whose properties (intensity, shape) vary significantly, the previous approaches are strongly impacted. As a result, an acceptable trade-off between TDR and FAR can no longer be obtained within the time constraints related to decommissioning operations. Moreover, considering a nuclear facility with a total surface to assess of a hundred thousand square meters, a measurement time of $3000\,s/m^2$, as required by such approaches, would result in a total investigation time of 20 years if a $100\,\%$ scanning is required. For these reasons, we are in need of alternative methods. The aim of this study is to introduce a new algorithmic approach, also based on statistical hypothesis tests, but reducing the measurement time while maintaining a sustainable trade-off between TDR and FAR.

In this framework, a candidate approach consists in modeling the contamination detection scenario using *a priori* knowledge of the expected signal, this to define the conditions for rejecting the non-contamination hypothesis. Roughly speaking, a statistical test is a mathematical procedure to decide, over a given acquisition time, whether we are:

- only in presence of the background radiation previously recorded over a reference acquisition time (a scenario we refer to as the "null hypothesis," labeled H_0),
- or whether an additional signal, in our case arising from uranium contamination, was detected (scenario named the "alternative hypothesis," H_1).

The availability of a physical model of the gamma-ray source term associated with uranium surface contamination, as well as multichannel spectra at the output of the HPGe spectrometer, indeed allows to build a hypothesis test incorporating a physics model of the observed phenomenon. CEA LIST then developed original Bayesian hypothesis tests with either absolute or relative properties. These are competing with frequentist CNB (absolute) and CBB (relative) tests. The Bayesian Absolute Multivariate Poisson Mix (BAM) test corresponds to a Bayesian alternative to the CNB test, while the Bayesian Multivariate Poisson Mixing (BRM) test can be used instead of the CBB test. The Bayesian approach characteristic property constitutes a way to inject a representation of the physical context in the form of probabilistic *a priori*. In this study, the performance evaluation of Bayesian statistical tests will still be performed by using classical receiver operating characteristic (ROC) curves, as well as the study of the TDR variation as a function of the integration time against a preset FAR. Our Bayesian approach, moreover, was based on *a priori* vectors built from the coupling of experimental data, acquired in a real nuclear facility, and simulated data that were obtained by a Monte Carlo method.

The paper is organized as follows: Section II presents an outline of some relevant works on this topic; Section III covers the technical description of the radiation detection problem; Section IV describes the formulation of our Bayesian proposed approach, and summarizes the theoretical foundations of the CNB and CBB frequentist tests; Section V presents the parameterization of the Bayesian BAM and BRM tests; and Section VI details the simulation results and performance evaluation of the different hypothesis tests.

II. Related Work

There are several significant contributions based on a Bayesian approach for detection and identification purposes in the frame of nuclear applications. We can identify two types of approaches reported in the scientific literature, each of them corresponding to dedicated applications.

The first approach aims to perform a complete deconvolution of spectra containing a large set of isotopic signatures. The main purpose is then to detect and quantify the activities of several radionuclides. In this regard, Zähringer and Kirchner (2008) [12] applied a Bayesian approach for detecting underground explosions prohibited by the Comprehensive Nuclear-Test-Ban Treaty (CTBT) in two specific cases: exclusion of Ba-140 when La-140 is measured, and identification of xenon isotopes (Xe-131m, Xe-133, Xe-133m and Xe-135). In both cases, the authors based their studies on the ratio calculation of radionuclide activity concentrations that are close to, or below the conventional detection limit. The results they obtained demonstrated that the Bayesian approach enabled to improve the gamma-ray spectrum analysis, as subsequently the decision-making process. Such results paved the way to an implementation of the Bayesian approach for xenon detection. Rivals et al. (2012) [13] developed a Bayesian method using specific priors for the detection of radioxenon isotopes in noble gas measurement stations of the International Monitoring System (also related to the Comprehensive Nuclear-Test-Ban Treaty Organization). Results showed that the detection capability benefited from the estimation of the radioactivity prior density probability, the confidence interval for the true radioactivity level, and the probability of no radioactivity being present.

The second approach aimed at detecting, and identifying the presence of an unknown radiation source. Candy et al. (2008) [10, 11] used such an approach for the detection and identification of radioactive contraband. The aim of their work was to improve the detection accuracy of nuclear contraband. Compared to conventional methods, the proposed solution enabled to discriminate more efficiently the signal from background level. The authors

demonstrated that, in a Bayesian framework, exploiting a richer time-energy information structure allowed to improve the detection limits of the radiation monitor in a low count rate configuration. From a different prospect, Bukartas et al. (2019) [14] used the likelihood of the Bayesian model to locate and estimate the activities of orphan radioactive sources, namely Cs-137, Ba-133 and I-131 in the activity range of 180–470 MBq, using a mobile gamma-ray spectrometer. Their results showed that the method allowed to simplify the search for lost gamma sources, as well as to quantify the source activity within 50 % of its real activity, and locate it within a 27 % spatial uncertainty.

In this perspective, the availability of a physical model of uranium surface contaminations, as well as of high-resolution spectra from an HPGe detector, encouraged us to develop novel Bayesian hypothesis tests to compete with the frequentist ones presented in [2]. We recall that the main purpose of this study is to ensure, with an acceptable measurement integration time, an operational trade-off between TDR and FAR in the context of nuclear decommissioning.

III. Technical Description of the Radiation Detection Problem

Before we proceed to an application of the Bayesian approach to the detection problem at hand, and the study of its compared merits with respect to a frequentist approach, we need to describe formally the uranium gamma-ray peaks of interest, alpha/gamma transfer function, and the expected radiological background. Such a quantitative description forms the topic of the current section.

A. The Gamma-Ray Signature of Uranium

As we mentioned in our introduction, we aim at detecting low levels of uranium contamination inside nuclear facilities that are being dismantled. In order to identify the presence of uranium by means of a gamma-ray spectrometer, we take advantage of the prominent photon emission lines from the decay chains of U-235 and U-238. An extensive list including tens of such lines was taken from the Nucléide-Lara database [15], and ranges from 10 keV up to 3 MeV. Nonetheless, the exploitability of any given emission line in the context of low SNR detection is governed by:

- The full energy peak efficiency (FEPE) of the detector, which itself depends mainly on the energy of the line (decreasingly in the absence of any window), on the volume of the sensor (increasingly), and on the shielding effect from the entrance window protecting the crystal;
- The amplitude of the continuum, or full energy peaks (FEP) adjacent to those of interest, from the radiological background.

For the study of concept we present in this paper, we selected an n-type coaxial HPGe diode commercialized by ORTEC under reference GMX35P4 [16]. The spectrometer has a 35 % relative efficiency, which represents an acceptable trade-off between handiness and FEPE at high energy (> 1 MeV). It is mounted with a 2 mm aluminium window to protect the crystal from irregularities that are to be expected when dealing with *in-situ* measurements. In such conditions, we recognized from prior acquisitions inside walls [2] that the gamma rays of interest for our detection case can be narrowed down to:

- From the U-235 chain, one main line at 185.7 keV, and one additional triplet at $\{143.8;\ 163.4;\ 205.3\}$ keV;
- From the U-238 chain, one main line at 1001.0 keV.

Moreover, let us consider the situation where the surface activity is in the vicinity of the detection limits stated in our introduction. In this case, we showed in [2] that all rays emitted with an energy superior to 1 MeV, will induce no significant count rate increase above the radiation background inside walls. As a consequence, we shall be able to use the entire spectral range above 1 MeV as a uranium signal-free control region when implementing relative detection tests (*cf.* Section IV below). As it will be discussed with further details throughout Sections IV to VI, the definition of selective ROI inside the detection spectrum over $X = [0 - 3000]$ keV, is key to improve the trade-off between TDR and FAR when dealing with low contamination-to-background scenarios. From the abovementioned arguments, such ROI will be centered on the main useful lines from U-235 and U-238. To set the spectral width of the ROI, we used the resolution calibration of the spectrometry chain that was carried out with a Eu-152 source. All analyses done, the ROI for uranium detection read as follows:

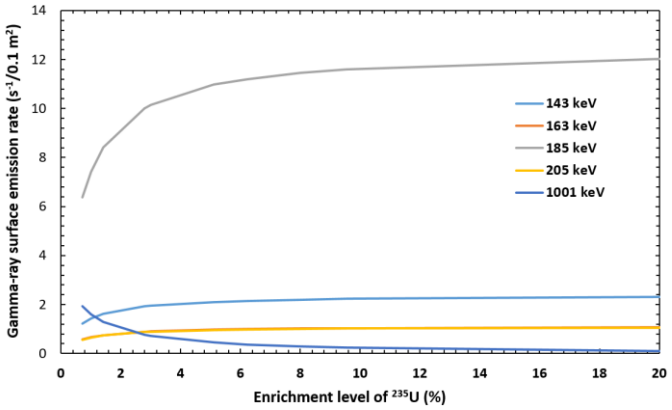
- $Z_1 = [140 - 148\,]$ keV, $Z_2 = [160 - 166]$ keV, $Z_3 = [182 - 188]$ keV, $Z_4 = [202 - 208]$ keV, and $Z_5 = [998 - 1004]$ keV for the signal regions in absolute and relative detection tests ;
- $W =]1004 - 3000]$ keV for the control region in relative detection tests.

B. Simulation of the Detector Response to the Minimum Surface Activities to be Detected

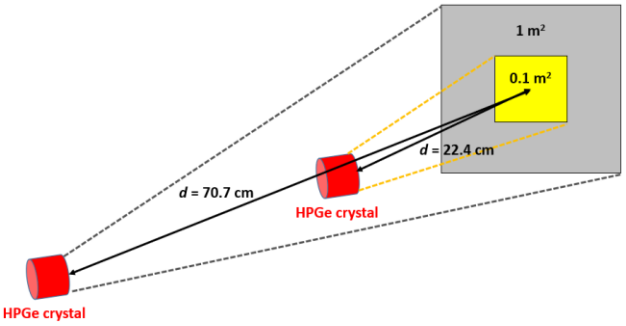
We state here the two decommissioning criteria that serve as minimum surface activities to be detected (MSAD) in presence of a uranium contamination:

- MSAD1: $2000\ \mathrm{Bq/m^2}/2\pi$;
- MSAD2: $500\ \mathrm{Bq}/0.1\ \mathrm{m^2}/2\pi$.

In order to assess quantitatively the performance of the dedicated algorithms that will be described throughout Section IV, we must primarily have knowledge of the spectral response, expected from a contamination corresponding to either MSAD1 or MSAD2, at the detector level. The first step is thus to build a transfer function relating the emission rates υ_γ (expressed in s^{-1}) of the five lines listed in Subsection III.A to MSAD1 and MSAD2. Now, the total activity A_U of the uranium contamination is expressed as the sum of the activities of the three main isotopes: $A_\mathrm{U} = A_{^{234}\mathrm{U}} + A_{^{235}\mathrm{U}} + A_{^{238}\mathrm{U}}$. Consequently, the transfer function will not only depend on the intensity I_γ of the line with respect to the decay of its mother isotope, but also on the level of enrichment (expressed as a mass fraction, wt%) in U-235. Under the assumption that secular equilibrium exists between U-238 and Pa-234m, whose decay gives rise to the 1001.0 keV gamma-ray, abaci of υ_γ were constructed for both MSAD, and mass fractions in U-235 varying between 0.7 and 8 wt%. Fig. 1. a) presents an example of such abaci for a uranium surface activity corresponding to MSAD2.



(a)



(b)

Fig. 1. Abacus of gamma-ray surface emission rate as a function of enrichment for a uranium contamination with a total surface activity of 500 Bq/(0.1 m^2) (a), and Simulation study diagram of homogenous surface activities MSAD1 and MSAD2 (b).

The next step is to determine the spectral count rate (s^{-1}) expected at the output of the spectrometry chain from a uranium contamination of a given surface activity, and enrichment level. To this end, we built a numerical model of the HPGe diode using the MCNP6.1 software [17]. The simulated spectral response of the detector is provided by tally 8 in photon-electron (PE) mode of the code. This macroscopic response was then calibrated experimentally through a series of measurements of uranium reference samples whose U-235-enrichment varied between 0.7 and 8 wt%. To this end, spectra were acquired on 8192 channels, for an energy range lying between 0 and 3 MeV approximatively. The signal was shaped using a symmetrical trapezoidal filter with a 2 s rise time and a 4 s plateau time. The procedure for this efficiency calibration was detailed in previous work [2], and showed noticeably that simulated and experimental FEP agreed in amplitude within two standard deviations for U-235, and three standard deviations for U-238. This agreement between simulated and experimental with uranium samples was also consolidated for several source-detector distances between 10 cm and 1 m. These distances correspond to lower and upper boundaries in target measurement configurations as we shall see below. As a result, and for any distance between 10 cm and 1 m, we used directly, and without any experimental calibration coefficient the tally 8 output $Y(E_\gamma)$ from MCNP6.1. This allowed us to compute the expected count rate $S(E_\gamma)$ (s^{-1}) under the FEP centered on energy E_γ as:

$$S(E_\gamma) = Y(E_\gamma) \cdot \upsilon_\gamma \left(A_U, \text{wt\%}_{^{235}U}\right) \quad (1)$$

The measurement protocol is simulated as realistically as possible by modelling the source term as a photon source homogeneously spread on the surface of a concrete block. The block dimensions are taken as $100 \times 100 \times 25$ cm^3. The photon flux corresponding to MSAD1 is generated over the entire square surface of the block, whereas the flux associated with MSAD2 is generated over a 0.1 m^2 square at the center of the previous one.

In both cases, the HPGe diode model is located so that its symmetry axis intersects horizontally the center of the photon source. As far as the distance d between the crystal front surface and the center of the surface source is concerned, we have set, following standard measurement protocol, $d = \frac{a\sqrt{2}}{2}$, where a is the length of side of the square emission surface that must be enclosed within the viewing cone of the detector. These settings lead to source-detector distances $d = 70.71$ cm, and $d = 22.36$ cm for MSAD1 and MSAD2 respectively. It should be noted, at this point, that the fact that MCNP6.1 tally 8 outputs could be used without further correction to compute the expected spectral count rates at the detector level, was only validated experimentally for "point-like," bulk uranium sources. In order to ascertain that this calibration result holds in the presence of a homogenously and extensively dispersed source, there exists a need for the manufacturing and characterization of uranium surface sources. Such sources would have to be laid down over a concrete bearing, and protected by as thin an entrance window as possible. As the homogeneity and radioactivity of said reference samples should be at least as metrogically reliable as the ones of the uranium sources that we used, this was left as an outlook in the course of the present study. Once we have shown that the influence of the U-235-enrichment levels was accounted for, we can indeed rely on the robustness of the particle-transport code to yield expected spectral responses. The latter are indeed accurate enough to investigate the interest of a Bayesian approach to large-surface decommissioning using gamma-ray spectrometry.

Regarding the U-235-enrichment levels that are simulated, we subdivided the decommissioning problem into three subunits (SU) of the same nuclear facility to be dismantled. A uranium contamination in any of these subunits would have an average expected level, determined from prior knowledge of the nuclear process that was taking place when the plant was in operation, or on the basis of previous sample measurements. For the sake of demonstration, we set these average enrichments to be 0.62 wt%, 2.03 wt%, and 6.84 wt% for SU1, SU2, and SU3 respectively. The spectral responses of the detector to these enrichments, estimated numerically, will represent the *a priori* signatures of uranium contaminations inside one SU when implementing Bayesian statistical tests (Sections IV to VI). Now, the enrichments are likely to vary widely around their average values. Consequently, instead of these *a priori* signatures, we used typical U-235 mass fractions to build tested spectra when simulating a real-time measurement inside the plant. These typical enrichments read as: Natural (0.7 wt%) and 1 wt% for SU1, 1 wt% and 3 wt% for SU2, and 3 wt%, and 8 wt% for SU3.

The simulation-based study of the main FEP response to the gamma-ray at 185.7 KeV ray shows that:

- For a photon emission rate, and measurement geometry corresponding to MSAD1, $S(185.7\ \text{KeV})$ varies between 0.01 s^{-1} (Nat wt%) and 0.03 s^{-1} (8 wt%) ;
- For a photon emission rate, and measurement geometry corresponding to MSAD2, $S(185.7\ \text{KeV})$ varies between 0.03 s^{-1} (Nat wt%) and 0.06 s^{-1} (8 wt%).

As the expected count rate for this line is in order of 10^{-2} s^{-1}, we see that the numerical study of the detection procedure that will be detailed throughout Section VI, will suppose integration times of the counting greater than or equal to 100 s. In the rest of this paper, the study of principle will be limited to the source terms, and measurement geometry corresponding to MSAD2, i.e., 500 Bq/ 0.1 m^2/2π to be detected within SU1, SU2, and SU3. Indeed, the purpose of this study is to indicate that, for a given minimum surface activity to be detected against natural background inside walls, and with a given measurement chain, some added value may be found in working within a Bayesian framework. Moreover, we privileged MSAD2 as a factor of merit in our investigations as the most relevant for the authorities in charge of supervising the decommissioning. This factor indeed encapsulates more features from the difficulties in dealing with "hot spots" than MSAD1, as an activity in the same order of magnitude is concentrated inside a ten-time smaller surface.

C. Quantitative Study of Background Radiation Spectra

The detection of low-level uranium contamination is to be performed within a closed environment, essentially made of concrete walls, floors, and ceilings. Now, in such an environment, the gamma-ray background is dominated by the signatures of K-40, and the

decay chains of U-238 and Th-232 [18]. It follows that we can expect some repeatability, at least as far as the normalized distributions (in other words, the shapes) of background spectra are concerned. Such an assumption with, however, need to be verified.

In order to challenge these assumptions, a first measurement campaign was conducted inside the three SU to be decommissioned. For each of SUi, with $i \in \{1; 2; 3\}$, acquisitions were carried out in two separate spots, the counting associated to these acquisitions being labeled m_{SUi1} and m_{SUi2} respectively. The empirical mean of m_{SUi1} and m_{SUi2}, for the three subunits, will serve:

- First, as a reference when dealing with frequentist detection tests;
- Second, as a prior signature of the background in the framework of Bayesian detection tests.

To quantify the variability of the background radiation signal, inside a given SUi, and over a given spectral ROI, we introduce the relative deviation between both measured spots as:

$$\forall\, i \in \{1; 2; 3\},\ \varepsilon_{SUi}[ROI] = \frac{m_{SUi1}[ROI] - m_{SUi2}[ROI]}{m_{SUi2}[ROI]} \qquad (2)$$

This factor of merit thus quantifies the amplitude variation between non-normalized countings $m_{SUi1}[ROI]$ and $m_{SUi2}[ROI]$. If we want to study exclusively the variability of the background radiation signal in terms of shape, we shall use instead the relative deviation defined as:

$$\forall\, i \in \{1; 2; 3\},\ \varepsilon'_{SUi}[ROI] = \frac{m'_{SUi1}[ROI] - m'_{SUi2}[ROI]}{m'_{SUi2}[ROI]} \qquad (3)$$

where $m'_{SUi1}[ROI]$ and $m'_{SUi2}[ROI]$ are the normalized countings acquired inside SUi, and integrated over spectral range ROI.

Let us note that other merit factors accounting for background vulnerability could have been selected, notably one using the summation of normalized (m'_{SUi1} and m'_{SUi2}) and non-normalized (m_{SUi1} and m_{SUi2}) countings in (3) and (2) respectively. The motivation for our choice comes from a practical implementation standpoint, envisioning a "two-step" process:

- First, a reference background measurement is carried out inside a room, or an extended open area that is known to bear no contamination;
- Second, a screening of unknown surfaces is carried out to find potential uranium contaminations under the assumption that the background does not vary significantly with respect to the reference measurement.

We have checked, moreover, that should we use the first campaign as the reference and the second one to compute the deviations, the orders of magnitude of the merit factors, especially the maximum absolute values of ε_{SUi} and ε'_{SUi} would not depart significantly from the ones we discuss below. These are the elements that we are interested in.

Tables I and II below present respectively the relative deviations $\varepsilon_{SUi}[ROI]$ and $\varepsilon'_{SUi}[ROI]$. They were calculated as a function of SU and ROI, with the same notations as in Subsection III.A to denote the spectral ROI.

TABLE I
RELATIVE DEVIATIONS $\varepsilon_{SUi}[ROI]$ BETWEEN NON-NORMALIZED BACKGROUND COUNTINGS AS A FUNCTION OF SU AND ROI

ROI \ SU	SU1	SU2	SU3
X	$-$ 0.6 %	$-$ 0.7 %	$-$ 0.2 %
Z_1	$-$ 0.4 %	$-$ 1.6 %	1.0 %
Z_2	$-$ 1.8 %	$-$ 0.2 %	$-$ 0.5 %
Z_3	$-$ 0.1 %	$-$ 0.4 %	$-$ 1.8 %
Z_4	$-$ 1.4 %	$-$ 0.5 %	0.1 %
Z_5	4.2 %	5.1 %	$-$ 2.7 %
W	$<$ 0.1 %	$<$ 0.1 %	$<$ 0.1 %

TABLE II
RELATIVE DEVIATIONS $\varepsilon'_{SUi}[ROI]$ BETWEEN NORMALIZED BACKGROUND COUNTINGS AS A FUNCTION OF SU AND ROI

ROI \ SU	SU1	SU2	SU3
Z_1	1.0 %	$-$ 0.9 %	1.1 %
Z_2	$-$ 1.2 %	0.6 %	$-$ 0.4 %
Z_3	0.7 %	0.3 %	$-$ 1.7 %
Z_4	$-$ 0.8 %	0.2 %	0.3 %
Z_5	4.9 %	5.9 %	$-$ 2.6 %
W	$<$ 0.1 %	$<$ 0.1 %	$<$ 0.1 %

From Table I, we observe that the relative deviations do not exhibit any significant variation as a function of the SU nor the spectral ROI. The average value of $\bar{\varepsilon} = (-0.1 \pm 1.1)$ % over the entire table. This led us to consider the value of 1 % as representative of the mean expected amplitude variation of background spectra during the measurements. From Table I, we can also identify the value of 5 % as representative of the maximum expected amplitude variation of the background between reference and current acquisitions.

The same analysis was carried out with the data presented in Table II, and led to similar conclusions:

- We denote no significant variation of ε' as function of the SU nor the spectral ROI;
- The average value of $\bar{\varepsilon'} = (0.4 \pm 1.3)$ % over the entire table. Hence the value of 1 % may also be taken as representative of the mean expected shape variation of background spectra during the measurements;
- The value of 6 % may be identified as representative of the maximum expected shape variation of the background between reference and current acquisitions.

Eventually, the comparison between Table I and Table II shows that, upon the basis of this first acquisition campaign, there is no justification to consider the gamma-ray background as more stable with respect to its shape than to its amplitude. As a consequence, we shall have no *a priori* reason to favor relative statistical tests (i.e., based on the normalized distributions of the reference and tested spectra) over their absolute counterparts (i.e., based on the non-normalized versions of said spectra). Let us eventually underline the fact that, over all the acquisitions that were performed inside walls, the loss fraction due to dead time was always kept between 0.5 and 1 %, while no summation peaks could be separated from the diffusion continuum. As a result, no correction of both effects was implemented. Should the measurement take place against a level of background radioactivity that departs from the target configuration (including an abnormally high-dose-rate natural background, or the presence of an artificial gamma-ray source such as Cs-137 or Co-60), our model would need to be consolidated with respect to count losses, or spectral distortions resulting from dead time and pile-up. As we are primarily looking for low-surface activity contaminations inside a

conventional gamma-ray background, such refinements were left as an outlook in the course of this study.

D. Generation of Random Tested Spectra and Receiver Operating Characteristics

As we shall see with further details in Section IV, the detection procedure of a uranium surface contamination is bound to determine, after an observation time labeled t, whether:

- The detector response is due solely to background radiations;
- Or the response reveals the presence of an additional signal for U-235 and/or U-238.

Given a preset time t, the conditions that lead to accept one of these two statements as true governs the trade-off between the FAR and TDR. However, parameter t plays a critical role in said trade-off once a decision procedure has been formalized. Consequently, the first prerequisite to carry out a comparative study of competing procedures is to build a random generator of representative tested spectra where t appears as a variable.

In order to assess the accuracy of a detection procedure, at least to two spectra are needed:

- one representing the signature of the radiation background only;
- and one representing the joint signatures of background and uranium signals.

To generate a uranium-free tested spectrum inside a given SU, we can use, as the time-independent probability density, either:

- one of both spectra that were acquired inside the said SU, being identical to the one selected as the reference spectrum for the entire SU (in which case, the simulator does not account for any variation in the expectation of the background, but only for statistical fluctuations of countings due to the stochastic nature of radioactivity);
- a random and uniform mixture of both spectra that were acquired inside the said SU, one of which only being selected as the reference spectrum for the entire SU (the simulator then accounts for amplitude and shape variations in the expectation of the background as quantified throughout Subsection III.C).

In both cases, the reference background spectra that we analyzed in Subsection III.C will serve as time-independent expectations.

As far as the generation of a tested spectrum with the presence of a uranium signal is concerned, we used, as the time-independent probability density, the sum of a so-constructed background expectation and the simulated signature of MSAD2 according to Subsection III.B.

The numerical reconstruction of a radiation measurement is then performed in two steps:

- First, the relevant expectation (with or without the addition of the time-normalized uranium signature) is multiplied by the time variable t;
- Second, a random draw following a Poisson law [19] is carried out within the so-constructed time-dependent expectation vector.

For every predetermined measurement configuration (including SU, background variability, uranium enrichment level, integration time t), a number of N spectra is generated in the absence, and the presence of MSAD2, on the basis of which factors of merit FAR and TDR may be evaluated.

As abovementioned, any parameterization of a chosen detection procedure will result in different FAR and TDR factors. The FAR/TDR trade-off quantifies the specificity (against background) and the sensitivity (with respect to MSAD2) of the detection procedure. To study the influence of any test parameterization on this trade-off, we used classical ROC curves [20]. For the record, ROC curves are performance graphs of a complete detection chain, from radiation sensor to decision procedure, in the form of a plot of TDR as a function of FAR. The underlying variable in such a graph then represents the parameterization of the detection procedure. An ideal detection chain would yield a ROC curve passing through points $(0;0) - (0;1) - (1;1)$, whilst a random response chain would yield a line crossing points $(0;0)$ and $(1;1)$. As a consequence, the closest the "elbow" of such a curve gets to the point of coordinates $(0;1)$ in the (FAR, TDR) space, the better trade-off between specificity and sensitivity the detection chain is likely to achieve.

In the frame of this study, ROC curves were constructed using a Monte Carlo method: N random draws are carried out in the absence of uranium, and the same number N in the presence of uranium with the help of the abovedescribed generation procedure. For every parameterization of the detection procedure, we calculated:

- the number of positives in the presence of uranium signal, corresponding to the number of true detections labeled N_{TD};
- and the number of positives in the absence of uranium signal, corresponding to the number of false alarms N_{FA}.

We eventually derived, for every parametrization of a given procedure, the couple $\left(TDR = \frac{N_{TD}}{N}; FAR = \frac{N_{FA}}{N}\right)$, to which a one standard-deviation uncertainty, related to repeatability, is attached as:

$$\left(\sigma(TDR) = \sqrt{\frac{1}{N} \cdot TDR \cdot (1 - TDR)}\,; \sigma(FAR) = \sqrt{\frac{1}{N} \cdot FAR \cdot (1 - FAR)}\right) \quad (4)$$

The construction of ROC curves, as well as the generation of random tested spectra, were performed using a Python-based platform. The number of iterations $N = 1000$ is set so that $\sigma(TDR)$ and $\sigma(FAR)$ do not exceed 0.5 %.

The building blocks and performance criteria of the study having been presented, the next section will be dedicated to the description of specific detection procedures that will operate over these blocks and will be qualified on the basis of these criteria.

IV. Mathematical Formulation of Frequentist and Bayesian Statistical Tests

We dedicate this section to the formulation of the absolute and relative Bayesian tests that we introduce to detect U-235 and U-238 inside a closed, concrete-based environment. As these tests, that form the core of this paper, are competing solutions with respect the frequentist tests discussed in our previous work [2], we shall first restate the essential features of the latter ones to allow quantitative comparisons.

Now, depending on the way the spectral information is used (including the degree of modeling of hypothesis H_1), and the external conditions of measurement (notably the stationarity or non-stationarity of the radiological background, in amplitude as well as in shape), a variety of tests can be constructed, and different performance obtained.

A. Frequentist Absolute (CNB) and Relative (CBB) Tests

We start with a few definitions that will be useful for our discussion when dealing with the respective performance of statistical tests.

A hypothesis test is *absolute* when it is sensitive to any variation of counting rate in the spectral ROI its formulation relies upon. Such tests are then vulnerable to amplitude variations of the background signal. On the contrary, a test is said *relative* when its result is only affected by a change in shape of the spectrum. On the other hand, hypothesis tests are classified on the ground of their laterality. The

conformity of test data to the null hypothesis, at least when dealing with frequentist tests, is assessed through the calculation of a $p-$value, labeled p. This value is then compared to a decision threshold, labeled α, and directly related to the risk of false positive (or false alarm). In a bilateral test, the null hypothesis is rejected (and therefore a contamination is detected) whenever the modulus of the $p-$value is superior to α, whereas in a unilateral test it is rejected solely when the $p-$value exceeds α. Consequently, unilateral tests provide a possibility to discriminate a significant increase from a significant decrease of counting in the spectral ROI, both having potentially different physical meanings. In our previous work [2], we showed that reaching an acceptable trade-off between TDR and FAR, within an integration time below $3000\ s/m^2$ and a surface activity of $2000\ \mathrm{Bq}/m^2/2\pi$, was not achievable by means of a relative and bilateral statistical test.

The test that we used was the classical Kolmogorov-Smirnov (KS) test [2], [25], suited to detect any change in the shape of the spectrum. In other words, this test enables its user to determine whether two statistical distributions, in this case the reference and tested spectra, are identically distributed. The number of counts associated to the reference and tested spectra are classified in histograms with regular energy bins $\mathrm{E}_\gamma \in [0-3\mathrm{MeV}]$. Here, the $p-$value equals the maximum of the absolute bin-to-bin difference between both spectra after these have been normalized. For a fixed $\alpha-$risk level, the KS hypothesis test simply reads as:

- If $p > \alpha$, then hypothesis H_0 is rejected, and a uranium contamination is detected;
- Otherwise, H_0 is accepted.

Facing the limitations that we have highlighted regarding relative and bilateral tests such as the KS test, we then studied the performance of two alternative frequentist tests that are both unilateral.

The first one is a unilateral and absolute test, whose simplicity lies in the fact that its formulation only requires a total counting in the spectral ROI containing the main signatures of U-235 and U-238 (a region labeled Z according to the notation from Section III). We refer to this test as CNB [2] for it is, as we shall see, based on the cumulative (C) function of a negative binomial (NB) distribution.

Let T_{ref} and T_{test} be the acquisition times of the reference and test spectra respectively (the two needing not to be taken equal), $m_{ref}[Z]$ and $m_{test}[Z]$ the counts recorded in region Z over the two previously defined measurement periods respectively. The counting $m_{ref}[Z]$ is supposed to follow a Poisson law, whose parameter $\rho[Z]$ is supposed itself to be governed by a gamma distribution of parameters $(a; b)$. These parameters were set to represent the absence of any *a priori* on the counting intensity $\rho[Z]$.

The a *posteriori* law of $\rho[Z]$ knowing $m_{ref}[Z]$ is then a gamma distribution of parameters $\left(a + m_{ref}[Z]\,;\, b + T_{ref}\right)$, and the posterior predictive law of $m_{test}[Z]$, knowing $m_{ref}[Z]$ and supposing H_0 , is a negative binomial distribution of parameters $\left(\frac{T_{test}}{b+T_{ref}+T_{test}}\,;\, r = a + m_{ref}\right)$. The cumulative function $C[Z]$, associated to this negative binomial law, is calculated to be:

$$C[Z] = 1 - I_p(m_{test}[Z], r) \tag{5}$$

where I_p denotes the regularized incomplete beta function of parameters $m_{test}[Z]$ and r. The $p-$value of the hypothesis test is eventually defined as:

$$p = 1 - C[Z] \tag{6}$$

Just as the KS test, the CNB test, under a given $\alpha-$risk, is expressed as:

- If $p > \alpha$, then H_0 is rejected, and a contamination is detected;
- Otherwise, H_0 is accepted.

As a unilateral test, the CNB test remains at least as sensitive as a KS test with respect to the presence of an increase of counting in region Z, while being more specific against any decrease of counting in the same region. Such a decrease would indeed be attributable to a variation of the radiological background amplitude between the reference and the test acquisitions. However, as it is an absolute test, its specificity may be impinged when the background increases significantly over Z between the two said acquisitions. Provided such an amplitude variation prevails over the shape variation of the spectrum (in other words, that the background change may be approached as a scaling factor over the entire spectrum), then it may prove useful to use a relative counterpart of this CNB test.

The second detection procedure we implemented was therefore a unilateral and relative test we refer to as CBB [2], for being based on the cumulative (C) function of a beta binomial (BB) distribution. The CBB test, like the CNB test, relies upon two spectral acquisitions over T_{ref} and T_{test} respectively. As a relative test, however, the CBB test does not only make use of region Z for uranium detection, but also of a second region where no significant signal from uranium is expected. The latter is named "control region," and labeled W.

Using the same notations as before, $m_{ref}[W]$ is defined as the counting in the control region of the reference spectrum, and $m_{test}[W]$, the counting in the same region of the tested spectrum. The CBB test relies on the introduction of a random variable ρ, following a beta distribution of prior-free parameters $(a; b)$. The counting $m_{ref}[Z]$ is no longer supposed to follow a Poisson distribution, but a binomial distribution of parameters $\left(m_{ref}[Z] + m_{ref}[W]\,;\rho\right)$. The *a posteriori* law of ρ, knowing $m_{ref}[Z]$ and $m_{ref}[W]$, is then a beta distribution of parameters $\left(a + m_{ref}[Z]\,;\, b + m_{ref}[W]\right)$.

Under hypothesis H_0, the $m_{test}[Z]$ counting associated to the tested spectrum follows a binomial law of parameter ρ. The predictive posterior law of $m_{test}[Z]$ knowing $m_{ref}[Z]$, $m_{ref}[W]$ and $m_{test}[W]$, is a beta binomial law, with a number of draws $k_{test} = m_{test}[Z] + m_{test}[W]$ and parameters $\left(a + m_{ref}[Z]\,;\, b + m_{ref}[W]\right)$. Finally, the cumulative function associated to this beta binomial law is expressed as:

$$C[Z,W] = \frac{\mathrm{B}\left(a+m_{ref}[Z]+m_{test}[Z]+1, b+m_{ref}[W]+m_{test}[W]-1\right)\cdot F_{k_2}\left(\vec{a},\vec{b},m_{ref}[Z]\right)}{\mathrm{B}\left(a+m_{ref}[Z], b+m_{ref}[W]\right)\cdot \mathrm{B}\left(m_{test}[W], m_{test}[Z]+2\right)\cdot\left(m_{test}[Z]+m_{test}[W]+1\right)} \tag{7}$$

where B is the beta function, and $F_{k_2}\left(\vec{a},\vec{b},m_{ref}[Z]\right)$ is explicited by means of a generalized hypergeometric function ${}_3F_2$, parametrized with a, $b, m_{ref}[Z], m_{ref}[W], m_{test}[Z]$, and $m_{test}[W]$. In the same manner as above, the $p-$value is defined as:

$$p = 1 - C[Z,W] \tag{8}$$

As previously, the CBB test under a given $\alpha-$risk is expressed as:

- If $p > \alpha$, then H_0 is rejected, and a contamination is detected;
- Otherwise, H_0 is accepted.

In a nutshell, the CNB and CBB tests are both frequentist tests, founded upon the calculation of a $p-$value, and comparison of this $p-$value to a rejection threshold of the null hypothesis. The determination of this threshold is based itself on the quantiles of a

cumulative distribution function (CDF) built under the assumption that the null hypothesis is verified. We showed in previous work [2] that these detection procedures yield a significant improvement of the trade-off between TDR and FAR when compared to a KS test. As it was stated in Section I, however, the so-obtained gain may be deemed insufficient when facing very low SNR conditions. The most relevant option then appears to embed more uranium-specific, spectral information at the core of the detection method. So far, indeed, the only prior information used in the procedure lies within the definition of the spectral regions of interest Z and W. However, the availability of a physical model of the gamma-ray source term associated with a uranium surface contamination, as well as of multichannel spectra at the output of the HPGe spectrometer, allow further prior modeling of the alternative hypothesis H_1. Practically speaking, such a modeling involves the exploitation of the signature from multiple energy channels across the spectrum in a bin-to-bin fashion. The difficulty that arises lies in the fact that the type of probability distribution underlying the counting becomes multivariate, and the quantiles of the cumulative distribution of such laws can no longer be obtained symbolically. As a consequence, if we aim at exploiting the posterior multivariate laws of both hypotheses H_0 and H_1, we are to turn to a Bayesian approach. Such an approach, indeed, does not require evaluating the cumulative function of a multivariate law, and subsequently deriving a $p-$value. It is thus this line of thought that we shall follow in the following subsections, introducing two Bayesian statistical tests that, to our knowledge, are new in the field of nuclear data analysis.

As mentioned above, one of these tests, the Bayesian Absolute test by Mixture of multivariate Poisson laws (BAM), represents an alternative to the CNB test, while the Bayesian Relative test by Mixture of multinomial laws (BRM) test can be used instead of the CBB test.

B. Bayesian Absolute Test by Mixture of Multivariate Poisson Laws (BAM)

The construction of a Bayesian statistical test, whether it is absolute or relative, requires some prior knowledge of the spectral signatures of both the radiological background, and an expected uranium contamination inside the investigated unit. On the one hand, and in agreement with the description from Section III, an *a priori* signature of the background is directly provided by the reference spectrum acquired over T_{ref}. On the other hand, the prior spectrum from a uranium contamination is obtained via Monte Carlo simulations. The keystone of the method is that the multichannel countings, obtained from the acquisitions over T_{ref} and T_{test}, are not seen as resulting directly from a sampling of the said *a priori* signatures. Instead, they are each seen as resulting from the sampling of a multinomial law with random parameters. These parameters are themselves supposed to follow a Dirichlet distribution (that faithfully captures the discrete bin-to-bin profile of experimental energy spectra) whose mathematical expectation is the one vector directly given by the abovementioned *a priori* spectral signatures. This introduction of a statistical level between the *a priori* and the measurement makes it possible to limit the detrimental impact on the detection procedure of small fluctuations in the background, as well as minor discrepancies between simulation and experiment. As far as the BAM test itself is concerned, the fact that it is an absolute statistical test comes from the use of a Poisson law assumption with a defined intensity parameter as underlying the counting model. This is actually the main characteristic the test shares with the frequentist CNB test described in Subsection IV.A.

We note $\boldsymbol{m_{ref}} = \left(m_{ref}[1], \dots, m_{ref}[K]\right)$ and $\boldsymbol{m_{test}} = (m_{test}[1], \dots, m_{test}[K])$ the multichannel counting vectors respectively associated with the acquisitions over T_{ref} and T_{test}, $m_{ref}[k]$ and $m_{test}[k]$ labeling the number of counts recorded in the k^{th}channel, and K the number of channels in the portion of the histograms that is used in the analysis. Moreover, we introduce m_{ref} and m_{test} as the total numbers of counts over the K exploited channels of both energy spectra respectively.

The mathematical law hypotheses upon which the procedure is founded are as follows:

- The counting in the k^{th} channel of a histogram obtained in the presence of the sole background radiation follows a Poisson law of parameter $\lambda_B \cdot x_B[k]$ where:
  - the counting intensity λ_B, associated with the background radiation, follows a gamma law of parameters $(a_B\,; b_B)$;
  - and vector $\boldsymbol{x_B}$ follows a Dirichlet distribution, whose parameter vector is given by the product $\boldsymbol{\gamma_B} = K \cdot \nu_B \cdot \boldsymbol{s_B}$, where vector $\boldsymbol{s_B}$ represents the normalized *a priori* signature of the background, and scalar ν_B a weighting factor that quantifies the confidence of the user in said *a priori*.
- In a symmetrical fashion, the counting in the k^{th} channel of a histogram obtained in the presence of the sole uranium contamination follows a Poisson law of parameter $\lambda_S \cdot x_S[k]$ where:
  - the counting intensity λ_S, associated with the contamination, follows a gamma law of parameters $(a_S\,; b_S)$;
  - and vector $\boldsymbol{x_S}$ follows a Dirichlet distribution whose parameter vector is given by the product $\boldsymbol{\gamma_S} = K \cdot \nu_S \cdot \boldsymbol{s_S}$, where vector $\boldsymbol{s_S}$ represents the normalized *a priori* signature of the uranium, and scalar ν_S a weighting factor that traduces the confidence of the user in said *a priori*.

When dealing with Bayesian statistical tests, one no longer aims at comparing a $p-$value to a decision threshold, but at determining, given the data sets $\boldsymbol{m_{ref}}$ and $\boldsymbol{m_{test}}$, which one of H_0 or H_1 is the most likely verified hypothesis. In other words, the quantity of interest may be represented by the ratio of posterior probabilities $\frac{\mathbb{P}\left(H_0 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right)}{\mathbb{P}\left(H_1 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right)}$.

Now, the application of Bayes' theorem leads to the following equality:

$$\frac{\mathbb{P}\left(H_0 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right)}{\mathbb{P}\left(H_1 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right)} = \frac{\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_0\right) \cdot \mathbb{P}(H_0)}{\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_1\right) \cdot \mathbb{P}(H_1)} \tag{9}$$

In the framework of Bayesian statistics, probabilities $\mathbb{P}(H_0)$ and $\mathbb{P}(H_1)$ are called *priors*. There are set by the user on the basis of previous information about plan project, security rules, and operation history within the facility. As a consequence, the ratio r expressed as:

$$r = \frac{\mathbb{P}(H_1)}{\mathbb{P}(H_0)} \tag{10}$$

is a constant that represents, for the decommissioning team, the likeliness of a uranium contamination being present.

By combing (9) and (10), we see that the detection procedure relies entirely on a ratio ρ of likelihood functions, known as Bayes' factor, and defined as:

$$\rho = \frac{\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| \mathrm{H}_1\right)}{\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| \mathrm{H}_0\right)} \cdot r \tag{11}$$

In addition, we recall that:

$$\mathbb{P}\left(H_1 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right) = 1 - \mathbb{P}\left(H_0 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right) \quad (12)$$

The combination of the last equations leads to a factorization of the posterior probability of H_0, which is the quantity that will be eventually compared to a decision threshold:

$$\mathbb{P}\left(H_0 \middle| \boldsymbol{m_{ref}}, \boldsymbol{m_{test}}\right) = \frac{1}{1+\rho} \quad (13)$$

In a nutshell, we showed that the formulation of the statistical test requires solely the calculation of conditional joint probabilities $\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_0\right)$ and $\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_1\right)$.

We shall now underline the procedure that we followed to compute these two probabilities:

- First, we must calculate the joint likelihood of $\boldsymbol{m_{ref}}$, $\boldsymbol{m_{test}}$, $\boldsymbol{x_B}$, and λ_B under H_0, and the joint likelihood of $\boldsymbol{m_{ref}}$, $\boldsymbol{m_{test}}$, $\boldsymbol{x_B}$, λ_B, $\boldsymbol{x_S}$, and λ_S under H_1;
- Second, we must marginalize the first of these joint likelihoods with $\boldsymbol{x_B}$, and λ_B;
- Third, we must marginalize the second of these joint likelihoods with $\boldsymbol{x_B}$ and $\boldsymbol{x_S}$ as well as with λ_B and λ_S.

The first marginalization leads straightforwardly to an expression of $\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_0\right)$ that is explicited by means of the gamma function Γ. The derivation of $\mathbb{P}\left(\boldsymbol{m_{ref}}, \boldsymbol{m_{test}} \middle| H_1\right)$ is more difficult to carry out. Indeed, marginalizing the joint likelihood with both $\boldsymbol{x_B}$ and $\boldsymbol{x_S}$ requires integrating, and therefore developing a multivariate polynomial in $\boldsymbol{x_B}$ and $\boldsymbol{x_S}$. The best way around the complexity of such a development takes advantage of the moment structure of the Dirichlet distribution, and the term-by-term identification of the powers of the said polynomial to instantiations of function Γ.

All calculations being made, Bayes' factor may be explicited as:

$$\rho = \frac{\Gamma(K\nu_S)}{\Gamma(a_S)} \cdot \left(\frac{b_S}{T_{test}+b_S}\right)^{a_S} \cdot \frac{\Gamma\left(m_{ref}+m_{test}+K\nu_B\right)}{\Gamma\left(m_{ref}+m_{test}+a_B\right)} \cdot \sum_{j=0}^{m_{test}} \frac{\omega_j \Gamma(j+a_S)\Gamma\left(m_{ref}+m_{test}-j+a_B\right)}{\Gamma(j+K\nu_S)\Gamma\left(m_{ref}+m_{test}-j+K\nu_B\right)} \left(\frac{T_{ref}+T_{test}+b_B}{T_{test}+b_S}\right)^j \cdot r \quad (14)$$

The calculation of coefficients $\left(\omega_j\right)_{1\le j\le m_{test}}$ is carried out in two steps. We first form a set of K vectors denoted $(\boldsymbol{V_k})_{1\le k\le K}$. The k^{th} element of this set $\boldsymbol{V_k}$ is a vector of dimensions $(m_{test}[k], 1)$, whose j^{th} coordinate is:

$$V_k[j] = \binom{m_{test}[k]}{j} \cdot \frac{\Gamma(j+\gamma_S[k])}{\Gamma(\gamma_S[k])} \cdot \frac{\Gamma\left(m_{ref}[k]+m_{test}[k]-j+\gamma_B[k]\right)}{\Gamma\left(m_{ref}[k]+m_{test}[k]+\gamma_B[k]\right)} \quad (15)$$

Second, we construct the discrete $K-$convolution product of $(\boldsymbol{V_k})_{1\le k\le K}$, noted $\boldsymbol{\Omega} = \left(\omega_j\right)_{1\le j\le m_{test}}$, as:

$$\boldsymbol{\Omega} = \boldsymbol{V_1} \otimes \ldots \otimes \boldsymbol{V_K} \quad (16)$$

where: $\forall\, (m,n) \in [\![1,K]\!]^2, \forall\, q \ge 1,$

$$(V_m \otimes V_n)[q] = \sum_{q=-\infty}^{+\infty} V_m[q] \cdot V_n[p-q] \quad (17)$$

In a similar fashion as we have found in Subsection IV.A, the BAM test under a given $\alpha-$risk is eventually expressed as:

- If $\frac{1}{1+\rho} < \eta_\alpha$ then H_0 is rejected and a contamination is detected;
- Otherwise, H_0 is accepted.

where the relation between the thereshold η_α and the $\alpha-$risk must be established numerically via a Monte Carlo method.

As it can be seen from (14), the probability ratio r acts a scaling parameter that will not modify the trend of the ROC curve of the test, but the underlying value of η_α yielding the best TDR/FAR trade-off ratio. In the absence of any prior knowledge of how likely the presence of a uranium contamination is, we chose non-informative priors $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 0.5$, so that $r = 1$ in the rest of the analysis.

C. Bayesian Relative Test by Mixture of Multinomial Laws (BRM)

The formulation of the BRM test is very close to the one the BAM test. Going from the absolute to the relative version of the statistical test is done in the same way as we have seen when describing the transition from the CNB to the CBB test, i.e., by removing the Poisson law assumption from the modelling of countings.

Keeping the notations from Subsection IV.B, hypotheses H_0 and H_1 admit the following reformulations:

- H_0 : spectra $\boldsymbol{m_{ref}}$ and $\boldsymbol{m_{test}}$ are formed of countings resulting from two drawing experiments, with respectively m_{ref} and m_{test} trials, inside the same multinomial law of vector parameter $\boldsymbol{x_B}$ as defined in the last subsection;
- H_1: spectrum $\boldsymbol{m_{ref}}$ is formed of countings resulting from a drawing experiment with m_{ref} trials inside the sole multinomial law of vector parameter $\boldsymbol{x_B}$. Spectrum $\boldsymbol{m_{test}}$, on the contrary, is formed of countings resulting from a drawing experiment with m_{test} trials inside a mixture of two multinomial laws of respective vector parameters $\boldsymbol{x_B}$ and $\boldsymbol{x_S}$.

The relative contributions to the mixture of the background radiation and uranium contamination are quantified through the introduction a random variable that is supposed to follow a beta distribution of parameters $(a;\ b)$.

The explicitation of Bayes' factor in this case is similar to the one derived in Subsection IV.B except this time, it is expressed by means of both the gamma function Γ and the beta function B:

$$\rho = \frac{\Gamma(K\nu_S)\Gamma\left(m_{ref}+m_{test}+K\nu_B\right)}{\mathrm{B}(a,b)} \cdot \sum_{j=0}^{m_{test}} \frac{\omega_j \mathrm{B}(j+a, m_{test}-j+b)}{\Gamma(j+K\nu_S)\Gamma\left(m_{ref}+m_{test}-j+K\nu_B\right)} \cdot r \quad (18)$$

where, again, the probability ratio r is set equal to 1.

The calculation of coefficients $\left(\omega_j\right)_{1\le j\le m_{test}}$ is performed according to the same discrete convolution procedure as in the BAM test construction.

Similar to BAM test, the BRM statistical test under a given $\alpha-$risk reads as:

- If $\frac{1}{1+\rho} < \eta_\alpha$ then H_0 is rejected and a contamination is detected;
- Otherwise, H_0 is accepted.

V. Parameterization of the BAM and BRM Statistical Tests

Contrary to the CNB and CBB frequentist tests, the implementation of the BAM and BRM statistical tests requires the setting of certain parameters. Said setting must be done with the aim of minimizing the acquisition time (and thus the decommissioning costs) while keeping the TDR/FAR as high as possible. In accordance with the aforementioned definitions of BAM and BRM, there are three parameters that we can be considered as key, namely the number of channels K used for the test, and the weighting factors ν_S and ν_B. In the current section, we present the procedure that was followed to adjust the values of these parameters.

A. Selection of the Number of Channels K

As stated above, the construction of the BAM and BRM tests was built upon a discrete $K-$ convolution product. As a result, and as be

inferred from Eq. (16) above, the computational complexity of these tests will depend essentially on the total number of counts in the selected ROI. Furthermore, one of the advantages of the Bayesian approach is that its effectiveness does not necessarily rely on the definition of adequate ROI (contrary to frequentist CNB and CBB tests). As a result, a rigorous study of the dependence of its performance as a function of K is in order.

To investigate the influence of parameter K on the performance of the BAM and BRM tests, we considered two extreme cases:

- First, we used the full spectrum with $K = 8192$ channels;
- Second, we retained the disjoint ROI: $[140 - 210] \cup [990 - 1010]$ keV, where the FEP of interest of uranium are expected, and amounting to $K = 90$ channels.

These settings are the reflection of two contradictory motivations. The first ROI aims at taking advantage of the full spectrum to exploit maximally any prior information on the uranium and background signals, with the potential disadvantage of increasing the execution time due to the number of counts recorded in the spectrum. The second ROI is restricted to regions surrounding the FEP of uranium in order to limit the execution time, with the potential downside of limiting the *a priori* information on the background that is available for the detection procedure. In order to assess these settings quantitatively, we calculated the merit factor $max_{ROC}(TDR - FAR)$ (i.e., the elbow of the ROC curve), and the execution time/ROC for both extreme cases where $K = 8192$ and $K = 90$. Tables III and IV present the obtained results for an integration time $T_{test} = 100$ s and an enrichment level of 1 wt%.

TABLE III
ESTIMATION OF THE FACTORS OF MERIT FOR THE BAM TEST ACCORDING TO THE NUMBER OF SELECTED CHANNELS K.

BAM test	$K = 8192$	$K = 90$
Execution time/ROC	11.7 s	1.85 s
$max_{ROC}(TDR - FAR)$	0.76 ± 0.01	0.78 ± 0.01

TABLE IV
ESTIMATION OF THE FACTORS OF MERIT FOR THE BRM TEST ACCORDING TO THE NUMBER OF SELECTED CHANNELS K.

BRM test	$K = 8192$	$K = 90$
Execution time/ROC	24.3 s	1.89 s
$max_{ROC}(TDR - FAR)$	0.76 ± 0.01	0.76 ± 0.01

From Tables III and IV, we observe that, for $K = 90$, the BAM and BRM tests yield a factor of merit $max_{ROC}(TDR - FAR)$ equal to, or higher than the one obtained using the full spectrum ($K = 8192$) with a five- to ten-times lesser execution time.

Regarding the comparison of the execution times associated to the CNB, CBB, BAM and BRM tests run over $K = 90$ channels, Table V presents the values that we obtained for an integration time $T_{test} = 100\ s$, and a simulated enrichment level of 1 wt%.

TABLE V
EXECUTION TIMES FOR CNB, CBB, BAM AND BRM TESTS, $T_{test} = 100\ s$ AND $K = 90$.

Statistical test	Execution time/test
CNB	6.88 ms
CBB	13.5 ms
BAM	1.85 s
BRM	1.89 s

We observe that the execution times of the CNB and CBB tests are much lesser than the ones of the BAM and BRM. This result directly follows from discrete convolution product that is used by two latter. This being said, we notice that the execution time for the four tests will remain negligible (below 2 %) when compared to a minimum acquisition time in the order of 100 s.

We conclude that the selection of a number of channels $K = 90$ is the most appropriate to our study in spite of the potential loss of *a priori* information on the background signal it involves (which we have shown to be limited so far). The latter issue will be addressed by tuning weighting factor ν_S and ν_B as it will be detailed in the next subsection.

Eventually, one may note that the execution time/ROC could be reduced by using a more powerful machine. The decision of using a spectral region limited to our ROI instead of the full spectrum thus depends on our prototype architecture, and may be subject to future change.

B. *Selection of Weighting Factors* ν_S *and* ν_B

The weighting factors ν_S and ν_B reflect the user's confidence in the uranium signal and background signal *a priori*. These factors thus constitute other parameters, that are specific to the BAM and BRM Bayesian tests, and that play a crucial role in both the construction and performance of these tests. Furthermore, the selection of these factors allows us to account for the limitation of the available prior information on the radiation background. This limitation itself results from the use of restricted ROI in the vicinity of the FEP of uranium.

In practice, any value of ν_S and ν_B above 1 may be selected. In order to select optimal values for these weighting factors, we studied systematically their respective influence on the performance of the BAM and BRM tests, using a random sampling of the $\{\nu_S; \nu_B\}$ 2D-space without any law hypothesis. To this end, we varied ν_S and ν_B separately and in a similar way, according to ranges $\nu_B \in \{1; 5; 10; 50; 100\}, \nu_S \in \{1; 5; 10; 50; 100\}$. As in the previous subsection, the performance of a test was quantified using the merit factor $max_{ROC}(TDR - FAR)$. The evolution of this factor was investigated as a function of ν_S and ν_B for a number of channels $K = 90$, an integration time $T_{test} = 100\ s$ and an enrichment level of 1 wt%. The results are synthetized in Tables VI and VII below.

TABLE VI
ESTIMATION OF THE FACTOR OF MERIT FOR THE BAM TEST ACCORDING TO WEIGHTING FACTORS ν_S AND ν_B.

ν_S \ ν_B	BAM test				
	1	5	10	50	100
1	0.78±0.01	0.76±0.01	0.75±0.01	0.73±0.01	0.74±0.01
5	0.75±0.01	0.74±0.01	0.75 ±0.01	0.76±0.01	0.75±0.01
10	0.74±0.01	0.74±0.01	0.75±0.01	0.74±0.01	0.74±0.01
50	0.74±0.01	0.76±0.01	0.74±0.01	0.76±0.01	0.74±0.01
100	0.75±0.01	0.76 ±0.01	0.76±0.01	0.78±0.01	0.74±0.01

TABLE VII
ESTIMATION OF THE FACTOR OF MERIT FOR THE BRM TEST ACCORDING TO WEIGHTING FACTORS ν_S AND ν_B.

ν_S \ ν_B	BRM test				
	1	5	10	50	100
1	0.76±0.01	0.73±0.01	0.68±0.01	0.64±0.01	0.61±0.01
5	0.59±0.01	0.74±0.01	0.75±0.01	0.70±0.01	0.70±0.01
10	0.27±0.01	0.74±0.01	0.75±0.01	0.70±0.01	0.68±0.01
50	0	0.57±0.01	0.58±0.01	0.76±0.01	0.74±0.01
100	0	0.12±0.01	0.54±0.01	0.74±0.01	0.76±0.01

Let us first analyze the results associated with the BAM test in Table VI. We introduce the notation $M(\nu_B, \nu_S) = max_{ROC(\nu_B,\nu_S)}(TDR - FAR)$ to refer to every entry in Table VI. The one-sigma uncertainty associated with every value of $M(\nu_B, \nu_S)$ is noted $u\big(M(\nu_B, \nu_S)\big) = u\left(max_{ROC(\nu_B,\nu_S)}(TDR - FAR)\right)$. Eventually, we label $\bar{M}$ the empirical mean of all factors $M(\nu_B, \nu_S)$ in Table VI, and $\sigma(M)$ the empirical dispersion of the said factors around the mean $\bar{M}$. We observe that:

$$\forall\ (\nu_B, \nu_S) \in \{1; 5; 10;\ 50; 100\} \times \{1; 5; 10;\ 50; 100\} \quad (19)$$

$$\big[M(\nu_B, \nu_S) \pm u\big(M(\nu_B, \nu_S)\big)\big] \subseteq [\bar{M} \pm \sigma(M)]$$

In other words, as far as the BAM test is concerned, all considered settings of weighing factors ν_B and ν_S are equivalent with respect to the trade-off between FAR and TDR.

Carrying out the same statistical analysis with the results associated with the BRM test, we observe conversely that all considered settings are not equivalent with respect to the trade-off between FAR and TDR. In fact, we note from the factors presented in Table VII that the closer the values of ν_B and ν_B are to each other, the higher $M(\nu_B, \nu_S)$ will be. In order to corroborate this observation, we performed once again the above described analysis while restricting its domain to the cases where $\nu_S = \nu_B$ (i.e., with the sole factors that belong to the diagonal of Table VII). In this situation, we thus have:

$$\forall\ (\nu_B, \nu_S) \in \{1; 5; 10;\ 50; 100\}^2 \,/\, \nu_S = \nu_B \quad (20)$$

$$\big[M(\nu_B, \nu_S) \pm u\big(M(\nu_B, \nu_S)\big)\big] \subseteq [\bar{M} \pm \sigma(M)]$$

From the perspective of the merit figure $max_{ROC}(TDR - FAR)$, it then follows that all symmetrical settings of ν_S and ν_B are equivalent.

Now, on the basis of the discussion in Subsection III.C, we remind that:

- The background signal is likely to vary in shape and amplitude;
- The priors form only responses to average expected enrichments so that, in practical cases, the uranium signal may depart from this prior due to a significantly different mass fraction in U-235.

As a consequence, the choice of the lowest weighting factors $\nu_S = \nu_B = 1$ is favored. This setting indeed does not impinge the performance of the Bayesian tests, while remaining conservative with respect to potential discrepancies between tested spectra and *a priori* signatures.

To conclude this section, we should note that modelling a nuclear counting with a Poisson distribution requires more assumptions to be valid than describing it with a binomial distribution. One of such conditions is that the number of observed decays in the population of unstable nuclei be large enough, and thus that the measurement time is above a certain limit. This should lead to some discrepancies between the behaviors of BAM and BRM as the measurement test T_{test} keeps decreasing. However, preliminary studies have shown that when T_{test} is the order of several tens of seconds, both Bayesian tests exhibit random behaviors. In other words, the ROC curve becomes a straight line between points $(0; 0)$ and $(1; 1)$ in the TDR/FAR space. This is due to the fact that the signal in the spectral ROI for uranium detection is then too tenuous to be of use. Consequently, the study of such discrepancies was left as an outlook in the course of this work.

VI. SIMULATION RESULTS AND DISCUSSION

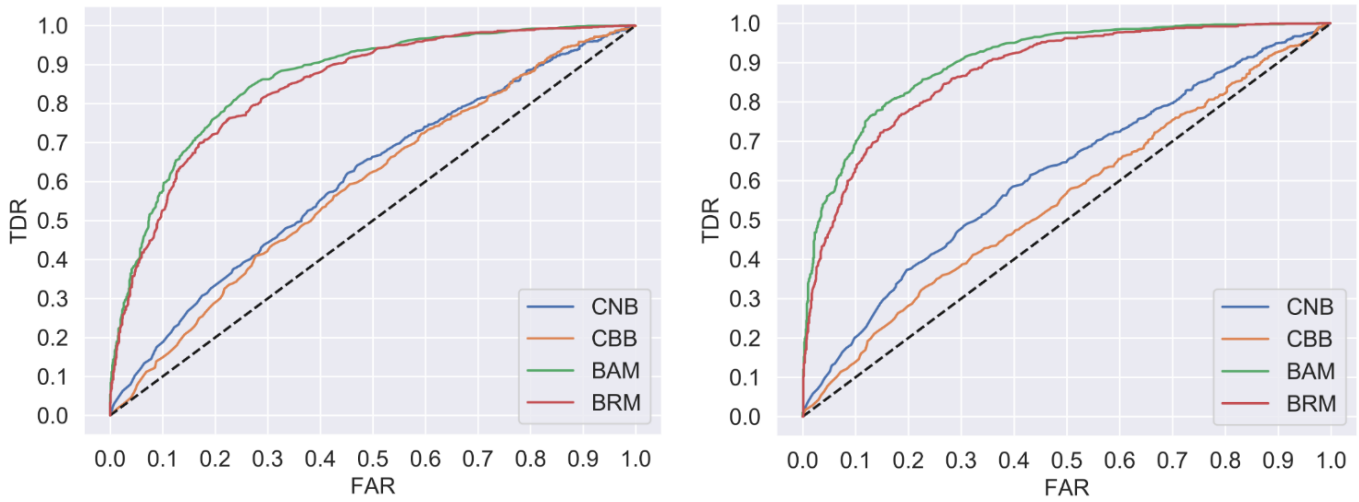
In this section, we evaluate the respective performance of the Bayesian and frequentist approaches by using ROC curves. Besides, we study the effect of the radiological background variations, both in amplitude and in shape, on the performance reached by any statistical test.

A. Comparative Study between Bayesian and Frequentist Tests: Stable Radiological Background

In order to investigate the effectiveness of the Bayesian approach to address uranium detection with low count rates, we carried out a comparative study of BAM and BRM with frequentist tests by using ROC curves. The study is based on the simulation of expected scenarios corresponding to different enrichment levels encountered within the three SU described in Subsection III.B, and the calculation merit figure $max_{ROC}(TDR - FAR)$. We conducted this study on the basis of the following configurations:

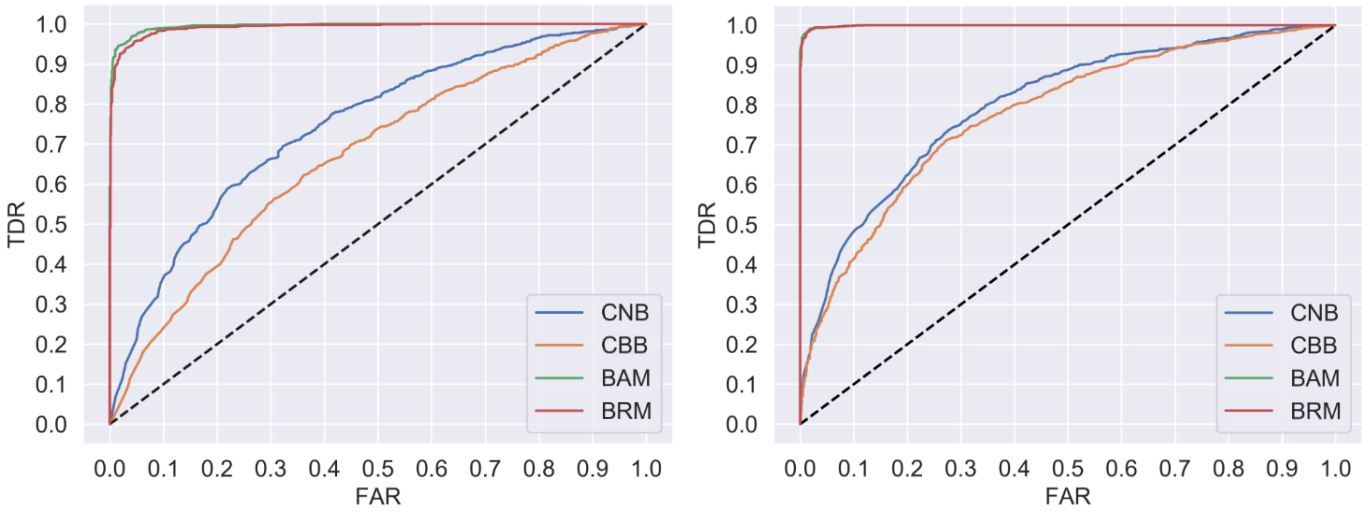
- Bayesian and frequentist hypothesis tests: BAM, BRM, CNB, CBB;
- Enrichment levels: Nat wt%, 3 wt%, 8 wt% ;
- SU of nuclear facility: SU1, SU2, SU3;
- Source-detector distance: $d = 22.36\ cm$ for 500 Bq/0.1 m^2/ 2π (MSAD2);
- Number of channels: $K = 90$;
- Weighting factors: $\nu_S = \nu_B = 1$;
- Integration times: $t \in \{300; 500; 1000; 3000; 5000\}\ s$;
- Number of iterations: $N = 1000$;
- Radiological background: stable, varied in shape and amplitude (Subsection VI.B).

Fig. 2 to 4 present the results obtained for an integration times $t \in \{3000;\ 5000\}$ s, and U-235-enrichments lying between natural level and 8 wt%.



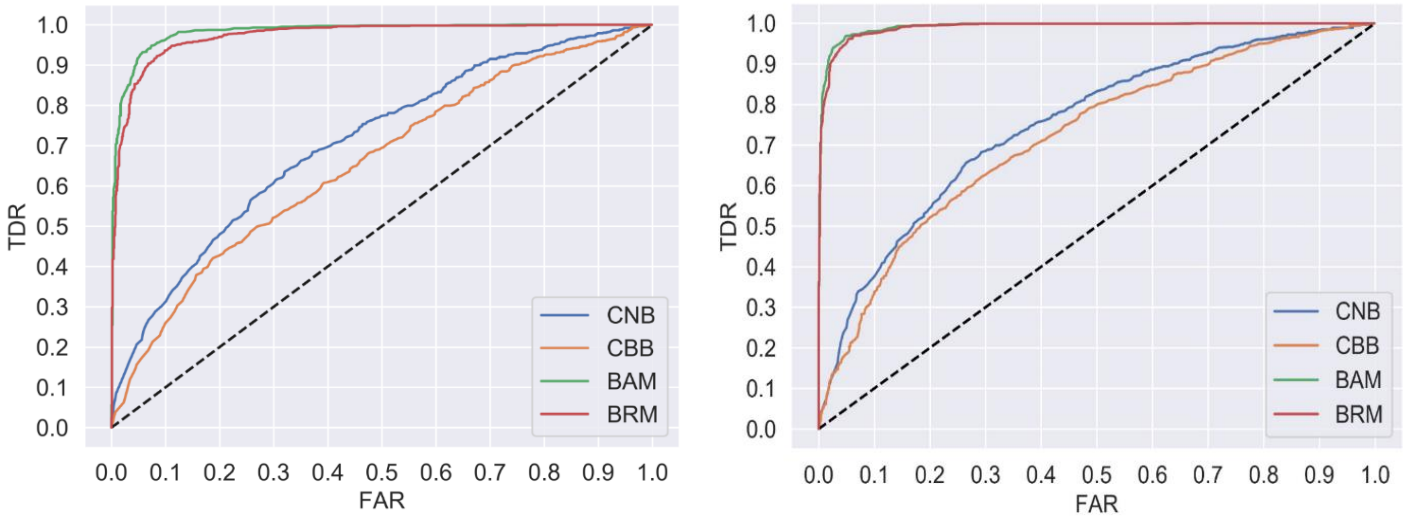
(a) (b)

Fig. 2. ROC curves for an enrichment level of Nat wt% and integration times: (a) t = 3000 s, (b) t = 5000 s.



(a) (b)

Fig. 3. ROC curves for an enrichment level of 3 wt% and integration times: (a) t = 3000 s, (b) t = 5000 s.



(a) (b)

Fig. 4. ROC curves for an enrichment level of 8 wt% and integration times: (a) t = 3000 s, (b) t = 5000 s.

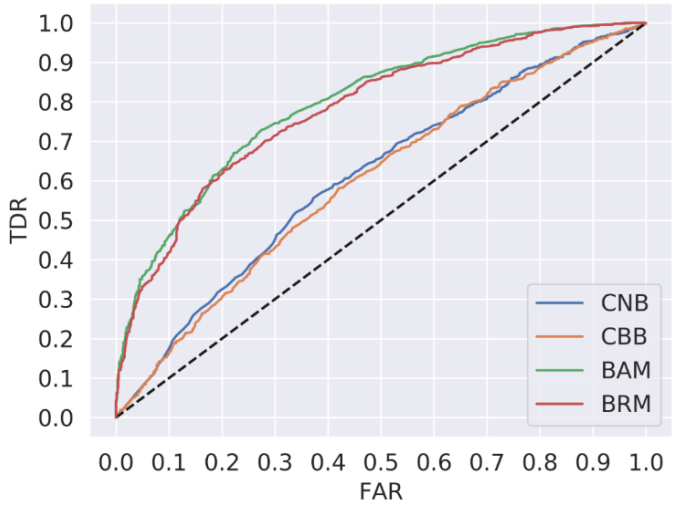
From Fig. 2 to 4, we observe that:

- The ROC curves for BAM and BRM Bayesian tests traduce a more statistically efficient behavior than the ones we drew for CNB and CBB frequentist tests. Indeed, these yield a factor of merit $max_{ROC}(TDR - FAR)$ closer to the point of coordinates $(0; 1)$ in the (FAR, TDR) space, as it can be seen in Fig. 3. In this regard, the BAM and BRM Bayesian tests show a superior performance by up to 50 % when compared to the CNB and CBB frequentist tests. As an illustration, taking the example of the BAM test, we obtained for both enrichment levels Nat wt% and 8 wt% at t = 5000 s, figures of merit equal to (0.63 ± 0.01) (Fig. 2. b) and (0.92 ± 0.01) (Fig. 4. b). These advantageously compare to (0.19 ± 0.03) (Fig. 2. b) and (0.40 ± 0.01) (Fig. 4. b) as yielded by the CNB test;
- The ROC curves for the absolute BAM and relative BRM tests traduce almost the same detection performance. The difference in their merit figures is indeed about 2 %, in contrast with the 7 % discrepancy we observe between the same figures for the absolute CNB and relative CBB tests. Moreover, BAM only shows a slightly higher performance than BRM when the enrichment is less than 3 wt%;
- The ROC curves of BAM and BRM show a large increase in the merit figure $max_{ROC}(TDR - FAR)$ when the enrichment level gets higher, and/or the integration time is augmented. For example, with an enrichment level of 8 wt%, BAM yields a merit figure of (0.85 ± 0.01) at t = 3000 s (Fig. 4. a), compared to a figure of (0.92 ± 0.01) (Fig. 4. b) at t = 5000 s.

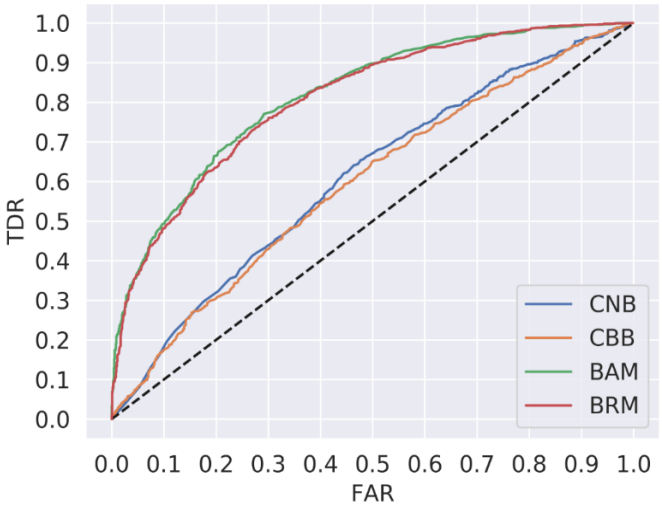
Eventually, we can conclude that the comparison between Bayesian and frequentist tests with a stable radiological background shows that BAM and BRM provide a significantly higher trade-off between specificity and sensitivity than CNB and CBB. Not only do they show an almost identical detection behavior, but their behaviors come close to the one of an ideal detection test when the enrichment level is high (above 8 wt%), and the integration time is long (above 5000 s).

B. Comparative Study between Bayesian and Frequentist Tests: Radiological Background Varied in Shape and Amplitude

In this section, we present the results of variability study on the background radiation signal with respect to shape and amplitude, comparing the effect of these variations on both Bayesian and frequentist tests. The following presentation focuses on the extreme enrichments levels Nat wt% and 8 wt%, and integration times $t \in \{3000; 5000\}\, s$.

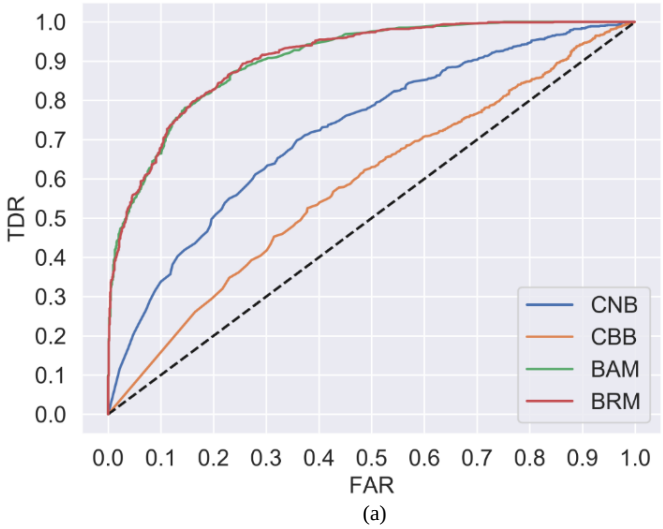


(a)

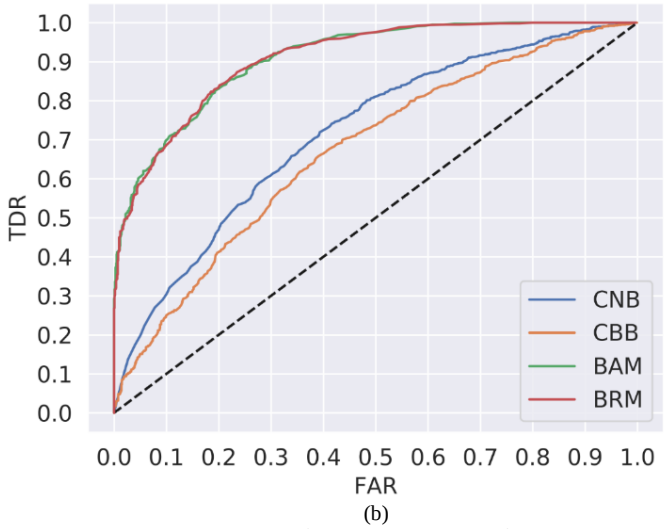


(b)

Fig. 5. ROC curves in the case of a varied background for an enrichment level of Nat wt% and integration times: (a) t = 3000 s, (b) t = 5000 s.



(a)



(b)

Fig. 6. ROC curves in the case of a varied background for an enrichment level of 8 wt% and integration times: (a) t = 3000 s, (b) t = 5000 s.

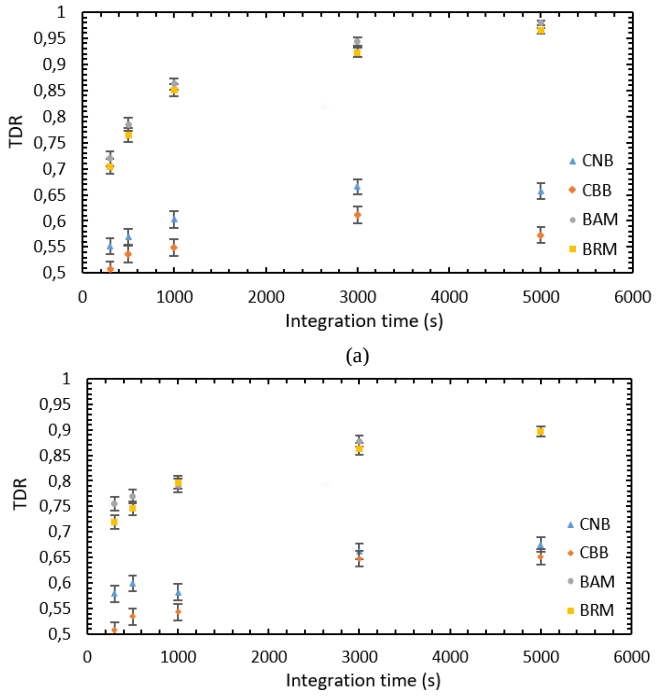
From Fig. 5 and 6, we observe that:

- The detection performance of the BAM and BRM Bayesian tests is still superior by approximately 30 % (with respect to CNB), and 37 % (with respect to CBB) to ones yielded by frequentist tests. However, the figure merits $max_{ROC}(TDR - FAR)$ are less than the ones we obtained with a stable background (Subsection VI.A). For example, for an enrichment level of 8 wt% and t = 5000 s, we obtained with the BAM test a factor of merit equal to (0.64 ± 0.01) (Fig. 6. b), to be compared to (0.92 ± 0.01) with a stable background (Fig. 4. b);
- In the meantime, the CNB and CBB frequentist tests retain the same detection performance as in the context of a stable radiological background. We found that, for an enrichment level of 8 wt% and t = 3000 s, the CNB test yields a factor of merit of (0.33 ± 0.02) (Fig. 6. a) equal to the one we get from Fig. 4a (0.32 ± 0.02);
- Regarding the absolute and relative nature of the statistical tests, the detection performance of BAM and BRM conserved the same behavior as in the case of the stable radiological background. The discrepancies between their factors of merit are indeed negligible within 1 %, in particular when considering high enrichment levels (from 3 wt%);
- Similarly, the background signal variation does not affect the respective behaviors of the CNB and CBB frequentist test. The difference between their factors of merit indeed remains more significant, as high as about 8 % when considering Fig. 6. a.

We can conclude that the variability of the background radiation signal, in both shape and amplitude, does affect the detection performance of the BAM and BRM Bayesian test. This performance that remains, however, remains far superior to the one attained with the CNB and CBB frequentist tests. In addition, the absolute BAM and relative BRM tests conserve equivalent detection behaviors. This confirms our indication in Subsection III.C that there is no reason to favor relative statistical tests over their absolute counterparts in the specific detection task that we are addressing.

C. Comparative Study between Bayesian and Frequentist Tests: TDR as a Function of Integration Time

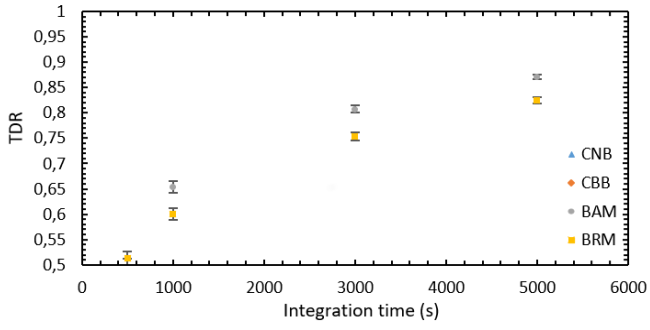
The driving goal of this work is not only to detect a low-activity uranium contamination, but also to do so with an integration time of the measurement that is compatible with project management. Based upon the last subsections, we can translate this objective into optimizing the TDR/FAR trade-off while minimizing the integration time. In this subsection, we thus compare the performance of the Bayesian and frequentist tests at hand with respect to the TDR variation of the integration time against a preset FAR. The study is performed considering a stable background, as well as a background that varied between reference and test acquisitions. The fixed FARs that we chose to tolerate are 0.5 and 0.25. These respectively represent risks of 50 % and 25 % of processing a non-contaminated concrete surface as nuclear waste. Fig. 7 to 10 present the results obtained for U-235-enrichment levels Nat wt% and 8 wt%.



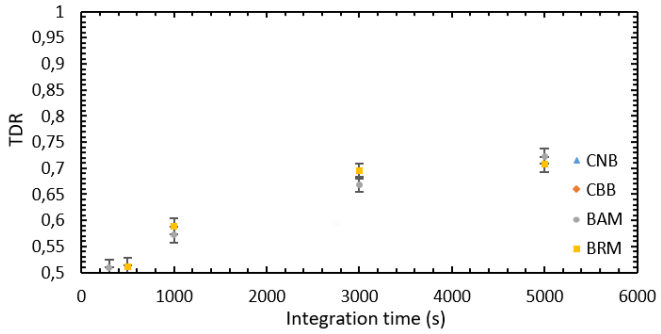
(a)

(b)

Fig. 7. TDR versus the integration time for: FAR = 0.5, enrichment level of Nat wt%, and stable (a) and varied (b) radiological background in shape and amplitude.

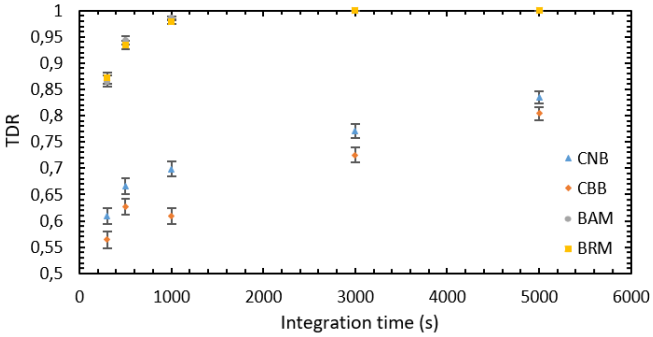


(a)

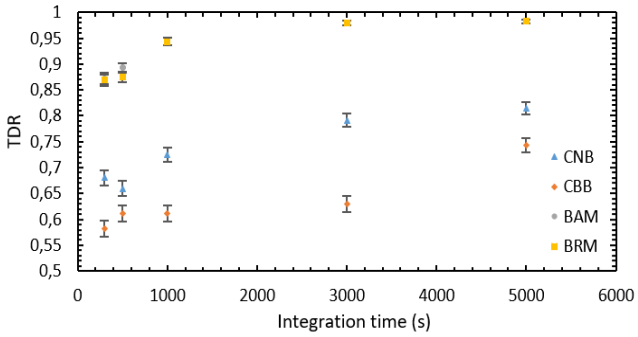


(b)

Fig. 8. TDR versus the integration time for: FAR = 0.25, enrichment level of Nat wt%, and stable (a) and varied (b) radiological background in shape and amplitude.

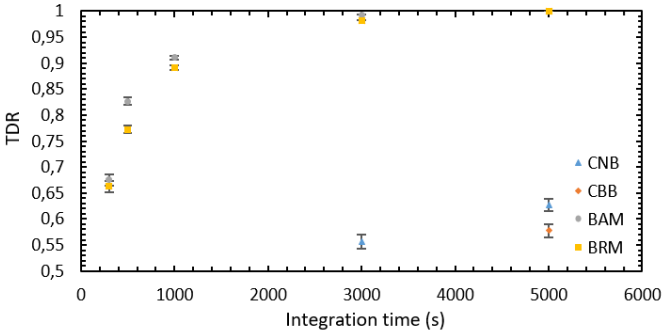


(a)

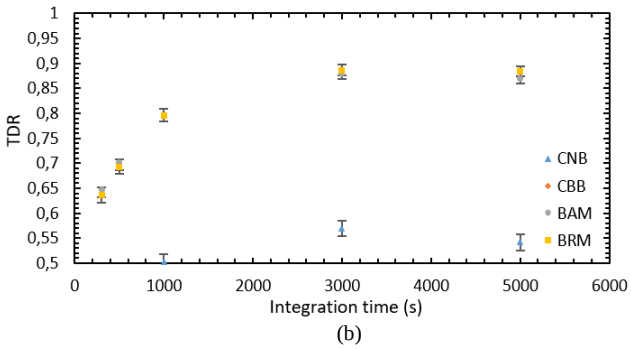


(b)

Fig. 9. TDR versus the integration time for: FAR = 0.5, enrichment level of 8 wt%, and stable (a) and varied (b) radiological background in shape and amplitude.



(a)



(b)

Fig. 10. TDR versus the integration time for: FAR = 0.25, enrichment level of 8 wt%, and stable (a) and varied (b) radiological background in shape and amplitude.

From Fig. 7 to 10, we observe that:

- The BAM and BRM Bayesian tests yield a higher TDR than the CNB and CBB frequentist tests, either with FAR = 0.5 or FAR = 0.25, and for a stable as well as a varied radiological background. As an illustration, for t = 3000 s, FAR = 0.5, and an enrichment level of 8 wt%, BAM provides TDRs equal to 1 (Fig. 9. a) and (0.98 ± 0.004) (Fig. 9. b) for stable and varied radiological backgrounds respectively, while its absolute frequentist counterpart CNB yields TDRs between 0.6 and 0.85;
- In the cases of FAR = 0.5 and FAR = 0.25, with an enrichment level of 8 wt%, and whether the background signal is stable or varied, the absolute BAM and relative BRM tests have almost the same detection performance. The difference between their TDR is indeed about 2 %, against 7 % when the absolute CNB and the relative CBB tests are compared to each other;
- For FAR = 0.5, an integration time t = 5000 *s* and whether the background signal is stable or varied, the BAM and BRM tests yield TDRs superior to 0.95, to be compared to 0.80 with the CNB and CBB tests. For example, for an enrichment level of 8 wt% and a varied background signal, the BAM and CNB tests provide TDRs equal to (0.98 ± 0.004) and (0.82 ± 0.01) respectively (Fig. 9. b);
- For FAR = 0.25, enrichment level of 8 wt% and a stable background signal, the detection performance of the BAM and BRM tests is superior to 0.98, compared to 0.50 with the CNB and CBB tests. For t = 5000 s , the BAM and CNB yield a TDRs equal to 1 and (0.63 ± 0.02) respectively (Fig. 10. a);
- For FAR = 0.25, enrichment level of 8 wt% and a varied background signal, the detection performance of the BAM and BRM tests is slightly reduced. For t = 5000 s, as an illustration, the BAM and CNB tests provide TDRs equal to (0.87 ± 0.01) and (0.54 ± 0.02) respectively (Fig. 10. b).

From these observations, we can conclude that the BAM and BRM Bayesian tests do present themselves as a promising alternative to frequentist detection procedures. With an application to nuclear decommissioning in view, the combination of high-resolution gamma-ray spectrometry and Bayesian statistics seems to pave the way, in particular, to a powerful “second level” detector.

We recall that a “first level” detector is used to characterize the surface to be dismantled before applying any other detector, and within a scalable response time. By contrast, a “second level” detector is operated to confirm the suspicion of a contamination that was detected in the first place by a lower response time detector (whether that is a beta counter, a gross alpha counter…).

For the purpose of this study, we shall consider that a "first level" detector must respond with a 2.5 % beta risk (i.e., a TDR above 0.975) within 1 to 10 min, whereas a "second level" detector will have to respond with the same beta risk within approximately 1 hour. From Fig. 9, we observe that the defining conditions of a "second level" detector are fulfilled when the FAR is set equal to 0.5, which will represent our default setting for such a deployment concept. Fig. 9 and 10, however, tend to show that, even with the implementation of high-performance statistical tests, the gamma-ray spectrometry setup that we have described is not suited to play the part of a "first level" detector.

VII. Conclusion and Future Work

This research work introduces a reliable method to detect low levels of radioactivity against a standard radiological background. The originality of the paper stems from using absolute (BAM) and relative (BRM) Bayesian hypothesis tests with high-resolution, gamma-ray spectra. These tests are competing with absolute and relative frequentist procedures that were used in the study to benchmark our original algorithms. Key to this benchmark is the trade-off between FAR, TDR, and the integration time of the measurement that is to be minimized. The driving goal is to develop a technological solution that is compatible with operational requirements in a large-scale decommissioning project.

The comparative study of BAM and BRM with frequentist tests was based on a minimum surface activity to detected of $500\ \mathrm{Bq}/0.1\ \mathrm{m}^2/2\pi$ within three subunits of a single facility. To simulate the online implementation of the uranium detection procedure, we first selected the number of channels and the weighting factors that parametrize the Bayesian algorithms. We then benchmarked the performance of frequentist and Bayesian tests in a variety of scenarios, involving different U-235-enrichments, and non-stationary radiological backgrounds.

The results of the study show that, against a stationary background, BAM and BRM give access to a significantly better TDR/FAR trade-off than to the one yielded by frequentist hypothesis tests. Said trade-off even comes close, within 1 %, to the behavior of an ideal detection test when the enrichment level is high (above 8 wt%), and the integration time is long (above 5000 s). Varying the radiological background between reference and test acquisitions, we observe a degradation of said trade-off with BAM and BRM, although to a much lesser extent than with frequentist detection algorithms. The absolute and relative versions of the Bayesian method, additionally, were shown to display almost identical detection behaviors as soon as uranium enrichment levels were increased above 3 wt%. Eventually, the evolution of TDR as a function of integration time, against a FAR set equal to 0.5, suggests that the assessed technological solution is suitable as a "second level" detector in the context of large surface decommissioning.

The promising results that were obtained encourage us to envision several outlooks of this research work. The first one lies in the design, assembly, and deployment of a multi-detector gamma-ray spectrometer, monitoring independently several spots inside a wide surface area. Another perspective is the definition and calibration of alternative Bayesian algorithms. One of such algorithms would present itself as a synthesis of the BAM and BRM tests. Its formulation would allow us, by means of a tunable coefficient, to adjust the detection behavior of the procedure between the two extreme features of an absolute and a relative statistical test. A second Bayesian algorithm would enable us to account for the interdependence between counts in nearby channels. On a more methodological level, the performance metric that we have used ultimately relies upon a notion of contamination signal to radiological background that is analogous to a classical SNR. This choice was justified from a nuclear engineering standpoint, for the selection of any technological solution in this field is essentially based on a TDR/FAR ratio, or a detection limit against a given SNR. However, whenever one wishes to compare hypothesis tests in a broader scope, as they succeed or fail in capturing the absolute or relative features of a spectral signal that can be modeled with Poisson or binomial distributions, they would turn to a more universal metric, such as the chi-square test [21]. Such a metric will thus be considered in future work, especially to benchmark novel Bayesian algorithms to BAM and BRM.

Eventually, and as it was stated in the article, there is a need for the manufacturing and characterization of homogenously-dispersed uranium sources according to metrology standards. Such a task is currently being undertaken in the framework of a partnership with Laboratoire National Henri Becquerel (LNHB, CEA List). Once available, the calibrated samples will allow us to consolidate the interpretation of simulation-based ROC curves, and TDR as a function of time in the challenging context of large-surface decommissioning.

Acknowledgment

The authors would like to thank Aurelie LIS, Jean-Philippe DANCAUSSE, Jose-Antoine JANICOT, Cyril CHAMBON, and Philippe GIRONES from CEA DES for funding this research, and helping in writing this paper.

References

[1] E. Neri, A. French, M.E. Urso, M. Deffrennes, G. Rothwell, I. Rehak, I. Weber, and S. Carroll, V. Daniska, "Costs of Decommissioning Nuclear Power Plants (NEA--7201)," Nuclear Energy Agency of the OECD (NEA), 2016.

[2] N. Dufour, J. Dumazert, E. Barat, G. H. V. Bertrand, F. Carrel, T. Dautremer, F. Lainé, and A. Sari, "Measurement of low-activity uranium contamination by gamma-ray spectrometry for nuclear decommissioning," *Nucl. Instrum. Meth. A*., vol. 951, pp.1-14, Jan. 2019.

[3] S. Aggarwal, "Alpha-particle spectrometry for the determination of alpha emitting isotopes in nuclear, environmental and biological samples: Past, Present and Future", Anal. Methods, vol. 8, pp. 5353-5371, May. 2016.

[4] P. Fichet, A. Leskinen, S. Guegan, F. Goutelard, "Characterization of Beta Emitters for Decommissioning," in Proc ICEM-ASME, Brussels, Belgium, Sep 2013, pp. 1-5.

[5] J. C. J. Dean, I. Adsley, P. H. Burgess, "Traceability for measurements of radioactivity in waste materials arising from nuclear site decommissioning," Metrologia, vol. 44, no.4, pp. S140–S145, August 2007.

[6] G. R. Gilmore, Practical Gamma-ray Spectrometry, 2nd ed., Hoboken, New Jersey, USA: John Wiley & Sons, 2008, pp.408.

[7] G. F. Knoll, Radiation Detection and Measurement, 4th ed., Hoboken, New Jersey, USA: John Wiley & Sons, 2010, pp.860.

[8] E. Rohée, R. Coulon, F. Carrel, T. Dautremer, E. Barat, T. Montagu, S. Normand, C. Jammes,"Benchmark of the non-parametric Bayesian deconvolution method implemented in the SINBAD code for X/γ rays spectra processing," Nucl. Instrum. Meth A., vol. 836, pp. 91-97, Nov. 2016.

[9] A. Wiens, B. Birkenbach, B. Bruyneel, J. Eberth, H. Hess, Gh. Pascovici, P. Reiter, D. Bazzacco, E. Farnea, C. Michelagnoli, F. Recchia, "Improved energy resolution of highly segmented HPGe detectors by noise reduction," Eur. Phys. J. A., vol. 49, no. 47, pp.10, Apr. 2013.

[10] Physics-based, Bayesian sequential detection method and system for radioactive contraband, by J. V. Candy et al. (2008, March 14). Patent US 8,676,744 B2. [Online]. Available: https://patents. google.com/patent /US20100030721

[11] J. V. Candy, E. Breitfeller, B. L. Guidry, D. Manatt, K. Sale, D. H. Chambers, M. A. Axelrod, A. M. Meyer, "Physics-based detection of radioactive contraband: a sequential Bayesian approach," IEEE T Nucl Sci., vol.56, no. 6 pp. 3694-371, Dec. 2009.

[12] Zähringer, G. Kirchner, “Nuclide ratios and source identification from high-resolution gamma-ray spectra with Bayesian decision methods,” Nucl. Instrum. Meth. A., vol.594, pp. 400-406, Sep. 2008.
[13] I. Rivals, C. Fabbri, G. Euvrard, X. Blanchard, “A Bayesian method with empirically fitted priors for the evaluation of environmental radioactivity: application to low-level radioxenon measurements,” J Radioanal Nucl Chem., vol. 292, pp. 141–153, Aug. 2011.
[14] A. Bukartas, R. Finck, J. Wallin, C.L. Rääf, “A Bayesian method to localize lost gamma sources”, Appl. Radiat. Isot., vol. 145, pp. 142–147, March. 2019.
[15] M.-M. Bé, V. Chisté, Table de radionucléides, LNE-LNHB/CEA, 2007.
[16] GMX Series Coaxial HPGe Detector Product Configuration Guide: https://www.ortec-online.com/-/media/ametekortec/brochures/gamma-x.pdf
[17] D. B. Pelowitz (Ed.), “MCNP6™ User’s Manual Version 1.0,” Los Alamos National Laboratory report, LA-807 CP-13-00364, Rev. 0, 2013.
[18] M. Tsutsumi, T. Oishi, N. Kinouchi, R. Sakamoto, M. Yoshida, “Simulation of the background for gamma detection system in the indoor environment of concrete buidings,” J Nucl Sci Technol. Vol. 38, pp.1109-1114, Feb. Dec. 2001.
[19] A. A. Jarrett, “Statistical methods used in the measurement of radioactivity with some useful graphs and nomographs,” Oak Ridge, Tennessee, USA, Tech. Rep. No. AECU-262, Jun 1946
[20] C. E. Metz, “Basic principles of ROC analysis,” Semin Nucl Med. Vol. 8, pp. 283-298, Oct 1978.
[21] M. Gaboardi, R. Rogers, “Local private hypothesis testing: chi-square tests,” in Proc ICML, Stockholm Sweden, July 2018, pp. 1626-1635.