Temporal spectral unmixing for rapid detection of radiological events by gamma ray spectrometry

Paul Malfrait

université
PARIS-SACLAY

Démélange spectral par fusion de données temporelles pour la détection rapide d'événements radiologiques par spectrométrie gamma

Temporal spectral unmixing for rapid detection of radiological events by gamma ray spectrometry

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580 : Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Sciences du traitement du signal et des images
Graduate School : Informatique et sciences du numérique
Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Département d'Electronique des Détecteurs et d'Informatique pour la Physique (Université Paris-Saclay, CEA)**, sous la direction de **Jérôme Bobin**, Directeur de recherche et le co-encadrement de **Anne de Vismes Ott**, Chercheuse

Thèse soutenue à Paris-Saclay, le 27 octobre 2023, par

Paul MALFRAIT

Composition du jury

Membres du jury avec voix délibérative

Thomas Rodet Professeur, ENS - Paris Saclay	Président
Abdel-Mjid Nourreddine Professeur, Université de Strasbourg	Rapporteur & Examinateur
Michel Paindavoine Professeur, Université de Bourgogne	Rapporteur & Examinateur
Marie Chabert Professeure, Université de Toulouse	Examinatrice
Fabrice Piquemal Professeur, Université de Bordeaux	Examinateur

THESE DE DOCTORAT
NNT : 2023UPASG055

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Démélange spectral par fusion de données temporelles pour la détection rapide d'événements radiologiques par spectrométrie gamma

Mots clés : Analyse de signal, spectrométrie gamma, analyse de tout le spectre, analyse conjointe, analyse en ligne, calibration

Résumé : Cette thèse vise à améliorer les algorithmes de démélange de spectres gamma afin de pouvoir accélérer la détection d'anomalie et l'estimation des contributions des radionucléides contenus dans les filtres aérosols analysés par l'Institut de Radioprotection et de Sûreté Nucléaire. Pour ce faire nous avons développé un algorithme de démélange de spectre qui analyse conjointement plusieurs mesures successives d'un même échantillon. Cet algorithme a ensuite été amélioré afin de pouvoir mener l'analyse en temps réel pendant que la mesure du spectre gamma est réalisée. Divers outils ont dû être développés en même temps que cet algorithme pour permettre d'estimer précisément les activités des radionucléides, notamment la calibration des signatures spectrales utilisées lors du démélange.

Les résultats de ces études sont présentés dans la thèse à la fois sur des spectres simulés et sur des spectres réels issus des analyses de routine du laboratoire. La détection d'une contamination de ^{123}I à 2 Bq dans un échantillon au bout de 2 minutes de mesure illustre bien l'accélération permise par les développements de cette thèse. En effet, cette détection n'aurait pas été possible avec les outils utilisés en routine.

Title : Temporal spectral unmixing for rapid detection of radiological events by gamma ray spectrometry

Keywords : Signal processing, gamma-ray spectrometry, full-spectrum analysis, joint analysis, online analysis, calibration

Abstract : The aim of this thesis is to improve spectral unmixing algorithms in gamma ray spectrometry in order to speed up anomaly detection and the estimation of radionuclide contributions in aerosol filter samples analyzed by the French Institute of Radiation Protection and Nuclear Safety. To this end, we have developed an unmixing algorithm that performs joint analysis of several successive measurements of the same sample. This algorithm has been improved to enable online analysis while the gamma-ray spectrum measurement is measured. Various tools had to be developed at the same time as this algorithm to enable radionuclide activities to be estimated accurately, including calibration of the spectral signatures used during the unmixing process.

The results of these studies are presented in the thesis both on simulated spectra and on real spectra from the laboratory's routine analyses. The detection of a contamination of ^{123}I at 2 Bq in a sample after 2 minutes of measurement illustrates the acceleration made possible by the developments made during this thesis. Indeed, this detection would not have been possible with the tools used routinely.

Table des matières

Remerciements

Les trois années qui ont constituées ma thèse n'ont pas été un long fleuve tranquille. Cependant j'ai eu de la chance car j'ai, tout au long de celles-ci, été très bien entouré.

Je commence par remercier Mathieu, mon encadrant d'alternance de M2 qui m'a permis de rencontrer Anne et Jérôme. Tu m'as donné le coup de pouce nécessaire pour que je me lance dans ce projet. Je tiens ensuite à remercier mes encadrants qui m'ont accepté comme doctorant, qui m'ont permis de partir en conférence, de publier des articles et surtout m'ont autorisé à donner des cours à l'université. En effet, l'enseignement et la pédagogie sont des choses qui me semblent importantes et enseigner me tenait à cœur pendant ces trois ans de thèse. Jérôme, tes relectures et tes mails à des heures impossibles montrent ton implication dans la réussite des thèses que tu encadres. Merci encore. Anne, merci d'avoir fait confiance à un matheux pour faire une thèse qui contenait quand même beaucoup de physique, merci pour ta patience et ta pédagogie pour que je puisse comprendre les bases de la spectrométrie gamma et que je puisse mener à bien cette thèse.

Merci aux membres de mon jury d'avoir accepté de lire mon manuscrit et de participer à l'évaluation de mes travaux de thèse. Particulièrement Thomas Rodet qui a participé au suivi à mi-chemin et a vu les progrès réalisés pendant ces trois ans, Abdel-Mjid Nourreddine et Michel Paindavoine pour avoir été rapporteurs de ma thèse et dont les retours ont été très positifs et ont accepté que je soutienne.

Je tiens aussi à remercier le Laboratoire de Métrologie de la Radioactivité dans l'Environnement, juché sur sa colline (ce qui m'a permis de rester en forme physique pendant ces trois ans par une marche d'approche que je ne qualifierais pas de facile). Un laboratoire où les gens savent la valeur du travail, se respectent et partagent des moments de complicité qu'on n'oublie pas. Merci donc Carole, éternelle pipelette et qui prend à cœur de mettre un sourire sur le visage de tous les jeunes du labo. Merci Vanessa pour m'avoir pris plusieurs fois en stop en bas de la côte, et surtout pour m'avoir m'accompagné (parfois à deux jours d'une conférence) dans toutes mes démarches administratives. Merci Sandrine, mon souffre-douleur de l'écologie, avec qui les débats sur l'avion ont toujours été passionnés. Merci Dianna, ton sourire et ta présence tout au long de cette thèse m'ont beaucoup aidé. Merci Nico et Maria-Stella les G.O. du

labo, vous mettez l'ambiance et gérez les fêtes dans la salle blindée comme personne. Merci Blaise et Cédric pour les éclats de rire et les bonnes (ou moins bonnes) blagues aux pauses-café, bonne continuation à Grenoble profite bien de la montagne pour le premier et au LMRE pour le deuxième ! Merci aux chefs de labo, Christophe, Jean-Louis et Eric, pour avoir tenu à continuer un projet sur le démélange spectral et pour m'avoir fait confiance pour le mener à bien. Merci à l'équipe Gamma, Neila, Sina, Sébastien, Lydia, Anne et Anne, faire partie de ce groupe permet de poser des questions sans peur d'être jugé, de se perfectionner dans un domaine technique précis et exigeant et d'être entouré de vrais pros qui mettent du cœur à l'ouvrage et prennent le temps de partager ce qu'ils savent. Merci à Béatrice et Claire pour avoir partagé les chocolats, biscuits, gâteaux, brownies, etc. Merci Maxime, chef de service, pour la confiance que tu as mise dans le projet démélange spectral et pour les échanges que nous avons eus, notamment autour des montres. Je tiens particulièrement à saluer ici mes colocs de bureau, Xavier dans un premier temps, Anne ensuite. Nos échanges sur la physique ont toujours été passionnants (bien que complexes pour le matheux que je suis) et vous m'avez permis de comprendre beaucoup de choses sur la spectrométrie gamma. Les relectures et les corrections jusque dans les moindres détails d'Anne m'ont permis d'améliorer mes articles et ma thèse et je te remercie beaucoup d'y avoir mis tant de cœur ! Pour finir avec le labo, merci aux jeunes, Candice, Mathurin, Johanna, Mélodie et Maëlle, les moments de complicité et les échanges avec chacun d'entre vous marqueront ma thèse et me font me dire que finalement la chimie ça a l'air intéressant aussi, mais ça n'a pas l'air facile !

Merci Anaëlle, amie et compagnon de galère dans la thèse. Grâce à toi j'ai pu voir ce qui pouvait m'attendre dans cette odyssée. J'ai aussi trouvé une personne ouverte, à l'écoute, intègre et compréhensive. Les discussions dans ton bureau ou autour d'un verre ont fait partie de la réussite de cette thèse et ton soutien, les beaux comme les mauvais jours me marqueront longtemps. Merci enfin d'avoir été là pour immortaliser le moment où on m'a donné le titre de docteur.

Merci Jiaxin, grâce à tes travaux sur le démélange spectral j'ai pu commencer la thèse dans de très bonnes conditions, avec un code clair et utilisable avant même que je n'aie compris le sujet. Merci aussi de m'avoir mis en co-auteur de tes deux derniers articles, la contribution à ces derniers m'a permis de comprendre directement les enjeux de la recherche dans ce domaine et d'avoir une première expérience dans la publication et dans le démélange spectral.

Merci Thomas, travailler avec toi a été une expérience riche et intéressante, je suis sûr que ton avenir sera brillant et j'espère que ton passage par le labo

restera un bon souvenir. Merci pour ta contribution à la thèse, je te dédie la partie sur la recalibration car tu y as beaucoup participé.

Pour finir avec le professionnel je tiens à remercier Louise, une amitié qui s'est forgée pendant cette thèse et qui a commencé un certain soir de journée des thèses, après un réveil très tôt et plusieurs heures de train, suite à un passage au micro lors d'une table ronde sur l'écologie. Toi aussi la thèse n'a pas été un long fleuve tranquille et je souhaite que la fin de celle-ci soit plus calme que la mienne. Ma thèse n'aurait pas la même saveur sans toi et je suis très heureux d'avoir fait ta connaissance et que notre amitié continue. J'ai hâte d'être assis dans la salle de ta soutenance.

Je vais maintenant passer aux remerciements plus personnels et commencer par mes soutiens de toujours, ma famille. J'ai la chance de faire partie de deux tribus soudées, aimantes et souriantes. Deux tribus qui font partie de moi comme je fais partie d'elles et qui ont donné un cap à ma vie dès le premier jour de celle-ci. Papa, Maman, promis, je m'arrête là pour les études, je vais passer à un vrai travail mais je vous remercie pour tout ce que vous avez fait pour que j'en arrive à un titre de docteur aujourd'hui, j'espère que vous êtes fiers de votre fils, je vous aime. Léa, je sais que tu ne m'appelleras pas docteur tous les jours, mais je sais qu'au fond, tu es fière de moi et je suis très content d'être ton frère même si mon occupation principale a toujours été de t'embêter. Merci d'avoir été là le jour de ma soutenance pour me soutenir et fêter ça avec moi, même si ça ne s'est pas passé comme prévu. Merci aussi à ma grand-mère Nicole, avec qui je partage maintenant un titre même si le mien ne me permet pas de sauver des gens.

Je tiens maintenant à remercier mes amis de presque toujours, Xavier, Bettina et Marin, si j'en suis là aujourd'hui c'est grâce à vous, je suis fier de vous et heureux de faire partie de ce groupe soudé que l'on a et qui évolue depuis le lycée. Xavier, j'imagine parfois ce que ma vie serait si tu n'avais pas été là et les conclusions sont toujours différentes mais jamais brillantes. Tu es un ami cher à mon cœur et te savoir à mes côtés depuis si longtemps et pour encore un bon bout de chemin me réconforte et me remplit de joie. Bettina, tout au long de ces trois ans de thèse tu m'as poussé, récupéré dans des états pas possibles et emmené en balade culturelles pour me changer les idées ou juste pour passer un bon moment entre amis. Ton canapé est un endroit où je me sens toujours bien, toujours écouté et aimé. Je suis très heureux d'être ton ami et même si parfois ton dosage de la tequila est un peu loupé, tu as su me faire grandir tout au long de notre amitié et j'ai hâte de voir ce que l'avenir nous réserve comme bonnes expériences. Marin, tu as été mon coloc à un moment clef dans mon histoire de matheux, sans toi je ne sais pas ce que

je ferais aujourd'hui mais je ne serais pas docteur en mathématiques. Toi aussi tu as changé ma vie à de multiples reprises et je ne te remercierai jamais assez de m'avoir fait découvrir l'escalade, d'avoir partagé cette année à Strasbourg avec moi et d'être mon ami. Vous me poussez tous les trois à me dépasser et à donner le meilleur de moi-même pour essayer d'être à la hauteur de votre amitié. Je vous aime fort.

Je vais maintenant remercier mes amis de l'escalade, Hervé, Floriane, Julie, Grégoire, Juliette, Marine, Hubert et Mona. Ce sport fait maintenant partie de mon ADN et vous aussi. Pendant ces trois ans vous avez été des compagnons géniaux et m'avez permis de déconnecter, le temps d'une séance, d'un (ou plusieurs) verre(s) ou d'une sortie à Fontainebleau. Sans vous je n'aurais pas tenu et je ne ferais pas des blocs noirs toutes les semaines. Hervé et Floriane, j'ai toujours hâte de vous revoir quand je prévois un retour à Strasbourg et je suis heureux que vous soyez dans ma vie depuis 4 ans, mon amour pour l'escalade ne serait pas le même sans vous. Julie et Grégoire vous m'avez permis de participer à une expérience que je n'oublierai jamais, de me dépasser physiquement comme je n'ai jamais fait, je parle évidemment du GR20 où vous m'avez convié. Merci de m'avoir fait confiance pour partir avec vous dans cette aventure géniale et merci pour tous les bons moments que l'on a partagés et que l'on va partager. Juliette merci d'accepter encore de grimper avec moi alors que je te fais souffrir sur les barres de traction, beastmakers et autres tapis d'abdos, je te rassure, j'ai mal aussi quand je te fais faire ces bêtises, mais j'aime toujours autant avoir des courbatures grâce à toi ! Marine, avant de faire partie du monde de l'escalade tu as fait l'ISUP avec moi, merci d'être encore là pour me supporter (j'imagine que ça va dans les deux sens puisque je te supporte aussi) j'ai hâte de la prochaine séance à Didot avec toi. Hubert et Mona, vous occupez une place spéciale dans la famille des grimpeurs, vous êtes aussi docteurs et partiellement mes colocs (je ne compte plus les soirées passées jusque tard chez vous quand vous étiez à Paris). Merci pour toutes les séances de grimpe décidées au dernier moment, merci pour les discussions autour d'une bière, d'un café, d'une tarte au citron ou les trois ensembles. Merci pour votre écoute, votre joie de vivre et votre dynamisme. Je ne remercierai jamais assez Xavier de m'avoir présenté Hubert un jour (« Tu verras, il est comme toi mais avec un physique de rugbyman et roux, vous devriez bien vous entendre » m'avait-il dit). Depuis, notre amitié a grandi et vous me manquez beaucoup maintenant que vous êtes à Munich. Mais je sais que vous n'êtes pas si loin et qu'on se reverra bientôt (même si ce n'est pas assez tôt à mon goût).

Si un grimpeur lit ces remerciements un dimanche : quand est-ce qu'on grimpe ?

Pour conclure je remercie les Suisses, Valentine et Nina en premier lieu, vous avez fait partie de mon chemin de matheux mais vous êtes aussi des amies chères à mon cœur et votre soutien à Lausanne ou à distance me réchauffe toujours le cœur. Ces remerciements ont déjà été longs mais je pourrais encore en ajouter beaucoup. Je terminerai par la maxime qui aura le mieux résumé ces trois ans : « une bonne thèse est une thèse soutenue », je suis très heureux que la soutenance ait, finalement, pu se faire. Merci encore à tous.

Résumé en français

Cette thèse présente le démélange spectral temporel appliqué à des spectres gamma. La spectrométrie gamma est une technique d'identification et de quantification des radionucléides contenus dans une source radioactive. L'Institut de Radioprotection et de Sûreté Nucléaire, dans sa mission de surveillance de l'environnement radiologique français, se sert de cette technique pour informer la population en cas d'incident ou d'accidents ou pour assurer la mesure en routine des radionucléides présents dans l'environnement. Le but de cette thèse est de rendre plus rapide l'estimation des activités dans un échantillon, notamment en améliorant les outils de démélange spectral développés dans une thèse précédemment encadrée au laboratoire. Ces développements rendront possible l'analyse conjointe de mesures successives de filtres aérosols.

Dans le chapitre 1, le contexte et l'état de l'art de la spectrométrie gamma sont présentés. Ce chapitre commence par un bref rappel des bases de la radioactivité nécessaires pour comprendre le fonctionnement de la spectrométrie gamma. Nous détaillons ensuite les mesures telles qu'elles sont faites au Laboratoire de Métrologie de la Radioactivité dans l'Environnement où ma thèse s'est déroulée. Ce laboratoire a développé de nombreuses techniques pour améliorer la mesure de spectres gamma. Une description des différentes méthodes d'analyse de spectre est faite dans l'état de l'art détaillé des trois principales méthodes utilisées aujourd'hui : l'une basée sur les pics, la seconde basée sur l'ensemble du spectre et la dernière utilisant les algorithmes de machine learning. Une comparaison des avantages et limites de chacune d'elles est présentée.

Dans le chapitre 2, l'analyse temporelle de spectres gamma permettant de réaliser l'analyse conjointe de mesures successives d'un même échantillon est présentée. Le modèle mathématique utilisé y est présenté en détail. De plus, deux modèles décrivant les chaînes de décroissances sont comparés pour comprendre les liens entre les activités de radionucléides en filiation. En effet, ces corrélations sont importantes dans l'analyse de spectres aérosols qui présentent 2 chaînes de décroissances, dues aux descendants particulaires du radon actifs dans les filtres. Les outils algorithmiques présentés sont ensuite testés sur des simulations réalistes et sur une vraie mesure d'un filtre aérosol. Les résultats montrent les performances de ces nouveaux outils et notamment la réduction du délai avant une première estimation des activités après que l'échantillon a été prélevé avec une première estimation des activités après seulement 30 min de mesure. La détection, à un niveau trace, du ^{137}Cs est possible après 1 jour

et demi de mesure permettant de réduire le temps de mesure total et de détecter des anomalies précocement durant la mesure d'un échantillon. Ce premier algorithme d'analyse conjointe de spectres gamma est utilisé sur 11 mesures successives d'un même échantillon comme preuve de concept sans se focaliser sur le temps de calcul necéssaire pour réaliser le démélange.

Dans le chapitre 3 de cette thèse sont présentés les développements réalisés afin de rendre possible l'analyse en ligne d'un spectre gamma. La segmentation de la mesure peut donc être choisie aussi fine que l'on souhaite, afin de réaliser une analyse des activités de l'échantillon au cours de la mesure. Cette segmentation impose des contraintes de rapidité quant au calcul des activités et de nouveaux outils tels que l'utilisation d'un buffer de mesures réduisant la taille des matrices impliquées dans le démélange spectral ont été développés. De plus, une régularisation se basant sur les estimations déjà faites et un arrêt de la mise à jour de certaines activités après que la contribution des radionucléides qui leurs sont liés dans le spectre aient disparues sont présentés afin de réduire le nombre d'itérations nécessaire pour arriver à une nouvelle estimation des activités. L'optimisation de ces nouveaux outils est détaillée et les résultats de l'analyse en ligne sont montrés à nouveau sur des simulations et sur une vraie mesure de filtre aérosol. L'analyse du spectre de la mesure d'un filtre d'aérosols ayant réellement été prélevé montrera la présence d'^{123}I après seulement 2 minutes de mesures à un niveau d'activité de 2.5 Bq, montrant bien les performances de l'algorithme présenté.

Finalement, le chapitre 4 présente la calibration des spectres gamma et des signatures spectrales utilisée pour réaliser le démélange. En effet, la solution parfaite pour obtenir ces dernières serait d'avoir accès à des sources pour l'ensemble des radionucléides que l'on souhaite étudier. Cependant l'accès à ces sources est impossible pour de nombreuses applications, en particulier pour les filtres aérosols où certains radionucléides ont une demi-vie très courte rendant l'utilisation de sources impossibles. Nous avons donc recours à des simulations pour produire les signatures spectrales utilisées dans cette thèse. La calibration de ces simulations et du détecteur est nécessaire afin d'obtenir un ajustement correct des signatures spectrales au spectre mesuré. La calibration que nous proposons est présentée en détail et comparée à la calibration telle que réalisée dans les travaux précédents. La nouvelle calibration se base sur les énergies caractéristiques des pics présents dans le spectre observé et sur une interpolation par spline pour réaliser la calibration en énergie permettant de réduire les écarts en énergie présents en utilisant un modèle linéaire durant cette calibration. Dans un second temps, la calibration en résolution est détaillée, utilisant une convolution par un noyau Gaussien afin de répliquer l'élargissement Gaussien des pics présents dans la mesure.

Summary

This thesis work focuses on the temporal unmixing of gamma-ray spectra. The gamma-ray spectrometry is a technique used in the identification and quantification of the radionuclides contained in a sample or source. The French Institute for Radiation Protection and Nuclear Safety (IRSN) in its mission of monitoring the French environment, uses this technique as it allows to inform the population in case of incidents or accidents or to perform the surveillance of the radiological environment. The aim of this thesis is to make the estimation of the activity in a sample faster by upgrading the algorithm tools already developed in a previous thesis, in order to perform a joint analysis of consecutive spectra obtained by the measurement of aerosol filter samples.

This thesis will thus present in chapter 1 the context and state of the art in gamma-ray spectrometry. In this chapter we will begin by some notions of radioactivity physics needed to understand gamma-ray spectrometry and detail the laboratory measurement as the Environment Radiation Metrology laboratory (LMRE) where my thesis has taken place. This laboratory has already developed a lot of tools to perform better gamma-ray spectrometry. A description of the analysis of a gamma-ray spectrum is presented and a detailed state of the art is given to understand the evolution and limitation of the 3 main analysis techniques for gamma-ray spectrum, namely, peak-based, full-spectrum analysis and finally machine learning techniques used in this field.

In chapter 2 we present the temporal spectral analysis allowing to perform the joint analysis of consecutive measurements of the same sample. The mathematical model used to perform the joint analysis will be detailed. Two models will be investigate considering the Bateman model usage for decay series of radionuclides allowing to model the link between the activity of a parent and daughter radionuclide during their decay. This is particularly useful in the framework of the analysis of aerosol filter sample as 2 decay chains are present in the radon progeny and thus appear in the spectrum. The algorithmic tools to perform the gamma-ray spectrum analysis are also presented and the performance are assessed on simulated data and on a real aerosol filter sample analysis. The results show the performance and reduced time between the sampling and the analysis with first estimations of the activity after 30 min of measurement. The detection of the trace level ^{137}Cs is achieved after a day and a half possibly allowing to reduce the measurement time and perform anomaly detection early on during the measurement process. This first algorithm

to perform the joint analysis of gamma-ray spectra is used on 11 time segment allowing us to produce a proof of concept application without focusing on the computation issues of a data-heavy segmentation of the measurement.

In chapter 3, we present the development made to allow to further decompose the measurement and perform quasi online estimation of the activity. In other words, the segmentation of the measurement can be chosen by the end user leading to any number of segments and imposing a rapid estimation of the activities during the measurement process. New tools are used such as the a buffer of spectra rather than the entire spectra collection allowing to reduce the size of the matrices involved in the unmixing algorithm presented in chapter 2. A regularization of the consecutive estimations is also proposed to allow to compute consecutive activity estimation in a timely fashion and reduce the number of iterations of the unmixing procedure. Finally, the update of the rapidly decaying radionuclide is stop early in the unmixing as soon as their contribution vanish from the spectrum allowing to further reducing the computation time of the unmixing procedure. The optimization of these new tools is detailed on simulation analysis and the performance of the new unmixing algorithm is presented on both simulation and real aerosol filter analysis. The results on the real aerosol filter sample shows the detection and quantification of ^{123}I after only 2 minutes of measurement with as low as 2.5 Bq of activity showing the capability of the presented algorithm in a case of low level contamination allowing to get a result in a timely fashion.

Finally, the chapter 4, focuses on the calibration of the spectral signatures used in the full-spectrum analysis. In fact, while the perfect solutions to get spectral signatures would be to have access to sources for each radionuclide that we could observe, is is usually not possible. For this reason we have to rely on simulated spectral signatures. The calibration of these simulations and the calibration of the detector have to be performed in order to get a good precision of the gamma-ray spectral unmixing. The calibration will be presented in detail and compared to the routine calibration of the detector in peak-based analysis and the previous calibration performed in the previous thesis work. The new calibration uses the information of the nominal energy of the gamma-ray peaks in order to perform the energy calibration of the spectrum. This calibration is done using spline interpolation rather than linear models in order to solve the shifting issues of the latter model. In a second time, the resolution calibration is performed with a non-stationary kernel convolution that mimic the Gaussian energy broadening of the observed spectrum in order to get a better agreement between the observed measurement and the spectral signatures.

1 - Gamma-ray spectrometry

Sommaire

In this chapter the context of my thesis work is explained, along with the physical knowledge useful to understand the gamma-ray spectrometry and the key principles used in this thesis work. In the first section 1.1 the basics on radioactivity are reminded, following are the specificities of the laboratory and the specification of the samples we analyse. Finally, the basics in gamma-ray spectrometry will be shown in section 1.3 and a review of the state of the art in the domain of spectral analysis will be made.

1.1 . General reminders of radioactivity

Radioactivity is a naturally occurring process. It is the decay of an out of equilibrium nucleus due to an excess of neutrons and/or protons to a more stable nucleus. In fact, a nucleus is composed of protons and neutrons, an excess of one of them causes an unstable state usually restored by disintegration and a change of the nucleus.

The atom is composed of the nucleus and its electron shell. The number of electrons, negatively charged, in the electron shell is equal to the number of protons, positively charged, so that the atom is electrically neutral. The number of protons of the nucleus is then defining the chemical element. The periodic table of elements (annex 5) presents the list of all the elements with the corresponding symbol M and number of protons Z.

The nucleus of an atom is characterized by its number of protons and neutrons. It is conventionally specified as ${}^{A}_{Z}M$, where A is called the mass number and is equal to the sum of the number of protons and neutrons. Two atoms of the same chemical element (*ie* same number of protons) may differ in their number of neutrons, they are called "isotopes". For instance, the carbon stable isotope is ${}^{12}_{6}C$, an isotope of carbon is ${}^{14}_{6}C$ due to the two added neutrons. This isotope is not stable and will tend to disintegrate into a more stable element, it is thus called "radioisotope". In this case the nucleus will change to regain a stable state, this is called a disintegration.

The decay from an unstable nucleus to a stable one can be of multiple forms :

— α decay : the nucleus emits a nucleus of ${}^{4}_{2}He$, called "alpha-particle", in the decay process we then have :

$$ {}^{A}_{Z}X \longrightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}He $$

— β^- decay : the nucleus emits an electron

$$ {}^{A}_{Z}X \longrightarrow {}^{A}_{Z+1}Y + e^- + \overline{\nu_e} $$

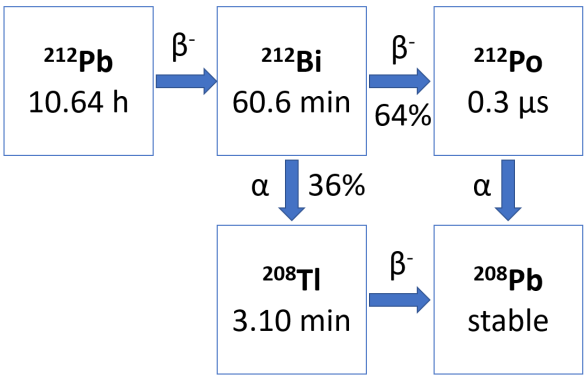


FIGURE 1.1 – *The decay scheme of* $^{212}_{82}Pb$ *to* $^{208}_{82}Pb$

— β^+ decay : the nucleus emits a positron

$$^A_ZX \longrightarrow ^A_{Z-1}Y + e^+ + \nu_e$$

— Electron capture : the nucleus captures an electron

$$^A_ZX + e^- \longrightarrow ^A_{Z-1}Y + \nu_e$$

These different forms of radiation from a parent to a daughter nucleus give place to the decay scheme of a radionuclide. Indeed, in order to decay to a stable element, a radionuclide can follow different paths, each with a different probability of occurrence. For example, in figure 1.1 we can see that $^{212}_{82}$Pb first decays into $^{212}_{83}$Bi by β^- radiation and then can take two different paths, the first with 36% of chance is to decay into $^{208}_{81}$Tl by α radiation and then to decay into $^{208}_{82}$Pb via β^- radiation. The other path is for $^{212}_{83}$Bi to decay by β^- radiation to $^{212}_{84}$Po and then decay to $^{208}_{82}$Pb via α radiation. The consecutive disintegrations are due to the fact that after each decay the nucleus is not yet in a stable state but needs some other disintegration in order to reach a stable element (in this case $^{208}_{82}$Pb).

While some decays produce a daughter nucleus at its fundamental energy level, some decays lead to an excited level. The de-excitation of the nucleus is done by gamma-ray emission, pair creation or conversion electron emission. The first is the emission of a high energy photon generated by the de-excitation of the nucleus, the second is a rare process that consists in the emission of a electron-positron pair, finally the internal conversion is a process where the energy available is transferred to an electron of the atom, which is then ejected. The rearrangement of the electrons surrounding the nucleus, leading to the shift from one of the surrounding electronic orbits to another, leads to the emission of an X-ray, which energy is the difference between the two electron energy levels. The difference between X-rays and gamma-rays is the origin of

the photon, if it comes from the nucleus it is a gamma-ray (usually at high energy >100keV) and if it comes from the rearrangement of the electron shell then it is an X-ray (usually at lower energies, <100keV). The energy of X-rays and γ photons are given in keV (1eV = $1.602176634.10^{-19}$J) following the law :

$$E = h\nu \tag{1.1}$$

Where E is the energy of the photon, h ($\approx 6.63 \times 10^{-34}$J.s) is the Planck constant and ν is the frequency (in s^{-1}) of the photon.

The de-excitation of a daughter nucleus after a disintegration follows a precise scheme representing the successive steps that lead to the de-excited state. An example of such a disintegration scheme is presented in figure 1.2. The disintegration scheme of each radionuclide can be found in the Evaluated and Compiled Nuclear Structure Data (ENSDF, [1]) or Lara-web data base ([2]).

In the decay scheme of the $^{60}_{27}$Co we can see that the β^- disintegration leaves a $^{60}_{28}$Ni nucleus in an excited state with 2505.748 keV to dissipate. This can be done via 2 intermediate states at 2158.61 keV and 1332.508 keV. The transition between the states leads to the emission of a gamma-ray which energy is the difference between the two energies. These transitions are not equally probable, the chance of each path is precised on top of the arrows in the figure. For example the emission of a gamma-ray of 2505.748 keV has a probability of 2.10^{-6}%, while the transition by the 1332.508 keV state is more likely, with the emission of a gamma-ray of 2505.748 - 1332.508 = 1173.228 keV with a 99.85% probability of emission, and a second gamma-ray is then emitted with the energy 1332.508 - 0 = 1332.508 keV with a 99.98 percent of emission probability. These gamma-rays emitted at known energy will then allow the analysis of the gamma-ray spectra as will be detailed in the next section.

Each radioisotope has a characteristic half-life noted $T_{1/2}$, it is the duration after which half of an initial quantity of such a radionuclide will have decayed. In other term an initial quantity $N(0)$ of radioisotope of half-life $T_{1/2}$ will decay following the formula :

$$N(t) = N(0)e^{-\lambda t}$$

where $\lambda = \frac{ln(2)}{T_{1/2}}$ is called the decay constant of a radionuclide and t is the time. For example we can see in figure 1.1 that $^{208}_{81}$Tl has a half-life of 3.10 min while $^{212}_{55}$Pb has a half-life of 10.64 hours.

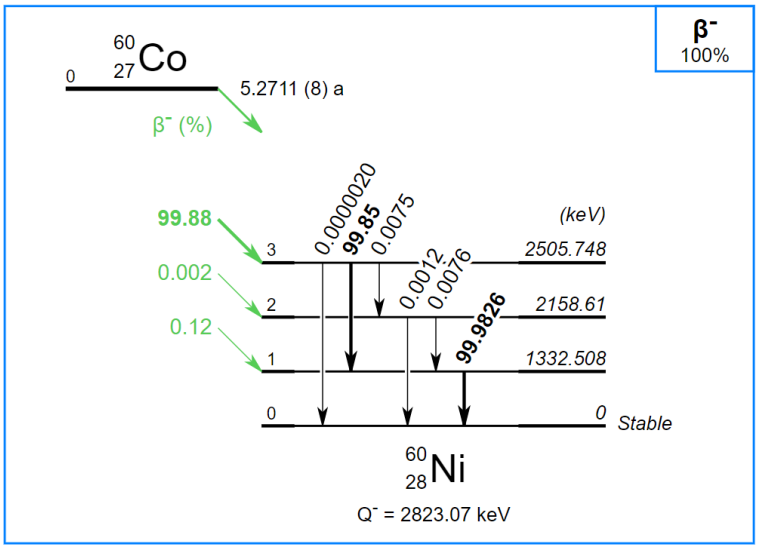


FIGURE 1.2 – *The decay scheme of the ^{60}Co to ^{60}Ni, presenting the various steps of the de-excitation of the atom, from the Lara-web site.*

In figure 1.3 we can see the exponential decay of a quantity N of radionuclide which half life would be 1 hour. At $t = T_{1/2}$ the number of unstable nuclides has been divided by 2, after two hours it has been divided by 4 and so on. The activity $A(t)$ in Bq is the number of radionuclides at time t multiplied by the decay rate λ of this radionuclide. We thus have :

$$\begin{aligned} A(t) &= \lambda N(t) \\ &= -\frac{dN(t)}{dt} \end{aligned} \tag{1.2}$$

The activity corresponds to the opposite of the derivative of the number of radionuclide with respect to time.

1.2 . Radioactivity measurements and laboratory specificities

1.2.1 . Radioactivity sources

As previously stated, the environment is naturally radioactive and, in most cases, not dangerous for health or the environment. But, if the natural radioactivity levels are low on earth, the artificial levels due to the human activity can be high and have an impact on health or environment. Moreover, some naturally occurring phenomena may lead to the accumulation of a radioactive element and have an impact on the population's health. The monitoring of

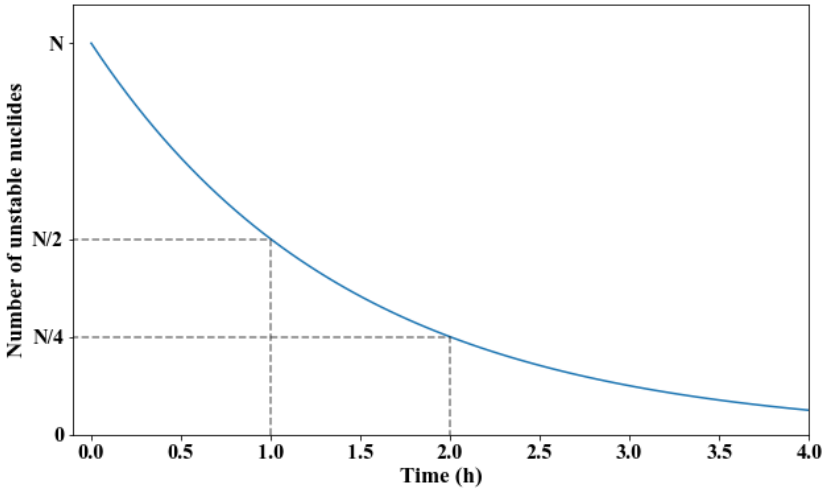


FIGURE 1.3 – *The decay of a quantity N of a radionuclide which half-life is 1h.*

the radioactivity is thus required to ensure that the quantity of radioactivity affecting the population is under control.

The naturally occurring radioactivity is dominated by the telluric and cosmogenic phenomena. In fact, some radioactive elements are naturally present on earth, mostly dominated by the decay products of ^{232}Th, ^{235}U and ^{238}U. Notably, the Radon (Rn) is a radioactive gas that has a radioisotope in each of the three decay series mentioned above. The ^{222}Rn is present in large quantities in the atmosphere and is known to be a cause of cancer in the population. Some other radionuclides are naturally present like ^{40}K present in the soil at a lower activity. On the other hand, the cosmogenic radioactivity is due to the interaction of cosmic flux with the atmosphere. This phenomenon leads, for instance, to the creation of ^{3}H, ^{7}Be, or ^{22}Na. These naturally occurring phenomena are mostly constant and constitute the radiological background.

The artificial radioactivity is the one that is produced by the human activities, it is also called "anthropogenic" radioactivity. It is dominated by the global fallout due to the atmospheric nuclear weapon tests beginning in 1951 and ending in 1963 and the nuclear accidents of Chernobyl and Fukushima in 1986 and 2011. The effect of these events can be seen in figure 1.4 that depicts the evolution of one of their radioactive product (^{137}Cs). Secondly, we have the authorized radioactive releases of nuclear facilities (e.g. ^{58}Co, ^{60}Co, ^{110m}Ag, ...) at low levels. Finally we could find traces of incident or accident releases at trace levels in case of a low magnitude incident or a higher magnitude but farther accident.

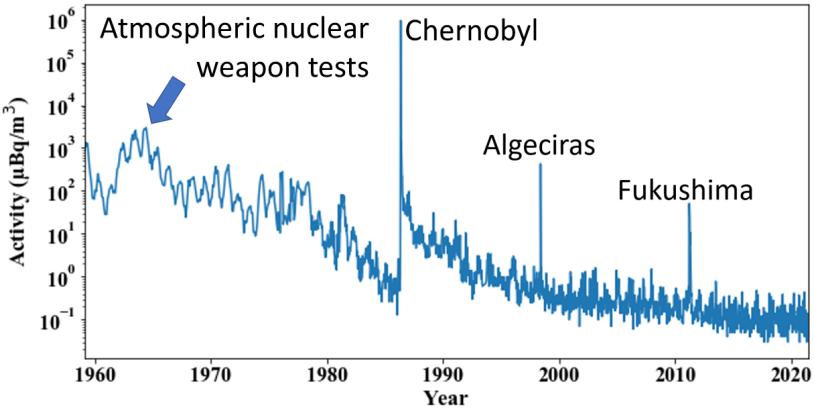


FIGURE 1.4 – *The time series of the* ^{137}Cs *concentration in the air from the analysis of the aerosol filter samples.*

1.2.2 . Radioactivity measurement

The measurement of the radioactivity allows to monitor the level of exposition of the population to the radioactivity. It can be done using different techniques such as alpha-spectrometry, liquid scintillation, mass spectrometry or gamma-spectrometry. Depending on the nature of sample, the application cases or the radioactive elements we focus on, some techniques are better suited than others.

The technique I focused on during my PhD is gamma-ray spectrometry. It allows the identification and quantification of the radionuclides contained in a sample provided that they emit a gamma-ray when they decay. It is a non destructive technique (no alteration of the sample or chemistry are required), with multiple applications from environment monitoring to the astronomical measurement. It revolves on a detector material's (*eg* NaI(Tl) or Ge) interaction with the high energy photon emitted by the sample or source (radioactive sample, observation of the gamma-ray emission of an object to understand its composition, ...).

In fact, a detector material under a high tension will react to incoming gamma-ray by generating an electrical pulse which amplitude is correlated to the deposited energy in the material (see [3] for details). The counting of the events that occur in the detector allows to create a histogram of the deposited energy in the detector, this is called a gamma-ray spectrum. Depending on the material of the detector the shape of the spectrum may vary. However, the analysis of gamma-ray spectra is based on the known energy of the gamma-rays emitted during a disintegration, the position and height of the peaks in a spectrum thus allow to determine the radionuclides present in the sample and

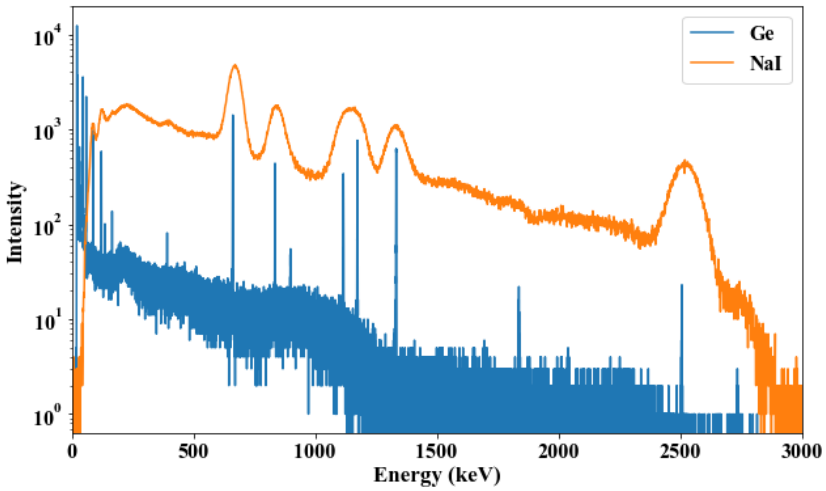


FIGURE 1.5 – *The gamma-ray spectrum of a radioactive source measured with a NaI(Tl) and a Ge detector.*

their activity. Some examples of gamma-ray spectra obtained by measuring a source containing various radionuclides are shown in figure 1.5. We can see that for a same sample the measured gamma-ray spectrum is different. This is due to the detection efficiency of the detector (the faculty of absorption of all the photons emitted by the source) and its resolution (width of the peaks in the spectrum). On the one hand, the resolution of the Ge detector gives place to distinct peaks whereas some overlaps can be observed on the NaI(Tl) detector. This phenomenon can be observed for the two peaks around 1000 keV, in the blue spectrum of figure 1.5, that are overlapping in the orange spectrum. On the other hand, due to the large volume of a NaI(Tl) detector and its higher intrinsic detection efficiency it is usually more efficient than a Ge detector.

In my thesis work I worked on Ge detectors to detect traces of radioactivity in the environment. The next section details the measurement as it is performed in the laboratory and the specificities that allows us to detect radioactivity at very low level.

1.2.3 . Environment Radioactivity Metrology Laboratory

The French Institute for Radiation Protection and Nuclear Safety (Institut de Radioprotection et de Sûreté Nucléaire, IRSN) is mandated to monitor the radioactive level in France and ensure that the nuclear facilities and installations respect the authorization in term of their releases of radioactive substances. As part of this mission the Environment Radioactivity Metrology Laboratory (LMRE) [1] is in charge of the analysis of various samples coming

1. https://en.irsn.fr/EN/Research/Scientific-tools/experimental-facilities-means/LMRE-Facility/Pages/default.aspx

from the French environment.

The missions of this laboratory are :

— The environment monitoring : watching at the level of radioactivity in multiple places all over the French territory and assess that the observed activities are consistent with the historical data or the authorisations of nuclear facilities or installations.

— The anomaly detection and characterisation : in case of incident or accident, our laboratory is able to detect the activity augmentation and the radionuclides that are in the environment. The characterization of such events comes with the determination of the radionuclide (such as ^{131}I that is particularly dangerous for humans) and the quantification of the contamination (activity, duration of the anomaly, ...).

— The study of the radioactivity transfers in the environment, in different matrices (soil, water, air, ...) and the mechanisms behind. For example, the characterization of the ratio between ^{135}Cs and ^{137}Cs is informative of the event that has originated the deposition of Cs (cf : Anaëlle Magré's Thesis [4]), the information of multiple aerosol samplers can be used to locate the origin of an incident with atmospheric dispersion models.

As the levels of radioactivity in France are decreasing, the activity estimation becomes more and more challenging. For this reason, the LMRE has developed throughout the years new techniques and analytical tools to allow the measurement of ever reducing radioactivity levels. Firstly, the LMRE is a unique installation in France due to the underground shielded room reducing the background radioactivity as low as possible and allowing to focus on the sample we observe, even at low levels of radioactivity. In fact, the measurement room is located under a 3m slab of borrated concrete allowing to reduce the cosmic ray induced background. Moreover, the room is shielded with a 10 cm lead wall covered by copper and each detector has his own lead shielding to prevent the telluric noise and the other samples interference. The ambient conditions have also been optimized to achieve the best performances. The air of the counting rooms is renewed (1000 m^3/h) and the fresh air taken on the laboratory roof is filtered by high efficiency filters to catch the radon's particle progeny. The nitrogen flushing of the detector allows to reduce and stabilize the residual radon contribution to the spectra. Moreover the air is also dried to prevent the resolution degradation of the Ge detectors caused by humidity. Secondly, the measurement techniques have been improved with the development

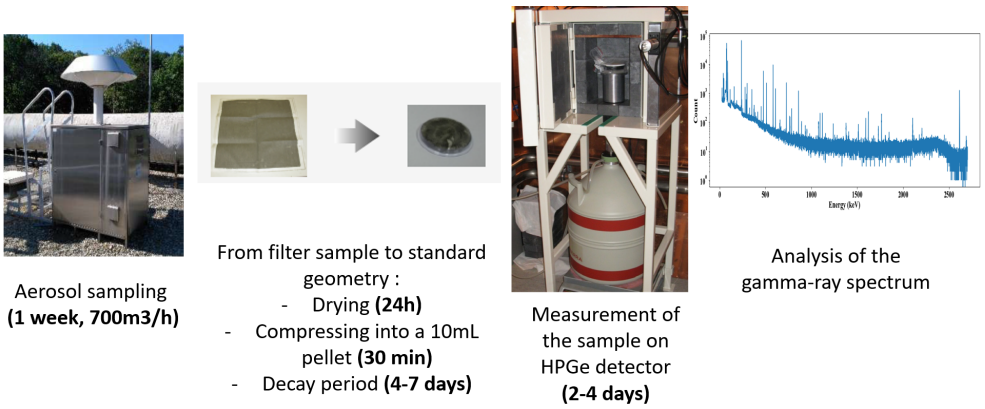


FIGURE 1.6 – *The framework for the aerosol filter sample analysis as performed in the LMRE.*

of better detector. For instance, the laboratory developed the Leda detector with 2 Ge crystal detectors and a surrounding NaI(Tl) detector in order to perform gamma-gamma coincidence measurement (cf Hugues Paradis' s thesis work [5]).

My thesis focuses on aerosol filter samples by gamma-ray spectrometry, they are particularly useful for the environment monitoring and in case of emergency. In fact, these samples are a rapid and reliable source of information of the possible incidents or accidents occurring near a sampling station for low magnitude incident or farther for high magnitude accident, allowing to detect and identify a radioactive contamination. The activity estimation of aerosol filter sample is difficult due to the complexity of the spectrum composed by numerous radionuclides and the various levels of activity observed (from few mBq to hundreds of Bq).

The measurement process for the aerosol filter samples has been optimized in the LMRE to allow to measure the activity at trace level. The aerosol filter sampling is performed with a high volume sampling station (300 - 700 m^3/h) to increase the radioactivity amount to measure, and the filters are put into small standard cylindrical geometry (10 mL) before the measurement is performed to optimize the efficiency of the measurement. Moreover, in order to analyse the low levels of artificial activity such as the one of ^{137}Cs, a decay period of 4 to 7 days is observed before the measurement is done to allow radon progeny to decay and their contribution to vanish from the spectrum. Finally, the measurement is processed on HPGe detectors in the shielded room of the laboratory further reducing the background of the measurement. The steps of this framework are presented in figure 1.6.

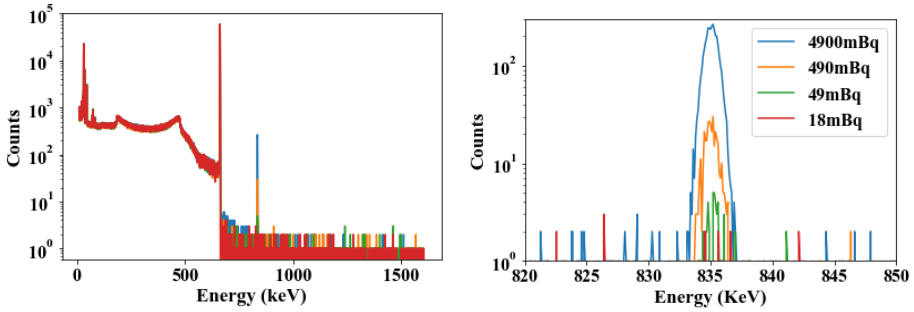


FIGURE 1.7 – *On the left we have 4 measurements of gamma-ray spectra of the same 3 radionuclides (^{54}Mn, ^{210}Pb, ^{137}Cs) with varying levels of ^{54}Mn activity. On the right is a zoom on the full energy peak of ^{54}Mn.*

After the optimization of the signal over noise ratio to the limit, new analysis tools have been developed allowing to perform gamma-ray spectrometry with low levels of activity (cf Jiaxin Xu's thesis [6]). The main issue remaining after these improvements is the rapid detection of an event which is impossible in the actual analysis framework due to the decay period and the long measurement time. In this work we will show how to improve the spectrum analysis in order to reduce the time between the sampling and the analysis of the gamma-ray spectrum.

In the next section I detail the gamma-ray spectrum analysis as it is performed in the laboratory and the new techniques developed in the recent years.

1.3 . Gamma-ray spectrometry

The analysis of gamma-ray spectra is a technique for the identification and quantification of the radionuclides contained in a sample. In fact, a gamma-ray spectrum as the ones in figure 1.7 is the histogram of the energies deposited by the gamma-rays that interact with the detector material. Each energy channel's content can be decomposed into the contributions of each radionuclides composing the spectrum.

The analysis of such spectra is twofold. Firstly, by looking at the energy of the peaks contained in the gamma-ray spectra we have information on the radionuclides composing it, in fact, each radionuclide emits gamma-ray at a specific energy according to its decay scheme (as previously seen in section 1.2), for example ^{137}Cs emitting at 661.7 keV. Secondly, by looking at the intensity of each peak which is proportional to the activity of the emitting radionuclide we can determine their quantity. This is illustrated in figure 1.7 which presents 4 gamma-ray spectra obtained by measuring 4 liquid sources spiked with ^{210}Pb, ^{137}Cs and ^{54}Mn. On the one hand the 4 spectra are composed of the same 3

radionuclides which can be identified by the 3 peaks present in each spectra at 46.5, 661.7 and 834.8 keV respectively. On the second hand, the difference between the 4 gamma-ray spectra as can be viewed on the right is the height of the ^{54}Mn peak varying according to its activity in the measured sample.

A gamma-ray spectrum can thus be decomposed in the different contributions of each radionuclides. As can be seen in figure 1.8 presenting the simulated spectra of 3 different radionuclides, each radionuclide has an individual characteristic spectral signature. The combination of them will form the observed spectrum. To this linear combination is added the background spectrum, it is the ambient noise around the detector, in other term it is the spectrum one would observe without any sample in the detector. It is composed of the radioactivity surrounding the detector, mainly from the building materials and from the shielding, and the intrinsic radioactivity of the detector itself due the natural radioactivity of its components. It is very low in our laboratory thanks to the underground shielded room where the gamma-ray detectors are all the implemented means described in section 1.2.3.

As can be seen in figure 1.8, the spectral signatures can be quite simple as for the ^{40}K or complex in case of multiple gamma-ray creation as for the ^{208}Tl or ^{214}Bi. These spectral signatures can be decomposed in different characteristic elements :

— Full energy peak : it is the peak related to the events when the photon deposes its whole energy. This peak is at the characteristic energy of the gamma-ray emitted after the decay of the radionuclide (*eg* 1460.82 keV for ^{40}K). The gamma-rays energy can be found in the decay scheme of each radionuclide as shown earlier in figure 1.2 for ^{60}Co.

— Compton continuum : this plateau is due to a partial energy deposit of the photons in the material of the detector when a Compton scattering occurs. The characteristic shape of this continuum can be seen in the signature of ^{40}K, in figure 1.8. The steps in the shape of the continuum appear at known energies, based on the model of the Compton scattering and multiple other physical phenomena. The energy of the steps are characteristic of the energy of the photon and thus of the radionuclide that has emitted it.

— Annihilation peak at 511 keV which is due to the annihilation of the positron of a β^+ decay or the annihilation of the positron emitted during a pair creation with an electron of the environment of the sample.

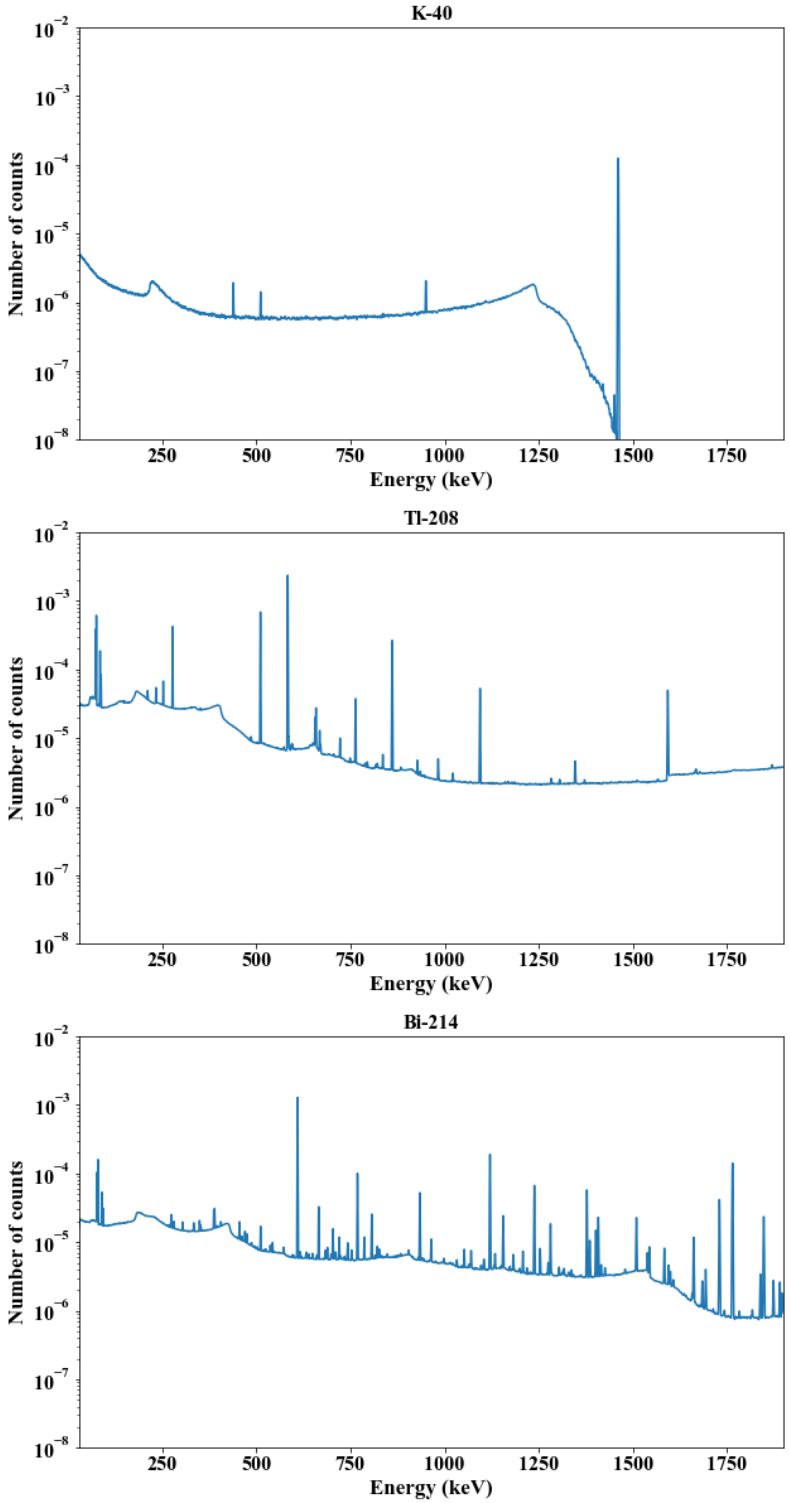


FIGURE 1.8 – *The spectral signatures of 3 radionuclides contained in aerosol filter samples. These are simulations of the spectra obtained on a Ge detector for one disintegration of the observed radionuclide.*

— Escape peaks at the characteristic energy of a gamma-ray less 511 or 1022 keV, due to the creation of an electron/positron pair mentioned above . This type of interaction between photons and the Germanium crystal can occur if the photon energy is higher than 1022 keV (i.e. twice the mass energy of the electron). It is followed by the annihilation of the positron, giving rise to 2 photons of 511 keV, which can escape the crystal without interaction if this annihilation occurs near its edge. In this case 2 peaks at the characteristic peak less 511 keV and less 1022 keV appear in the spectrum as observed in the signature of ^{40}K.

— Low-energy X-ray peaks : These are photons due to the reorganisation of the electron shell following some radionuclide decays or de-excitation or some photon interactions with matter. The energy of the emitted X-rays are characteristic of the atoms.

— A possible sum peak : In the case of rapid consecutive emission of 2 gamma-rays, there is a possibility for the detector to detect both at the same time giving place to a sum peak at the sum of both energies. For example as seen in the ^{208}Tl spectral signature at 1093.9 keV. In fact, it is rather unlikely (0.44% of apparition probability) to see a gamma-ray emitted at this energy but, due to the rapid consecutive emission of a 510.7 and a 583.2 keV gamma-rays, a peak is present at the sum energy 1093.9 keV in the spectral signature resulting of the summing effect.

The estimation of the activity of a radionuclide is allowed by quantifying the contribution of its spectral signature in the spectrum as explained in details in the next section.

1.4 . State of the art of gamma-ray spectrum analysis

The gamma-ray spectrum analysis allows to identify and quantify the radionuclides that compose the measured spectrum. This identification is performed thanks to the characteristic energy of the gamma-rays emitted by each radionuclide composing the sample. The quantification is done thanks to the proportionality between the activity of a given radionuclide and the contribution of the spectral signature in the spectrum, the more active an element is in the sample, the higher will be its contribution in the spectrum. The gamma-ray spectrum analysis can be performed using various data present in the spectrum. In this section we will describe the main ways to analyse such spectra.

1.4.1 . Peak-based analysis

The first technique that I will explain to analyse gamma-ray spectra is the peak-based method. It is the most used technique and softwares are available to perform this analysis (*eg* Genie 2000 from Canberra [7]). The principle is to seek the peaks present in the spectrum, this gives the radionuclides contained in the sample by comparing the observed energies to the characteristic energy of the gamma-ray emitted during a disintegration. For each radionuclide, the activity is then computed via the following equation :

$$A = \frac{n}{\epsilon I t} \tag{1.3}$$

Where :

— A, in Bq is the activity of the radionuclide of interest.

— n, is the net number of counts in the region of interest around the characteristic energy : its determination is detailed later.

— ϵ, is the detection efficiency at the characteristic energy, it is the ratio of detected events over the number of emitted photons. Its determination is mostly empirical.

— t, in second is the duration of the measurement. The radionuclide is considered active throughout the entire measurement.

— I is the intensity of emission of the gamma-ray, it can be found in the de-excitation scheme as previously seen in figure 1.2, it is the percentage of apparition of a given gamma-ray for a disintegration.

The determination of the number of counts n is performed as follows (see figure 1.9). Let e be the characteristic energy of a given radionuclide, the Region Of Interest (ROI) $[e \pm \delta]$ around this energy is computed to encapsulate the peak and a part of the local background. The background continuum is computed as the mean of the counts on the left and right regions of the ROI (orange in the figure). This background is then subtracted from the number of counts in the peak region to obtain the 'net number of counts' n (blue in the figure). The ROI is usually determined as a function of the width of the observed peak, usually the full width at half maximum (FWHM) value is used. For example, the peak region can be $[e \pm 2FWHM]$ and the full ROI be $[e \pm 4FWHM]$ as in the example of figure 1.9.

This method is easy to understand but some issues may arise. First, the

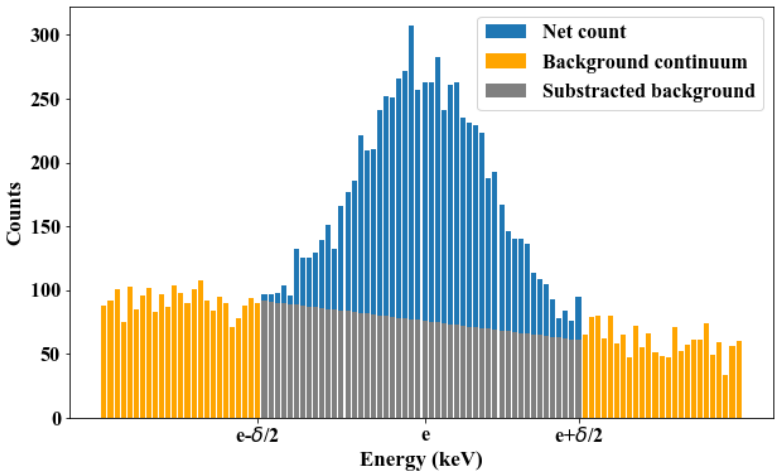


FIGURE 1.9 – *The net number of count is computed by subtracting the background to the number of count in the region of interest around the characteristic energy.*

correct statistics of the spectrum is not taken in account, indeed the computation of the activity using the net number of counts does not use the knowledge of the Poisson statistic of the gamma-ray spectrum. Moreover, in case of multiple peaks in the same area the computation of the net number of counts can be challenging and the deconvolution of the mixed signals may be difficult to achieve if the only information used is the ROI. Finally, in case of low statistics (peak that is at the level of the background) the determination of the net count using this method is not possible, the contribution of some radionuclides can thus be underestimated or missed in the spectrum.

Some authors have generalized the peak-based analysis (see [8]) to account for the exact Poisson statistics of the data. It is interesting to point out that this study demonstrates the need to account for the exact statistics of the measurement. Nevertheless, these methods do not account for the whole information carried by the full spectrum since they rely on region of interest at the vicinity of related peaks. Moreover, the overlapping of the peaks in the ROI is still an issue with this technique, the analysis of complex spectra such as the one we want to analyse in my thesis work is not doable this way. The fact that a radionuclide is supposed active during the whole measurement time is also an issue for some applications such as aerosol filter analysis where the activity of the radon progeny vanish as the decay proceeds. Indeed, in the presence of short-lived radionuclide the background composition of a peak may evolve during the measurement and is more complex than the approximation used in the peak-based analysis.

1.4.2 . Full-spectrum analysis

A second way to analyse gamma-ray spectra is to consider them as a sum of the contribution of all the radionuclides composing the sample. Indeed, as mentioned earlier, a gamma-ray interacting with the detector not only contributes to the full energy peak region but to the whole spectrum leading to a characteristic spectral signature. The full-spectrum analysis uses the entire spectral signature of the radionuclides to determine the presence and activities of the radionuclides.

To perform such analysis the mathematical model used to describe a gamma-ray spectrum y on C channels, with N active radionuclides is :

$$\forall c = 1, \cdots, C; \quad x_c = \sum_{n=1}^{N} \phi_{nc} w_n + \delta b_c, \tag{1.4}$$

$$y_c \sim Poisson(x_c) \tag{1.5}$$

where $\phi_n = [\phi_{n1} \cdots, \phi_{nc}]$ stands for the spectral signature of the n-th radionuclide and b_c for the background per second in channel c, δ is the measurement time in second. The numbers of disintegrations during the measurement time, $\{w_n\}_{n=1,\cdots,N}$, are estimated from the analysis of a single spectrum of a long measurement. The problem of estimating the activity thus resumes to solving an inverse problem as we know the spectral signatures, the background spectrum and the observed spectrum. This will be detailed in chapter 2.

Some author as, in [9] or [10], have proposed other methods for the decomposition of the spectral contributions using stripping methods or singular value decomposition. The presented results are ways to estimate the solution of the inverse problem without actively trying to solve it and without using the whole information contained in the spectral signature. In this thesis work, we focus on the unmixing using the spectral signatures knowledge and thus seek to solve the presented inverse problem.

Further accounting for the full-spectrum information in gamma-ray spectrum analysis has been advocated in [11]. In this article, the authors demonstrate that, compared to the standard peak-based analysis, using the full spectrum improves the sensitivity and reduces the time of measurement. However, these investigations make use of a re-weighted least squares method, which is not fully adapted to account for the Poisson statistics of the data.

Activity estimation problem in gamma-ray spectrometry has been also stu-

died in [12], considering activity estimation as a sparse regression problem. In this article, the authors propose to estimate the number of individual electrical pulses and their arrival times. Using the full-spectrum information allows reducing the measurement time and gaining in sensibility as proven in [11]. In [13] a new method based on the full spectrum analysis is proposed to deal with complex spectra containing a lot of radionuclides. Following these works the method has been improved to include a model selection via a sparse algorithm that allows to choose in a list of possible radionuclides the ones that contribute to the observed spectrum (see [14], [15], [15]). In [16] a new method is proposed to estimate the activities as the measurement is processed by adapting the gamma-ray spectrum model to temporal analysis allowing to get a quicker estimation of the activity of the radionuclide composing the spectrum.

1.4.3 . Machine learning in gamma-ray spectrometry

In the last years, new developments were made in the field of gamma-ray spectrometry using Artificial Intelligence tools such as Convolutional Neural Networks ([17], [18]) or Generative Adversarial Networks ([19]). In these articles the focus is on the creation of a learning set of data via simulation and the performance assessment on detection of radionuclides in different conditions (presence of shielding, in a complex or laboratory controlled environment). These new methods are easy to compute once the training is done, leading to online applications that are fast without more development. On the other hand, these algorithms are very sensitive to variations from the learning data sets to real measurements, for example if the presence of a shielding affects the observed spectra or for low statistics such as low activity or short measurements.

The inherent problem for these methods is the need for a creation of a database of example spectra. Indeed the only way to build such a database is through Monte-Carlo simulations based on spectral signatures (simulated or measured on standard sources). This process is long and can lead to a few endpoints, *eg* the great divergence in the case of spectral signature variability or deterioration of the signatures with a shielding or a background noise causing a bad signal/noise ratio. Moreover these approaches are, for the time being, agnostic to the physics behind the observed data. Indeed, in one hand we have the peak-based and full-spectrum analysis that are building on the physical model of a gamma-ray spectrum to estimate the contributions of the various radionuclides, whereas, on the other hand, AI-driven analysis build upon the correlations in the database of examples without any *a priori* information on the physical processes that take place to generate the spectrum. For example, if a decay chain is present in the spectrum, the correlation between the two is not taken into account except if one has simulated the decay chain in the learning data-base.

1.5 . Conclusion

In a nutshell, the analysis of gamma-ray spectra revolves on the physical information such as the spectral signatures or the energy of the peaks present in the spectrum and the characteristic energy of the radionuclides to identify and quantify the radionuclides contained in a sample. The following table (1.1) sums up the pros and cons of each methods :

At the beginning of my thesis a small experiment was carried to compare the performances of both the spectral unmixing and the peak-based approach using Genie 2000. The experiment is based on the spectra shown in figure 1.7 with 2 radionuclides (namely ^{210}Pb and ^{137}Cs) at a stable level and fewer and fewer ^{54}Mn levels of activity. The detailed results are published in [20] and are reported in annex C.3.1. In a nutshell, the full-spectrum analysis performed better for the really low level of activity and performed as well as the routine analysis using Genie 2000 for the high level of activity. In the context of trace level measurement such as the one performed in the LMRE for the aerosol filter samples the spectral unmixing is thus allowing to decrease the limit of the achievable detection.

On the other hand, the process to allow us to get such simple spectra (only 3 radionuclides and no overlap or low statistics in the peaks) is long, after a week long measurement a decay period has to be observed followed by a 4 days measurement leading to a first activity estimation a week after the sampling at best. The main objective of my thesis is to lower this time between the sampling and the correct estimation of the activity in the sample. This will require to analyse complex spectra before the radon progeny can decrease and to develop new tools to analyse the spectrum as it is measured. These new tools will be presented in the next two chapters 2 and 3.

Moreover, as seen in this introduction, the full-spectrum analysis relies on precise spectral signature simulations to process with the activity estimation. The precise calibration of the spectral signature will be presented in chapter 4 to present the calibration pipeline I developed during my thesis.

Method	Pros	Cons
Peak-based	Easy to perform Software readily available Tolerant to calibration	Hard to use in case of complex spectra Based on Gaussian noise Hardly copes with mixed peaks Problems with low statistics
Full-spectrum	Allows to use the physical model Use the entire information of the spectrum Adapted to the Poisson noise	Makes use of known spectral signatures Can be computationally heavy
AI-driven	Fast once the training is done Software solutions that are ready-to-use and available	Agnostic to any physical model Out of the database examples cannot be analysed Slow and computationally heavy to build the database Only used on simulated spectra for the moment

TABLE 1.1 – *The pros and cons of the different methods used in gamma-ray spectrometry.*

2 - Temporal spectral unmixing

Sommaire

2.1 . Introduction

As seen in chapter 1 gamma-ray spectrometry is one of the major techniques used to measure the activity concentrations of radionuclides in environmental samples. The main challenges of gamma-ray spectra analysis lies in : i) the rapid detection and identification of the radionuclides which can be detected from the sample and ii) the accurate estimation of the radionuclides' activities. This is a particularly complex challenge when the statistics of the counting rate is low. This has attracted a lot of attention in the field of rapid detection and rapid characterization of sources under emergency conditions.

In this chapter, we present the results of a new approach for the estimation of the activity of radionuclides from multiple gamma-ray spectra. Firstly, it builds upon the joint analysis of gamma-ray spectra measured in consecutive time intervals. This allows to account for the radionuclides' activity decay, which bears information to better distinguish between them. Secondly, we introduce a novel spectral unmixing algorithm to tackle multi-temporal measurements and which is based on a recently introduced method ([14]). In contrast to standard methods that perform on single measurements, we propose to investigate two distinct models for the time dependency of the radionuclides' activity : i) the first one assumes that all the radionuclides are in equilibrium, and ii) the second one allows accounting for the presence of radionuclides that are not in equilibrium. In the latter, the method allows taking into consideration the correlation between parent and daughter radionuclides activities, as described in [21]. These models are detailed in Section 2.2.1. Section 2.3 details the proposed multi-measurements spectral unmixing method. In Section 2.4, the proposed method is applied to simulated spectra, to assess the accuracy of the proposed models and algorithms. Next, experimental results on real spectra are presented in Section 2.5. We particularly show that accounting for time-dependency in multiple gamma-ray spectra allows for a faster detection of the radionuclides present in the data while preserving a good accuracy of the estimation of their activities.

2.1.1 . The challenge of the rapid analysis of aerosol filter samples

As seen in chapter 1 the aerosol filter sample analysis is a long process. Firstly, the sampling process, on a high volume air sampler takes a week after which time the sample is weighted and pressed into a standard 10 mL cylindrical geometry. A decay period is then observed, varying from 4 to 7 days, in order for the radon progeny to decay. The effect of this decay period on the spectrum can be observed in figure 2.1, in fact the spectrum after the decay period presents only a reduced number of peaks when compared to the spectrum measured right after the sampling process. After which the measurement

is processed during 2 to 4 days and up to a week in order to get more statistics and a better estimation of the remaining radionuclides and to provide reliable information for the low level of activity we search in the sample (*eg* ^{137}Cs to the mBq level). The whole process is long and the first estimation of the activity of the sample are only given at the end of the measurement, in case of contamination or incident the time between the end of the sampling and the estimation of the activities has to be shorten.

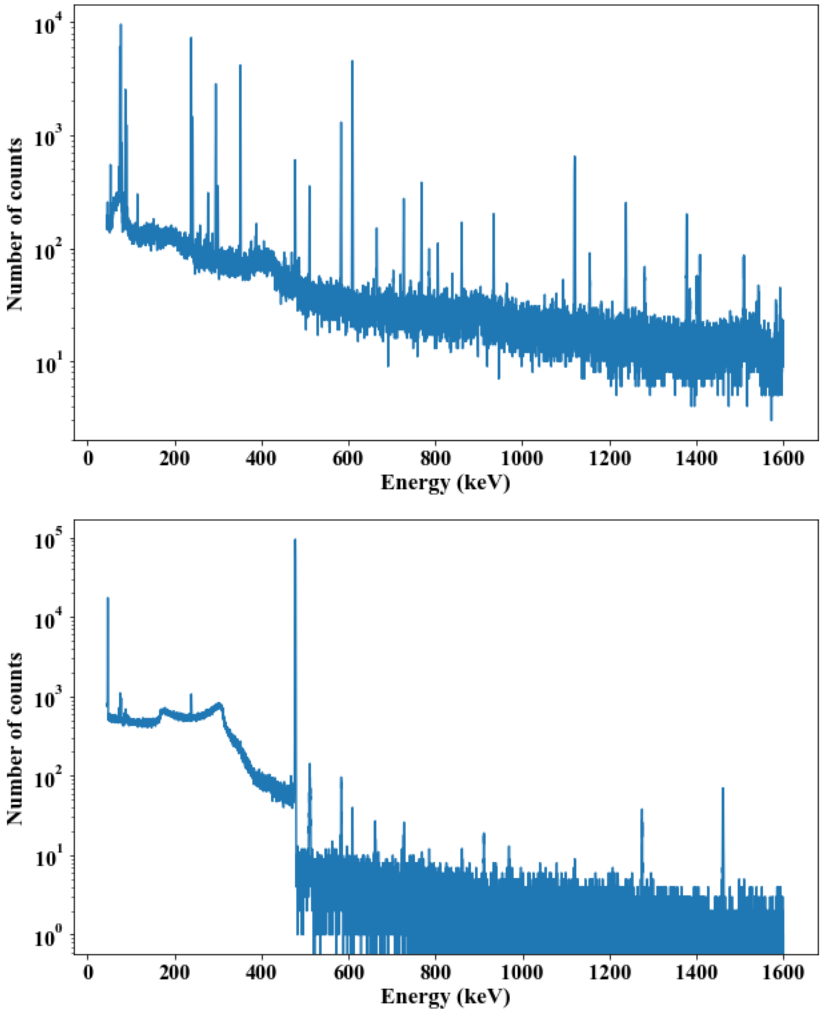


FIGURE 2.1 – *Examples of gamma-ray spectra of two measurements of an aerosol filter : half-an-hour long measurement performed half an hour after sampling (on the top), and a 3 days and 17h long measurement performed 4 days after sampling (on the bottom).*

The problems carried with the radon progeny are twofold. On the one hand, the multiplicity of the radionuclides (^{212}Pb, ^{212}Bi, ^{208}Tl, ^{214}Pb and ^{214}Bi) add complexity to the spectrum. Indeed, the signatures of these radionuclides are

complex and pollute the spectrum with numerous peaks that a peak based analysis will struggle with. Moreover, the activity of the radon progeny are large compared to the activity of the radionuclides of interest that the laboratory wants to estimate, namely ^{7}Be, ^{22}Na, ^{40}K, ^{137}Cs and ^{210}Pb (cf table 2.1 for an example of the activities we usually observe in aerosol filter samples) further hampering the activity estimation of the low level of activity. On the second hand, the half-life of these radionuclides is short (from 20 min to 10.6 hours), however the analysis based on a single measurement is based on the hypothesis that a radionuclide has been active during the whole measurement in order to provide an activity estimation. This second problem is alleviate with the decay period only leaving the long-lived radionuclides in the spectrum but it is problematic in case of rapid measurement of the aerosol filter after it has been sampled.

Radionuclide	Half life	Activity (Bq)
^{7}Be	53.22 days	420
^{22}Na	2.60 years	0.10
^{40}K	1.265 10^{9} years	1.9
^{137}Cs	30.05 years	0.020
^{208}Tl	3.06 minutes	219
^{210}Pb	22.3 years	27
^{212}Bi	60.54 minutes	635
^{212}Pb	10.64 hours	481
^{214}Bi	19.9 minutes	2164
^{214}Pb	26.8 minutes	861
^{228}Ac	14.02 10^{9} years (period of ^{232}Th)	0.075

TABLE 2.1 – *List of radionuclides typically observed in an aerosol filter sample with their half-life and an example of measured activity. These activities will be used in the simulations of Section 2.4.*

In the next section I will present the new framework we propose to resolve these problems and proceed with the measurement without observing a decay period. The successive measurement of the aerosol filter sample allowing to perform the analysis of the spectra containing the radon progeny and still correctly quantify the activity of all the radionuclides contained in the sample.

2.2 . Modelling multi-measurements gamma-ray spectrometry data

The traditional measurement process consists in acquiring a single spectrum which is obtained as an integrated spectrum during a certain amount of time. This largely hampers the ability to rapidly detect artificial radionu-

clides which could trace for incidents or accidents. The aim of this chapter is to present a temporal model that copes with these short-lived radionuclides and is able to detect the low-level radionuclides earlier than the usual week of measurement.

In fact, making use of the time decay of the radionuclides should provide extra information to better discriminate between the radionuclides to be identified. Therefore, we propose a different strategy where the measurements are acquired on multiple, shorter time intervals. This leads to multiple snapshots of the same sample taken in different consecutive time intervals. This further allows taking advantage of the time dependency between these multiple measurements to perform a joint analysis on a sample measured multiple times.

More formally, the variation in time of each of the radionuclides can be described by the following physical time decay model :

$$A_n(t) = A_n(0)e^{-\lambda_n . t}, \tag{2.1}$$

where $A_n(t)$ is the the number of nuclei of the n-th radionuclide at time t, $A_n(0)$ the number of nuclei at the beginning of the measurement and λ_n the decay constant of the radionuclide. Let's define by w_{sn} the total number of disintegration of the n-th radionuclide during the s-th segment ($s = 1, ..., S$) between time t_{s-1} and t_s as follows :

$$\begin{aligned} w_{sn} &= \int_{t_{s-1}}^{t_s} \lambda_n A_n(0) e^{-\lambda_n t} dt & (2.2) \\ &= a_n(0)\psi_{sn} & (2.3) \end{aligned}$$

where $a_n(0) = \lambda_n A_n(0)$ is the activity of the n-th radionuclide at $t = 0$ and ψ_{sn} is defined as :

$$\psi_{sn} = \int_{t_{s-1}}^{t_s} e^{-\lambda_n t} dt \tag{2.4}$$

In the next, and for simplicity, the activity $a_n(0)$ of each radionuclide will be denoted by a_n, which will be the quantity of interest to be estimated from the measurements.

A gamma-ray spectrum measured between time t_{s-1} and t_s is then defined as the linear combination of each of the N radionuclides spectral signatures ϕ_n (*i.e.* the detector's response in energy for a single radionuclide, see [14]). A signature for a radionuclide is obtained by normalizing the spectrum by the measurement time and the activity of the source. A background term b is added (of dimension $1 \times C$, C being the number of channels of the spectrum).

The spectrum for the s-th segment and energy channel c is then described as follows :

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, C : \\ &x_{sc} = \sum_{n=1}^{N} a_n \psi_{sn} \phi_{nc} + b_{sc} \end{aligned} \tag{2.5}$$

where $b_{sc} = (t_s - t_{s-1}) b_c$

Finally, the spectrum is subject to Poisson noise due to the counting process of the detection. Defining y_{sc} as the measured spectrum in the s-th time segment and c-th energy channel, the final model is :

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, C : \\ &y_{sc} \sim Poisson\Big(\sum_{n=1}^{N} a_n \psi_{sn} \phi_{nc} + b_{sc} \Big) \end{aligned} \tag{2.6}$$

where the measured spectrum is a random realization following a Poisson distribution with mean x_{sc}.

Without loss of generality, these models will be illustrated with a sequence of gamma-ray spectra, for which both simulations and real measurements are available. These data originate from an aerosol filter sample which has been measured with a High Purity Germanium (HPGe) detector. They are composed of 11 measurements covering a total time of almost 8 days (670 800s). As displayed in Figure 2.2, time intervals of various lengths have been chosen. The starting point has been taken 30 minutes after the end of the sampling process, this is due to the sample preparation process. More details about the detector setup is given in 2.5.1. The detailed table of the length of every time segment is presented in table 2.2. The segmentation has been done so that the first 7 spectra contain approximately the same number of counts ($\approx$1 000 000), the last 5 segments have been chosen so that we can start and stop each time segment during a standard work day while still having a high number of counts ($>$1 000 000 each) even with the complete decay of the radon progeny.

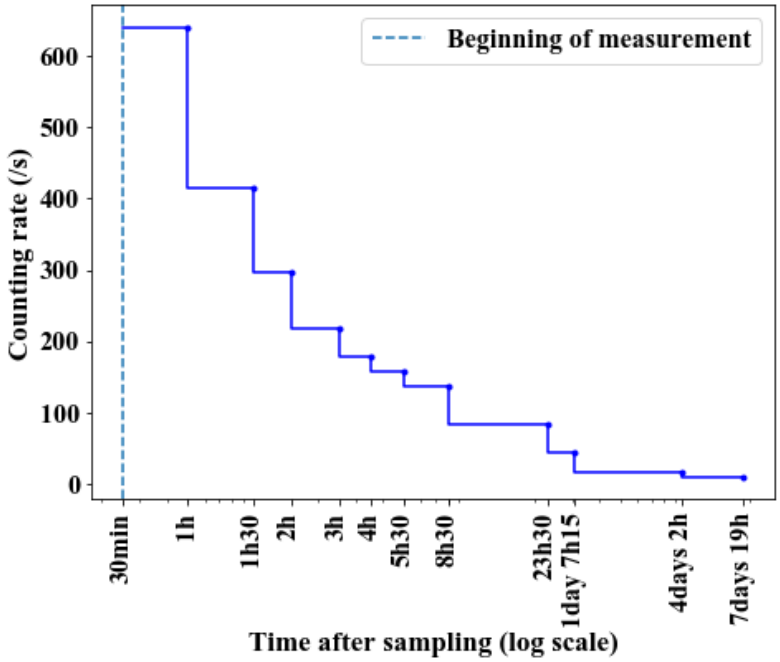


FIGURE 2.2 – *Temporal scheme of measurement used in this paper.*

Segment	Start time	Duration
1	30min	1 800s (30min)
2	1h	1 800s (30min)
3	1h30	1 800s (30min)
4	2h	3 600s (1h)
5	3h	3 600s (1h)
6	4h	5 400s (1h30)
7	5h30	10 800s (3h)
8	8h30	54 000s (15h)
9	23h30	28 000s (∼7h45)
10	∼1 day 7h15	240 000s (∼2 days 19h)
11	∼4 days 2h	320 000s (∼3 days 17h)

TABLE 2.2 – *Temporal scheme of the measurements used throughout this paper.*

2.2.1 . Modelling the decay series

Accounting for temporal information in spectral unmixing first requires designing dedicated models to describe multi-temporal gamma-ray spectra. To that purpose, we hereafter focus on various models for multi-temporal gamma-ray spectra, with increasing complexity.

In fact, in aerosol filter samples one of the issues that arise when the measurement is processed early after the sampling is the radon progeny polluting the measured gamma-ray spectrum. In order to properly model the time correlations between consecutive gamma-ray spectra we have to get a proper understanding of the activity of the radioactive series of radionuclides. More precisely, when radionuclides are in the same radioactive sequence and the half-life of the parent radionuclide is longer than the half-life of the daughter radionuclide, the parent's activity affects the daughter's. In fact, any radionuclide sequence is described by Bateman's equation which relates the number of nuclei of a daughter radionuclide in terms of its decay constant λ_d, its original quantity at $t = 0$, $A_d(0)$, its parent decay constant λ_p and original quantity $A_p(0)$, and the branching ratio r for the parent radionuclide to disintegrate into the daughter radionuclide (see [21] for more details) :

$$A_d(t) = A_d(0)e^{-\lambda_d t} + r\frac{\lambda_p}{\lambda_d - \lambda_p}A_p(0)\left(e^{-\lambda_p t} - e^{-\lambda_d t}\right) \tag{2.7}$$

As an illustration, let us consider the ^{214}Pb - ^{214}Bi chain :

$$^{214}Pb(27min) \longrightarrow ^{214}\ Bi(20min)$$

In this case, the half-life of the parent radionuclide is of the same order of magnitude as the daughter's half-life : $T_{parent} \approx T_{daughter}$. For this disintegration chain, the correlation between the activity of the parent and daughter radionuclide can be described with Bateman's equation and the activity of ^{214}Pb will participate in the activity of ^{214}Bi (in this decay series the branching ratio is 1).

As a second example, let us consider the radioactive series composed of ^{212}Pb, ^{212}Bi and ^{208}Tl, which can be described as follows :

$$^{212}Pb(10.6h) \longrightarrow ^{212}\ Bi(61min) \longrightarrow ^{208}\ Tl(3min)$$

In this chain, each radionuclide disintegrates in a descendant with a much shorter period. In this case we observe that the descendants ^{208}Tl and ^{212}Bi are decreasing with the period of their parent ^{212}Pb. It is a state that is known as equilibrium. This phenomenon occurs when the disintegration of the daughter

cannot be faster than its creation. Usually it is accompanied by a constant ratio of activity between the parent and the daughter (the ratio is $\frac{A_d(t)}{A_p(t)} = r\frac{\lambda_d}{\lambda_d - \lambda_p}$). In our study we observe that each radionuclide of the decay series have the decay constant of ^{212}Pb but the ratio is not achieved, in the paper we will call this state 'equilibrium'.

In the next sections, we describe the different multi-temporal models, following the Bateman's equation or considering the decay series at equilibrium.

Model for multi-temporal gamma-ray spectra in equilibrium

The first model to be considered is dedicated to the decay series which are at equilibrium. In each decay series the half-life of the daughter radionuclides are considered to be equal to the half-life of the parent. This model is described as in Equation 2.5, and is composed of different quantities, which are summarized below :

— a_n : the activities $a_n(0)$ for each of the N radionuclides (a vector of size N)

— ϕ_{nc} : the spectral signature of the n-th radionuclide in the c-th energy channel.

— ψ_{sn} : the temporal signatures of the n-th radionuclide in the s-th time segment, described in Equation 2.4.

— b_{sc} is the background in the c-th energy channel. It originates from cosmological or telluric gamma activity. It is assumed to be constant during the acquisition time so that the background for a given time interval is defined as $(t_s - t_{s-1})b_c$. The background measurement is the spectrum which is obtained by leaving the detector empty during the acquisition process.

Throughout this article, the spectral signatures $\{\phi_{nc}\}_{n,c}$ are assumed to be known as it represents the energy signature of the radionuclide. It can be derived from measurements of a standard source containing a single radionuclide. However, this can only apply to radionuclides for which a standard source can be made. This is not the case for some of the radionuclides that can be identified in aerosol samples. For that purpose, we rather compute the spectral signatures from simulations (see 2.5.1).

Model for multi-temporal gamma-ray spectra with Bateman's equation

In the previous section, the temporal signatures do not model for possible out of equilibrium radionuclides. To further deal with the radionuclides which are not in equilibrium, it is necessary to introduce an extra term : ψ_s^B, modelling the interaction between the daughter and its parent radionuclide, such as with the ^{214}Pb/^{214}Bi family. This extra factor takes into account the activity of the parent radionuclide but will have an impact on the daughter's activity. This is described by Bateman's equation 2.7. To get this equation in term of activity we apply multiply by λ_d :

$$a_d(t) = a_d(0)e^{-\lambda_d t} + r a_p(0)\frac{\lambda_d}{\lambda_d - \lambda_p}\left(e^{-\lambda_p t} - e^{-\lambda_d t}\right) \tag{2.8}$$

which models for the relationship between the parent and daughter radionuclides. The expression of the extra term ψ_s^B is then defined as follows :

$$\psi_s^B = \frac{\lambda_d}{\lambda_d - \lambda_p}\int_{t_{s-1}}^{t_s}\left(e^{-\lambda_p t} - e^{-\lambda_d t}\right)dt \tag{2.9}$$

Without loss of generality, the model in Equation 2.5 for a disintegration family of two radionuclides, a parent and its daughter, reads as :

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, C : \\ &\quad x_{sc} = \sum_{n=1}^{N} a_n \psi_{sn}\phi_{nc} + a_p \psi_s^{(B)}\phi_{dc} + b_{sc} \end{aligned} \tag{2.10}$$

Where a_p is the activity of the parent radionuclide, $\psi_s^{(B)}$ is the Bateman term as defined in equation 2.9 and ϕ_{dc} is the spectral signature of the daughter radionuclide. The second term in equation 2.10 is added for each decay family which is out of equilibrium and needs the Bateman equation to be properly modelized.

All the models are compared in figure 2.3 in the case of 1000 nucleus of ^{214}Bi and 1000 nucleus of ^{214}Pb at $t = 0$. In blue, we consider the case where the parent and daughter are at equilibrium, the decay rate of the parent ^{214}Pb

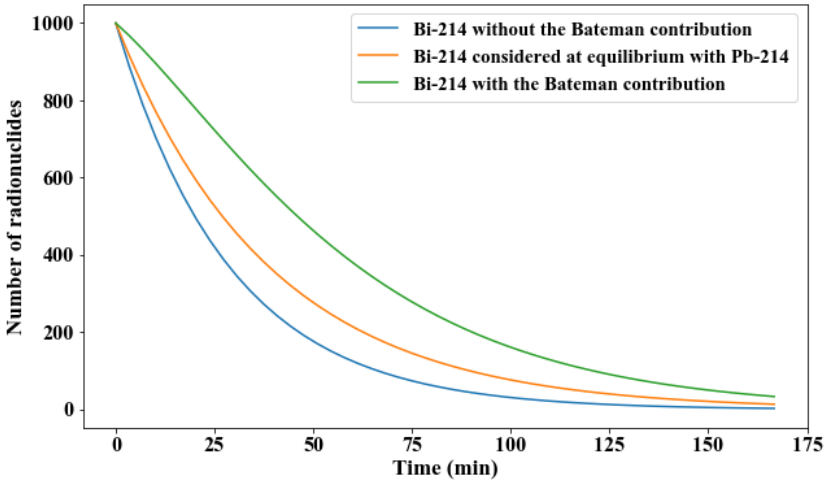


FIGURE 2.3 – *The comparison of the number of nucleus of the daughter radionuclide between the models presented in this section for 1000 nucleus of* ^{214}Bi *and* ^{214}Pb.

has thus been adopted by the daughter radionuclide resulting in a 20 min half-life. In orange, we consider the simple decay of ^{214}Bi with its own half-life of 27 min. Finally in green we consider the Bateman equation in which ^{214}Bi decay with its own period and an added component coming from the decay of the parent ^{214}Pb is observed.

In the next section, we will investigate how these models can be used to design spectral unmixing algorithms that are specifically tailored to analyse multi-temporal measurements.

2.3 . Spectral unmixing algorithms for multi-temporal measurements

2.3.1 . Statistical modeling

In this section, we introduce a new spectral unmixing algorithm to analyse multi-temporal measurement in gamma-ray spectrometry. For that purpose, we first need to define an estimator for the mixing weights of the mixture model that describes gamma-ray spectra. Let us recall that, for each channel $c = 1, ..., C$ and time segment $s = 1, ..., S$, the general expression for the mixture model is of the form :

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, C; \\ &x_{sc} = \sum_{n=1}^{N} a_n \psi_{s,n} \phi_{nc} + b_{sc} \end{aligned} \tag{2.11}$$

This expression gives the average number of counts per energy channel and time interval. The actual measurement follows a Poisson distribution, which can be formalized as follows :

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, C; \\ &y_{sc} \sim Poisson(x_{sc}) \\ &\sim \frac{x_{sc}^{y_{sc}} e^{-x_{sc}}}{y_{sc}!} \end{aligned} \tag{2.12}$$

Since the channels and time segment are statistically independent, the likelihood with respect to the complete multi-temporal measurements is given by :

$$\mathcal{P}(\{y_{sc}\}_{s,c}|\{x_{sc}\}_{s,c}) = \prod_{s=1}^{S}\prod_{c=1}^{C} \frac{x_{sc}^{y_{sc}} e^{(-x_{sc})}}{y_{sc}!} \tag{2.13}$$

It is then customary to take the log to simplify the expression by defining the negative log likelihood :

$$L(\{y_{sc}\}_{s,c}|\{x_{sc}\}_{s,c}) = \sum_{s,c} x_{sc} - y_{sc}\log(x_{sc}) + \log(y_{sc}!) \tag{2.14}$$

Since the model only depends on the unknown activities $\{a_n\}_n$, and following [14], we take the negative log likelihood and add an extra constraint to enforce the non-negativity of the activities :

$$\{\hat{a}_n\}_n = \text{argmax}_{a\geq 0} L(\{y_{sc}\}_{s,c}|\{a_n\}_n) \tag{2.15}$$

$$= \text{argmax}_{a\geq 0} \sum_{s,c} x_{sc} - y_{sc}\log(x_{sc}) \tag{2.16}$$

where $\log(y_{sc}!)$ is omitted since it is constant and each term x_{sc} is a function of the activities $\{a_n\}_n$ according to models 2.5 or 2.10.

The problem in Equation 2.15 does not admit a closed-form expression and must be evaluated using a minimisation algorithm. The main challenge is that the presence of the non-negativity constraint makes this problem non-differentiable. A traditional gradient descent algorithm cannot be adopted. Fortunately, different types of algorithms can be used, such as primal-dual

algorithms (see [22, 14]) or the multiplicative update algorithm (see [23]). Primal-dual algorithms offer a highly flexible framework at the cost of the need for tuning extra hyperparameters (see [14] for more details). For its simplicity, we rather focus on extending the multiplicative update algorithms to multi-temporal measurements.

2.3.2 . The multiplicative update algorithm and its extension to multi-temporal data

In this section, we introduce a spectral unmixing algorithm, based on the multiplicative update algorithm, which is tailored to the three following models :

— Model "without time" : Segment by segment analysis, no time correlation, we analyse the whole spectrum and estimate the activity in each temporal segment using the algorithm developed in [14].

— Model "equilibrium hypothesis" : we benefit from temporal information with the joint analysis of the multi-temporal measurements. This amounts to use the temporal signatures to model for the time dependencies. In this model, every disintegration family is assumed to be at equilibrium.

— "Bateman" model : This model makes use of Bateman's equation to model for the temporal dependencies between radionuclides that are not in equilibrium.

In each model the algorithm is based on the non-negative matrix factorization ([23]). The difference lies in the update of the activities. We will now present the most complex case of "Bateman" model as the others deduct directly from this one.

The multiplicative update algorithm first builds upon a gradient-based minimisation scheme. The derivative of the negative log likelihood with respect to each parameter a_n is given by :

$$\frac{\partial x_{sc}}{\partial a_n} = \psi_{sn}\phi_{nc} + \phi_{nc}\psi_{sn}^{(B)} \tag{2.17}$$

If the n-th radionuclide is not in a radioactive series, then $\psi_{sn}^{(B)} = 0$ and the second term disappears. This then builds down to :

$$\frac{\partial L(\{y_{sc}\}_{s,c}|\{x_{sc}\}_{s,c})}{\partial a_n} = \sum_{s,c}\left(\psi_{sn}\phi_{nc} - \frac{\psi_{sn}\phi_{nc}y_{sc}}{x_{sc}}\right) \tag{2.18}$$

Following [23], the update rule of the multiplicative update algorithm can be obtained by zeroing the derivative of the negative-log-likelihood with respect to each parameter. The main difference between all the mixture models lies in the structure of the temporal signatures, which embeds the temporal information and is dependent on the temporal scheme of the measurements. The update rule for each model is defined as follows for each step k of the algorithm :

— Model "without time" : each time segment is processed independently, the activities of the sought-after radionuclides are then dependent on the time segment s, which yields multiple estimations of the activities $\{a_{sn}\}_{s,n}$. The resulting update rule then reads :

$$a_{sn}^{(k+1)} = a_{sn}^{(k)} \frac{\sum_{c=1}^{M} \phi_{nc} y_{sc}/x_{sc}}{\sum_{c=1}^{M} \phi_{nc}}, \tag{2.19}$$

where $x_{sc} = \sum_{n=1}^{N} \phi_{nc} a_{sn}^{(k)} + b_{sc}$. This is similar to the update rule of the algorithm proposed in [13].

— Model "equilibrium hypothesis" : does include the temporal signatures, but sets the Bateman's term at 0 :

$$a_{n}^{(k+1)} = a_{n}^{(k)} \frac{\sum_{s,c} \psi_{sn}\phi_{nc} y_{sc}/x_{sc}}{\sum_{s=1}^{S} \psi_{sn} \sum_{c=1}^{M} \phi_{nc}} \tag{2.20}$$

where $x_{sc} = \sum_{n} \psi_{sn}\phi_{nc} a_{n}^{(k)} + b_{sc}$. The activities are estimated considering that every decay series are at equilibrium.

— "Bateman" model : This model includes Bateman's equation for decay series which are not in equilibrium (*i.e.* for instance ^{214}Pb/ ^{214}Bi parent/daughter chain). The multiplicative update rule is then defined as follows :

$$a_{p}^{(k+1)} = a_{p}^{(k)} \frac{\sum_{s,c} \left(\psi_{s,p}\phi_{p,c} + \psi_{s,d}^{(B)}\phi_{d,c}\right) y_{sc}/x_{sc}}{\sum_{s=1}^{S} \psi_{s,p} \sum_{c=1}^{M} \phi_{p,c} + \sum_{s=1}^{S} \psi_{s,d}^{(B)} \sum_{c=1}^{M} \phi_{d,c}} \tag{2.21}$$

where a_p (resp. $\psi_{p,c}$ and $\phi_{p,c}$) stands for the parent activity (resp. its temporal and spectral signatures), $\psi_{s,d}^{(B)}$ stands for the corresponding Bateman temporal signature and $\phi_{d,c}$ is the associated daughter spectral signature. The current estimate of the mixture model is given by

$x_{sc} = \sum_n \psi_{sn}\phi_{nc}a_n^{(k)} + b_{sc} + a_p^{(k)}\psi_{s,d}^{(B)}\phi_{d,c} + b_{sc}$. The activities of the radionuclides that are in equilibrium are updated as in Model "Equilibrium hypothesis".

These update rules are performed sequentially until convergence is reached. The stopping criteria we impose to our algorithm is to stop when the relative variation of $a^{(k)}$, which is defined by :

$$\frac{\sum_n (a_n^{(k+1)} - a_n^{(k)})^2}{\sum_n {a_n^{(k)}}^2} \leq \epsilon \tag{2.22}$$

is smaller than value $\epsilon = 10^{-6}$. The maximum number of iterations is fixed to $K = 2000$ iterations.

2.4 . Numerical experiments with simulated data

2.4.1 . Description of the simulations

In order to simulate a spectrum similar to the one we obtain in the environmental monitoring, we use the previous models and simulate the most common radionuclides in the samples we usually gather. The radionuclide dictionary will then be : ^{7}Be, ^{22}Na, ^{40}K, ^{137}Cs, ^{210}Pb, ^{228}Ac, ^{212}Pb, ^{212}Bi, ^{208}Tl, ^{214}Pb and ^{214}Bi. As previously seen (Section 2.2.1), the ^{212}Pb, ^{212}Bi, ^{208}Tl series is considered at equilibrium. Bateman's equation is used to model the activities of ^{214}Pb and ^{214}Bi. These simulations will be used to illustrate the differences between the unmixing algorithms and the importance of the theoretical knowledge carried out by the choice of the temporal models.

In order to simulate gamma-ray spectra, we make use of the mathematical model described in equation 2.10, and precisely described as follows :

$$x_{sc} = \sum_n a_n\psi_{sn}\phi_{nc} + a_{^{214}Pb}\psi_s^{(B)}\phi_{^{214}Bi,c} + b_{sc} \tag{2.23}$$

where,

— the term ϕ_{nc} stands for the spectral signatures of the n-th radionuclide in the channel c, which are computed thanks to MCNP-CP simulations (see [24]). This software allows simulating the response of the HPGe detector used for measurements, as if standard sources were used for each radionuclide. We refer the reader to 2.5.1 for more details about

the simulations.

— The activities $\{a_n\}_n$ of the radionuclides are chosen to be close to a real aerosol sample. The levels are presented in Table 2.1.

— the temporal signatures $\{\psi_{sn}\}_{s,n}$ are computed according Equation 2.4. The duration of the time intervals are presented in Table 2.2.

— The term $\psi_s^{(B)}$ corresponds to the the Bateman term as defined in Equation 2.9.

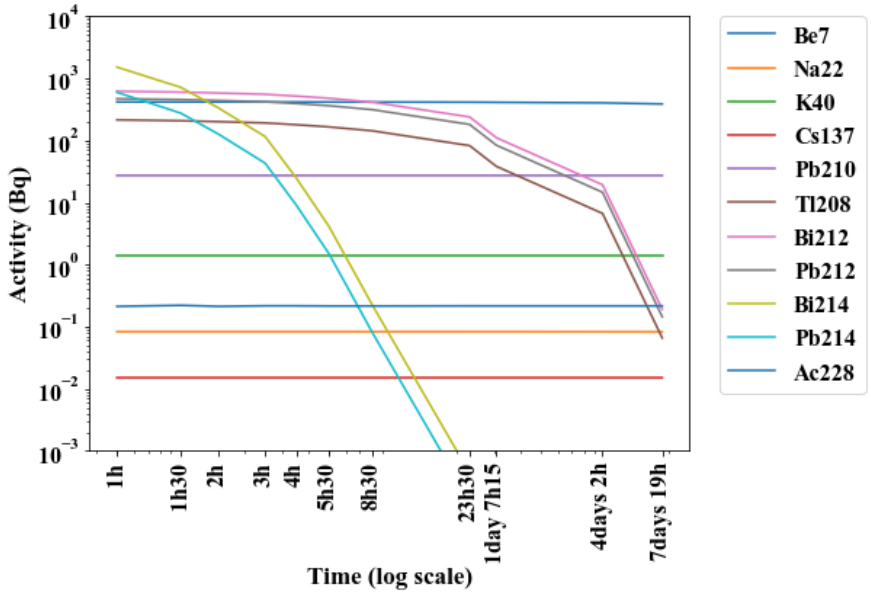
— The background b_{sc} is a real measurement, realized with the empty HPGe detector used in the laboratory over a long time (typically one week). It is customary to take such a spectrum as background for the analysis that we carry in the laboratory. The background spectrum is then reduced to the background per second (by dividing by the counting time) and multiplied by the counting time of the segments we are using.

As it can be seen in top-left panel of Figure 2.4, the simulated radionuclides have different behaviours through time. Three categories can be distinguished. The first one corresponds to the long-lived radionuclides :^{7}Be, ^{22}Na, ^{40}K, ^{137}Cs, ^{210}Pb, ^{228}Ac. The activities of these radionuclides are mostly constant over time since their half-lives are much greater than a week. The activities of these long-lived radionuclides vary quite significantly from few mBq for ^{137}Cs to 420 Bq for ^{7}Be allowing us to compare the performances of the models at low and higher activities. The second category is composed of the ^{212}Pb, ^{212}Bi, ^{208}Tl decay chain which is considered at equilibrium. The daughter radionuclides of this series have short half-lives (3 minutes up to an hour). Finally, the third one is related to the ^{214}Pb and ^{214}Bi decay chain that decreases quickly. It is well described by Bateman's equation. It is expected that the impact of the model will be very important on the estimation of their activities. Monte-Carlo simulations of gamma-ray spectra are then obtained by drawing Poisson realizations.

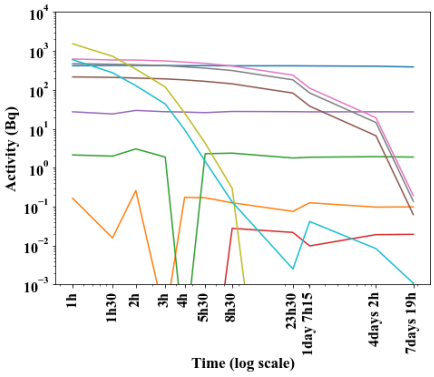
2.4.2 . Comparisons between the different models

In this section, we aim at evaluating the impact of the temporal models on the estimation of the activities. Figure 2.4 shows the estimated activities in Bq for each time segment and each model, which allows drawing the following conclusions :

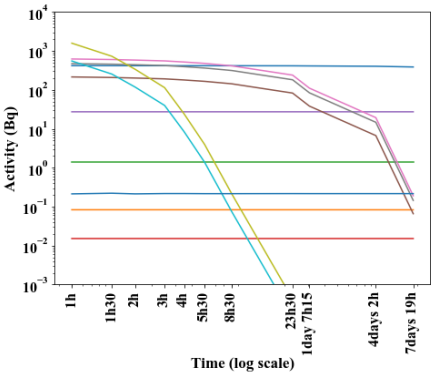
— The model "without time" (b in Figure 2.4) provides accurate activity estimations for the long-lived radionuclides within the last 3 intervals,



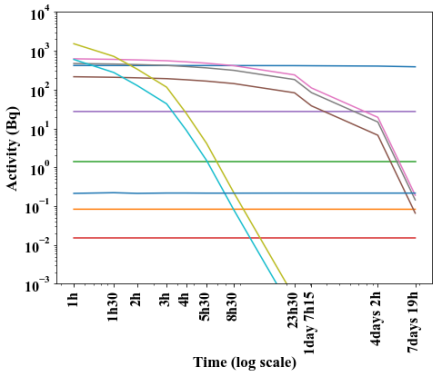
(a) Simulated activities



(b) Model "without time"



(c) Model "Equilibrium hypothesis"



(d) "Bateman" model

FIGURE 2.4 – *Simulated and estimated number of counts per second for the different models and time intervals.*

where the short-lived radionuclides have vanished. In contrast, it is not able to correctly retrieve the activities of the short-lived radionuclides after the third or fourth intervals, when their activities have very low levels. It is important to point out that the estimated activities may vary significantly in time, since no information about the time dependencies of the activities is used in the unmixing procedure. For example, the activity of ^{214}Bi (light green in Figure 2.4) is estimated correctly in the first 6 or 7 intervals but is completely wrong in the last ones since the radionuclide have almost completely disintegrated.

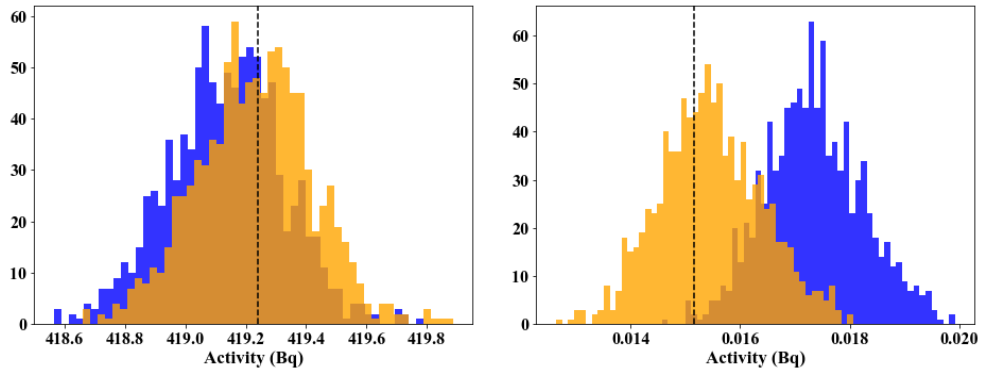
— The two models "Equilibrium hypothesis" and "Bateman", which account for time dependencies (c and d in Figure 2.4), yield similar unmixing results. Both give activities' estimates which are close to the ground truth activities.

In order to precisely assess the differences between the unmixing algorithms based on the two temporal models, we propose evaluating experimentally the estimator bias and variance of each method. For that purpose, we applied the unmixing algorithms to 1000 Monte-Carlo simulations of the same mixture.

Figure 2.5 shows the distribution of the estimated activities for all the 1000 Monte-Carlo simulations, for both Model "Equilibrium hypothesis" (in blue) and "Bateman" model (in orange). These distributions are displayed for 4 different radionuclides. For short-lived radionuclides (c and d in Figure 2.5), Model "Equilibrium hypothesis" yields a clear bias, while "Bateman" model provides an unbiased solution. This is especially true for ^{214}Bi. This discrepancy is clearly related to the ability of the "Bateman" model to take into account the correlation in time of the activities of this radionuclide. More interestingly, for long-lived radionuclides, such as ^{7}Be and ^{137}Cs (a and b in Figure 2.5), the "Bateman" model provides numerically unbiased solutions while Model "Equilibrium hypothesis" yields a slight but statistically significant deviation from the ground truth values. Since these radionuclides have longer half-lives, we would have expected no impact of the choice of a given time models on the estimated activities. This reveals that i) both models "Equilibrium hypothesis" and "Bateman" use information in all time segments to estimate the activities, and ii) correctly modelling the activity in time of short-lived radionuclides also impacts the activity estimation for radionuclides with longer lives iii) we note that the estimation is not biased if the correct model is taken.

2.4.3 . The role played by temporal information

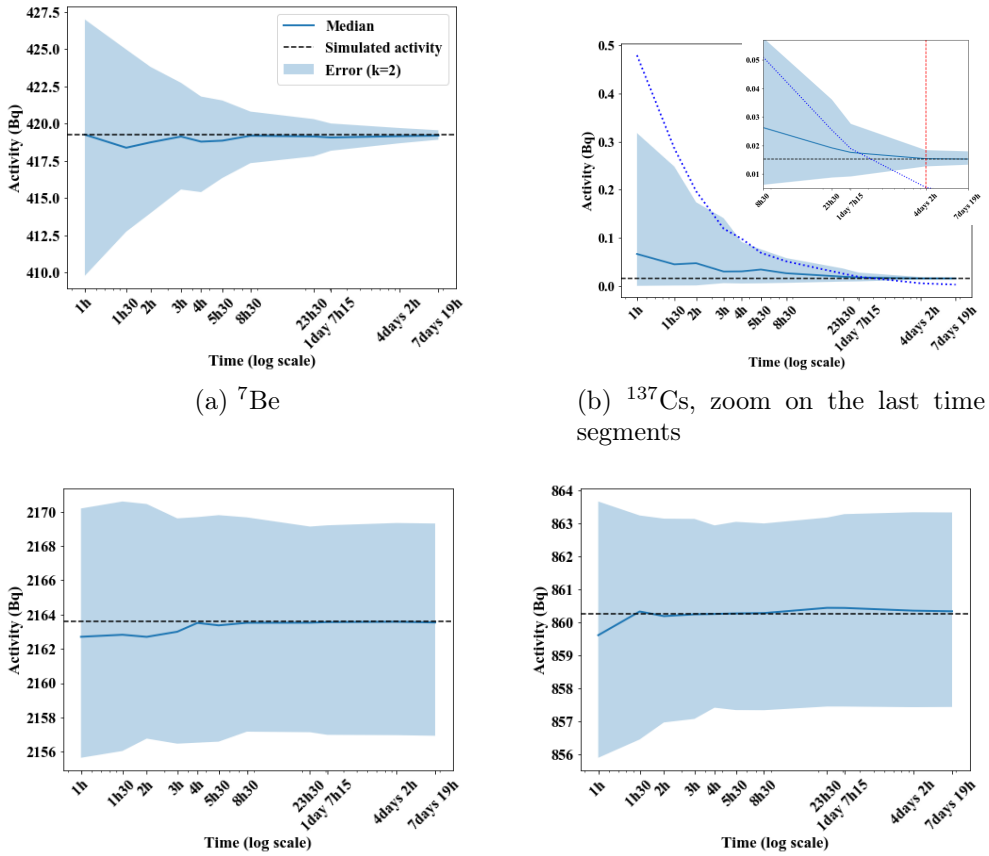
In this section, we focus on the role played by the temporal information and its impact on activity estimation. For that purpose, we specifically emphasize



(a) ^{7}Be (b) ^{137}Cs

(c) ^{214}Bi (d) ^{214}Pb

FIGURE 2.5 – *Estimated activities by both Model "Equilibrium hypothesis" and "Bateman" for different radionuclides. These graphs are obtained after* 1000 *Monte-Carlo simulation.*



(a) ^{7}Be

(b) ^{137}Cs, zoom on the last time segments

(c) ^{214}Bi

(d) ^{214}Pb

FIGURE 2.6 – *Results of the "Bateman" model on 1000 Monte-Carlo simulations. For ^{137}Cs the decision threshold is presented in dotted blue line.*

on the "Bateman" model, which provides a precise description of the time dependencies of the radionuclides' activities. In the following experiment, we propose evaluating the precision of the estimated activities when an increasing number of time intervals are used in the unmixing process. To do so, we start by using only the first interval (i.e. which is similar to take the Model "without time" only for this single interval), then the first two segments, up to all the 11 available time segments. This procedure is also meaningful in the context of crisis, as it allows evaluating how well the radionuclides' activities can be estimated from few and early measurements.
The results are obtained from 1000 Monte-Carlo simulation.

Figure 2.6 shows the estimated activities of the four radionuclides ^{7}Be (a), ^{137}Cs (b), ^{214}Bi (c) and ^{214}Pb (d). It is interesting to notice that the evolution of the estimated activity of the radionuclides is quite different depending on whether it is a short or long-lived radionuclide.

In case of short-lived radionuclides, such as ^{214}Bi and ^{214}Pb, the relevant information for spectral unmixing is mainly contained in the very first intervals. It entails that the estimated activities of ^{214}Bi and the ^{214}Pb quickly reach a final state. This is also the case for ^{208}Tl, ^{212}Bi and ^{212}Pb, the half lives of which do not exceed 10 hours.
For long-lived radionuclides, the results are slightly different and depends on the activity level. It also depends on how much the presence of short-lived radionuclides impacts the precision of their activity estimation. For instance, ^{7}Be has a rather high activity and seems less impacted by the presence of short-lived radionuclides. Consequently, its activity is very well estimated already from interval 1 only. In this case, the joint analysis of consecutive time intervals mainly helps improving the uncertainty of the estimation : it naturally decreases as the number of time segments used for the estimation increases.
The case of ^{137}Cs is different since its level is lower (*i.e.* 3 orders of magnitude below the level of ^{7}Be), which makes its activity estimation more sensitive to the presence of short-lived radionuclides. As a consequence, its activity estimation is significantly biased when few time intervals are used. The use of an increasing number of time intervals helps enhancing both the precision of the estimation and its uncertainty. The estimation of ^{137}Cs is below the decision threshold while the activity of the short-lived radionuclides are predominant but becomes significant at $t = 1$ day 7h15. It is a significant improvement with respect to previous work ([14]), where detection can be claimed at $t = 4$ days 2h as depicted in red on the zoomed plot of Figure 2.6.

In the cases of ^{7}Be and ^{137}Cs, it is important to note that the uncertainties decrease with the number of intervals used. Indeed, for these long-lived radionuclides the longer the time segment is and the later it begins, the easier it is to determine their activities. As a result, the statistical uncertainty shown in Figure 2.6 decreases for ^{7}Be and ^{137}Cs. On the other hand, the uncertainty for short-lived radionuclide like ^{214}Pb and ^{214}Bi is stable from the first segments to the last ones, as their contribution disappears from the spectra after 8 hours of measurement.

2.5 . Application to experimental aerosol measurements

In this section, we analyse a real aerosol sample measurement. This sample is composed of the exact same 11 time intervals we described in Table 2.2 and illustrated in Figure 2.1. To that purpose, we apply the proposed unmixing algorithm with the "Bateman" model, being the most realistic when out-of-equilibrium decay series are likely to be present. The experimental protocol is exactly the same as the one used in the simulations. The instrumentation and the radionuclides' signatures are also the same as in the simulations, as

described in 2.5.1. The background used in the analysis has been measured with the same detector a few weeks after the measurement of the sample, which is the standard laboratory routine.

2.5.1 . Description of the experimental setup

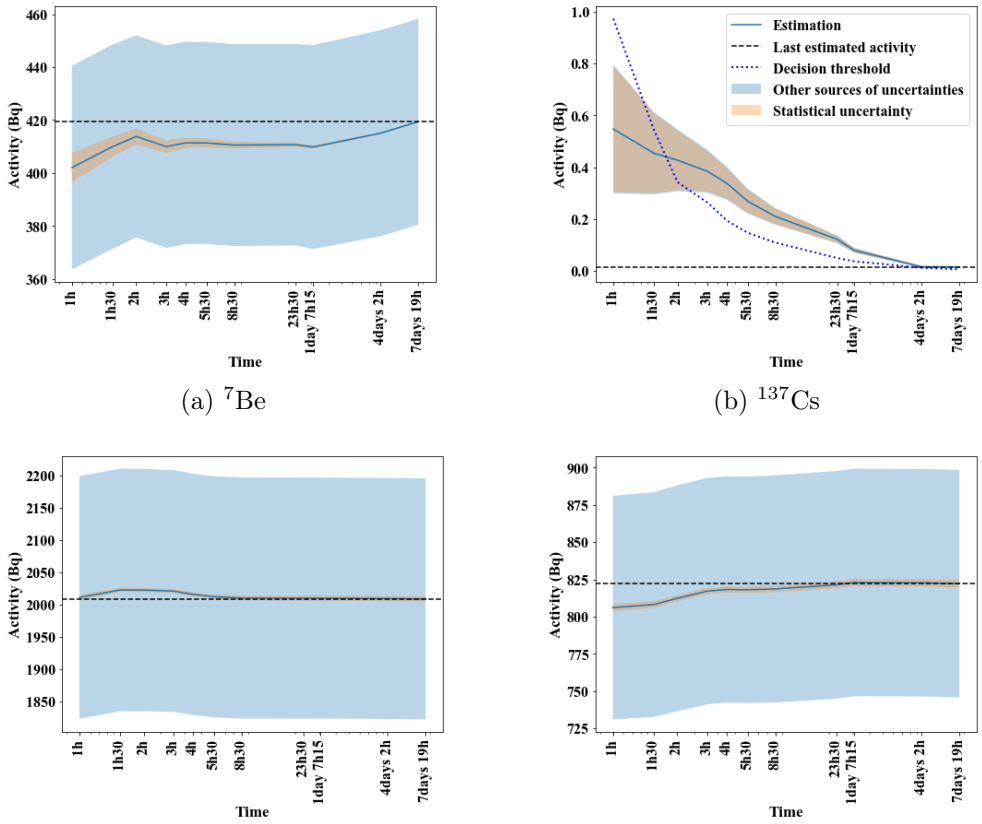
The experimental set-up consists of a gamma-ray spectrometer with a BEGe 5030 (Broad Energy Germanium, MIRION-Canberra) detector (crystal dimensions : $\emptyset$ = 80mm, h = 30mm) and a DSA-1000 (MIRION-Canberra) multi-channel analyzer based upon digital signal processing using 16384 channels for energies ranging from 20keV to 1700keV. The system has a relative efficiency of 61% and a resolution of 0.54 keV, 1.2keV and 1.7keV at 46keV, 622keV and 1460keV, respectively. The detector is surrounded by a 5cm thick lead shield, and equipped by an anti-cosmic set-up consisting in 5 scintillating plastic plates (5 cm thick). Furthermore, the whole system is installed in a shallow shielded room made of 10 cm lead bricks and internally lined by 5 mm oxygen-free copper, in the second basement of the laboratory, under 3 m borated concrete slab. Finally, the inner measurement chamber is flushed by the gaseous nitrogen escaping from the liquid nitrogen dewar to reduce and stabilize the radon induced background. The sampled aerosol filters are pressed into pellets packaged in 10mL cylindrical counting geometries (dimensions : $\emptyset = 52$mm, $h = 4.7$mm) and measured directly on the detector endcap.

The modeling process of the measurement configuration (detector/counting geometry) as well as the simulations performed to obtain the spectral signatures used by the spectra unmixing algorithms have been thoroughly described in [20]. We have also used the new calibration tools presented in chapter 4 to fit to the observed spectrum better.

The background spectrum used in the unmixing is considered as constant and perfectly known, and has been experimentally obtained by a background measurement for 560 000s using the same detector as the aerosol measurement.

2.5.2 . Multi-temporal spectral unmixing of an aerosol measurement

In this section, the analysis we performed is exactly the same as in the simulations. Consequently, Figure 2.7 depicts the evolution of the estimated activities across the different time intervals for ^{7}Be, ^{137}Cs, ^{214}Bi and ^{214}Pb. In this Figure, the orange shaded area stands for the statistical uncertainty, which originates from the Poisson statistics of the measurements (every uncertainties are taken at 2σ). It is similar to the uncertainty featured in Figure 2.6. In the previous section (2.4) this uncertainty was computed thanks to



(a) ^{7}Be (b) ^{137}Cs

(c) ^{214}Bi (d) ^{214}Pb

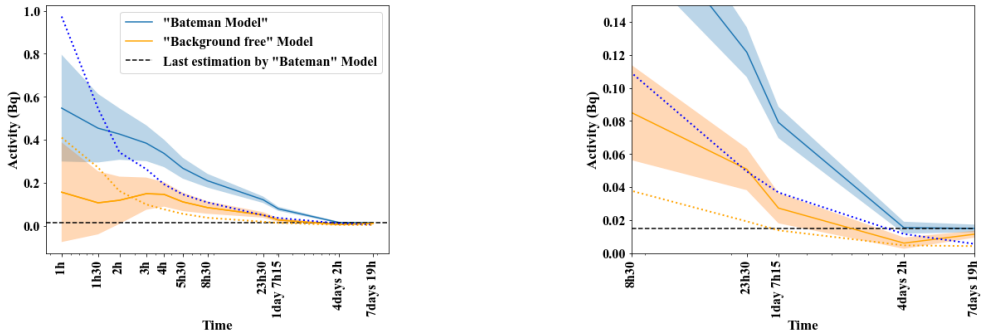
FIGURE 2.7 – *Evolution of the estimation of 4 radionuclides' activity. For ^{137}Cs, the decision threshold is indicated in dotted blue line.*

Monte-Carlo simulations, this cannot be done on real measurements as the true value of the activities are unknown. The computation of the statistical variability of the estimator is done following the work of [15], which resorts to the inverse of the Fisher's information matrix of the estimator. If we focus on the estimation of ^{7}Be and ^{137}Cs, the statistical uncertainties for simulations and the real data are similar ($\pm$10Bq for ^{7}Be and $\pm$0.1Bq for ^{137}Cs in the first segment and decreasing as more times segments are added to the estimation) in both figures. For ^{214}Pb and ^{214}Bi the statistical uncertainty is small and stable from the first time segment to the last as the activity of these short-lived radionuclides is estimated in the first segments and not affected by the last (the activity of ^{214}Bi, for example, is divided by 512 after only 3 hours of measurement as its half life is 20 minutes). In conclusion we can state that the statistical uncertainties behave similar to the ones measured in the simulations.

Following [20], the other sources of uncertainty come from the slight variation of the geometry of the sample, its placement on the endcap of the detector, the calibration of the detector. It is represented by the blue shaded area.
A first observation is that the estimated activities, in all time intervals, are compatible with the last estimate within the total uncertainty budget. However, the case of ^{137}Cs is more peculiar as the estimated activity is about 0.6mBq in the first time intervals, which is about 6 times higher than in the simulations. Furthermore, in these time intervals the estimated activities are not compatible with the final estimate, and also way beyond the detection threshold. This suggests that the estimated activity for ^{137}Cs is significantly biased, at least for the first day of measurements.
As this phenomenon is not observed in the simulations, it is very likely that this originates from a deviation between the observed data and the model that describes these data. This points to the accuracy of the simulated spectral signatures $\mathbf{\Phi}$ and the background; both may deviate from the actual spectral signatures and background. Indeed, in contrast to the other radionuclides displayed in Figure 2.7, ^{137}Cs has a low level, and is likely to be more impacted by variabilities of the spectral signatures and the background signature.

2.5.3 . Investigating the impact of the background level

In this section, we propose a simple experiment to highlight the impact of the background level on the estimated activities. For that purpose, rather than fixing the level of the background, it is estimated during the unmixing process. The background is then described, as any radionuclide, with a spectral signature that is added to the library of spectral signatures. A temporal signature is considered, based on the durations of the time intervals. The "Bateman" model now reads as :



(a) ^{137}Cs (b) Zoom on the last segments

FIGURE 2.8 – *Results for ^{137}Cs with the background treated as a signature (and zoom on the last 4 time segments). The total uncertainty and decision threshold are depicted in light colour and dotted line, respectively. The black dashed line is the last estimation by the "Bateman" model.*

$$\begin{aligned} &\forall s = 1, \cdots, S; \forall c = 1, \cdots, c; \\ &x_{sc} = \sum_{n+1} a_n \psi_{sn} \phi_{nc} + a_{^{214}Pb} \psi_s^{(B)} \phi_{^{214}Bi,c} \end{aligned} \tag{2.24}$$

Where ϕ_{n+1} is the normalized background spectrum (*i.e.* corresponding to a 1 second measurement), $\psi_{s,n+1}$ is equal to the duration of the s-th time interval and a_{n+1} is the level of the background. This model is called "Free-background". In the unmixing process, the update rules are similar to the ones detailed in Section 2.3.

Figure 2.8 shows the evolution in time of the estimated ^{137}Cs activity for both the standard "Bateman" model and the "Free-background", where the background is estimated. This experiment shows that the bias observed in the first time intervals is significantly reduced. The estimated activities during the three first hours of measurements are now consistent (with respect to the uncertainties) with the final estimated activity. While being reduced, the discrepancy between the final estimate and the ones between 3 hours and 1 day becomes prominent as the statistical uncertainty decreases with time.
The estimated activity of the background is largely superior to 1 on the first 4 segments (estimated to 30 on the 1st-segment-only estimation and decreasing towards 1 between the 1st and the 8th segment).
This first highlights that letting free the background allows capturing some deviation between the data and the original, that could either come from some mis-modelling of the background or the radionuclides' spectral signatures. A simple correction based on the background level is not enough, which sug-

gests that the shape of the spectral signatures should be estimated during the unmixing process.

2.6 . Conclusion

In this chapter, we introduced the full spectrum analysis on multiple spectra which allows accounting for the temporal information of the activity decay in time within the unmixing procedure. For that purpose, we introduce models which are composed of a spectral dictionary and a temporal dictionary. The former contains information about the detectors' response in energy with respect to each radionuclide. The latter is composed of the time decay of each of the radionuclide in the time intervals of the measurements. Two distinct models have been considered : the first one assumes that all the radionuclides are in equilibrium, while the second one allows taking care of parent/daughter dependencies as described by Bateman's equation. We further extended the multiplicative update algorithm to minimise the likelihood of the activities under non-negativity constraints. The proposed algorithms and models have been evaluated and validated on simulations of aerosol samples. These results show that such a method should theoretically yield earlier results for the detection of ^{137}Cs, about one day and a half after sampling, while previous methods required about four days ([14]). The proposed methods have been applied to real measurements of multi-temporal aerosol samples. Though these experiments show the applicability of multi-temporal unmixing for fast activity estimation, the detection and quantification of low-level radionuclides requires understanding the role played by the uncertainty with respect to the background and the radionuclides' spectral signatures.

In the next chapter we will see how to adapt this new measurement scheme and the proposed unmixing algorithm to an online estimation of the activities. In fact, thanks to another acquisition electronic it is possible to get the spectrum as it is measured in an online fashion. The refinement of the activity estimation to be able to analyse these spectra as fast as they are measured is a challenge but will provide even faster results than the one presented in this chapter.

3 - Online spectral unmixing of gamma-ray spectra

Sommaire

3.1 . Introduction

In this chapter I present the generalisation of the joint analysis on consecutive measurements of an aerosol filter sample so that an online activity estimation can be performed. The model is mostly the same as the one presented in the previous chapter but multiple problems occur when the number of time segments grows. This chapter is, in essence, what I presented at the ICRM (Internation Conference on Radionuclide Metrology and its applications) conference in Bucharest in 2023. The algorithm we used in the previous chapter is updated so that the computaion time is greatly reduced and we can perform an estimation as the measurement is performed.

In this chapter we propose a new algorithm that allows online estimation of the activities as the measurement of aerosol filter samples is done. Thus drastically reducing the time to correctly estimate the activity of the radionuclides composing the spectrum (1 min to detect a few Bq of activity). This new analysis is based on the full spectrum unmixing and the multiplicative update algorithm showed in [16] with a few tweaks allowing to deal with large matrices of data.

Based on the previous work the aim is to generalize the algorithm to any number of time segments and to get an estimation of the activities in the sample as soon as possible during the measurement process. To achieve this, we seek to reduce the size of the data matrices involved in the unmixing process and to reduce the number of iterations needed to converge.

The improvement to the algorithm will be presented on simulations and on a real aerosol filter sample measured in our laboratory. The detection of a few Bq of ^{123}I in a filter sampled for 1 week and an air volume of 90 000 m^3 after only a minute of measurement is achieved thus proving the performances of the unmixing in the case of fast detection. For ^{137}Cs, which is hard to detect as its activity is at the level of the decision threshold in the French environment, the detection is achieved after one day and 12 hours, which is almost four times faster than previous results based on full spectrum analysis [14].

3.1.1 . Challenge of the online analysis of gamma-ray spectra

In the context of radiological surveillance and radioecology studies, the analysis by gamma-ray spectrometry of aerosol filter samples allows to detect radionuclides at very low level (lower than 1μBq/m^3 in the air *ie* less than 1 Bq per sample) that are present in the environment among large concentrations of natural radionuclides. The rapid analysis of a gamma-ray spectrum as it is produced is a challenge in gamma-ray spectrometry as the measurement process can be quite long (varying from 1 to 4 days). In fact, if one wants to analyse

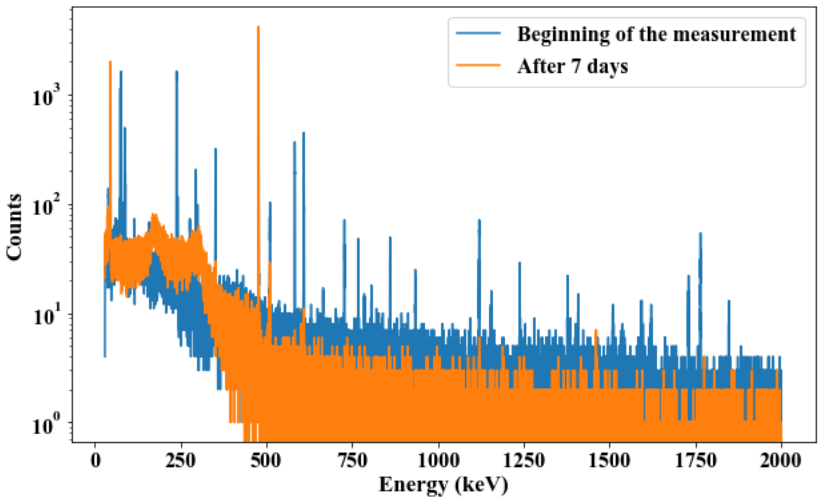


FIGURE 3.1 – *The spectrum obtained at the beginning of the measurement (first 20 minutes after the aerosol filter has been sampled and put into standard geometry) and the end of the week of measurement (last 4 hours).*

the spectrum in the minutes after the aerosol filter is sampled the gamma ray spectrum is dominated by the radon progenies that are short-lived radionuclides. The routine method to analyse gamma ray spectrum, which generally relies on peak-based analysis, struggles to cope with the complexity of such spectra. For a precise analysis of low-level radionuclides, activity estimation is generally performed at least one day after the filter has been sampled. This is illustrated in figure 3.1, the blue signal is the spectrum of an aerosol filter sample time-integrated over the first 20 minutes after it has been sampled, it presents a lot of peaks, mainly due to radon progeny. The orange spectrum is the one of the same sample for a 4 hours long measurement after a week of decay.

This highlights that postponing the measurement long after the filter is sampled allows for a more accurate analysis of low-level long-lived radionuclides. However, estimating the activity of the radionuclides composing the spectrum as soon as possible is key in the context of radiation protection and emergency preparedness, and we will illustrate later in this article that the observed delay before measuring the samples and getting the first activity estimations can be problematic in case of air contamination.

Getting an activity estimation of the radionuclides composing the spectrum as soon as possible is key in the context of radioprotection and emergency preparedness and we will show that the observed delay before measuring the samples and getting the first activity estimations can be problematic in case

of contamination.

As for the precedent chapter, we will focus on aerosol filter measurement as it is done in our laboratory and produces gamma ray spectra that contains a high complexity due to the radon progeny. The surveillance of the radionuclides levels in the environment is carried out by the IRSN (French Institute for Radiation Protection) and the analysis of the aerosol filter sample is part of the emergency preparedness and routine survey of the European nuclear environment.

3.1.2 . Recall of the model for temporal analysis of gamma-ray spectra

As stated in the introduction, a gamma ray spectrum is an histogram of the deposited photons in C channels representing the different energies. Following [13], it can be modelled as the linear combination of radionuclides spectral signatures. The content of each of the C channels of the spectrum can be written :

$$\forall c = 1, \cdots, C; \quad x_c = \sum_{n=1}^{N} \phi_{nc} w_n + \delta b_c, \tag{3.1}$$

$$y_c \sim Poisson(x_c) \tag{3.2}$$

The signal x_c is the model spectrum without the Poisson noise coming from the counting process of the detection. y is the measurement, ϕ is the matrix that contains the N signatures of each radionuclides in column. The vector w_n is composed of the weights of each of the radionuclides. The scalar δ is the duration of the measurement in seconds and b is the background spectrum per second , it is the spectrum obtained for a 1-second measurement on an empty detector, obtained by normalizing a weak long measurement. As we focus on laboratory measurements the background is considered stable through the entire measurement (see 1.2.3 for details on the experimental setup that provides the background stability). The mixing weights, w_n, are proportional to the activity of the radionuclides in the sample. In fact, the mixing weights can be split in two factors following the radioactive decay formula :

$$\forall n = 1, \cdot, N; w_n = a_n(0)\psi_n \tag{3.3}$$

$$\psi_n = \int_{t_0}^{t_1} e^{-\lambda_n t} dt \tag{3.4}$$

$$\tag{3.5}$$

w_n is the accumulated number of disintegrations obtained for a measurement from t_0 to t_1. $a_n(0)$ is the activity of the n-th radionuclide at $t = 0$, in the following, it will be noted a_n, it is the quantity of interest we want to estimate. λ_n is the decay constant of the n-th radionuclide.

Following [16] the temporal model to explain the gamma-ray spectra obtained on consecutive time segments is the following :

$$\begin{aligned} X &= \Phi(\Psi \text{diag}(a)) + B \\ Y &\sim \text{Poisson}(X) \end{aligned} \tag{3.6}$$

With :

— $diag(a)$ being the matrix containing all the activities $(a_1, ..., a_N)$ of each radionuclides at time $t = 0$ on its diagonal and 0 outside of its diagonal.

— X a $C \times S$ matrix, it is the base gamma ray spectrum without the Poisson noise, where S is the number of time segments considered.

— Φ a $C \times N$ matrix composed of the spectral signatures of the radionuclides composing the spectrum. A spectral signature is the gamma ray spectrum obtained with the measurement of an activity of 1 Bq for 1 s. It is the response of the detector with respect to the emission of a given radionuclide. These signatures can be obtained from the measurement of single radionuclide sources. However, these sources being hardly available for some radionuclides, they are generally derived from simulations. In this article, they have been simulated with MCNP-CP (see [24]), the details of the detector and the simulations can be found in 3.3.1.

— Ψ a $N \times S$ matrix containing the temporal information. It is the integral of the radioactive decay model of the n-th radionuclide over each time segment :

$$\psi_{n,s} = \int_{t_s - 1}^{t_s} e^{-\lambda_n t} dt$$

— B is a $C \times S$ matrix representing the background spectrum of the detector. This matrix can be decomposed into : $B = \delta.b$ where δ is the vector of the segment duration in second and b is the normalized back-

ground spectrum of the detector we use.

— Y is the Poisson-noised version of X, the Poisson noise comes from the counting process of the gamma-ray measurement.

In the last chapter 2, we introduced a time-dependent model for the full spectrum analysis of gamma ray spectra with a corresponding algorithmic scheme. We showed that accounting for the time dependency of consecutive measurement of the same aerosol filter sample fasten the estimation time, with reducing uncertainties as the data are collected in time. This was achieved by including the radioactive decay model in the spectral unmixing procedure. This way, activity estimation can be performed from multiple time measurements at once. Moreover, we showed that finely describing the decay chains could help in the estimation of low level radionuclides that would otherwise be slightly biased. In fact as we are dealing with the radon progeny before its decay we have to take into account multiple decay chains namely ^{214}Pb/^{214}Bi and ^{212}Pb/^{212}Bi/^{208}Tl. In this article both chains are considered at equilibrium and each daughter radionuclides adopts the period of its parent.

The mathematical model presented in equation 3.6 allows to tackle the activity estimation of the radionuclides as an inverse problem that can be solved using the multiplicative update algorithm (see [23]).

While satisfactory on a small number of time segments (*e.g.* 11 consecutive measurements used in [16]), applying this algorithm to a very large number of segments in an online estimation scheme is problematic as the data to be processed increase with time, and spectral unmixing has to be performed each time a new measurement is collected. A dedicated algorithm is needed to perform spectral unmixing using temporal correlations on a fine time sampling of the measurement.

3.1.3 . Full spectrum analysis with time dependence

To account for the time decay of the radionuclides to perform time-dependent processing, a new algorithm has been introduced in [16], it is an extension of the full-spectrum analysis method proposed in [13]. This algorithm is based on a multiplicative update scheme so that, at each iteration k to $k+1$, the activities of the radionuclides are updated simultaneously as follows :

$$a_{sn}^{(k+1)} = a_{sn}^{(k)} \frac{\sum_{c=1}^{C} \phi_{nc} y_{sc} / x_{sc}}{\sum_{c=1}^{M} \phi_{nc}}, \tag{3.7}$$

Where $Y_s = (y_{s1}, ..., y_{sC})$ is the observed spectrum during the s-th segment and $X_s = (x_{s1}, ..., x_{sC})$ is the theoretical model as viewed in equation 3.6. The multiplicative update is performed until convergence, which we define as :

$$\frac{\sum_n (a_n^{(k+1)} - a_n^{(k)})^2}{\sum_n {a_n^{(k)}}^2} \leq \epsilon_{diff} \tag{3.8}$$

ϵ_{diff} is usually taken small (10^{-6} in this paper).

This update allows us to minimise the following loss function in term of the activities :

$$L(\{y_{sc}\}_{s,c}|\{a_n\}_n) = \sum_{s,c} x_{sc} - y_{sc}\log(x_{sc}) + \log(y_{sc}!) \tag{3.9}$$

$$x_{s,c} = \sum_{n=1}^{N} \phi_{c,n}\psi_{n,s}a_n + \delta_s b \tag{3.10}$$

The algorithm thus aims at reducing the difference between the theoretical spectrum (X_s) and the observed one (Y_s) in terms of the Kullback-Leibler divergence.

The main aspects that we will focus on to reduce the computation time are : (i) reduce the number of iterations while keeping a correct estimation of the activities and (ii) reduce the size of the matrix involved in the update so that the computation can be done even with a really fine time segmentation and a huge data matrix.

3.1.4 . Reducing the size of the data matrix involved at each iteration

The main limitation of the time-dependent algorithm introduced in [16] is that, for s time segments, the size of the data matrix to be analysed is equal to $s \times C$. For a typical aerosol sample with about 16000 energy channels, jointly analysing a 100 time segments requires performing spectral unmixing on matrix with more than a million entries. This first requires a significantly important computational cost, and for a large number of time segments, an increasingly large matrix to be stored. In the context, both the data matrix X and temporal model matrix Ψ grow in size.

In order to make online time-dependent spectral unmixing tractable, the

proposed approach consists in lowering the size to the matrix to be handled each time a new time segment needs to be processed. For that purpose, we chose to set a maximum size to the matrix and to process the data as a buffer of fixed size r. More precisely, the data matrix describing the observed spectra is :

$$\begin{pmatrix} | & | & ... & | \\ x_1 & x_2 & ... & x_s \\ | & | & ... & | \end{pmatrix}$$

The buffer we propose to use instead is :

$$\begin{pmatrix} | & | & ... & | \\ \sum_{k=1}^{s-r+1} x_k & x_{s-r+2} & ... & x_s \\ | & | & ... & | \end{pmatrix}$$

If we were to build a small size buffer from few past time segments, the resulting procedure would suffer from two main drawbacks :

— It would only account for the time decay of the radionuclides on a short duration, which would limit the efficiency of the unmixing process.

— For short-lived, the activity estimation procedure would become unstable as their activity would quickly vanish.

To mitigate these two pitfalls, we add to the buffer the sum of the past measurements from time segment 1 to $s - r + 1$. Formally, it will be stored in the first column of the buffer. This procedure allows limiting the instability of the activity estimation of the short-lived radionuclides. This is illustrated in figure 3.2 : if only the segments x_{s-r+1} to x_s are used the activity estimation explodes due to the vanishing activity of the short-lived radionuclides.

3.1.5 . Regularising the activity estimation in time

As we pointed out in the previous paragraph, performing spectral unmixing on a small size buffer allows avoiding manipulating large data matrices but at the cost of reducing the time interval on which time decay is exploited. From a statistical viewpoint, processing numerous short-duration time segment also entails that spectral unmixing will have to capture time-dependency from few time segments with potentially a low number of counts. This could also result in an unstable estimation of the activity in time.

To limit this source of instability, we propose to regularise the estimation of the radionuclides in time by imposing a time-dependent regularisation : loss function is altered so that the estimation of the activity for time segment s does not deviate too much from the last estimation produced at time $s - 1$. In

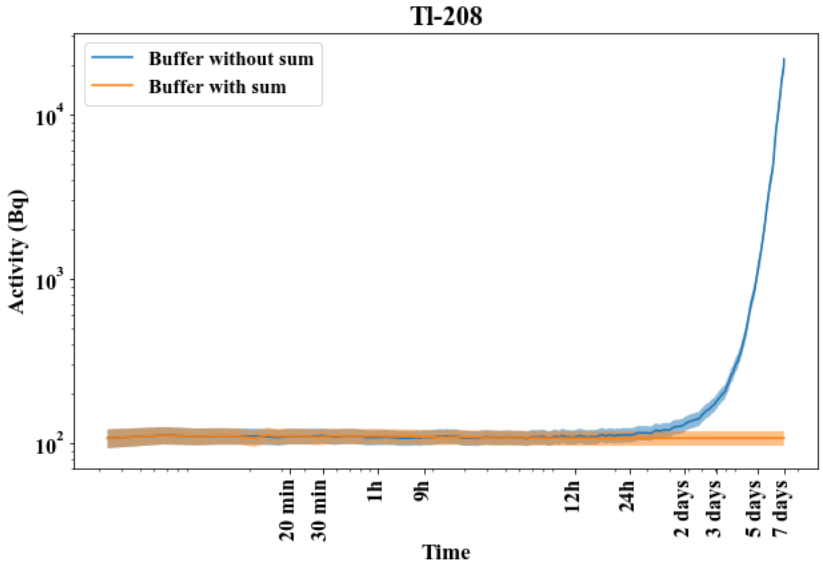


FIGURE 3.2 – *The activity estimation of* ^{208}Tl *with respect to the measurement time, which half life is 10.6h (because it is considered at equilibrium with* ^{212}Pb*)*

details, we add an extra term that limits some distance between the sought-after activity at time s and its previous value at time $s-1$. This results in the following loss function :

$$L(\{y_{sc}\}_{s,c}|\{a_n\}_n) = \sum_{s,c} x_{sc} - y_{sc}\log(x_{sc}) + \log(y_{sc}!) + \beta\ \mathrm{dist}(a_{s-1}, a) \quad (3.11)$$

The distance *dist* is chosen as the Kullback-Leibler divergence between the estimated spectrum at time $s-1$ and the estimated spectrum at time s this leads to :

$$\mathrm{dist}(a_{s-1}, a) = \sum_{s,c} x_{sc} - x_{s-1,c}\log(x_{sc}) + \log(x_{s-1c}!)$$

where $x_{s-1,c} = \sum_{n=1}^{N} \phi_n \psi_{n,s} a_{s-1,n} + \delta_s b$ is the estimated spectrum using the $(s-1)$-th estimation of the activities.

As a consequence, this choice allows for a simple multiplicative update step that is similar to the one we implemented in the original time-dependent spectral unmixing algorithm.

The factor β is a regularisation parameter that controls the trade-off bet-

ween the new information carried by the measurement at time s and the past value at time $s-1$. In fact, if we set $\beta = 0$, the weight given to the past estimation is zero and the update is the same as in [16]. This would lead to a less stable estimation of the radionuclides activity. If, on the other hand, β is too large, the estimation will remain almost stationary from one time step to another. This will affect the quality of the activity estimation since the estimation will poorly benefit from the long measurement process and will focus only on the first time segment. This would result in a very slow and biased estimation procedure.

The value of β needs to be optimised to get a good compromise between the speed of the algorithm to the correct activity and the stability of the estimation scheme in time. This will be detailed in section 3.2.
In order to make explicit the resulting multiplicative update scheme with the new regularised loss, we need to compute its derivative with respect to each a_n is :

$$\frac{\partial L(\{y_{sc}\}_{s,c}|\{a_n\}_n)}{\partial a_n} = \sum_{s,c}\left(\psi_{sn}\phi_{nc} - \frac{\psi_{sn}\phi_{nc}y_{sc}}{x_{sc}}\right) + \tag{3.12}$$

$$\beta\sum_{s,c}\left(\psi_{sn}\phi_{nc} - \frac{\psi_{sn}\phi_{nc}x_{(s-1)c}}{x_{sc}}\right) \tag{3.13}$$

where $x_{(s-1)c} = \psi_{sn}\phi_{nc}a_{(s-1)n} + b_{sc}$. To obtain the update rule we have to zero the derivative of this function with respect to each parameter as in [23]. As the chosen quasi-distance is the Kullbach-Leibler divergence we can see that the structure of both terms in equation 3.13 are similar, leading to a compact update step that reads as :

$$a_n^{(k+1)} = a_n^{(k)}\frac{\sum_{s,c}\psi_{ns}\phi_{cn}.(y_{cs} + \beta x_{cs}^{(s-1)})/x_{sc}}{(1+\beta)\sum_{s,c}\psi_{ns}\phi_{cn}} \tag{3.14}$$

This new update step allows keeping the last activity estimation as a good starting point for the computation of the new one and forces the algorithm to converge quicker to a new estimation.

3.1.6 . Updating only the radionuclides that are still active in the incoming data

A last source of instability could originate from the estimation of the short-lived radionuclides, when their activity becomes very low. In this case, keeping

updating their activity is very likely to yield noise overfitting. To mitigate this effect, we propose adopting an early stopping procedure for short half-lives.

In fact, on a week-long measurement with some radionuclides that have short half-lives (*eg.* ^{214}Pb with a half-life of 26.8 min), their contribution to the spectrum tends quickly to 0. Adding new time measurements does not bring more information to update their activity. Moreover, updating their activity will tend to fit noise rather than actual physical contributions. Hence, we proposed adding a stopping rule to estimation procedure of the radionuclides's activity for which we know that there are no counts in the incoming spectrum.

This stopping rule is defined as follows : if the expected contribution of a radionuclide is less than a fixed number of counts (ϵ_{counts}), then the estimation is not updated. The expected number of counts is obtained through the following formula for any radionuclide n and any time segment s :

$$count_{n,s} = ||\phi_n||_1.\psi_{n,s}.\max(a_{n,s-1}, 1) \tag{3.15}$$

This way of computing the expected number of counts ensures that the radionuclides that have a very small activity (*eg.* ^{137}Cs) are still being updated even if the number of counts that are due to them is low while leaving the short-lived radionuclides out of the new estimation. Indeed, the contribution of these radionuclides can be small but they are still valuable, and we want to estimate correctly their activities. Moreover, we will see in the next section (3.2) that the estimation of the radionuclides that have a low activity can be null if we analyse only the first time segments and increase as time goes, and more data are acquired.

This criterion allows us to focus only on the radionuclides that contribute to the new measured spectrum while keeping the estimation fixed for the radionuclides that do not contribute any more. This combined, with the stopping criterion of the algorithm (see 3.8), contributes to keep the number of iterations as low as possible as the number of coordinates of the activity estimation that are updated is low and thus the difference between the estimation at time $s-1$ and time s is null for some of them.

The different new tools that we developed are summarized in the following algorithm (1). If the buffer are used the matrices Ψ, Y and B are changed to the corresponding reduced matrices.

Algorithm 1 The proposed unmixing algorithm

Require: $Y = (y_1, ..., y_s)$, $\Phi = (\phi_1, ..., \phi_N)$, $\Psi = (\psi_1, ..., \psi_s)$, $a_{s-1}, \epsilon_{\text{count}}$, ϵ_{diff}, $B = (\delta_1 b, ..., \delta_s b)$, β, K

$E = \emptyset$

$x^{s-1} \leftarrow \Phi(\Psi \text{diag}(a_{s-1})) + B$

Diff $\leftarrow \epsilon_{tol} + 1$

$\hat{a} = a_{s-1}$

$a = a_{s-1}$

for $i = 1, ..., N$ **do**

  count $\leftarrow ||\phi_i||_1 \psi_{is} \max(a_{i,s-1}, 1)$

  if count $\geq \epsilon_{\text{count}}$ **then**

    $E \leftarrow E \cup \{i\}$

  end if

end for

$k = 0$

while Diff $\geq \epsilon_{\text{diff}}$ and $k \leq K$ **do**

  $k = k + 1$

  $x_s \leftarrow \Phi(\Psi \text{diag}(a)) + B$

  for i in E **do**

    $\hat{a}_i \leftarrow a_i \frac{\sum_{s,c} \psi_{ns} \phi_{cn} . (y_{cs} + \beta x_{cs}^{s-1}) / x_{sc}}{(1+\beta) \sum_{s,c} \psi_{ns} \phi_{cn}}$

  end for

  Diff $\leftarrow \frac{\sum_{n=1}^{N} (\hat{a}_n - a_n)^2}{\sum_{n=1}^{N} {a_n}^2}$

  for i in E **do**

    $a_i \leftarrow \hat{a}_i$

  end for

end while

return a

3.2 . Numerical evaluation of the online spectral unmixing algorithm

In this section, we focus on evaluating the performances of the proposed online spectral unmixing algorithm on simulated data. We first start with the optimisation of the value of the regularisation parameter β, which plays a key role to balance between estimation bias, speed and stability.

Indeed, as we can see in table 3.1, we simulate spectra with very close activities to the one we observe on real aerosol filter samples. This way we can assess the performances of the unmixing algorithm on realistic simulations. The activity of the different radionuclides varies from really low levels (few

mBq) to high activity (hundreds of Bq) as shown in table 3.1.

Radionuclide (subgroup)	Half-life	Simulated activity (Bq)	Estimated activity after 1 min (Bq)	Estimated activity after a week (Bq)
^{7}Be (i)	53.22 d	150	159 ± 29	157 ± 16
^{22}Na (ii)	2.60 y	0.02	0.49 ± 0.79	0.0284 ± 0.0033
^{40}K (ii)	1.265 10^{9} y	0.5	13 ± 10	0.511 ± 0.079
^{123}I (iii)	13.22 h	2	2.13 ± 0.71	2.48 ± 0.27
^{137}Cs (ii)	30.05 y	0.01	0.61 ± 0.66	0.0213 ± 0.0033
^{208}Tl (iii)	3.06 min	100	107 ± 14	107 ± 11
^{210}Pb (i)	22.3 y	30	30.0 ± 9.9	30.9 ± 3.2
^{212}Bi (iii)	60.54 min	300	339 ± 56	323 ± 33
^{212}Pb (iii)	10.64 h	200	237 ± 28	241 ± 24
^{214}Bi (iii)	19.9 min	300	328 ± 39	395 ± 42
^{214}Pb (iii)	26.8 min	100	114 ± 15	124 ± 14
^{228}Ac (ii)	14.02 10^{9} y	0.1	5.6 ± 2.6	0.132 ± 0.017

TABLE 3.1 – *The 12 simulated radionuclides, their subgroup, half-life and simulated activity. The estimated activities after 1 min and a week are presented for the real aerosol filter measurement with associated uncertainties (at k=2).*

Thanks to the gamma-ray spectrum mathematical model (equation 3.6) we can simulate any mixture of signatures and apply a random Poisson noise to create realistic simulations of gamma-ray spectrum on which we apply our algorithm. The size of the buffer is fixed to 5 so that the computations are quick to perform and we can simulate a great number of Monte-Carlo repetitions for each experiments. An energy and efficiency calibration step are done to better simulate the detectors response to each of the radionuclide of interest (details on the detector can be found in 3.3.1 or [25]).

3.2.1 . Optimisation of the regularisation parameter β

In section 2.2.1, we highlighted that β is important to stabilise the activity estimation procedure in time as it controls how much the estimated activity at time s can deviate from its previous estimate.

The impact of β is twofold : i) a large value will tend to slow down the estimation in time and bias the estimates by giving too much weight to past time segments, and ii) a small value will tend to favor newest measurements in the activity update step leading to potential instability.

The optimization of the value of β also allows us to assess the convergence and performances of the algorithm as a function of β. To choose an efficient trade-off for β, we performed simulations and found a good compromise between the correct estimation of the activities and a rapid convergence of the algorithm at each time step.

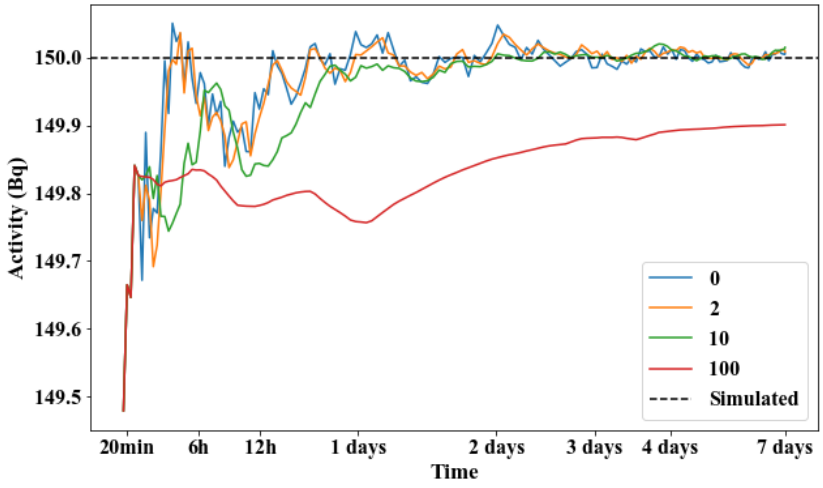


FIGURE 3.3 – *The median of the activity estimation of* ^{7}*Be with respect to the value of beta on 1000 Monte-Carlo simulations.*

First, we can see in figure 3.3 the effect of the value of β on the estimation. As the value of β grows, so does the bias between the simulated activity and the estimation provided by the algorithm. Indeed, a large value of β implies that the update step only focuses on reducing the distance between the past estimation and the new one, the new estimation will then be equal or near the previous one. This is bad as the new information carried by the spectrum entering the analysis are not taken into account in the new estimation of the activities. As it can be observed in figure 3.3 the larger β is the larger the bias is at the end of the week-long measurement. For small values of β the effect is only to smooth the consecutive estimations, indeed the memory kept in the estimation leads the estimation at time s to be close to the new one at time $s+1$ while still allowing to converge to the correct estimation.

After this first test we can focus on small values of β, the results of this second study are presented in figure 3.4. In this figure we can see that the bias of estimation does not vary a lot from one value of β to another as long as β is reasonably low ($ie : \beta \leq 2$). This allows to focus only on the computation time required to converge, indeed we want to find the smallest value of β that allows to reduce the computation time. The computation time will be seen as the number of iterations the algorithm needs at each time steps to converge (we recall that the stopping criteria is the distance between two consecutive iterations of the algorithm as viewed in 3.8).

Figure 3.5 displays the evolution of the computation time with respect to the value of β. One can observe that when the value of β increases, the computation time reduces. Indeed, the loss function requires the algorithm to

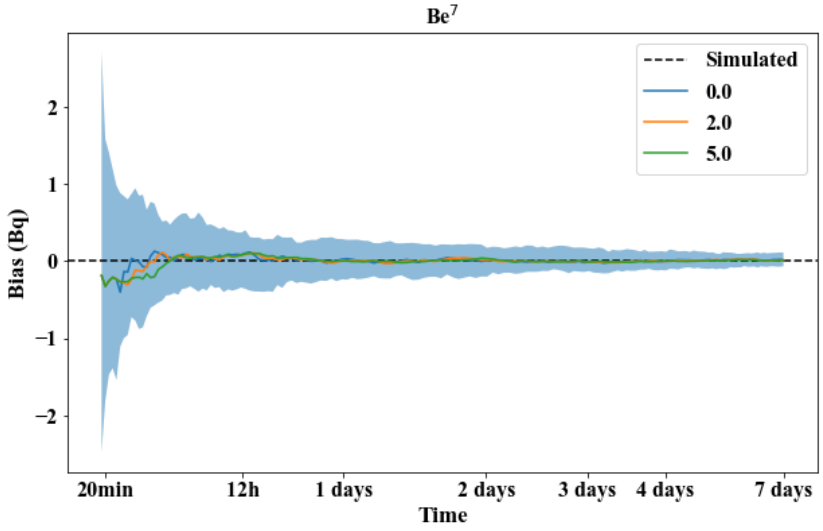


FIGURE 3.4 – *The bias on the activity estimation with respect to the value of β. We focus on small values of this parameter. The results are based on 1000 Monte-Carlo simulations.*

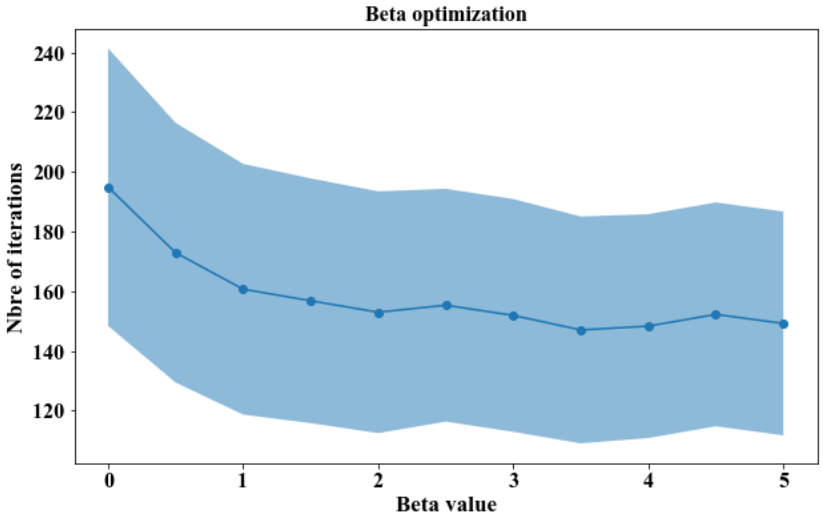


FIGURE 3.5 – *The mean number of iteration needed to converge (over the week of measurement) with respect to the value of β.*

remain near the estimation of the previous time segments and reduces the searching range to a vicinity of this past estimation. The higher the value of β the smaller the vicinity, thus reducing the number of iterations of each time step. The results of figure show that a good compromise between the convergence and the computation time is $\beta = 2$. In fact, increasing the value of β does not reduce the computation time as much after $\beta = 2$. On the other hand, we have seen that allowing β to be high can create a significant bias on the estimation.

3.2.2 . Performances of the unmixing algorithm

In this section, we further explore the simulations with $\beta = 2$ and the results of the estimation on high and low activities. The results of this section show the performances of the proposed online spectral unmixing algorithm to estimate the activity of 12 simulated radionuclides.

The radionuclides we simulate, and their activities are presented in table 3.1. We can split these radionuclides in three groups for analysis purposes : (i) the high activities with long half-life, (ii) low activity and long half-life, (iii) short half-life radionuclides. We focus on an example of each group (i) and (ii) and on two examples for group (iii) as it contains high and low activities. The results of our algorithm on simulations are shown in figure 3.6.

For the first group, an example of which is the ^{7}Be, the activity estimation is not difficult, the activity is high and the long period ensures a stable presence of the signature of these radionuclides in the spectrum. The correct estimation is achieved after only 2 minutes of measurement and can probably be even quicker (for computation time of the simulation we did not test the algorithm to a much finer segmentation than 2 minutes). The estimation of the activity of the second group, such as ^{137}Cs, is much harder, in fact we observe activities that are close to the detection limit, the contribution of the radionuclides of this group to the spectrum is thus close to zero. The estimation can be null in the first segments and be achieved only after a short measurement time is observed. This is the case for ^{137}Cs, for which the detection limit is achieved only after 37h of measurement due to its really low simulated activity (0.01 Bq). For the radionuclides such as ^{123}I or ^{214}Bi, belonging to the third group, we note that the activity estimation is stable through time and is correct after only a few minutes. This shows the performances of the temporal unmixing to perform even for low level (1Bq of activity) radionuclides with short half-life (*eg* 13.22h in the case of ^{123}I). The stable performances for these short half-lived radionuclides are allowed thanks to the type of buffer that we adopted, that keeps track of the past time segments as previously seen in section 3.1.4. The other radionuclides' estimation behave the same as these examples depending on their activities and half-life.

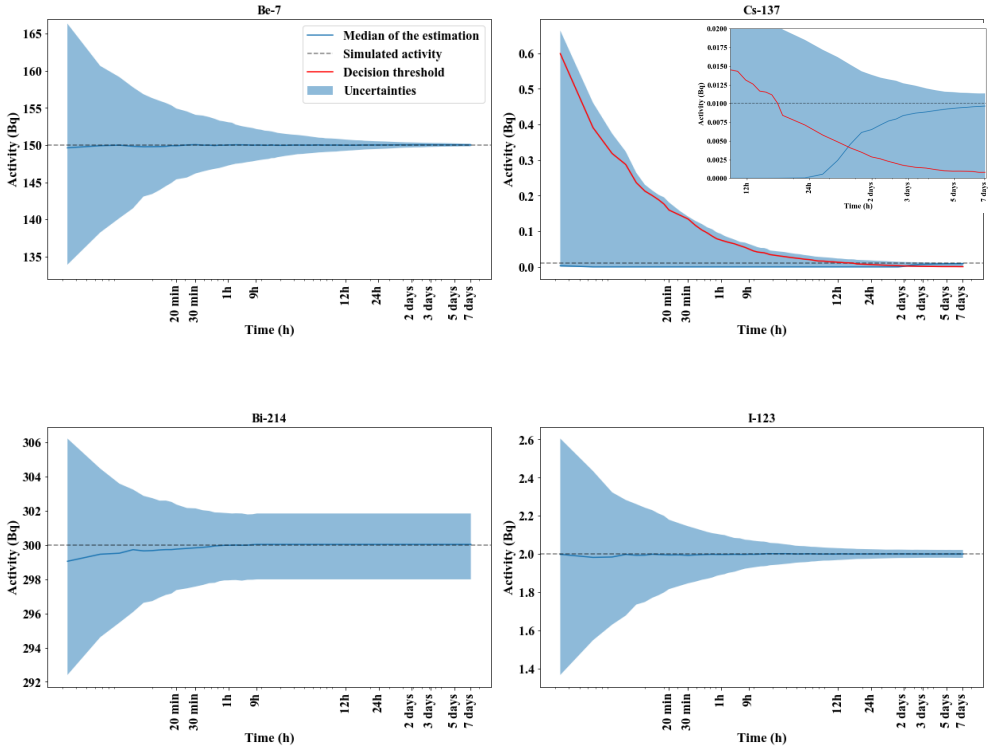


FIGURE 3.6 – *Results on simulations of the unmixing algorithm. The error bars are the 10% and 90% quantiles of the estimations based on 1000 Monte-Carlo simulations.*

Concerning the uncertainties, it is important to note that the main factor that impacts them is the half life of the radionuclides. In fact, if the period is short then the radionuclides contribution to the new time segments tends to 0, the uncertainty regarding its activity does not vary after this effect. On the other hands, for long-lived radionuclides the uncertainties decrease as long as the measurement is done and more and more statistics are taken into account. The uncertainties presented in this section only concern the statistical variability as the only random factor is the Poisson noise of the simulated spectra. The metrological uncertainties (around 10% of the estimated activity) must be added to this in order to get the total uncertainty we show on the laboratory results. This will be further explored in the next section where we will analyse the results of the algorithm on real aerosol filter samples and real detector from our laboratory.

3.3 . Results on real aerosol filter sample

In this section, we show the results of the unmixing on real aerosol filter sample. We show that it performs very well on this sample, allowing to detect a few Bq of ^{123}I with a correct uncertainty after only a few minutes of measurement. After the spectral signatures have been calibrated in energy and resolution to fit the observed spectrum we can estimate the activities of the 12 radionuclides of interest. The number of channels of the spectra is $C = 27\ 000$ going from 30 keV to 2 730 keV.

3.3.1 . Experimental setup

The HPGe detector used in this paper is a Broad Energy Germanium detector (BEGe5030, Mirion-Canberra). Its resolution is 0.5 keV at 59 keV, 1.2 keV at 662 keV and 1.8 keV at 1332 keV. Its relative efficiency is 51%. This setup is used to measure aerosol filter collected by a high volume air sampler (700-900 m^3/h). The 45 x 45 cm^2 filters are compressed into standard 10 mL cylindrical geometries. The detector is placed in the shallow shielded room in the second basement of the laboratory under 3 meters of borated concrete to reduce the cosmic ray induced background. This 20m^2 room is made of 10cm thick lead bricks internally covered by 5mm copper - so that the telluric radioactive background is reduced by 2 orders of magnitude. The detector is connected to a digital electronics (Pixie-4, XIA) allowing us to proceed the data in list mode. This acquisition mode provides a list of all the detected events with time stamps and channel allowing us to separate the whole acquisition into time segments of the duration we chose.

The background spectrum is considered to be constant throughout the measurement. It is measured with the empty detector during a week and standardized to a second by dividing the intensity of each channel by the duration

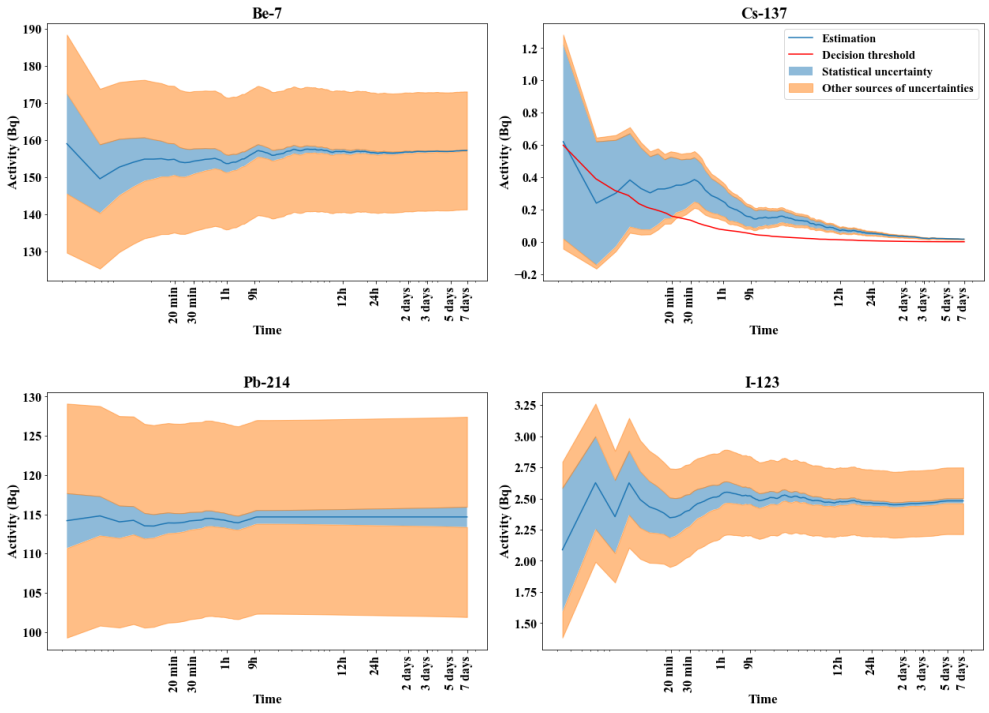


FIGURE 3.7 – *The results of the activity estimation on real aerosol filter samples for* ^{7}Be, ^{137}Cs, ^{214}Pb *and* ^{123}I. *The statistical uncertainties are computed following [20] and the metrological uncertainties represent 10% of the estimated activities.*

of the measurement. The background measurement has been measured a few weeks before the aerosol filter measurement so that it is very close to the real background during the sample measurement.

3.3.2 . Results

The results of the algorithm are presented in figure 3.7. This figure presents the activity estimation of the same radionuclides we showed in the simulations as they are representative of each group of radionuclides of interest. These estimations are obtained using every new elements we presented in the past section. We have adopted a buffer of size 5 thus reducing the size of the matrix. We have chosen $\beta = 2$ as seen in section 3.2 so that we reduce the number of iterations while keeping a correct activity estimation. Finally we chose to stop the update of the radionuclide if their expected number of counts was lower than 10 in the whole spectrum. The total number of segments is 355, we chose to have a duration of 2 minutes for the first segment and to keep the same number of counts for every time segment (each time segment thus having ~ 26000 counts).

As we can see in figure 3.7 the estimation of every radionuclide is quick.

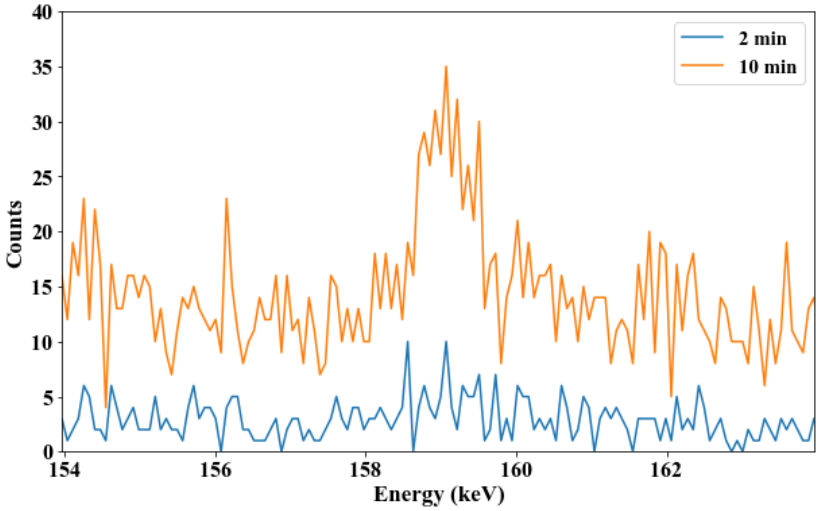


FIGURE 3.8 – *The cumulative spectrum of the peak region of the ^{123}I, we can see that the peak growing as the measurement proceeds. We can note that the first estimations are based on a really low number of counts showing the performances of the unmixing.*

Indeed except for ^{137}Cs the estimation is stable after only 20-30 minutes of measurement. The new algorithm allows us to estimate the activities as soon as possible and to cope with a large number of spectra, which was impossible with the first version of the data processing. This allows to detect ^{123}I as soon as 2 minutes after the measurement started. This radionuclide is not usually observed in the environment and shows the interest of our method for the early detection of contamination. Indeed if the routine methods was used, observing a decay period before the measurement starts and only estimate the activities after the week-long measurement the contamination by ^{123}I would have been missed or given very late. These results show the performances of our new algorithm as it is able to detect contamination out of a very complex spectrum in a very short time and with an algorithm that is fast to compute and reliable. Figure 3.8 presents the peak of interest of ^{123}I at the beginning of the measurement, showing how minute it is in the first segments. The algorithm is still able to detect the radionuclide and to estimate its activity correctly. Figure 3.9 shows the number of counts observed in the peak of ^{123}I. This further shows that the identification of this radionuclide would have been impossible if we had observed the 4 days of decay before the measurement. Indeed the ^{123}I would have totally decayed before we could have observed it in the spectrum.

The estimation of ^{137}Cs is the hardest of the 12 radionuclides we are focusing on in this paper, as it is at very low level (only a few mBq) it is the only one where the estimated activity is of the same order of magnitude as the decision threshold. We saw in the simulations that the correct estimation can

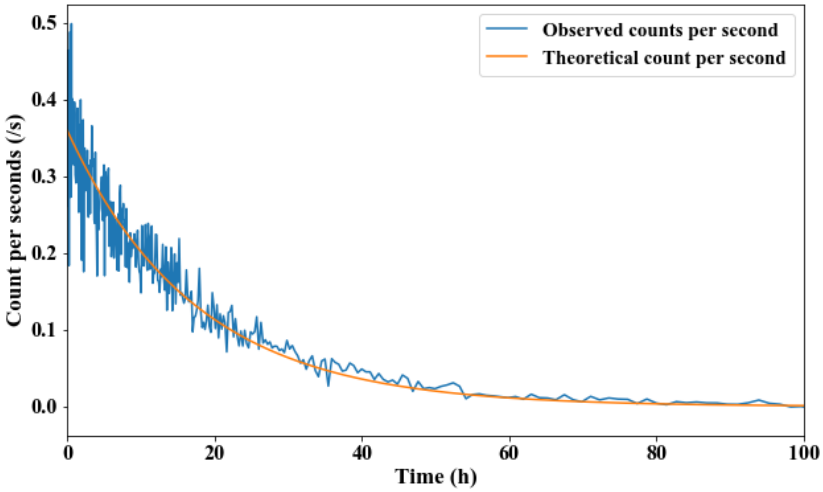


FIGURE 3.9 – *Net number of counts per second in the I^{123} region and the theoretical decreasing (with a period of 13.2h) assessing the presence of the radionuclide.*

be late if the level is too low (0.01 Bq in the simulation). In the aerosol filter sample we analysed it seems that the activity of ^{137}Cs is high enough so that the detection is achieved after 20 minutes. But we can see a bias between the estimation on the first few time segments (0.4-0.6 Bq) and the final estimation (0.02 Bq). The estimation of radionuclides that are at the level of the decision threshold is always problematic and the results must be given within the correct uncertainties.

The statistical uncertainties are computed via the Fisher information as advocated in [20]. The metrological uncertainties are composed of a few different factors summing to about 10% of the estimated activity as in chapter 2 the factors are the slight variation of the geometry and density of the sample, its placement on the endcap of the detector and the efficiency calibration of the detector. To obtain the decision threshold for a radionuclide we compute simulations with its simulated activity equal to 0 and take the quantile of the activity estimation corresponding to the α risk that we want (here $\alpha = 5\%$ is taken).

3.4 . Conclusion and perspectives

In this paper, we introduce an online spectral unmixing algorithm which allows achieving radionuclide activity estimation from temporal gamma spectroscopy measurements. Following [16]. accounting for time information in the unmixing procedure provides early estimation of the radionuclides composing the spectra earlier in the measurement process at the cost of large uncertainties

at the beginning of the measurement reducing as the measurement processes. Time segmentation allows taking into account the time correlation in the unmixing algorithm and thus reduce the uncertainties and allows to deal with short-lived radionuclides such as the radon progeny. However, processing measurements with increasingly shorter time segments entails cumbersome computation issues if the algorithm is not adapted to a large number of time segments. To alleviate this bottleneck, the solutions we propose allows to reduce the number of iteration of the algorithm we presented in [16] and allows adapting the loss function so that the activity estimation at time $s-1$ is used at time s as a guideline so that the new estimation is not too far from the precedent one. The results on real aerosol filter samples show that contaminations with a few Bq of ^{123}I can be correctly estimated after a few minutes of measurement. This could not have been achieved in the routine measurement scheme where a decay period of 2-4 days is observed so that the radon progeny decay and the spectrum simplifies.

4 - Calibration for gamma-ray spectra analysis

Sommaire

In this part of the thesis I present the calibration tools used for the analysis of measured gamma-ray spectra. In fact, as seen in the introduction chapter 1, in the routine gamma-ray spectra peak-based analysis, a local calibration is processed during the peak fitting of each individual detected peak. The only calibration needed in order to use the peak-based analysis is the energy calibration allowing to know the position of a peak in the spectrum. This is done once and for all with a pre-calibration using a multi-gamma source that allows to locate known peaks in the spectra giving an energy calibration that can be later used for other measurements. Since the peak-based analysis processes each peak present in the spectrum locally it does not require the same precision in the calibration as we do for full-spectrum analysis. For this kind of analysis, the energy calibration must be performed, not only in the region of interest around the peaks, but on the entire spectrum and Compton continuum so that the spectral signatures match the entire spectrum. Moreover, the spectral signatures simulation has to be precisely tuned to reproduce the detector's response to the photons emitted by the sample. These simulations mainly rely on the efficiency and resolution calibrations that will be presented.

In the first section I present the different kinds of calibrations for gamma-ray spectrum analysis, namely energy, resolution and efficiency calibrations. In the second section of this chapter I present the calibration that are usually done in order to perform peak-based and full-spectrum analysis. We point at some issues regarding the latter and in a last section I will present a new framework for the simulation of the spectral signatures and the calibration of gamma-ray spectrum for full spectrum analysis that allows to reduce the discrepancies observed in the second section.

4.1 . Instrument and simulation calibration for gamma-ray spectrum analysis

A gamma-ray spectrum is an histogram of the counting of the amplitudes of the electrical voltage pulses produced by the detector when the photons interact with the germanium crystal. Analog or digital pulse processing electronics classify these pulses in channels (16 384 or even 65 536) which can be related to the energy deposited by the photons after an energy calibration step since the pulse height is proportional to the deposited energy. As a result, the outcome of the detector consists in a simple vector of the counts collected in each channel.

The calibration of a spectrum is needed in order to analyse the composing elements of the spectrum and quantify the active radionuclides that a sample contains. The calibration is three-fold : the energy calibration of the output

of the detector, the resolution calibration of the spectral signatures and the efficiency calibration of the MCNP-CP model of the detector.

4.1.1 . Energy calibration

The energy calibration is basically the fitting between the channels of the detector output and the energy of the spectra. It focuses on matching the peaks x-positions observed in the measured spectrum and the nominal energies of the associated photons. This calibration allows to match the peaks present in the spectrum with the corresponding radionuclide. In the context of peak-based analysis, only the peaks has to be correctly calibrated in energy, whereas in the full-spectrum analysis, the entire spectrum (peak and Compton continuum) needs to be calibrated. The detector output is the only element that needs to be calibrated in energy since the spectral signatures are simulated with the nominal energy as the output x-positions.

This task relies on the knowledge of the nominal energies of the gamma-ray spectrum and the correct fitting of a few peaks in order to get the fitting function between the channel number and the energy, we will denote this fitting function :

$$E = f(c) \tag{4.1}$$

Where E is the energy at channel c, f is thus the function that links the channel number and the correct energy in keV. A typical gamma ray spectrum, measured in the laboratory, ranges from 20-30 keV to 1.7-2.7 MeV, the number of channels varies from one electronic to another from 16 384 to 65 536.

As can be seen in the first plot of figure 4.1 the raw spectrum is presented in term of channels that has to be matched with the correct energy. The aim of the energy calibration is to find the best matching between the energy and the channel number. The function that links the two will be based on the analysis of a standard source in which we know the energy of the peaks.

Once the x-position of the peaks in the observed spectrum has been found the function f that links the channel number to the energy can be fitted. As can be seen in the second plot of figure 4.1 the entire energy range is used for this calibration and it is important during this process to have access to information of the total spectrum in order to calibrate the spectrum at low and high energy. The fine tuning of the energy is complex as the energy bin of the observed spectrum is not regular throughout the entire spectrum, this will be seen in detail in the next section.

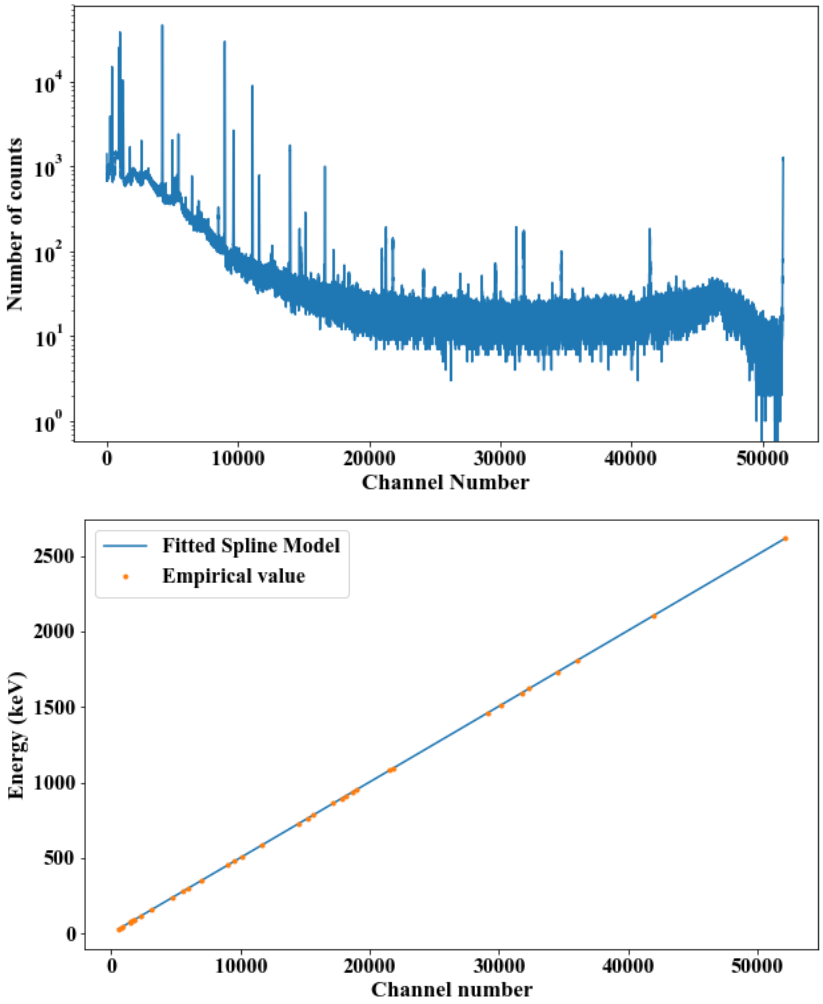


FIGURE 4.1 – *On the top, the raw spectrum out of the detector. On the bottom the model built on the empirical position of the peaks and their known energy.*

The problems that can arise from a bad energy calibration are two-fold. Firstly, it can lead to non identifications in the context of the peak based analysis of a gamma-ray spectrum : if a peak is not correctly matched in energy the radionuclide generating the peak cannot be identified. In the case of a simple gamma-ray emitting radionuclide (such as ^{7}Be of ^{137}Cs) only one peak is present in the spectrum their identification thus being particularly sensible to the correct energy calibration. If these peaks are not correctly matched in energy this will lead either to a false identification if another radionuclide emits a photon at close energy, or if the peak cannot be linked to any radionuclide in the library of possible emitters, it is listed as "unidentified peak" and the emitting radionuclide is missed.

Secondly, in the case of the full-spectrum analysis the identification and quantification depends on the correct modeling of the spectral signatures, if there are some mismatch between the observed spectrum and the spectral signatures this will lead to some leakage in the activity estimation. The entire energy range has to be correctly matched in this latter case as the Compton continuum is also used in the full-spectrum analysis.

4.1.2 . Resolution calibration

The resolution calibration focuses on the peaks' width of the spectral signature simulations. In fact, the resolution is related to the variance of the number of charge carriers produced in the detector and to the preamplifier noise that sum up, leading to a Gaussian broadening of the peaks around the characteristic energy. The simulated spectral signatures are generated with perfect Dirac distribution at the standard characteristic energy, the resolution calibration is performed so that the simulations reproduce the correct Gaussian broadening.

Formally, this broadening can be viewed as a Gaussian distribution of the counts around the characteristic energy of the emitted photon. The standard deviation of the Gaussian broadening depends on the energy (e, in keV), the greater the standard energy the larger the broadening. The resolution calibration is the fitting of the function $\sigma(E)$ that allows to pass from the energy to the broadening observed in the spectrum.

The peak width is used to process the net count in the region of interest around a peak in the peak-based analysis and thus get the activity of each radionuclide. In this framework, if this calibration is not done properly the identification of the region of interest around each peak will be hampered which can lead to a bad estimation of the activity of the sample. In the context of full-spectrum analysis the calibration must be performed so that the simulation of the spectral signatures are done with the correct Gaussian energy broadening. In fact, precisely matching the the measured spectrum leads to a

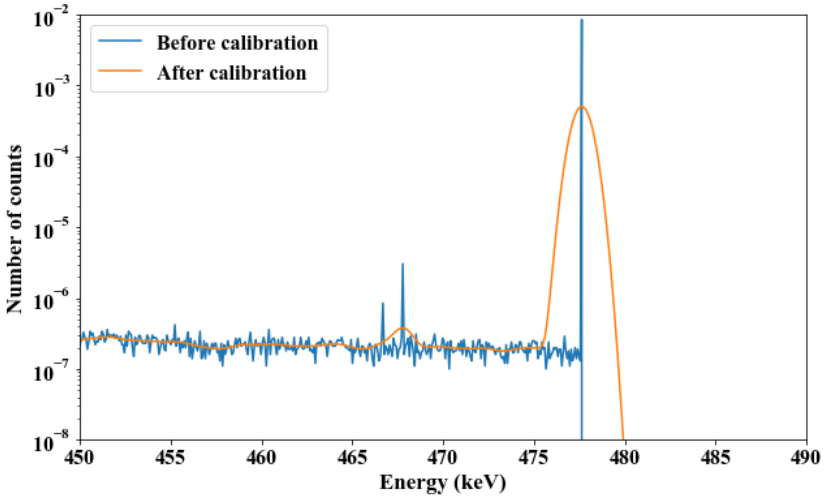


FIGURE 4.2 – *The spectral signature of* ^{7}*Be before and after the resolution calibration.*

better estimation of the activities and, once again, prevent leakage phenomena during the unmixing.

The Full Width at Half Maximum is usually the value used in gamma-ray spectrometry to deal with the resolution calibration. In the next section I will rather refer to the σ value of a Gaussian peak as it is directly present in the function that we fit inside the peak-fitting algorithm that I will present in section 4.3.3. The two values are virtually the same as the function that links one to the other is :

$$FWMH = 2\sqrt{2ln(2)}\sigma \approx 2.355\sigma \tag{4.2}$$

4.1.3 . Efficiency calibration

Finally, the efficiency calibration is the computation of the ratio of gamma-ray photons that interact with the detector over the entire number of emitted ones. The efficiency is mainly affected by the geometry of the detector and the sample. To achieve a correct matching of the observed efficiency and the simulated one the calibration needs to be performed and the model of the detector has to be tweaked accordingly. It is important, in order to quantify precisely the activity of a radionuclide present in a sample, to have a good efficiency calibration.

More precisely, the spectral signatures are the standard detected energy of a gamma-ray in the detector material. The efficiency can be seen as the number of photons that have been detected during a measurement for a given

number of emitted ones in the sample. The efficiency for an energy E reads as :

$$\varepsilon(E) = \frac{n(E)}{A\ I(E)\ t} \tag{4.3}$$

where,

— $n(E)$ is the net count at the energy E (in keV).

— A is the activity (in becquerel, Bq) of the radionuclide of interest (the one that emits photons at the energy E, the activity is considered to be constant throughout the measurement).

— $I(E)$ is the intensity of emission of the photon at energy E, it is a probability of emission, during a given disintegration for a photon to be produced $I(E)$ percent (for example the intensity of the gamma-ray of ^{137}Cs at 661.7 keV is 84.99%).

— t (in seconds, s) is the measurement duration.

The efficiency is the ratio of emitted photons that have been counted by the detector during the measurement process over the entire number of emitted photons. This ratio has to be empirically observed at various energies using a standard source.

The main factors that have an impact on the efficiency are the dimensions of the Ge crystal of the detector and particularly the width of the dead layer, the geometry of the sample and its position on the detector. The MCNP-CP models of the detectors are built to be close to the empirical efficiency observed with the standard source. An example of model is shown in figure 4.3 in which we can see the dead layer, it is the top part of the Ge crystal that is not active and does not participate to the observed spectrum. The efficiency calibration is done in stages, first the model of the detector is done with the technical specifications of the manufacturer of the detector in terms of materials and dimensions. Then a multi-gamma standard source measurement allows to compared the measured spectrum to the theoretical spectrum obtained with MCNP-CP, in shape and intensity. This measurement allows also to determine experimental detection efficiency for various energies, and the simulation's efficiencies are then tweaked to the experimental ones by modifying some elements of the MCNP-CP model so that the simulations closely match the measured spectrum.

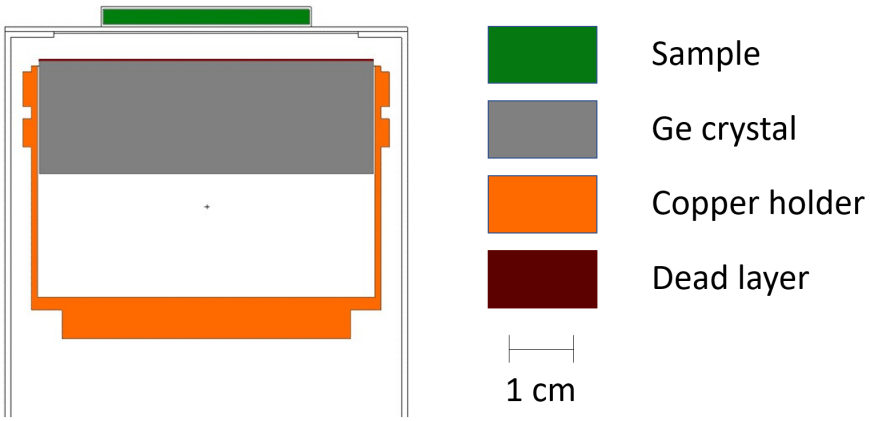


FIGURE 4.3 – *The basic model of a detector that has been optimized to produce the same efficiency as the empirical one.*

The efficiency for the two detectors used during my PhD has been computed with a standard source and the simulation process using MCNP-CP allows to take the correct efficiency into account. This calibration leads to a maximum error between the simulated and observed efficiency of ±5%.

4.2 . State of the art for the calibrations

In this section, I present the routine calibrations performed in the context of peak-based analysis and the first elements of calibration developed by Jiaxin Xu in her thesis work regarding the full spectrum-analysis.

4.2.1 . Peak-based analysis

In the context of a peak-based analysis the only required calibrations are the energy and efficiency. In fact, the basic for this analysis is to focus on specific peaks of known energy in order to quantify the radionuclides. The resolution is not relevant as it is computed each time a peak is fitted during the analysis of a spectrum.

To achieve this analysis, a pre-calibration is done once and for all on each detector. This pre-calibration allows to locate, more or less precisely, the peaks in the spectrum and to match the channel number and the energy. The function f that links the energy and the channels is usually a linear model between the nominal energy of the peaks and the channel number at which they appear in the spectrum. This calibration is done with a multi-gamma source of known activity and composition. The radionuclides of the source are chosen so that the peaks that are in the spectrum are scattered in the whole energy range

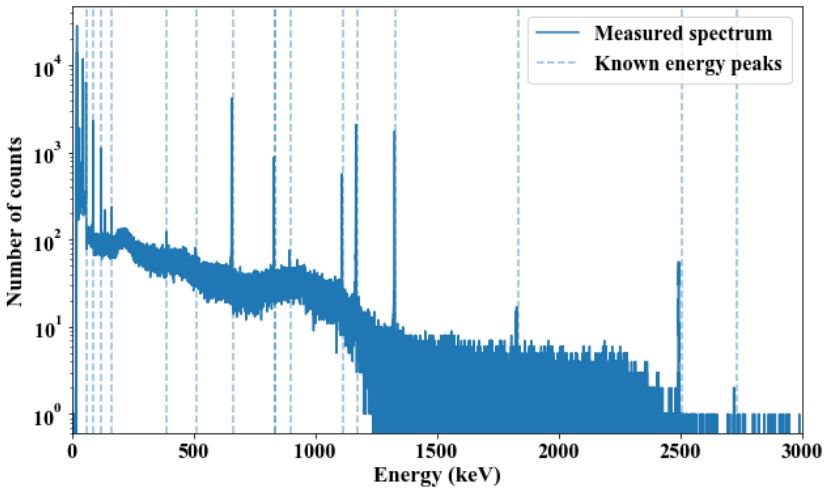


FIGURE 4.4 – *Spectrum of an hour of measurement of the gamma-ray source used to calibrate our detector.*

(an example of the composition of a multi-gamma source is found in table 4.1 and the associated gamma-ray spectrum can be found in figure 4.4). Once the peaks are correctly identified a peak-search algorithm can be used to find the channel number of each one in the measured spectrum. This algorithm performs in three steps : (i) it locates the peak precisely by finding μ, the x-position of the peak, in order to match it with a radionuclide, (ii) it separates the peak from the background by fitting a linear by part background, (iii) it computes the number of counts in the peak during the Gaussian peak-fitting algorithm by finding the amplitude A and width σ of the Gaussian-type peak.

A peak matching algorithm will be detailed in the next section (algorithm 3) in which our method for the Gaussian fitting of the peaks is explained. The gamma-ray analysis software do not provide the in depth explanation for the peak fitting algorithm. Problems may arise in the case of a low activity, for which the peak present in the spectrum will not provide a shape sufficiently well defined to perform the peak-fitting. Another problem may occur in case of overlapping peaks, this issue is alleviate by using a multi-gamma source which radionuclides has been chosen to have different peak region.

The energy matching is done with a linear model of order 3, based on the P peaks of the multi-gamma source of known nominal energy e_k. The function that links energy to the channel number via a polynomial of order 3 reads as :

$$E = f(c) = r_1 c^3 + r_2 c^2 + r_3 c + r_4 \tag{4.4}$$

Peak of interest (keV)	Radionuclide
46.5	Pb-210
59.54	Am-241
88.03	Cd-109
112.06	Co-57
165.86	Ce-139
320.08	Cr-51
391.7	Sn-113
514	Sr-85
661.66	Ce-137
834.84	Mn-54
1115.54	Zn-65
1173.23	Co-60
1332.49	Co-60
898.04	Y-88
1836.05	Y-88
2505.69	Co-60
2734.07	Y-88

TABLE 4.1 – *Composition of the multi-gamma source used in the laboratory.*

The 4 parameters r_1, r_2, r_3 and r_4 are optimized via least squares minimization. The cost function is then :

$$L(r_1, ..., r_4) = \sum_{k=1}^{P} \Big(E_k - f(c_k) \Big)^2 \tag{4.5}$$

This leads to the energy calibration of the spectrum. This calibration is performed on every detector in the laboratory and the parameters r_1, r_2, r_3 and r_4 are entered in the gamma-ray analysis software, it then uses the mapping in energy to locate the peaks in new measured spectrum and perform the peak-based analysis. The efficiency is then used to correctly quantify each detected radionuclide in the measured spectra.

As stated earlier, the efficiency calibration is done empirically. The efficiency is observed with a multi-gamma source of known activity, once the peak have been found in the spectrum, the efficiency is computed via the equation 4.3. In fact observing the spectra allows to determine the net count in a peak, the activity is known in the multi-gamma source, and the intensity of emission of a gamma photon can be computed and can be found in nuclear databases such as Laraweb ([2]). Once the efficiency is observed we build a MCNP-CP

model of the detector to match the observed efficiencies. The model is then used to simulate the response of the detector to radionuclides that are not in the multi-gamma source allowing to get the efficiency of the detector for each radionuclide we could need, over a wide energy range.

Once the efficiency and energy calibration are done, the peak-based analysis can be processed. This analysis only works locally around each peak, The computation of te width of each peak is performed locally and the resolution calibration does not need to be performed in this analysis framework. Moreover, if the energy calibration is not precise over the entire spectrum it won't affect the performance or correct identification of the radionuclides that are present in a given spectrum. In case of overlapping peaks a peak separation algorithm will be used in order to estimate the energy of the peaks and the area of each one.

4.2.2 . First model for full-spectrum calibration

In the framework of the full-spectrum analysis, the entire spectrum needs to be calibrated in energy, resolution and efficiency in order for the simulated spectral signatures to be used in the unmixing. In fact if a spectral signature is shifted or different when compared to the real spectrum some leakage or bad estimation may occur during the analysis. In the following sections I explain in details the different calibrations to build spectral signatures that matches the observed spectra and to calibrate a gamma-ray spectrum to perform full spectrum-analysis. As previously seen for the peak-based analysis, the calibration can be performed by using a multi-gamma source. However, with the need of a perfect matching between the spectral signatures and the observed spectrum, we will see that a fine energy calibration is needed in this framework, while the efficiency and resolution calibration can be carried out with the single multi-gamma source study.

Multi-gamma source analysis for efficiency and resolution calibration

The simulation of the spectral signature is performed with MCNP-CP. As previously seen, these simulations need the perfect modeling of a detector in order to precisely match the empirical efficiency over the entire energy range of the spectrum and for every radionuclide of interest.

In a first time, the analysis of a multi-gamma source is used to process the efficiency calibration that is used to tweak the simulation model of the detector to match the empirical efficiency as for the peak-based analysis. The results of this study is shown in figure 4.5, as can be seen the ratio between the empirical efficiency and the simulated efficiency is near 1 and inside a 5%

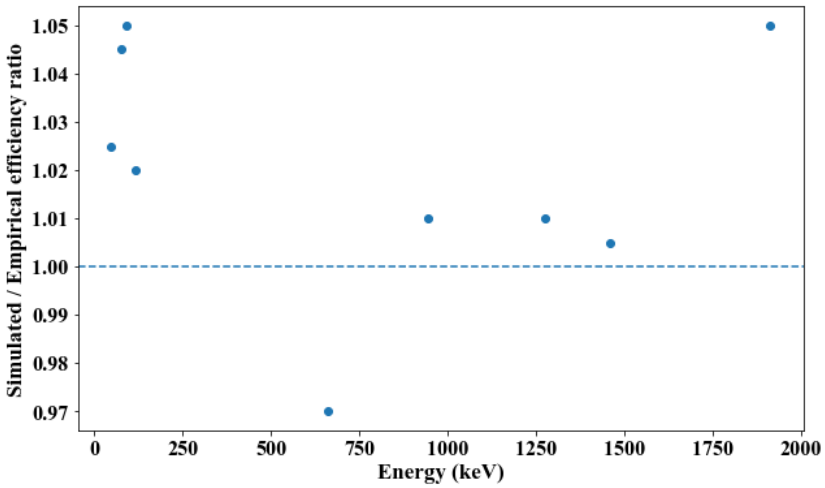


FIGURE 4.5 – *The ratio between the simulated and empirical efficiency resulting of the study of a multi-gamma source.*

error interval for each of the observed peaks.

In a second time, the Gaussian broadening effect of the resolution calibration can also be done in the simulation process. In fact, the model of the Gaussian broadening is empirically known, its equation reads as :

$$\sigma(E) = p_1 + p_2\sqrt{E + p_3E^2} \tag{4.6}$$

This equation is the one used in the GEB function of MCNP-CP so that we can compare to the standard simulation process (see annex C and [24] for details).

In [26] the advised function to be fitted is the following :

$$\sigma(E) = \sqrt{p_1 + p_2E + p_3E^2} \tag{4.7}$$

The two formula being near from one another and no difference between the two being observed on the energy range we are working on, we will focus on the first one (equation 4.6).

To achieve the resolution calibration, the width of the gamma-ray peaks present in the spectrum are computed on source spectrum and a model for the resolution with respect to the energy is fitted. The model is fitted by mean square minimisation to the empirical value as can be seen in figure 4.6.

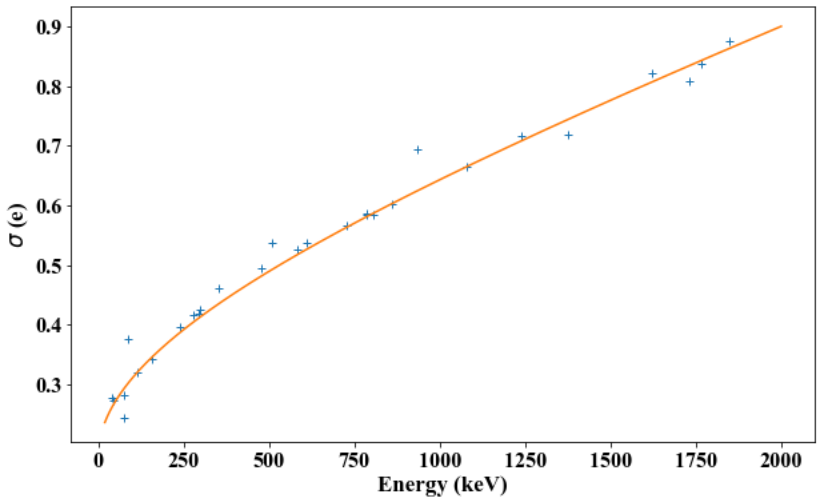


FIGURE 4.6 – *The equation 4.6 fitted with the empirical values gathered in a spectrum of a multi gamma-source.*

These parameters are optimised by minimizing the standard error between the function at the empirical points and the model. The function we want to minimise thus reads as :

$$L(p_1, p_2, p_3) = \sum_{k=1}^{P} \left(\sigma_k^{obs} - \sigma(p_1, p_2, p_3, E_k) \right)^2 \tag{4.8}$$

where,

— P is the number of peaks in the spectrum of the multi-gamma source.

— σ_k^{obs} is the empirical standard deviation of the observed Gaussian peak.

— E_k is the nominal energy of the gamma-ray peaks.

The peak width is computed with the peak-search algorithm that will be detailed in the next section.

Once the parameters p_1, p_2, p_3 are found we can use them to simulate the spectral signatures via MCNP-CP. In fact, during the simulation process of the spectral signatures it is possible to specify the broadening effect of the peak throughout the spectrum via the specific function called GEB (Gaussian Energy Broadening. More details are given in the annex C. The spectral signatures are then simulated with a correct efficiency and a broadening that is comparable to the one in the observed multi-gamma source.

The resolution and calibration is supposed stable through time and the calibration made on a source is supposed to give sufficient information on the efficiency of the detector to be able to build a MCNP-CP model that yields correct spectral signatures for each radionuclide we want to simulate. As will be seen in the next section the resolution calibration can be refined to produce a better agreement between the simulation and an observed spectrum.

Energy calibration of an observed spectrum

The energy calibration for the full-spectrum analysis is complex as the spectral signatures usage is sensible to any shift in energy. In fact, contrary to the peak based analysis, the full-spectrum analysis builds on the entire spectrum, Compton continua and full absorption peaks, to perform an analysis. The energy calibration thus needs to be extremely precise at the energy peaks and globally correct throughout the entire spectrum. Finally, the energy bin of the spectral signatures and the energy bin of the observed spectrum needs to be harmonized in order to perform the unmixing algorithm using full-spectrum analysis.

For the basic matching of the channels to the energy the first approximation presented in the peak-based analysis is used. This leads to a pre-calibration that can be used to locate the peaks in any sample measurement. The fine tuning of the energy calibration as proposed by Jiaxin Xu in her thesis ([6]) an in her article ([20]) proceeds as follows : (i) a first rebinning of the spectral signatures is processed in order to match the energy bin of the pre-calibration obtained on a multi-gamma source, followed by (ii) a first unmixing process to produce an estimation of the activity and a selection of the active radionuclides contained in the measured spectrum. After this first unmixing, (iii) a peak-search algorithm is performed on the peaks that are present in the sample's spectrum. The energy and channel number thus found are used to fit a better model function between the energy and the channel number. Finally, (iv) the spectral signatures are once again rebinned to match the new model. These rebinned spectral signatures are then used to unmix the measured spectrum leading to a better estimation of the activities.

The rebinning of the spectral signatures is processed in two stages. A first step is needed to produce a fine mesh size for the spectral signature after which the rebinning algorithm (2) is used to distribute the counts into the new bins. The interpolation into a fine bin is performed via interpolation of the spectral signatures and normalisation so as to keep the efficiency correct. The interpolation algorithm thus reads as :

In this algorithm, $x_{new}[left]$ and $x_{new}[right]$ are respectively the closest

Algorithm 2 The linear interpolation algorithm allowing to evaluate the spectral signature at a new bin.

Entries :

— x original energy bin
— y original channel content
— x_{new} : the new energy bin at which we want to interpolate y

$y_{new} \leftarrow \text{zeros}(\text{len}(x_{\text{new}}))$
for $i = 1, ..., \text{len}(x_{\text{new}})$ **do**
  $\text{left} \leftarrow \max(k \text{ s.t. } x_{new}[k] < x[k])$
  $\text{right} \leftarrow \text{left} + 1$
  $\lambda \leftarrow \frac{\text{right} - x[i]}{\text{right} - \text{left}}$
  $y_{\text{new}} \leftarrow \lambda * y[\text{right}] + (1 - \lambda) * y[\text{left}]$
end for
$y_{new} \leftarrow y_{new} / \text{sum}(y_{new}) * \text{sum}(y)$

Return :

— y_{new} the new channel content at the energy bin x_{new}

interpolation points to which we need to put a point. Finally the last step of the algorithm is used to normalise the spectrum and conserve the efficiency calibration valid after the rebinning.

After the spectral signatures have been rebinned to a finer mesh size they are aligned to the observed spectra by using the position of the main peaks of the measurement. This is done by sing a peak search algorithm on a few peaks in the measured spectrum and by allowing a new linear model to fit the empirical energies to the energies of the spectral signatures (as the spectral signatures are simulations, the peaks they contain are at nominal energy).

This second energy calibration uses the gamma-ray peaks of the measured spectrum rather than the ones of a multi-gamma source (for example ^{208}Tl peaks that are present in most of the aerosol filter sample measured in the laboratory but cannot be put into a source due to its short half-life). The aim of this process is thus to build an energy calibration that better fits the actual energy bins of the observed spectrum. Once the model is fitted the spectral signatures are rebinned once more to fit the new energy calibration.

Issues with these calibrations

Both techniques shown above, lead to some issues, mainly, the existence of some discrepancies between the spectral signatures and the observed spectrum as can be viewed in figure 4.7. The two main issues are the shift between the

nominal energy of the peak and the position of the peak in the spectrum, this comes from the energy calibration. Secondly, the slight difference in width between the spectral signature and the observed spectrum showing problems on the resolution calibration. The shape of the residuals between the observed spectra and the spectral signatures also enlightens these issues, the peak of the residuals being slightly off from the nominal energy of the peak and the positive/negative alternation being the result of the width difference. These discrepancies lead to some problems in the unmixing procedure as the model tries to match some spectral signatures that are not close enough to the actual spectra leading to leakage phenomena in the activity estimation.

The origin of these issues is the energy calibration being performed via a linear model. In fact, some residuals are observed after the model fitting that lead to some energy shifts between the spectral signatures ant the observed spectrum. Even if the linear model reduces the residuals in the most efficient way, the residuals are still showing after the recalibration processed. We want to fit a model that precisely matches the anchor points of the model to a nominal energy. In the next section, we will use spline interpolations to fix this issue and have a perfect alignment of the peaks in the spectral signatures and the observed spectrum.

Moreover, the problems on resolution are due to the rebinning process being performed on the broadened spectral signatures. In fact, the spectral signatures are simulated once and for all and then rebinned and interpolated to fit the energy model. This leads to some small distortion of the Gaussian energy broadening that has been specified during the creation of the spectral signatures as can be seen in figure 4.7. A solution will be to perform the Gaussian broadening after the rebinning so that there is no deformation of the peaks during the rebinning process.

Finally, it is important to note that the global process of calibration is long. In fact, if one wants to change a parameter on the resolution of the spectral signature the whole simulation process has to be done using MCNP-CP, as the broadening is done during the simulation process. The same can be said for the energy calibration, as this process involve a fine mesh rebinning and some peak search in the spectrum. In the next section we will use a kernel smoothing procedure described to apply the broadening at the end of the calibration procedure and we will reduce the time to perform the spectral signatures rebinning.

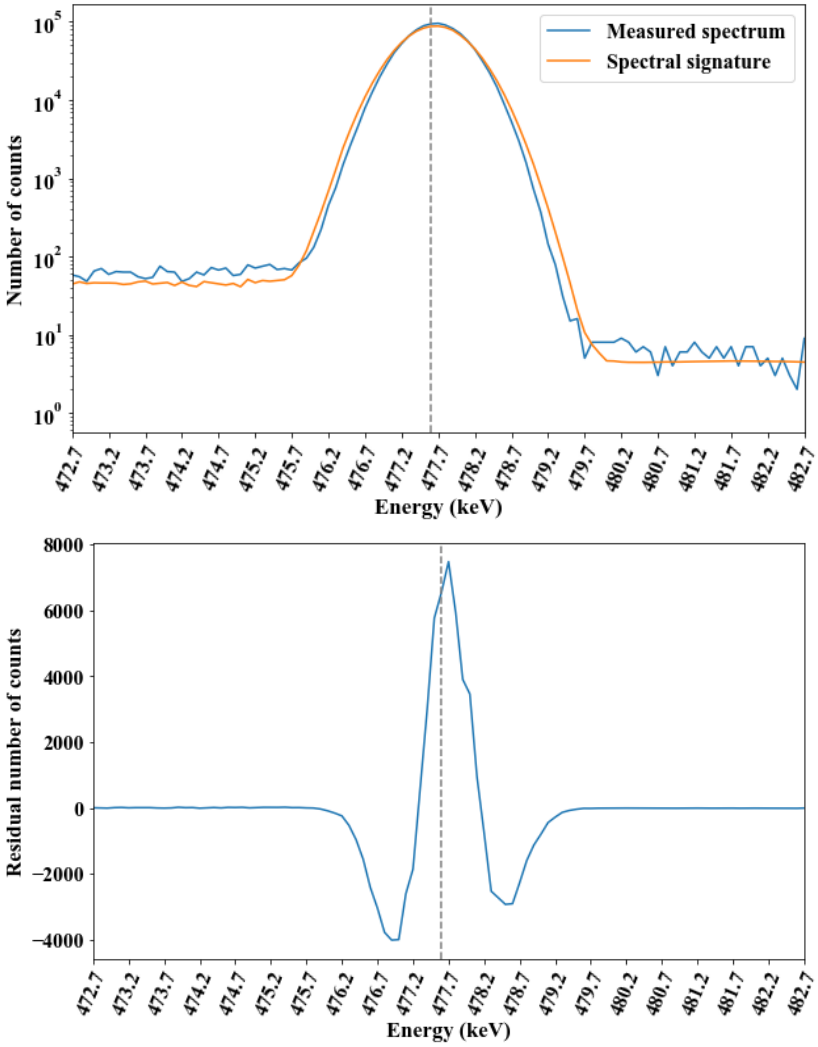


FIGURE 4.7 – *On top, the resulting energy shift on the peak of the ^{7}Be of less than 0.1 keV and the slight resolution problems after calibration. On the bottom, the residuals between both spectra.*

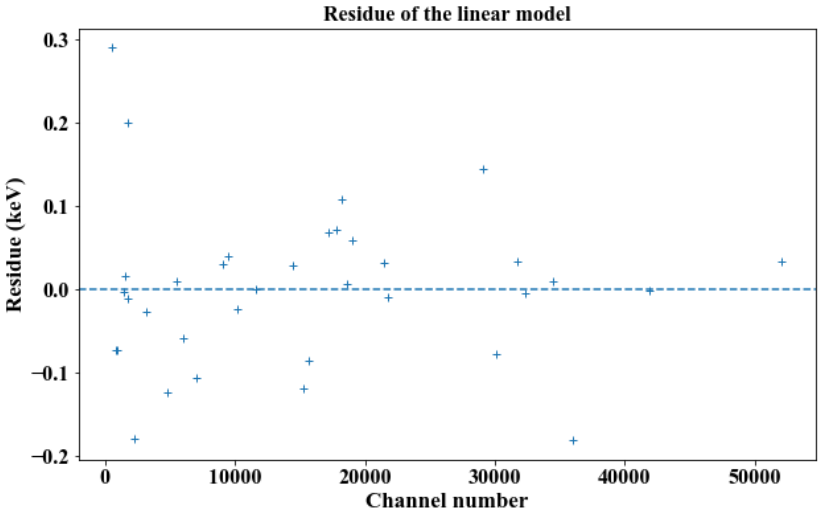


FIGURE 4.8 – *The residues of the linear model used to fit the energy.*

4.3 . Improvements in energy and resolution calibration

In the full-spectrum analysis framework, it appeared necessary to make the calibration more precise in order to perform spectral unmixing in the most efficient way possible. As previously seen, the calibration of the spectral signatures in energy, resolution and efficiency are important in order to get a proper estimation of the activities of the radionuclides present in a sample. Moreover, the usage of the Léda detector for which no pre-calibration were available and a larger energy range than the usual detector used in the laboratory is achieved, pushed us to develop a new calibration process.

However, the efficiency can only be optimized with standard sources, for the application to aerosol filter sample activity estimation it is not possible to get standard sources for the radionuclides of interest (radon progeny with a short half-life, overlapping peaks, ...). Moreover the efficiency calibration performed with the multi-gamma source leads to satisfactory precision (5% of maximum discrepancy between empirical and simulated efficiencies). As a result, we will focus only on the energy calibration along with the resolution calibration leaving the efficiency (and thus the MCNP-CP model of the detector) as it was after the analysis of the multi-gamma source.

The new calibration I developed for energy and resolution works as follows : (i) first the simulation of the spectral signatures is done with an infinite resolution (without GEB) and a fine energy bin. As a result, the peaks in the spectral signatures are perfect Dirac at the nominal energies and the rebinning is allowed without relying on a high resolution interpolation. (ii) Secondly, a

peak detection in the observed spectrum is performed on a family of gamma-ray peaks of known nominal energy. This peak detection allows to gather the channel number of each peak and perform a pre calibration in energy. (iii) A second peak search analysis is performed this time on a big number of peaks, this allows to get the position and the peak width of a large number of gamma-ray peaks in the entire spectrum energy range. (iv) A spline regression to match the channel number and the energy is then performed to get a perfect alignment of the peaks to the nominal energy in the observed spectrum. The observed spectrum is then rebinned to obtain a regular bin in energy following the fitted spline interpolation. (v) The spectral signatures are also rebinned to match the bin of the observed spectrum.

After the energy calibration the resolution calibration is performed with the peak width information gathered by the second peak search algorithm performed on a large number of peaks. The model of the Gaussian broadening with respect to the energy is then applied to the spectral signatures using a convolution of the spectral signatures with a non stationary Gaussian kernel in order to finely match the observed resolution in the measured spectrum.

4.3.1 . Detailed peak fitting procedure

The proposed energy and resolution calibrations only builds on the knowledge of the radionuclides composing the spectrum and no pre-calibration with multi-gamma source. In fact, during my PhD the only measurements I analysed were aerosol filter samples dominated by radon progeny. I thus have access to the spectral signatures and nominal energies of a lot of gamma-ray peaks throughout the spectrum as can be seen in the table 4.2. This information will be sufficient to process the calibration that will be presented in this chapter.

As stated above, the MCNP-CP simulations are performed with a model that correctly matches the efficiency of our detector. The simulations are processed without Gaussian energy broadening and on a fine energy bin (compared to the observed spectra). We thus have simulated spectral signatures at a fine bin size and infinite resolution (Dirac form peaks) the bins are evenly spaced every 0.1 keV from 10 keV to 3 000 keV.

The observed spectrum is produced in term of channels (from 16 384 to $\sim$ 60 000 channels) and we want to match these channels to the correct energy via the determination of f, a function, linking E the energy in keV and c the channel number, such that :

$$E = f(c) \tag{4.9}$$

The first step is to find a few peaks as anchor points for a pre-calibration of the energy. To achieve that, we have selected 3 peaks that are easy to find and use a peak-fitting algorithm to get the position (channel number) of the peaks. The 3 peaks we are using for the pre-calibration are at 46 keV (^{210}Pb), 477 keV (^{7}Be) and 2614 keV (^{208}Tl) as they cover the entire range of the spectrum. Moreover, no other peaks are present in their vicinity, allowing to identify them easily.

The peak-fitting algorithm proceeds as follows :

Algorithm 3 The algorithm performing the peak fitting on a given region of interest.

Entries :
- x : the energy bin of the spectrum
- y : the channel content of the spectrum
- [low, top] : the region of interest where we want to fit a peak

$m \leftarrow \text{mean}(y[\text{low} : \text{top}])$
$\mu \leftarrow (\text{low}+\text{top})/2$
$A \leftarrow \max(y[\text{low} : \text{top}])$
$\sigma \leftarrow 1$
$\mu, \sigma, A, m \leftarrow \text{argmin}_{\mu,\sigma,A,m} \sum_{k=low}^{top} \Big(y[k] - G(\mu, \sigma, A, m, x[k])\Big)^2$

Return :
- μ : the Gaussian peak centre
- σ : the Gaussian peak peak width
- A : the amplitude of the peak
- m : the background considered as a constant

The aim of the algorithm is to fit a function G depending on multiple parameters in the [low, top] region of interest. G can take different forms, the simplest consists in a constant background m and a Gaussian peak of amplitude A, mean μ and standard deviation σ giving the following function :

$$G(\mu, \sigma, A, m, x) = m + A\frac{1}{\sqrt{2\pi}\sigma} exp\Big(\frac{-(x-\mu)^2}{2\sigma^2}\Big) \tag{4.10}$$

The optimization is carried out using scipy solver ‘curvefit‘.

The results of this peak-fitting algorithm is shown in figure 4.9.

The model we chose is a standard Gaussian peak and the background is

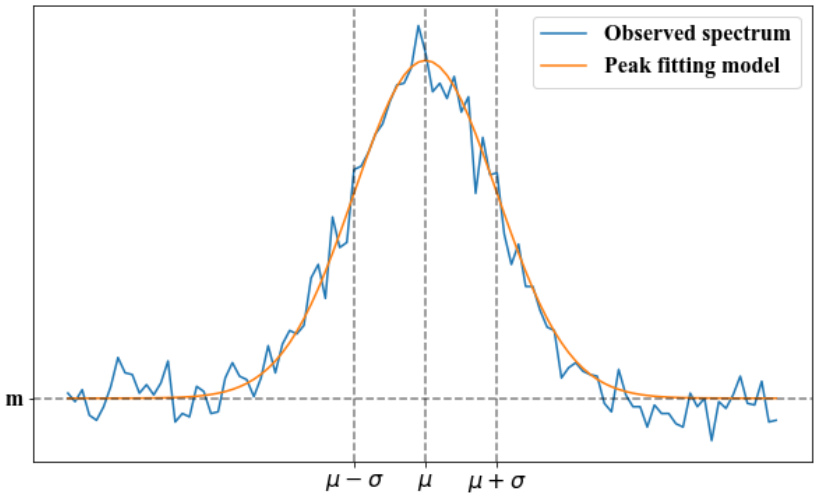


FIGURE 4.9 – *The result of the peak fitting as presented in the algorithm 3*

supposed constant but more complex shapes could have been used. In fact following the Laboratoire National Henri Becquerel ([27]), the Gaussian peak is a good approximation of the peak's shape in a gamma-ray spectrum. If a more precise model has to be applied, an exponential tail can be added to the left or the right side of the peak the formula for this exponential tail is :

$$T(x, A, T, \tau) = \frac{AT}{2} \exp\Big((x-\mu)\tau + \frac{\sigma^2\tau^2}{2}\Big)\text{erfc}\Big(\frac{1}{\sqrt{2}}(\frac{x-\mu}{\sigma} + \sigma\tau)\Big) \tag{4.11}$$

where erfc is the complementary error function defined as :

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} \exp(-t^2)dt \tag{4.12}$$

As can be seen in figure 4.10 no improvement in the fit of the Gaussian peaks with the exponential tailing is observed adding the cost of 2 more parameters to be estimate and a complex model to be fit. The basic Gaussian peak model is thus chosen as it is sufficient to estimate the value of μ and σ for the peak of interest of our dictionary.

On the other hand, still following the recommendation of Laboratoire National Henri Becquerel ([27]) a linear by part background that better estimates the background contribution can be used. The algorithm stays the same but fits a constant background on the left and right parts of the peak region contrary

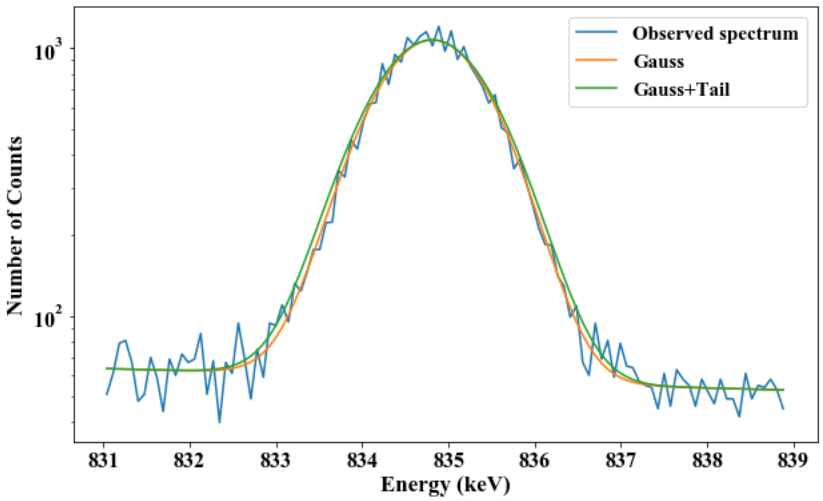


FIGURE 4.10 – *The comparison between the peak fitting algorithm performed with the simple Gaussian peak and the Tailed version.*

to 3 that fits a constant background. The result of this fit is shown in figure 4.11 and a better agreement is produced between the fitted model and the observed spectrum while keeping a simple model.

As a conclusion, the peak-fitting algorithm that will be used get the information on the position μ and the width σ of the peak of the spectrum. The model we fit on the data is the following :

$$
\begin{aligned}
&G(\mu, \sigma, A, 0, x) + B(l, r, x) \\
&\text{where :} \\
&B(l, r, x) = \begin{cases} l \text{ if } x < \mu - 3\sigma, \\ l \times \frac{\mu+3\sigma-x}{6\sigma} + r \times \frac{x-\mu-3\sigma}{6\sigma}, \text{ if } x \in [\mu \pm 3\sigma] \\ r \text{ otherwise} \end{cases}
\end{aligned} \tag{4.13}
$$

This model simply states that inside the area of the peak ($\mu \pm 3\sigma$) the background is the linear interpolation between the left and right vicinity (as can be seen in figure 4.11) and outside of this peak region the background is constant with value l on the left region and r on the right region.

The result of this algorithm are the parameters μ, σ, A, l and r. While remaining easy to process, this model allows a better fit of the peaks that are present in the spectrum.

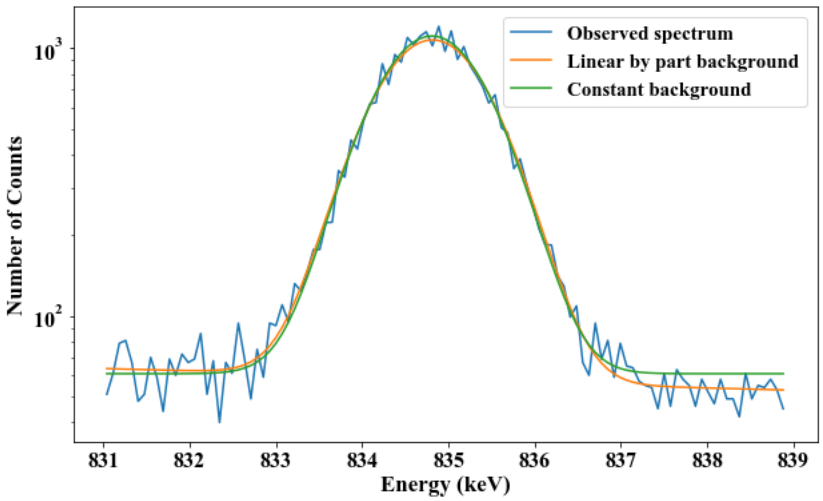


FIGURE 4.11 – *The comparison between the constant and the linear by part background.*

4.3.2 . Energy calibration process

With the peak-fitting algorithm we can get the x-position of the 3 peaks (46keV ^{210}Pb, 477keV ^{7}Be and 2614keV ^{208}Tl) that we want to perform a pre-calibration. This calibration uses simple linear function :

$$E(c) = Ac + B \tag{4.14}$$

Where $E(c)$ is the pre-calibrated energy of the c-th channel. This pre-calibration allows to perform the peak-fitting of all the peak dictionary automatically without having to get a region of interest for each peak. In fact after this pre-calibration the region of interest can be specified within the algorithm as $[(E_k \pm \delta - b)/a]$. The interval is not precisely the peak region but it is sufficient to perform the peak fitting automatically.

Once this pre-calibration is achieved the precise energy calibration begins. The peak-fitting algorithm is used once again on the full peak-dictionary. This time 35 peaks are found and the channel number of the centre of each one is found. The peak dictionary I use is presented in table 4.2. These are the peaks that are easy to identify in the aerosol filter sample I worked on.

With the information of μ and σ obtained for each one of these peaks, a spline interpolation is performed in order to get the function f that links the channel number and the energy. The function f is chosen to be a cubic spline function based on the 35 anchor points. The spline interpolations have the ad-

Energy	Radionuclide	Intensity (%)
27.4726	I-123	45.98
39.858	Bi-211	1.07
46.539	Pb210	4.252
74.8157	Pb-211	10.07
77.1088	Pb-211	16.9
87.347	Pb-211	5.77
89.8087	Bi-211	0.0223
115.183	Pb-211	0.624
158.97	I-113	83.25
238.632	Pb-211	43.6
277.37	Tl-208	6.6
300.089	Pb-211	3.18
351.932	Pb-214	35.6
452.98	Bi-211	0.34
477.6035	Be-7	10.44
583.187	Tl-208	85.0
727.33	Bi-211	6.65
763.2	Na-22-511	NA
785.37	Bi-211	1.11
860.53	Tl-208	11.4
893.408	Bi-211	0.38
911.196	Ac-228	26.2
934.061	Bi-214	3.1
952.2	Bi-214	0.0059
1078.63	Bi-211	0.55
1093.9	Tl-208	0.44
1174.537	Na-22	99.94
1460.822	K-40	10.55
1511.7	Bi-211	0.29
1592.511	Tl-208-1022	NA
1620.738	Bi-211	1.51
1729.595	Bi-214	2.844
1805.96	Bi-211	0.11
2103.511	Tl-208 - 511	NA
2614.511	Tl-208	99.755

TABLE 4.2 – *Radionuclides and peaks (energy and intensity) used for the energy and resolution calibration.*

vantage to pass by all the anchor points used for the interpolation thus leading to a 0 residual in the peak x-positions. More information on the definition of spline can be found in the annex A.

An issue arises from the use of the spline interpolation, in fact, it is only well defined inside the range of interpolation (*ie* the lowest and highest anchor points). The extrapolations of the spline function outside of the anchor points region does not provide stability and thus cannot be used. But, this problem is alleviated by cutting the lowest and highest energy of the spectrum. In fact, it is usual to cut low energy under 30 keV so that the electronic noise that can arise at low energy is not kept in the spectrum. For high energy, 1700 keV and higher, the gamma-ray peaks are more and more scattered and the majority of the information is contained under the 1700 keV threshold, for the majority of the detectors used in the laboratory the working range of energy is chosen to be 30-1700 keV. In my work the lowest energy peak is a 27.5 keV and the highest is at 2614 keV, the cut has been chosen at 30 and 2300 keV so that we only focus on the interior of the interpolation range and do not lose to much information that the cut parts of the measured may contain.

Once the interpolation is done the peaks perfectly match the simulations in term of energy/position of the peaks a rebinning process is then performed so that each channel of energy of the measured spectrum has the same energy width as the spectral signatures. The re-binning algorithm proceeds as follows :

The algorithm simply performs a binomial randomization of the contents of any channel of the observed spectrum to the new bin. This randomization depends on the proximity of the old bin to the nearest new ones. This re-binning has no impact on the spectrum other than to get an even binning throughout the spectrum and potentially reduce the dimension of the spectrum by reducing the number of bins of the new spectrum. The rebinning process conserves the integer nature of the content of each channel, indeed the Poisson noise originating from the counting process is conserved by performing the rebinning this way. It would not have been the case with a typical repartition of the count by interpolation.

Another possibility to perform the rebinning while not impacting the measured spectrum would be to perform the rebinning on the spectral signatures rather than the observed spectrum. This would change the even binning of the spectral signatures to follow the uneven energy bin of the observed spectrum and require more computation later during the resolution calibration.

It is important to note that the algorithm sums every counts that are

Algorithm 4 The algorithm performs a rebinning by finding the nearest bins in the new abscissa vector and split the content of the channel to each one with a random binomial value.

Entries :

- x : the initial bin in energy
- y : the initial channel content
- x_{new} : the new energy bin at which we want to interpolate the spectrum

for $i = 1, ..., \text{length}(x)$ **do**
  $l \leftarrow \max(x_{new}\ s.t.\ x_{new} <= x[i])$
  $r \leftarrow \min(x_{new}\ s.t.\ x_{new} >= x[i])$
  $p \leftarrow \frac{x[i]-l}{r-l}$
  $counts \leftarrow \text{Binom}(y[i], p)$
  $y_{new}[\text{where}(x_{new} = l)] \leftarrow y_{new}[\text{where}(x_{new} = l)] + count$
  $y_{new}[\text{where}(x_{new} = r)] \leftarrow y_{new}[\text{where}(x_{new} = r)] - count + y[i]$
end for

Return :

- y_{new} : the new channel contents

lower than the lowest new bin to the lowest new bin. The same is performed for the higher bins. This can lead to some issues in the extreme bins of the new spectrum y_{new} that is processed. But as stated above, the spline interpolation in itself needs to cut the extreme parts of the spectrum because it is well defined only inside the anchor points range. The problem is then alleviated by first rebinning from 29 keV to 2600 keV at the same bin width as the spectral signatures (0.1 keV/channel) and then cut to the chosen energy range 30-2300 keV.

The calibration in terms of energy is done at this point, the spectral signatures and observed spectrum being aligned in energy and on the same energy bin. In the next section we focus on the resolution calibration of the spectral signatures. The same rebinning process is applied to the background measurement in order to get the same energy calibration and the same binning for this spectrum that will be used during the unmixing.

4.3.3 . Resolution calibration

For the resolution calibration we use the information gathered during the finding of the 35 gamma-ray peaks of the energy calibration. In fact thanks to the peak-search algorithm we have access to the energy broadening of the peaks on the complete energy range via the σ value gathered for each peak.

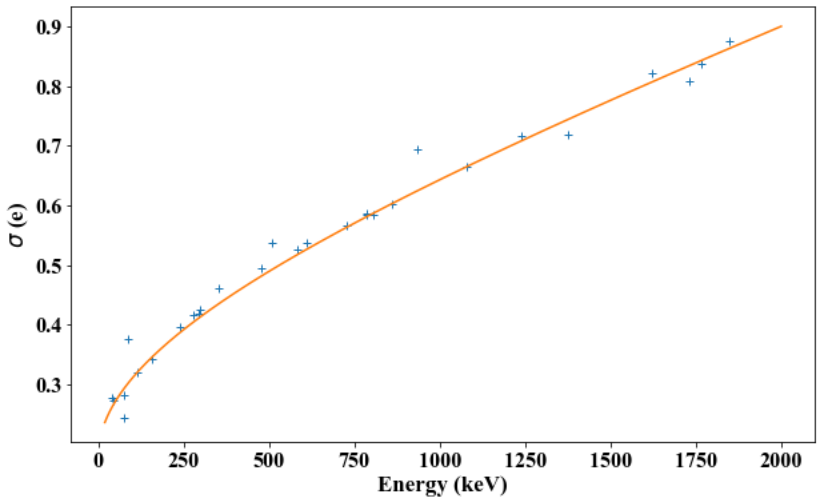


FIGURE 4.12 – *The empirical values of σ found via the peak-finding algorithm and the model of $\sigma(e)$ fitted via least square.*

With this information we fit the polynomial function previously introduced :

$$\sigma(E) = p_1 + p_2\sqrt{E + p_3E^2} \tag{4.15}$$

by least square regression. The empirical values and the fitted function can be seen in figure 4.12. The model, while remaining mostly empirical as stated in [26] fits correctly the empirical points. One does now apply the proper Gaussian energy broadening to the spectral signatures that are for the time being, at infinite resolution, presenting Dirac peaks.

Once the fit has been performed and we have the proper Gaussian energy broadening with respect to the energy we must apply the broadening to the spectral signatures. This can be done by the convolution of the spectral signatures with Gaussian kernels of correct width $\sigma(E)$.

The kernel used is a Gaussian kernel defined as :

$$G(\sigma(e), x) = \frac{1}{\sqrt{2\pi}\sigma(e)}\exp\Big(\frac{-(x-e)^2}{2\sigma^2(e)}\Big) \tag{4.16}$$

This kernel is convoluted with the spectral signatures that are simulated with infinite resolution (Dirac peaks at nominal energy) so that we reproduce the Gaussian broadening effect of the detector. An example of the effect of

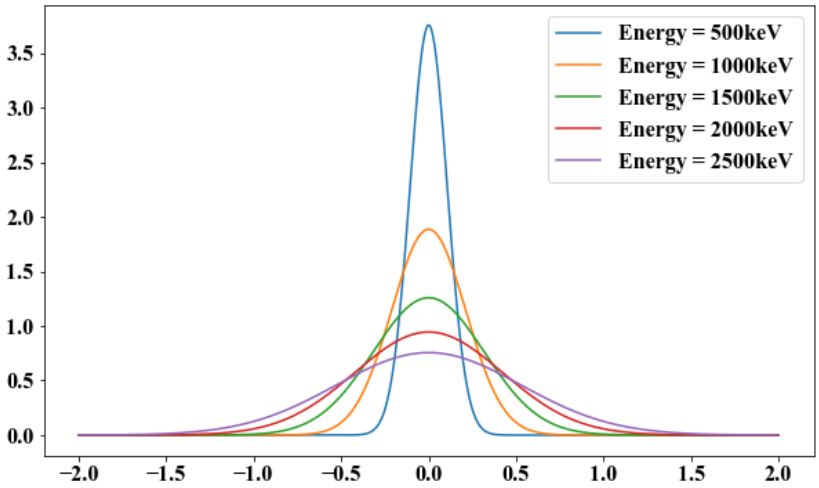


FIGURE 4.13 – *Some examples of the Gaussian kernels with the resolution calibration used with the Léda detector at various energies.*

this convolution can be seen in figure 4.2, which shows the application of the Gaussian broadening on a spectral signature focusing on the full-energy peak region. The shape of the Gaussian kernel can be seen in figure 4.13. In this figure we can see that the broadening is not the same at low and high energy. In fact, the Gaussian broadening depending on the energy of the gamma-ray, the higher the energy, the larger the broadening will be.

This convolution can be performed by matrix multiplication in the framework of discrete vectors of data leading to the following operation :

$$\Phi_{calib} = K\Phi \tag{4.17}$$

Where K is the matrix containing the Gaussian kernel. It is a matrix of dimension $C \times C$. Each column of the matrix is built as follows :

$$\begin{aligned} &\forall c = 1, ..., C \ and \ \forall i = 1, ..., C \\ &K[c,i] = exp\Big(\frac{-(E[c] - E[i])^2}{2\sigma(E[i])^2}\Big) / \sqrt{2\pi\sigma(E[i])^2} \end{aligned} \tag{4.18}$$

Because of the size of the matrices involved in the matrix multiplication the convolution is a time consuming process, hence a development has been done in order to speed up the process. A sparse approach has been used to create the

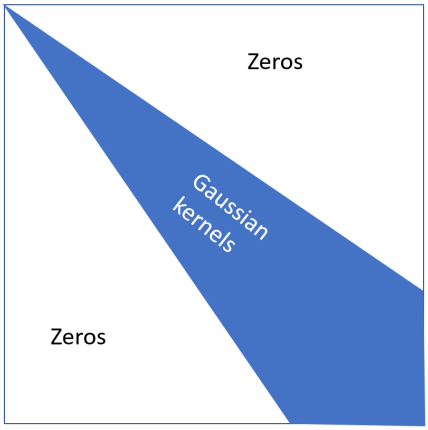


FIGURE 4.14 – *Scheme of the structure of the matrix K containing the Gaussian kernels defined around diagonal of the matrix.*

kernel matrix that uses only a few entries to get the proper broadening effect. This development is based on the light tailed property of the Gaussian kernel, in fact the Gaussian distribution is mainly distributed around the mean, after 3σ the contribution of the kernel can be considered as being negligible. The matrix K can thus be redefined only around its diagonal as :

$$\begin{aligned} &\forall c = 1, ..., C \text{ and } \forall i = 1, ..., C \\ &K[c,i] = \exp\Big(\frac{-(E[c]-E[i])^2}{2\sigma(E[i])^2}\Big)/\sqrt{2\pi\sigma(E[i])^2}\chi(E[i],E[c]) \end{aligned} \tag{4.19}$$

where χ is the characteristic function that reads as :

$$\chi(E[i],E[c]) = \begin{cases} 1 \text{ if } |E[i]-E[c]| < 3\sigma(E[i]), \\ 0 \text{ else} \end{cases} \tag{4.20}$$

This definition of K, provides us the sparse matrix without losing information as the tails are near to the value 0 in a Gaussian distribution. The structure of this matrix is summarized in the scheme 4.14.

The Gaussian energy broadening is applied on the spectral signatures giving the proper resolution calibration as can be seen in the following figure 4.15. The green spectrum is correctly calibrated both in energy (the peaks in the observed spectrum and the peaks in the signatures are perfectly aligned) and in resolution (the peak width is correctly reproduced in the spectral

signatures).

It is important to note that the kernel is applied on the entire spectral signature and not only on the peaks, as a result we observe a slight smoothing of the continuum. Apart from the peaks, the rest of the signature is mainly composed of the nearly constant parts, they are not heavily impacted by the convolution.

4.4 . Conclusion

In figure 4.16 we can see the better resolution and energy calibration of the process as it is presented in this thesis. In fact looking at the old calibration we observe a slight shift in energy at the peak energy (it can be seen on the comparison between the observed and reconstructed spectra or by looking at the asymmetry of the residuals). Moreover the resolution calibration is off due to the distortion produced by the calibration after the simulation of the resolution. On the other hand, with the calibration procedure as presented in this thesis there is no shift and the resolution calibration is good, in fact the reduced residuals show that the only remaining issue is located in the peak height. This is due either to a bad efficiency calibration or an under estimation of the radionuclide causing the gamma-ray peak. The presented peak is one of the ^{212}Pb's, as no standard source can be obtained for this radionuclide due to its short half-life (3.085 min) we can only rely on the efficiency calibration around its peak performed with the multi-gamma source.

I have presented the way we improved on the energy calibration and on the resolution calibration during my PhD, the scheme of the entire calibration process as it is presented in my thesis can be seen in figure 4.17. Some issues left by the previous tools of the calibration, namely, the energy shifts due to the energy calibration performed via a linear model and the resolution calibration improvements via the Gaussian broadening performed after the spectral signature simulation to properly fit the measured spectrum have been fixed or improved upon. The remaining issue in the framework of online gamma-ray spectrometry is the fact that the spectrum is not available at the beginning of the measurement. In fact, during my work I worked on the observed spectrum in order to improve the calibration considering that I have access to the final measured spectrum. In the laboratory setting the information is gathered during the week long measurement and the calibration needs to be performed beforehand or during the actual measurement.

During an internship directed in the laboratory some developments were achieved to perform online calibration. I will present these first results in the

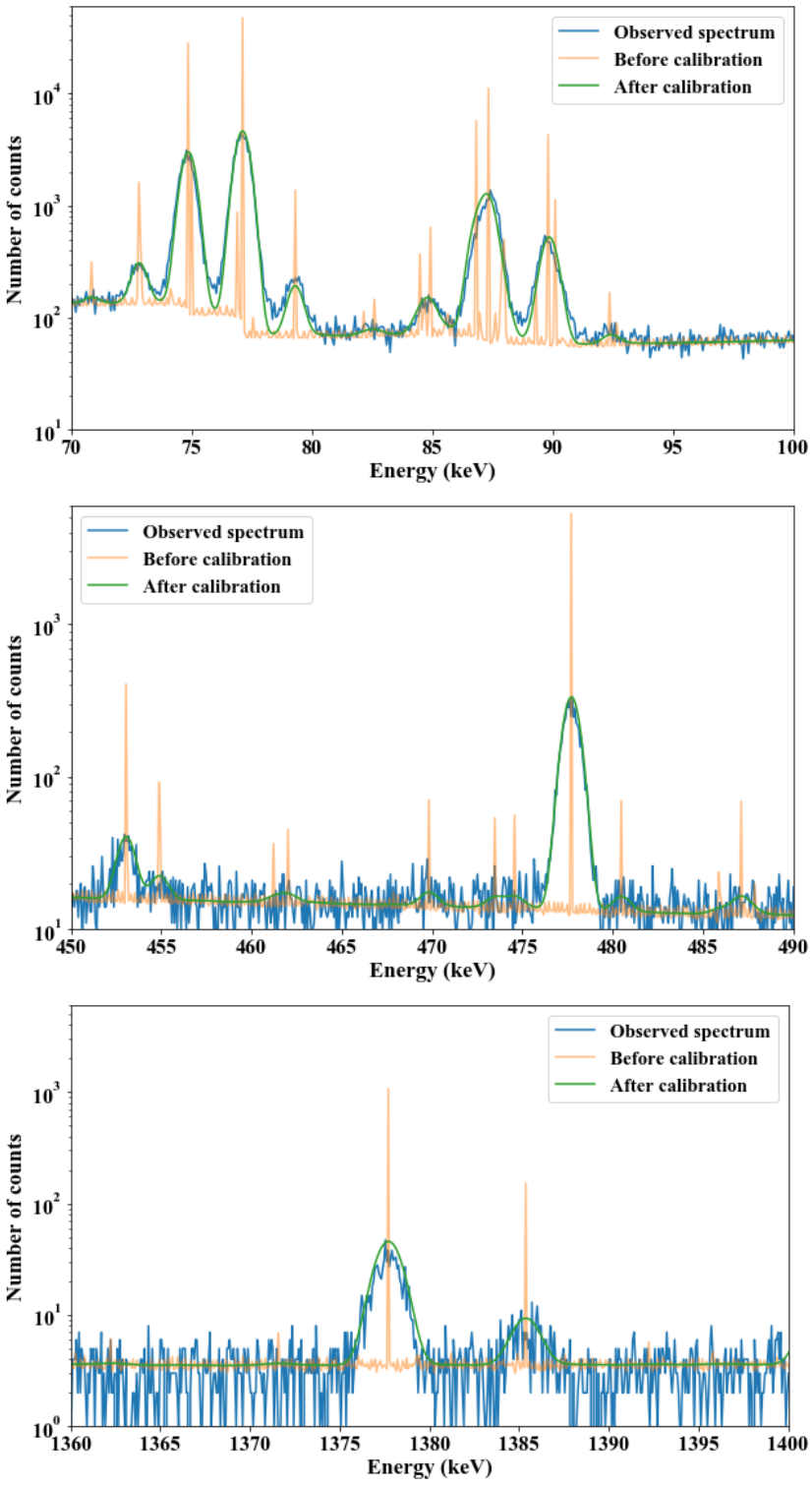


FIGURE 4.15 – *Three examples of the resolution calibration processed on the spectral signature compared to the observed spectrum at various energies.*

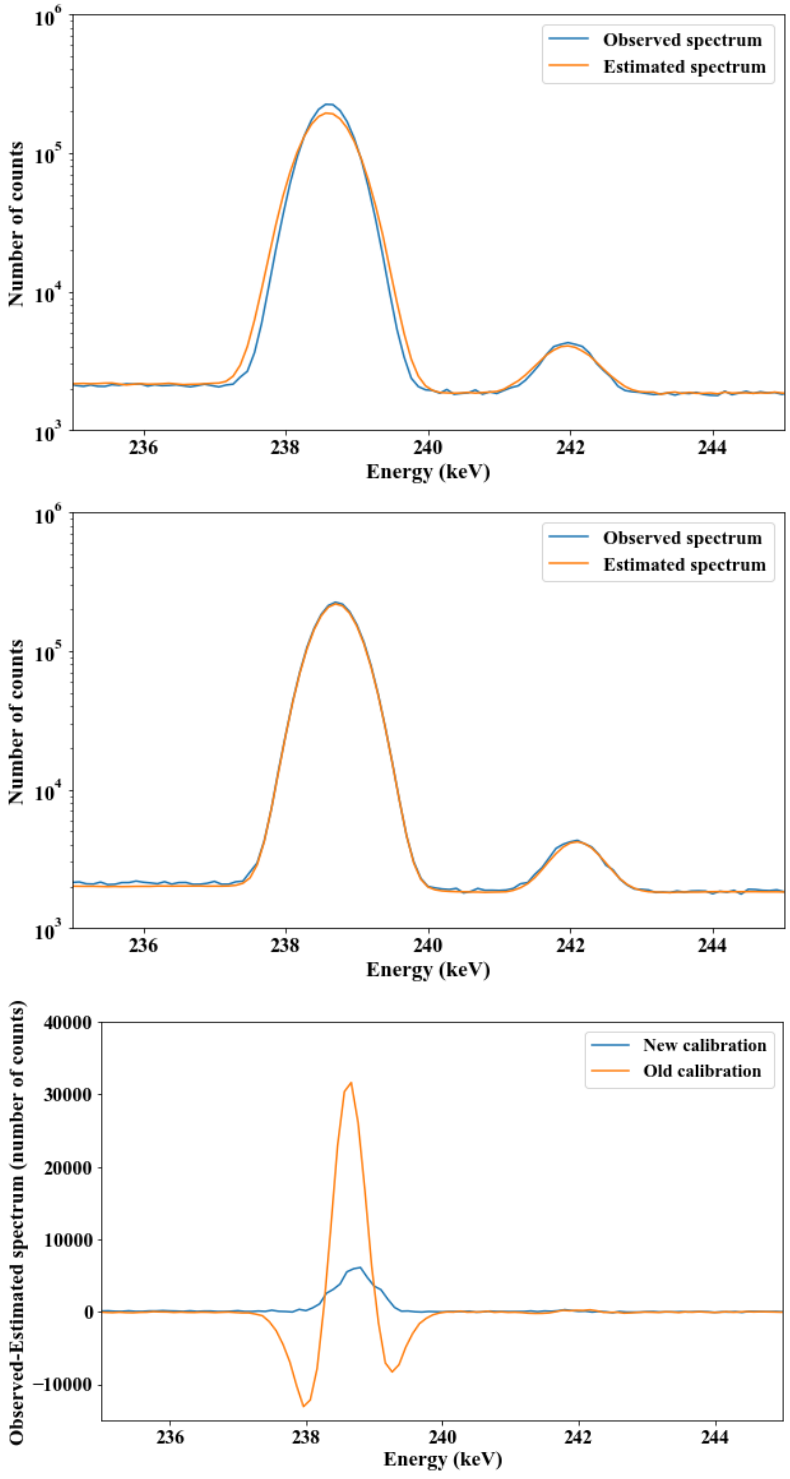
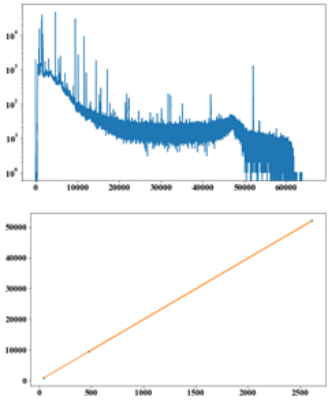


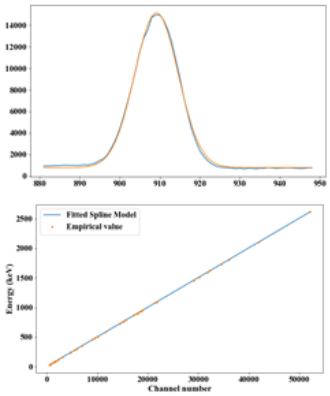
FIGURE 4.16 – *Comparison between the results of the calibration as performed by Jiaxin Xu and the calibration presented in this work. On top we see the observed and estimated spectrum using the old method, in the middle is the results of the procedure presented in this section. On the bottom is a comparison of the residuals of both methods.*

next chapter 5. On the other hand, the calibration performed as I described previously can be used as a precalibration for other measurements and still provide improvements on the previous tools. Moreover, the tools I presented thus far were done on complex spectra that presents a lot of gamma-ray peaks allowing the energy and resolution calibration even if the measurement is not finished.

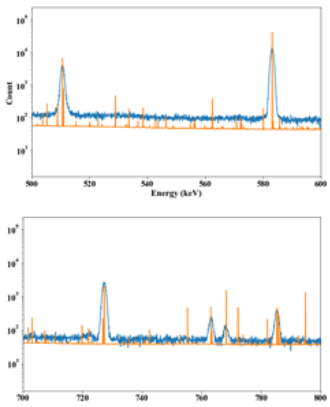
Pre calibration using the position of 3 easy to find peaks
Entry :
Measured spectrum in term of channel number
3 peaks and their ROI in channel number
Return :
Pre-calibration based on linear model ax+b



Location of 35 peaks and spline interpolation
Entry :
Observed spectrum, precalibrated in energy
Dictionary of 35 peaks
Return :
Position and width of the 35 peaks
Energy to channel number function



Rebinning of the spectral signatures and spectrum
Entry :
Measured spectrum,
Spectral signatures,
Channel number to energy interpolation function
Return:
Perfectly matched energy wise signatures and observed spectra on the same energy bin



Resolution calibration by fitting the GEB funcion with respect to energy and Gaussian kernel convolution
Entry:
Infinite resolution spectral signatures
Width of the 35 peaks in the measurement
Return:
Broadened signatures

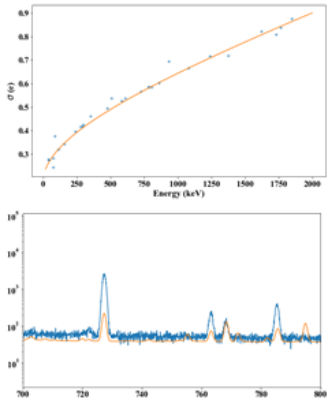


FIGURE 4.17

5 - Conclusion and perspectives

Sommaire

5.1 . Conclusion

In this thesis work, I introduced and studied new tools to perform gamma-ray spectrometry that account from time dependencies from consecutive measurements of the same sample. The aim of the present study was to achieve the rapid detection of possible contamination in the sample or the trace level radionuclides contained in the aerosol filters. This task is difficult due to a high complexity of the spectra contaminated by the radon progeny after the sampling process. Building on the decomposition of an observed gamma-ray spectrum into characteristic spectral signatures, the known physical model of the radioactive decay and the decay chains that can arise in the observed spectra, the proposed methods lead to the estimation of the activities contained in a sample in a timely fashion. The time information carried by consecutive measurements of the sample allows to improve the accurary of the unmixing process, as we showed in chapter 2 and 3. While focusing on aerosol filter sample measured on HPGe detectors, the same framework can be applied to other measurements on different detectors.

In chapter 1, we introduced a new framework for temporal spectral unmixing. We presented an unmixing algorithm that allow the joint analysis of consecutive measurement. In fact, analysing consecutive spectra allows to use temporal correlation to cope with complex spectra containing high level of activity from the radon progeny without hampering too much the estimation of the radionuclides of interest. We have shown that using the joint analysis allows to provide earlier results than the independent analysis of each spectra, reducing the uncertainties and the bias of the estimation of the activities as the measurement proceeds. The proposed method has been evaluated on both simulated data and real aerosol measurement providing a first estimation of the trace levels of ^{137}Cs after only 1 day and a half of measurement without observing a decay period.

In chapter 3, an extension of the temporal spectral unmixing algorithm has been introduced so as to estimate the activities in the sample in an quasi-online fashion providing first estimations after 1 minute of measurement. To that purpose, the unmixing algorithm has been optimized and tuned to speed up the computation of the activities in the sample. It has been evaluated on both simulations and real aerosol filter sample measurement. The size of the data that needs to be processed in an online framework being problematic, we opted for a buffer approach allowing to keep the matrix's size under control as well as keeping the memory of the past segment in the analysis of the new time segment. The number of iterations of the algorithm has also been reduced thanks to the regularisation we developed, allowing to take the past estimation into account to proceed to a new one, and to the sparse update of the activities

only taking into account the radionuclides that are still active in an incoming time segment to proceed with the activity estimation. Thanks to this new algorithm we detected a contamination of ^{123}I after 1 minute of measurement at the activity of 2 Bq in a spectrum dominated by the radon progeny. This would not have been possible if the usual decay period of 2 to 4 days had been observed before proceeding with the measurement as the half-life of this radionuclide is about 11 h.

Finally, new tools for the proper calibration of the spectral signatures have been presented in chapter 4, allowing to reduce the discrepancies between the simulated spectral signatures and the observed spectra, further allowing the rapid and precise unmixing of gamma-ray spectra. This new calibration process allows to precisely match the energy of the peaks between the simulated spectral signatures and the observed spectra thanks to a spline interpolation of the energy calibration function. Moreover, the resolution calibration is not impacted by the energy calibration thanks to a Gaussian broadening that is performed after the energy calibration so that no distortion are done in the latter process. The results of this new framework were compared to the passed calibration in order to show the fidelity of the spectral signatures to the observed spectrum.

The proposed temporal analysis framework can be applied to routine measurement in order to reduce the time needed to perform an activity estimation from a week to a few days or to perform analysis automatically with new sampling station such as the Cinderella automatic air sampling station from Selya. This automatic sampling station performs a day of sampling followed by a day of decay and a day of measurement with a HPGe detector. The joint analysis of the consecutive samples may be performed with the online algorithm presented in this thesis to further automatize the process from the sampling to the analyse.

5.2 . Perspectives

5.2.1 . Performing the estimation while the sampling proceeds

While this thesis work only focuses on laboratory measurement, once the aerosol filter has been sampled, we could extend the analysis to the aerosol filter during the sapling process. In fact as can be seen in [28] the model of the deposition of the radionuclides on the aerosol filter can be used along with the known radioactive decay model to be able to cope with a measurement of the activity during the sampling process.

On the other hand this would require to handle a measurement in the

sampler which will be accompanied by an augmentation of the background (no more shielded room or borated concrete slab to protect the detector from cosmological or telluric noise). Moreover, the radon has a day/night cycle and is subject to weather phenomena such as the rain. Finally, the fact that the measurement is carried outside means that the detector won't be as stable as in the controlled environment of the laboratory, this means that the calibration of the detector may be needed to be carried during the measurement as the temperature and other factors may change slightly the calibration.

5.2.2 . Online calibration

A seen in chapter 4, the precise calibration of the detector and the spectral signatures is a problem that we face in the case of spectral unmixing because we are using spectral signatures to estimate the activities in the measured sample. In order to correctly calibrate the energy and resolution for the analysis we are using the spectrum measurement of the sample. In the online analysis framework, this information is not available, the estimation being performed as the measurement proceeds. New developments of online calibration would thus be needed in order to perform the joint estimation of the activities and correct calibration parameters. During an internship, which I co-supervised, we have explored preliminary solutions to this problem.

In order to perform the online calibration we first need to plant the mathematical framework for the optimization of the parameters of each element of calibration. Namely, we will try to fit the model for the energy with respect to the channel number as a linear model, rather than a spline interpolation, so that the model only depends on the parameters r_1, r_2, r_3 and r_4. The energy function f is then :

$$E = f(c) = r_1 c^3 + r_2 c^2 + r_3 c + r_4 \tag{5.1}$$

These 4 parameters will be the first element of the online calibration noted $R = (r_1, r_2, r_3, r_4)$. It is important to note that contrary to the method proposed in this thesis work, the spectrum won't be changed in the online method of calibration, only the spectral signatures will be processed, to fit the spectrum. Even if the peaks in the measured spectrum are not at the exact nominal energy, the signatures will be shifted to fit the spectrum.

The second calibration we parameterise is the resolution, as previously seen the function we need to optimize is the Gaussian broadening width, $\sigma(E)$ with respect to the channel number (or the energy). The function reads as :

$$\sigma(E) = p_1 + \sqrt{p_2 E + p_3 E^2} \tag{5.2}$$

This second set of parameters will be noted $P = (p_1, p_2, p_3)$.

The optimization of the efficiency is not possible as it requires the MCNP-CP modelling and simulation to be optimized. Moreover, it requires the exact activity of the source which is not available in routine measurements as it is the value we want to estimate. The efficiency calibration will thus not be part of the online calibration we will present.

As a result we have 2 sets of parameters, R, P to be optimized, each leading to the calibration of the spectral signatures. Finally a way of determining the correct parametrization is needed, the likelihood is chosen as the model of a gamma-ray spectrum is well known reading as :

$$\begin{aligned} X &= \Phi(\Psi.a) + \delta.b \\ Y &\sim Poisson(X) \end{aligned} \tag{5.3}$$

The estimated spectrum depending on the different sets of parameters and the activity estimation thus reads as :

$$\hat{X}(R, P, \hat{a}) = \Phi(R, P)(\Psi.\hat{a}) + \delta.b \tag{5.4}$$

Where Ψ is the matrix that contains the spectral signatures after the re-binning in energy depending on R and the broadening has been done according to the parameters P. The likelihood depending on every set of parameters thus reads as :

$$\mathcal{L}(Y|R, P, \hat{a}) = \sum_{c=1,t=1}^{C,T} \hat{X}_{c,s}(R, P, \hat{a}) - Y_{c,s} log(\hat{X}_{c,s}(R, P, \hat{a})) \tag{5.5}$$

The optimization of each parameters will be performed in stages with a Block Coordinate Descent algorithm, following the scheme detailed in 5.1.

The idea is to perform each calibration/optimization of a set of parameters with the others being fixed, at each optimization the Ψ is updated accordingly,

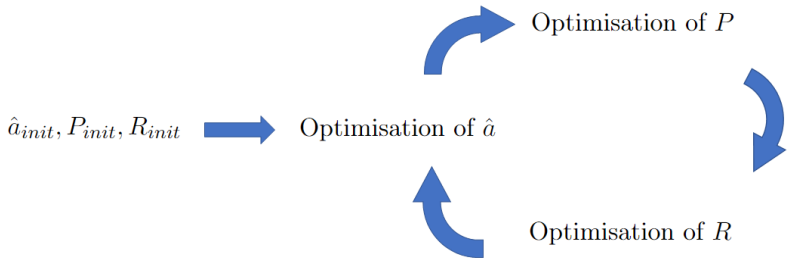


FIGURE 5.1 – *The scheme of the Block coordinate descent algorithm. Each step considers that the two other set of parameters are stable and to perform its optimisation.*

in the case of the efficiency, the correct matrix is selected, in the case of the resolution, the correct broadening is applied and finally for the energy calibration, the correct rebinning is processed. Once the calibration has been set, the unmixing algorithm estimates the activities and the cycle of optimization can begin again until convergence is achieved.

The problem we face in this framework are notably the computation time, in fact, each step of the optimisation for the parameters R and P require to compute the spectral signatures after calibration which is a long process. Moreover, some biases were still observed on the final results of both calibrations while still leading to correct results for the activity in the sample. Finally, each optimisation is a non-convex problem, the task is thus difficult. Indeed if we only want to carry the optimisation of the activity estimation we are requiring a lot of computation and complex algorithm. The optimisation of the two other sets of parameters would need some refinement to be able to optimize them properly.

5.2.3 . Spectral variability

Another way one could upgrade the full-spectrum analysis is by modeling the spectral variability of the signatures used in the unmixing process. In fact as can be viewed in chapter 4, we need to precisely match the spectral signatures to the spectrum in order to perform precise estimation of the activities in the sample. If a model for the variability of the spectral signatures was available this would allow more flexibility in the process and lead to an acceleration of the calibration. This would be another solution for the online calibration presented above. In fact, if one has access to a "basic" spectral signature database and a model of the spectral variability, the calibration can be processed using machine learning algorithm or the model of the variability.

5.2.4 . Online uncertainties

Finally, as can be seen in the annex B, the uncertainty quantification and the detection limits for the online analysis is a difficult task for which we still lack computationally efficient methods. For the time being, only Monte-Carlo simulations allow to tackle the uncertainty computation in case of low activity in the early estimation framework. In fact, due to the noise and relative contribution of the low activity radionuclides to the spectrum the Fisher's approximation with a Gaussian does not provide satisfactory results. The online quantification of the uncertainties and decision threshold is still an open problem.

Publications

Conferences and workshops :

— Poster : "Joint multitemporal spectrum analysis for low-level radioactivity in gamma-ray spectrometry" at International Conference on Radionuclide Metrology - Low Level Radioactivity Measurement Techniques 2022, Gran Sasso, Italy

— Oral presentation : "Online spectral unmixing of gamma-ray spectra" at Scidoni workshop, IRSN, Paris

— Oral presentation : "Online spectral unmixing of gamma-ray spectra" at International Conference on Radionuclide Metrology 2023, Bucharest, Romania

— Oral presentation : 'Temporal spectral unmixing for rapid detection of radiological events by gamma-ray spectrometry', Data-Analysis Group meeting, CEA Saclay, 2023

Articles :

— "Spectral unmixing of multi-temporal data in gamma-ray spectrometry", Paul Malfrait, Jérôme Bobin, Anne de Vismes-Ott, In : Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment 1045 (2023)

— "Online spectral unmixing in gamma-ray spectrometry", Paul Malfrait, Jérôme Bobin, Anne de Vismes-Ott, in Applied Radiation and Isotopes, accepted in June 2023

— "Analysis of gamma-ray spectra with spectral unmixing, part 1 : Determination of the characteristic limits (decision threshold and statistical uncertainty) for measurements of environmental aerosol filters", Jiaxin Xu, Jérôme Bobin, Anne De Vismes-Ott, Christophe Bobin, Paul Malfrait in Applied Radiation and Isotopes 182, 110109

— "Analysis of gamma-ray spectra with spectral unmixing, part 2 : Recalibration for the quantitative analysis of hpge measurements", Jiaxin

Xu, Jérôme Bobin, Anne De Vismes-Ott, Christophe Bobin, Paul Malfrait, in Applied Radiation and Isotopes 156, 108903

A - Spline

Sommaire

In this appendix we will presents the definition of spline interpolation. The splines has been used in the calibration section (4) to interpolate the energy function linking the channel number and the energy in a spectrum. In the first section we will give the definition and an example of the spline interpolation, in the second section we will detail the cubic splines and the interesting properties they have leading to the choice of this interpolation for energy calibration.

This annex is based on the book and course of Alfio Quarteroni [29] for more details see page 347 to 353.

A.1 . Definition

Splines are a way of interpolating a function using a basis of polynomials. Let $x_0 < x_1 < ... < x_n$ be distinct nodes on an interval $[a, b]$, $x_0 = a$ and $x_n = b$. The function $s_k(x)$ defined on the interval $[a, b]$ is called a spline of order k if, on the interval $[x_j, x_{j+1}] : s_k \in \mathbb{P}_k \; \forall j = 0, 1, ..., n$ and $s_k \in C^{k-1}[a, b]$.

where $\mathbb{P}_k$ is the space of the polynomials of order k and $C^{k-1}[a, b]$ is the set of the $k - 1$ times continuous function defined on the interval $[a, b]$. In other words, a function $s_k(x)$ is called a spline of order k if, on each interval between to nodes $[x_j, x_{j+1}]$, the function is a polynomial of order k and the $k - 1$ derivatives of the function are continuous in each of this nodes (the $k - 1$-continuity inside each interval between two nodes being given by the polynomial nature of the function).

For example let us build the spline interpolation of order 1 for the function $f(x) = cos(x)$ on the interval $[0, 2\pi]$. Given $n - 1$ distinct nodes inside the interval $[0, 2\pi]$ and the value of f on the total nodes : $0 = x_0 < x_1 < ... < x_{n-1} < x_n = 2\pi$ (for this example I chose 10 uniformly distributed values on the interval). We want to build a C^0 function based on straight lines (*ie* polynomials of order 1) between each interpolation nodes. This is exactly the definition of the spline interpolation of order 1 as shown in figure A.1.

This basic example shows two properties of the spline interpolation, on the one hand, in each of the interpolation nodes, the interpolation function $s_k(x)$ is equal to the approximated function $f(x)$. On the other hand, the spline is smooth and we will see in the next section in details what we mean by smooth.

If the definition of spline of order 1 is trivial, the definition in case of greater order of k is more complex. In fact, if we count the number of degrees of freedom of our interpolation, on the one hand, we have a polynomial of order k on each of the n interval giving $(k + 1)n$ degrees of freedom ($k + 1$ parameters for each

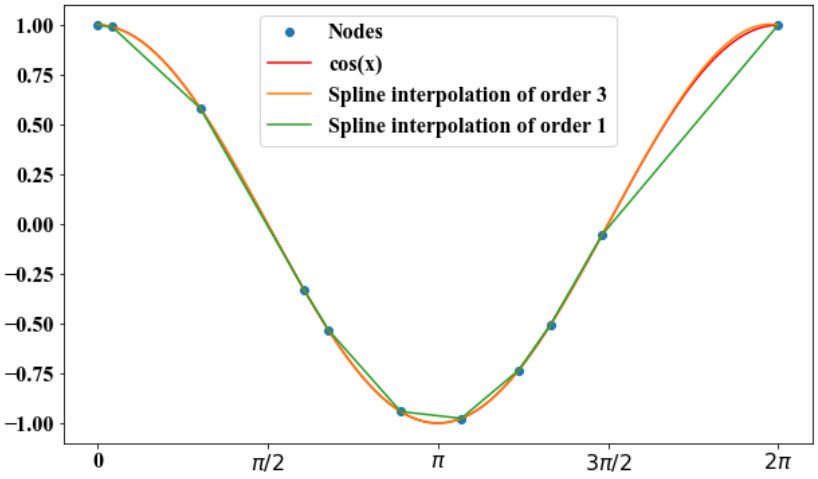


FIGURE A.1 – *The spline interpolation of order 1 and 3 for the function cos(x).*

polynomials, which are defined on the n intervals $[x_j, x_{j+1}]$). On the other hand, the continuity condition gives some constraints at each nodes being :

$$\begin{aligned} &\forall\ l = 0, 1, ..., k \text{ and } \forall\ j = 0, 1, ..., n : \\ &\lim_{x \to x_j} s_{k,j}^{(l)}(x) = \lim_{x \to x_j} s_{k,j+1}^{(l)}(x) \end{aligned} \tag{A.1}$$

where $s_{k,j}(x)$ is the polynomial defined on the interval $[x_j, x_{j+1}]$ and the superscript (l) is the l-th derivative of the function. These constraints are imposing $(n-1)k$ conditions on the coefficients. Leading to $(k+1)n-k(n-1) = k+n$ degrees of freedom based on the given $n+1$ interpolation nodes.

Conditions are thus usually added to the actual definition. These conditions reads as :

$$s_k^{(l)}(a) = s_k^{(l)}(b),\ \forall\ l = 0, 1, ..., k-1 \tag{A.2}$$

which is know as the periodic constraint and :

$$s_k^{(l+j)}(a) = s_k^{(l+j)}(b) = 0\ \forall j = 0, 1, ..., l-2 \text{ and some } l > 2 \tag{A.3}$$

which essentially means that the last $l-1$ derivatives of the spline interpolation have to be null at the extreme points (giving some control of the explosion outside the interpolation interval).

In the next section I will detail the cubic spline properties and application to the energy calibration interpolation.

A.2 . Cubic splines and its properties

Let us focus on the cubic spline application, in this section, $s(x)$ will denote a spline interpolation of the function $f(x)$, based on the given of $f(a) = f(x_0), f(x_1), ..., f(x_n) = f(b)$, at each nodes $a = x_0 < x_1 < ... < x_n = b$. First, let us observe that for each interval $[x_{j-1}, x_j]$, the spline interpolation $s(x)$ second derivative is linear :

$$\begin{aligned} &s^{(2)}(x) = s^{(2)}(x_{j-1})\frac{x_j - x}{x_j - x_{j-1}} + s^{(2)}(x_j)\frac{x - x_{j-1}}{x_j - x_{j-1}} \\ &\text{on the interval } [x_{j-1}, x_j] \end{aligned} \tag{A.4}$$

Integrating this two times gives :

$$\begin{aligned} &s(x) = s^{(2)}(x_j)\frac{(x_j - x)^3}{6(x_j - x_{j-1})} + s^{(2)}(x_{j-1})\frac{(x - x_{j-1}^3)}{6(x_j - x_{j-1})} + c_{j-1}(x - x_{j-1}) + d_{j-1} \\ &\text{on the interval } [x_{j-1}, x_j] \end{aligned} \tag{A.5}$$

Where c_{j-1} and d_{j-1} are imposed by the conditions at x_{j-1} and x_j : $s(x_{j-1}) = f(x_{j-1})$ and $s(x_j) = f(x_j)$. The parameters are then :

$$\begin{aligned} &\forall j = 1, 2, ..., n-1 \\ &c_{j-1} = f(x_{i-1}) - s^{(2)}(x_{j-1})\frac{x_j - x_{j-1}}{6} \\ &d_{j-1} = \frac{f(x_i) - f(x_{i-1})}{x_j - x_{j-1}} - \frac{x_j - x_{j-1}}{6}(s^{(2)}(x_j) - s^{(2)}(x_{j-1})) \end{aligned} \tag{A.6}$$

Let us now use the continuity of the simple derivative of $s(x)$ at the node x_j, we have :

$$\begin{aligned}\lim_{x\to x_j^-} s^{(1)}(x) &= \frac{x_j - x_{j-1}}{6}s^{(2)}(x_{j-1}) + \frac{x_j - x_{j-1}}{3}s^{(2)}(x_j) + \frac{f(x_j) - f(x_{j-1})}{x_j - x_{j-1}} \\ &= \frac{x_{j+1} - x_j}{6}s^{(2)}(x_j) + \frac{x_{j+1} - x_j}{3}s^{(2)}(x_{j+1}) + \frac{f(x_{j+1}) - f(x_j)}{x_{j+1} - x_j} \\ &= \lim_{x\to x_j^+} s^{(1)}(x)\end{aligned} \tag{A.7}$$

Which leads to the following system :

$$\begin{aligned}&\forall j = 1, 2, ..., n-1 \\ &\mu_j s^{(2)}(x_{j-1}) + 2s^{(2)}(x_j) + \lambda_j s^{(2)}(x_{j+1}) = \nu_j \\ &\text{where :} \\ &\mu_j = \frac{x_{j+1} - x_j}{x_{j+1} - x_{j-1}} \\ &\lambda_j = \frac{x_j - x_{j-1}}{x_{j+1} - x_{j-1}} \\ &\nu_j = \frac{6}{x_{j+1} - x_{j-1}}\Big(\frac{f(x_{j+1} - f(x_j))}{x_{j+1} - x_j} - \frac{f(x_j) - f(x_{j-1})}{x_j - x_{j-1}}\Big)\end{aligned} \tag{A.8}$$

This system is still ill-posed as only $n-1$ conditions are imposed over $n+1$ parameters. The usual choice is to impose either $\nu_0 = \nu_n = 0$ leading to the "natural" spline ($s^{(2)}(a) = s^{(2)}(b) = 0$), or to impose $\nu_0 = \nu_1$ and $\nu_{n-1} = \nu_n$ leading to a smoother spline.

The natural spline has the following properties, if $f \in C^2[a, b]$ then :

$$\int_a^b [s^{(2)}(x)]^2 dx \leq \int_a^b [f^{(2)}(x)]^2 dx \tag{A.9}$$

with equality if and only if $f(x) = s(x)$. This essentially means that the spline interpolation is smoother than the function we try to interpolate, in the application of the energy function with respect to the channel number this is a great property to have as we want a smooth interpolation and make the hypothesis that the function of energy calibration is itself reasonably smooth. We can see in figure A.1 that cubic spline are smooth interpolations and that the approximation s is really close to the function we try to reproduce based on the nodes' information.

B - Characteristic limits and confidence interval for the spectral unmixing

Sommaire

In this annex I will present the way the confidence intervals and decision limits are computed in my thesis work. In fact, the mathematical model we use to represent a spectrum allows us to compute the characteristic limits with Monte-Carlo simulation but this kind of computations are long and in a framework of online estimation of the activity it impossible to do the simulations in a time effective manner. The Fisher information matrix wil thus be used to approximate the repartition of the estimator we give to the estimation with a Gaussian distribution. I will present the base work on which the new tools to estimate the uncertainties in a joint analysis framework in a second section. Finally we will see the limits of this approximation when we get to the online estimation and a low number of count in the spectrum to estimate a radionuclide. In this final section we will try to estimate the uncertainties via a truncated Gaussian and compare the analytical results to simulations.

B.1 . Definitions

The first definition we need to understand when we talk about confidence interval and detection limits are the statistical errors that one can make during a estimation process or when making a decision based on statistics. In the statistical framework we try to make a decision between two hypothesis, H_0 and H_1. In this work the two hypothesis are as follows :

— H_0 : The radionuclide is not in the sample (its activity is null), this is called the "null hypothesis".

— H_1 : the radionuclide is in the sample (its activity is > 0), called the "alternate hypothesis"

The statistical framework allows to decide which hypothesis is more likely given the observed data, in our case, is it more likely that the radionuclide is present in the spectrum or not. The two errors thus consist in the following bad decisions :

— Consider that a radionuclide is present when it is not (*ie* wrongly rejecting H_0), this is the type 1 error and we fix the maximum probability of this error α when we create a statistical test discriminating H_0 and H_1.

— Consider that a radionuclide is not in the mixture while it is (*ie* wrongly rejecting H_1), called the type 2 error its probability is β and needs to be the lowest possible given the α-probability of type 1 error.

In fact, the two types of error are bound to one another. Minimising α is allowing β to grow and the converse is also true. If we look at a test that always rejects H_0 or a test we have the following :

$$\mathbb{P}(\text{reject } H_0 | H_0 \text{ is true}) = \alpha \tag{B.1}$$

This is equal to 0 in the case of a test that always rejects H_0, maximising α. On the other hand we have :

$$\mathbb{P}(\text{reject } H_1 | H_1 \text{ is true}) = \beta \tag{B.2}$$

One can see that, with any other test, even one that is completely random and have no correlation with the observe data, if sometimes the null hypothesis is rejected then β will be smaller than the value of β in this extreme case. The same applies to the converse test that always rejects H_0, maximising β lowers the value of α.

It is for this reason that the statistical framework orders the two types of error with respect to the gravity of the consequence of the error. In our case the type 1 error, missing a radionuclide in a sample, can lead to cancer in the population or an accident or incident that is not detected, this is the type of error we want to minimise first. If we now focus on the type 2 error, detecting a radionuclide while it is not present, can lead to a false alarm, it will enhance the surveillance and we will carry other tests to determine if the radionuclide can have an impact or not. In the worst case we will alert on a problem that does not exists, by following the safety principle this error can be minimised in a second time.

Following the extreme example that I have shown above the statistical framework proceeds by fixing a α probability of type 1 error and then optimise the test to reduce the β probability of type 2 error. The ISO norm 11929 ([30]) gives the ISO framework for the determination of the confidence intervals and detection limits in the case ionizing radiation measurement. This text gives the definition of Decision Threshold (DT) and Detection Limit (DL). We will see that these definition are linked to the type 1 and type 2 error that we introduced earlier.

The Detection threshold is defined as the limit DT above which should be the estimation of the activity in order to consider that a radionuclide is present in a sample. Noting $\hat{a}$ the estimation of the activity, we thus have :

$$\begin{aligned}\mathbb{P}(\hat{a} \geq DT | a = 0) &= \mathbb{P}(\text{reject } H_0 | H_0 \text{ is true}) \\ &= \alpha\end{aligned} \tag{B.3}$$

The Detection Limit is defined as the type 2 error limit. It is the smallest observable activity $a = DL$ such that the test with a given DT would reject the hypothesis of the presence of a radionuclide in the spectrum. In other terms we have :

$$\begin{aligned}\mathbb{P}(\hat{a} \leq DT | a = DL) &= \mathbb{P}(\text{reject } H_1 | H_1 \text{ is true}) \\ &= \beta\end{aligned} \tag{B.4}$$

This notion of detection limit is not used in this thesis preferring the decision threshold as the lowest observable activity given a measurement.

Finally a confidence interval for a given probability p, is an interval surrounding the estimation $\hat{a}$, based on the statistic distribution of the value we want to measure. In other terms it reads as :

$$\mathbb{P}(a \in [low, top] | \hat{a}) = p \tag{B.5}$$

as will be seen in the next section, we will often consider the activity estimation for a given radionuclide to follow a Gaussian distribution centered around the estimation $\hat{a}$ with a certain variance σ^2 that we have to determine. The confidence interval will then be taken as $\hat{a} \pm k\sigma$. It is usual to take $k = 2$ leading to $p = 09545$ in the previous equation.

B.2 . Monte-Carlo simulation for confidence intervals

The mathematical model of a gamma-ray spectrum we use that allows us to perform the spectral unmixing may be used to simulate any mixture of radionuclide at any activity we want. This allowed us to assess the performance of the temporal spectral unmixing via Monte-Carlo simulation and in this section we will show how the same simulation framework can be used to determine the confidence interval and decision threshold of the activity estimation. Let us first recall the mathematical model of a gamma-ray spectrum :

$$
\begin{aligned}
x &= \sum_{n=1}^{N} \phi_n \psi_n a_n + \delta b \\
y &= Poisson(x)
\end{aligned}
\tag{B.6}
$$

where ψ_n and δ would be defined as in chapter 2 for a given measurement from t_0 to t_1. We thus consider a mixture of N radionuclide and we want to compute a confidence interval of the estimated activities $a_1, ..., a_N$ and the decision threshold for each radionuclide.

B.2.1 . Confidence interval computation

For the computation of the confidence interval around the estimated activity a_n of a given radionuclide we will perform Monte-Carlo simulations of the estimated mixtures. The simulation will build a model spectrum $\hat{x}$ with the estimated activity estimation $\hat{a}_1, ..., \hat{a}_N$ and the Poisson noise will be applied to this base spectrum in order to simulate random spectra mimicking the observed one. This leads to :

$$
\hat{x} = \sum_{n=1}^{N} \phi_n \psi_n \hat{a}_n + \delta b \tag{B.7}
$$

The Poisson noise allows us to access to multiple repetition of the same base spectrum with different noise. Performing the spectral unmixing on these repetition will give us the statistical noise of the estimation, in other terms, the response of the estimator of the activity subject to a statistical Poisson noise on the spectrum. This allows to get the statistical uncertainty shown in chapter 2 and 3. In figure B.1 are the result of the Monte-Carlo simulation process for 2 radionuclides, the ^{7}Be and the ^{137}Cs. We can see that both activity estimation at low and high level, follow a Gaussian distribution. Allowing us to get the variance of the estimator and estimate the proper 2σ confidence interval.

To this statistical uncertainty is added some other sources of uncertainties that we cannot simulate or are more complex to estimate properly. This uncertainty can be expressed in term of percentage of the estimated activity by the following equation :

$$
u_{stat} = 2\sigma / \hat{a} \tag{B.8}
$$

Throughout this thesis we have called these sources of uncertainties the

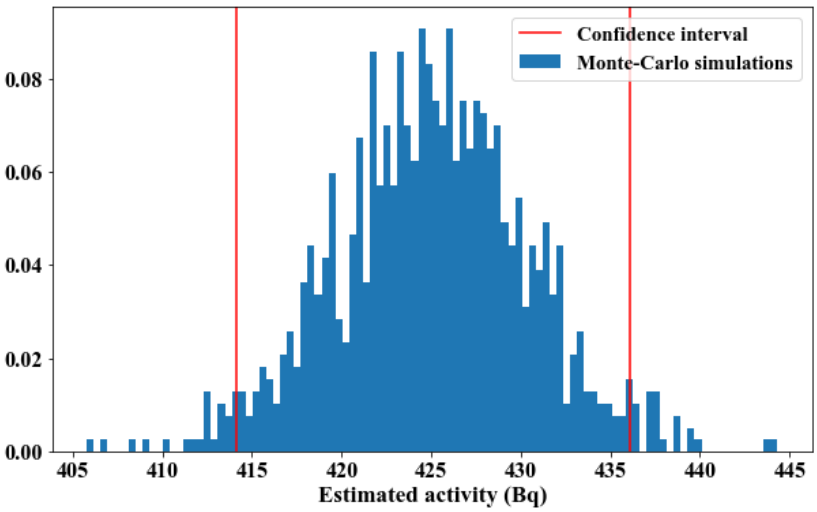


FIGURE B.1 – *The result of 1000 Monte-Carlo simulations for the estimation of* ^{7}Be *and the selected quantiles for the 95% confidence interval.*

metrological uncertainty. The components of these uncertainties are as follows :

— the use of simulated spectral signatures rather than sources
— the position of the sample on the detector
— the detector's variation
— the sample's variation

These sources of variability are considered to be equal to 10% of the activity in term of uncertainties. The total uncertainty on the activity estimation will then be :

$$u_{tot} = \sqrt{u_{stat} + u_{metro}} \tag{B.9}$$

B.2.2 . Decision threshold computation

The computation of the decision threshold in the Monte Carlo framework is similar to the computation of the confidence interval. In fact the mathematical definition of DT as seen in the previous section (B.3) leads us to compute the same simulations as for the confidence interval but considering that the n-th radionuclide's activity is equal to 0.

The estimation of the activity on these simulations will give us a representation of the estimator in the case of an absent radionuclide which is exactly what we defined in the definition according to the ISO norm 11929.

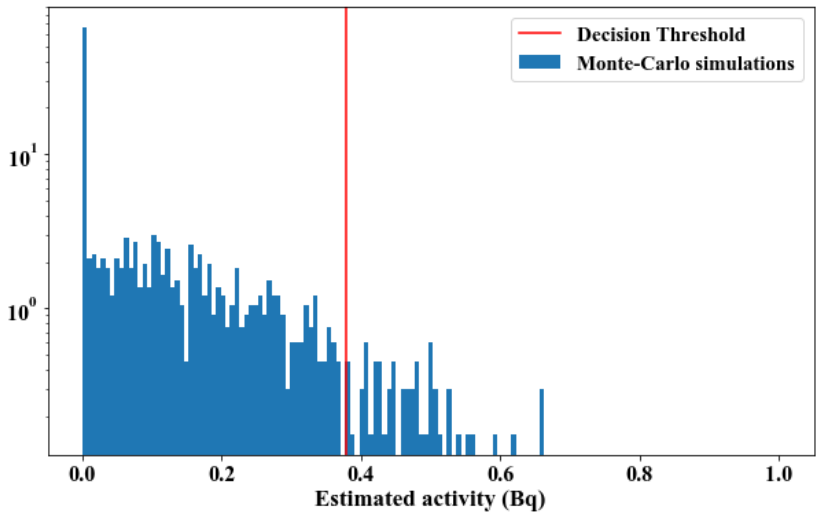


FIGURE B.2 – *The simulation of the estimation of ^{137}Cs on 1000 repetitions where the simulated activity of this radionuclide is null. The log-scale is used because a lot of estimations are at 0. The decision threshold is the empirical 95%-quantile.*

Suppose we want to quantify the DT for the k-th radionuclide. We will thus consider Monte-Carlo simulation of the form :

$$\begin{aligned} \hat{x}_k &= \sum_{n=1}^{N} \phi_n \psi_n \hat{a}_n + \delta b - \phi_k \psi_k \hat{a}_k \\ &= \sum_{n \neq k} \phi_n \psi_n \hat{a}_n + \delta b \end{aligned} \tag{B.10}$$

The spectrum considering that the contribution of the k-th radionuclide is null. The Poisson noise is applied to give us the Monte-Carlo repetition and allowing us to have the decision threshold as the $1 - \alpha$ quantile of the repetitions.

The major issue with the Monte-Carlo simulation to obtain the confidence interval and the decision threshold is the time consuming computation. In fact, in order to get proper statistical significance we want to carry a lot of simulations (1000 simulations were performed throughout this work) and the unmixing of each of the simulated spectrum takes time. In the case of the confidence interval the simulations are carried for the determination of the interval of every radionuclide but in the case of the decision threshold we have to repeat the simulations for each radionuclides for which we want to have a value. In this thesis the only radionuclide which activity is near the decision

threshold is the ^{137}Cs and thus is the only one for which the simulations have been performed.

In the next section we will present another way to perform the computation of the confidence interval using the Fisher information matrix.

B.3 . Uncertainty quantification with the Fisher information matrix

Rather than relying on heavy computation for the determination of the confidence interval we can rely on the Fisher information. In fact following Fisher ([31]) we have the following property : the distribution of the maximum likelihood estimator can be approximated by a Gaussian distribution

$$\mathcal{N}(\theta, I(\theta)^{-1}) \tag{B.11}$$

where θ is the parameter we want to estimate and $I(\theta)$ is the Fisher information. The Fisher information is defined as :

$$I(\theta) = \mathbb{E}_\theta\Big[\frac{\partial^2 log(f(x|\theta))}{\partial^2 \theta}\Big] \tag{B.12}$$

With $f(x|\theta)$ being the distribution of the observed values $x_1, x_2, ...$ with the parameter θ. The θ value can be evaluated using its estimator's value $\hat{\theta}$.

In the spectral unmixing framework θ the parameter of the distribution of the observed values $x_1, x_2, ...$, are the activities $a_1, a_2, ..., a_N$ of the N radionuclides composing the observed spectrum, the distribution of the observed values is :

$$y = Poisson(\Phi(\Psi.a) + \delta b) \tag{B.13}$$

Considering that the probability for a random value X following a Poisson distribution of parameter λ is :

$$\mathbb{P}_\lambda(X = x) = \frac{e^{-\lambda}\lambda^x}{!x} \tag{B.14}$$

the content of each channel of a spectrum being considered independent from one another we have, noting $y = y_1, ..., y_C$ the observed spectrum, $x = x_1, ..., x_C$ the random representing the content of each channel and $\hat{x} = \hat{x}_1, ..., \hat{x}_C$ the estimated spectrum :

$$\begin{aligned} \mathbb{P}(x = y) &= \Pi_{c=1}^{C} \mathbb{P}_{\hat{x}_c}(x_c = y_c) \\ &= \Pi_{c=1}^{C} \frac{e^{-\hat{x}_c} \hat{x}_c^{y_c}}{!y_c} \end{aligned} \tag{B.15}$$

Taking the derivative of the log of this expression with respect to each $a_1, ..., a_n$ leads to :

$$\begin{aligned} &\forall k = 1, ..., N \\ &\frac{-\partial^2 log(\mathbb{P}(x = y))}{\partial a_k^2} = \sum_{c=1}^{C} -\frac{y_c(\phi_k \psi_k)^2}{\hat{x}_c^2} \end{aligned} \tag{B.16}$$

The Fisher information being the opposite of this expression. Following the approximation given earlier (equation B.11) we have a way to efficiently compute the statistical confidence interval of the activity estimation in the framework of the spectral unmixing.

In [20] Jiaxin Xu has shown comparison between the Fisher approximation and Mote-Carlo results proving the validity of this approximation for large activities or after a long time has passed and the radon progeny have decayed. In figure B.3 is an example showing the pertinence of this approximation. As can be seen the Monte-Carlo and Fisher approximation lead to some confidence intervals that are comparable. In the results of simulation and real aerosol filter samples analysis of the chapter 2 and 3 we have further improved the comparison with the added temporal component. In the following section we will demonstrate how to adapt the Fisher approximation to the joint analysis of consecutive measurement.

It is important to note that for the computation of the decision threshold the approximation using Fisher is not viable, in fact, we have a truncated estimator of the activity in the extreme regime of a null activity. In other words, the fact that our estimator of the activity is always positive leads to a non-Gaussian behavior of the estimator when the estimated activity is low. This will be further demonstrated in the following section wherein the temporal estimation leads to some extreme behaviors of the estimator.

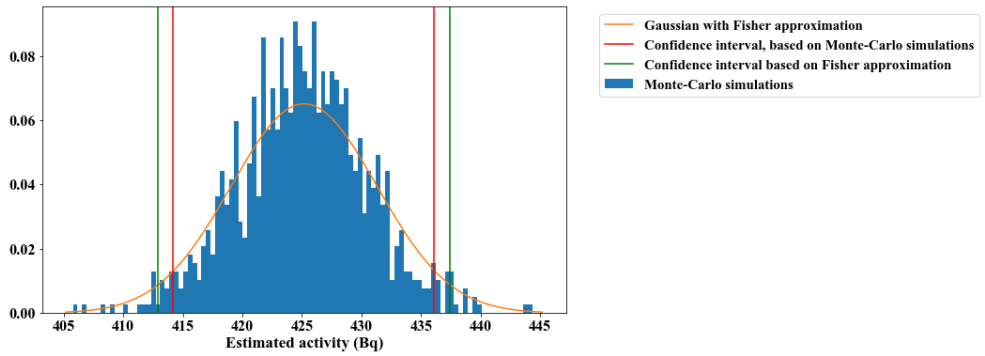


FIGURE B.3 – *The result of the estimation of ^{7}Be of 1000 Monte-Carlo simulations and the corresponding Fisher Gaussian approximation. The confidence interval using both the Monte-Carlo and Fisher methods are shown.*

B.4 . Confidence interval and decision threshold in the joint estimation framework

In this section I will present how to build the confidence interval in the temporal frameworks of the chapter 2 and 3 from the Fisher information procedure presented above. But in a first time I will present the Monte-Carlo simulation process one would have to do if we were to construct the confidence interval this way. In the final section we will look at the decision threshold and compare the results of Fisher and simulation in extreme cases to show the limits of the Fisher approximation in the case of temporal unmixing.

B.4.1 . Monte-Carlo simulations

In the temporal analysis framework presented in this thesis, the same procedure of Monte-Carlo simulation can be applied to build the confidence interval. The only difference lies in the fact that, rather than considering a single spectrum, we simulate collections of consecutive spectra.

During the simulation process we will thus mimic the consecutive spectrum acquisition by simulating the S time segments independently and perform the unmixing beginning with the first segment alone, then the first 2 segments and so on. This leads to multiple repetition of the joint analysis of consecutive measurement of a sample.

The same applies to the online framework, by using the buffer presented in 3 we are able to perform the simulations with reasonably small matrices even if the time steps are small. The main issue of the Monte-Carlo simulation being the computation time. In fact, the adding of the temporal dimension to the simulation leads to a multiplication of the computation time by the number of time steps we take. Even with the buffer and algorithmic optimization added

in chapter 3 the computation time is a major issue for the computation of the confidence interval and the decision threshold.

B.4.2 . Fisher's approximation in the temporal framework

To cope with the temporal segmentation and the increasing of the number of simulations needed in the Monte-Carlo framework, the Fisher approximation has been extended to the joint analysis of multiple consecutive gamma-ray spectra. In fact by looking at the equation B.15 we can see that we can apply the same equation with multiple gamma-ray spectra that depend on the same activities $a_1, ..., a_N$. The following derivations are the same leading to the new Fisher information :

$$\begin{aligned} &\forall k = 1, ..., N \\ &\frac{-\partial^2 log(\mathbb{P}(x_1, ..., x_S = y1, ..., y_s))}{\partial a_k^2} = \sum_{c=1}^{C} \sum_{s=1}^{S} -\frac{y_{cs}(\phi_k \psi_{ks})^2}{\hat{x}_{cs}^2} \end{aligned} \tag{B.17}$$

where $x_1, ..., x_S$ are the consecutive spectra model, $y_1, ..., y_S$ are the observed realisations and $\hat{x}_1, ..., \hat{x}_S$ are the estimated spectra using the estimation of the activity.

The same applies to the online buffer presented in chapter 3. In fact, replacing S in the above equation with r the size of the buffer, and using the Ψ buffer instead of the Ψ matrix used previously leads to the correct Fisher's information and approximation of the variance of the estimation of the activities.

But some issues arise when we deal with low activity for which a low number of counts is present in the measured spectrum (*eg* ^{137}Cs for a 2 minute measurement). In fact the estimation is so low that the approximation using a Gaussian distribution is no more valid as can be seen in figure B.4. In fact, the approximation using Fisher's information to compute the standard deviation of the estimator cannot cope with such low activity in a framework where little to no information are present in the spectra. In fact, the low activity of ^{127}Cs, when compared to the high background and other radionuclides contribution, hampers the estimation of its activity as seen in chapter 2 and 3.

As a result, the only solution leading to some correct estimation of the uncertainties and decision threshold, in the case low level of activity is the Monte-Carlo simulation method. The only issue being the computation time required in order to get enough statistics to give a correct result. In the online framework it is mandatory to get quicker ways of estimating the uncertainties.

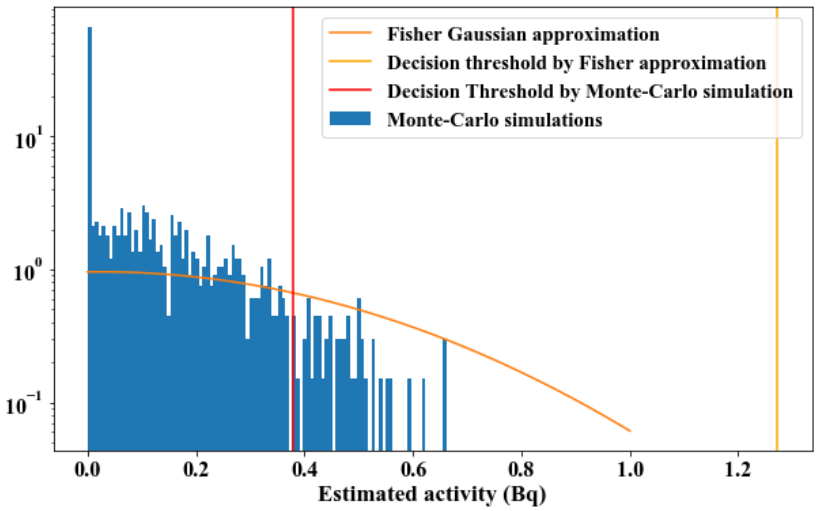


FIGURE B.4 – *Comparison between the Fisher approximation and the 1000 Monte-Carlo simulation for ^{137}Cs decision threshold and uncertainty estimation. The simulated activity of ^{137}Cs is 0.002 Bq.*

Some work were done to estimate the distribution of the Monte-Carlo simulations using a truncated Gaussian distribution. The results can be seen in figure B.5 and shows that a truncated Gaussian may lead to some correct estimation of the uncertainties and decision threchold given the concordance between the estimated distribution and the Monte-Carlo simulation distribution. Further investigation need to be performed in order to give a heuristic or synthetic model to the mean and variance of this truncated Gaussian Model.

B.5 . Conclusion

The rapid estimation of the uncertainties and decision threshold is a difficult problem. For the time being, the only reliable way to compute these two quantities is to perform Monte-Carlo simulations which is a long and computationally heavy process. This is not compatible with the Online framework developed in this thesis.

Some new ways to compute these quantities can rely on Truncated Gaussian distributions to estimate the theoretical distribution of the Monte-Carlo simulations. These estimation yet need further investigation to give satisfactory results.

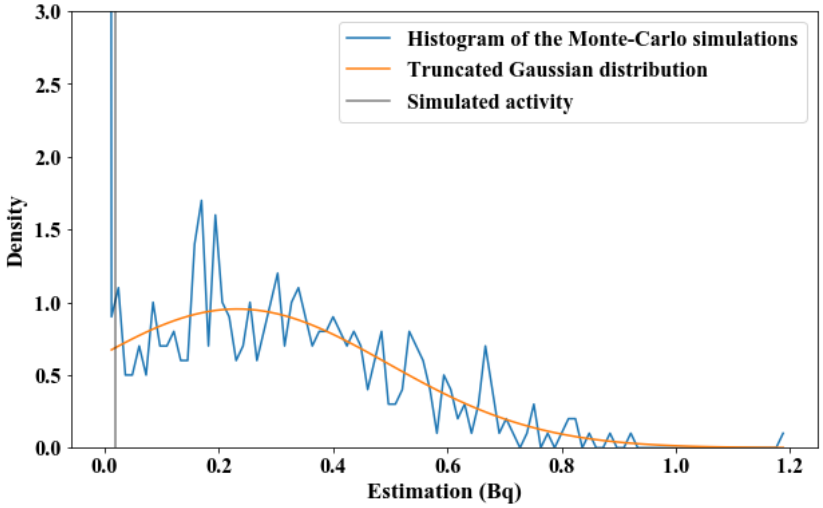


FIGURE B.5 – *The estimation of the distribution of the estimator using a truncated Gaussian. The 1000 Monte-Carlo simulation shows the estimation with a simulated activity of 0.002 Bq of* ^{137}Cs.

C - MCNP-CP simulation

Sommaire

C.1 . Description of MCNP

The simulations of the spectral signatures used throughout the thesis is done via the Monte Carlo N-Particules transport code (MCNP). This program allows to simulate the trajectory of a particle (in our case photons) in the mater. More precisely the interaction between the particle and the mater of the simulated model can be precisely followed. These codes were developed by the Los Alamos National laboratory.

The record of the path and interaction can be generated for various configurations and various outputs, the one that interests us the most is the tally F8 that represents the spectrum of the simulated particle in a specified material (in our case the Ge detector). This tally is essentially the normalized histogram of energy deposited in the detector for a given simulated particle. The simulation process is very precise and in case of physical phenomena te effects will be simulated as well. For example, in case of pair production of electrons the simulated spectrum will contain the 511 keV peak along with the peak at the nominal energy E and the peak at E-511 keV.

A extension to this program is the MCNP-CP codes (Correlated Particles) that allows to generate an atom rather than a particle. This allows to get the entire disintegration scheme with respect to the Evaluated Nuclear Structure Data File (ENSDF). For example for the simulation of ^{208}Tl, rather than simulating every photon according to its nominal energy, we only simulate the disintegration of a ^{208}Tl atom and get the proper spectral signature of the disintegration as a result. The ENSDF encapsulate the disintegration scheme, emission intensity, energy and joint emission for the radionuclides that we observe in the sample of the laboratory.

As viewed in 4 the model of the detectors of the laboratory are done in two steps. Firstly, we create the model of the detector according to the manufacturer standards as a base for the model. Secondly, thanks to a multi-gamma source, we tweak the model of the detector to get the correct efficiency calibration at each peak of the standard source spectrum. The model is then used to simulate the spectral signatures that I used in my thesis work.

In a first time the Gaussian Energy Broadening were computed via the analysis of the same multi-gamma source spectrum and given to the simulation code to generate the broadening. After the development of our own algorithm to perform the broadening we used the MCNP-CP codes with 0 GEB specification.

The generated spectrum is really close to the output of the detector at

the only detail that the energy bin is very regular in the simulation. In fact we have to rebin the spectrum or the simulated spectral signature to get the proper binning (see 4 for more information).

In the next section I will detail an input file for the simulation and detail the important parameters for the study presented in this thesis. In a second section, I will present the comparison between simulation and source measurement to see the effect of the different parameters and assess the performances of the simulations to reproduce the correct spectral signatures.

C.2 . Example and important parameters for our study

The input file for MCNP-CP for one of our detector proceeds as follows :

— The construction of the model based on simple geometric objects (mostly planes, and cylinders), precising the composition of each elements.

— The physical data, mainly the mode of simulation P, E or PE (for photo, electron or both). Finally a list of flags for various options in the IMP line, such as the decay gamma-ray emission, the beta-particle emission, the gamma-gamma angular correlation, ...

— The definition of the emitting source and the emitted photon or photons, in fact as stated already, MCNP-CP allows to simulate not only photons of a given energy but a particle, and its entire disintegration scheme. The emission is defined in the SDEF line with the particle's ZAM number (Z = atomic number on 3 digits, A mass number on 3 digits as well and finally the isometry M as a flag 0 or 1), followed by the emitting cell and the distribution of the position of the disintegrations within the given cell (in our case we simulate a uniform distribution inside the aerosol filter sample). The only parameter used in my thesis is the GLECS for which we will detail the effect with the example.

— The definition the GEB parameters (a, b and c of the Gaussian broadening in term of FWHM as seen in section 4), if the spectral signatures are to be simulated as perfect Dirac we set the parameters of the GEB at 0.

— The definition of the output we need and the cell in which we want the output to be followed, for our example we look at the spectrum in the Ge crystal (F8 in the cell defined as the Ge crystal).

— The definition of the materials used in the model (Ge for the crystal,

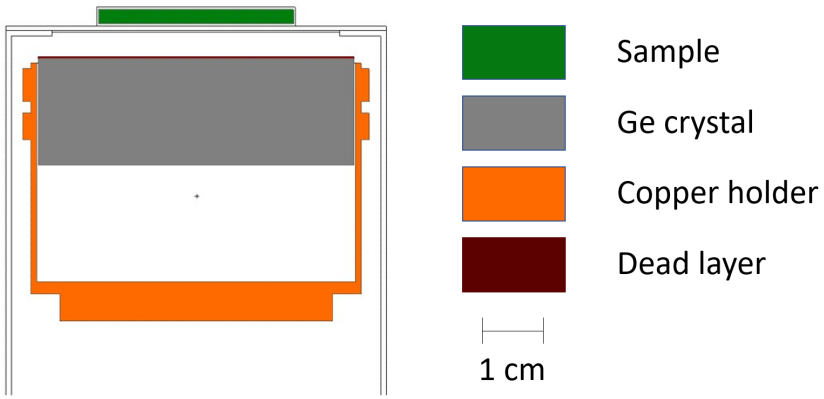


FIGURE C.1 – *On top, the model of the detector used in chapter 2.*

copper shielding, lead, ..

— Finally, the NPS input is the number of particles that are simulated, the more the longer the simulation will be but the better the results will be as for any Monte-Carlo simulation.

In the example input we see the different elements mentioned (C.2). Once the input file is properly defined a Python code can be used to generate batches of simulations for different radionuclides only changing the ZAM number in the input file. The model of the detector we use is shown in figure C.1 along with the realisation of the trajectory of a few particles.

C.3 . Comparison between the simulations and a source measurement

The different options of the physical specifications has been explored during my thesis. Mainly the GEB option, this parameter is linked to the resolution of the spectrum and the calibration of this parameter is no more performed during the simulation process as seen in chapter 4. As can be seen in figure C.3 this parameter mainly affects the peaks by applying a Gaussian broadening. If no GEB parameters are specified or if the 3 parameters are set to 0, then the simulated spectrum will have infinite resolution, the peaks will be simulated as Dirac peaks at the nominal energies (orange and blue in the pictures).

The second parameter that will impact the spectral signature simulation is the GLECS (GEANT Low-Energy Compton Scattering) parameter. This will mainly impact the compton edges of the spectrum as can be seen in the last picture of figure C.3 in blue. The activation of the GLECS effect in the

```
simulation detector 61C

C Cells
1 -5.323 6 -12 -25  $ Ge Crystal
7 -0.77 18 -19 -21  $ source geometry (aerosol)
[...]

C Surfaces
PZ    0.15   $ Endcap material thickness = 1.5 mm
PZ    9.9030 $ dead layer thickness = 50 µm
[...]

C Physical specifications
Mode  P
IMP:P 1 1 1 1 1 1 1 1 1 1 1 1 0
PHYS:P 100 0 0 0 0       $ GLECS option with 1 1
SDEF  CEL=10 RAD=D1 EXT=D2 AXS=0 0 1 ZAM=0551370 PAR=2
SI1   0 2.5
SP1 -21 1
SI2   10.87 11.246
SP2 -21 0
CPS 50 1 1 1 1 1 1 1 1
F8:P  1
FT8 GEB 0.0003406 0.000865 0.51115   $ Infinite resolution by putting only zeros
E8   0 1E-5 1E-3 0.02 17999i 1.9200       $ Definition of the binning
CUT:P 2J 0 0

C Material definition
32000 1                  $ Germanium (5.33 g/cm3)
29000 1                  $ Copper (8.96 g/cm3)
[...]

NPS   1E8
```

FIGURE C.2 – *An example of MCNP-CP input, the principal parameters on which we focused in the thesis are in color.*

simulation leads to a more realistic simulation of the spectra in fact as can be seen in the measurement of the ^{137}Cs source in figure C.4 the Compton edges are not sharp as in the orange and green simulations but more alike the blue simulation.

As a result in this thesis and thank to the new resolution calibration the spectral signatures are simulated with infinite resolution and the GLECS option activated (as the blue simulation in figure C.3).

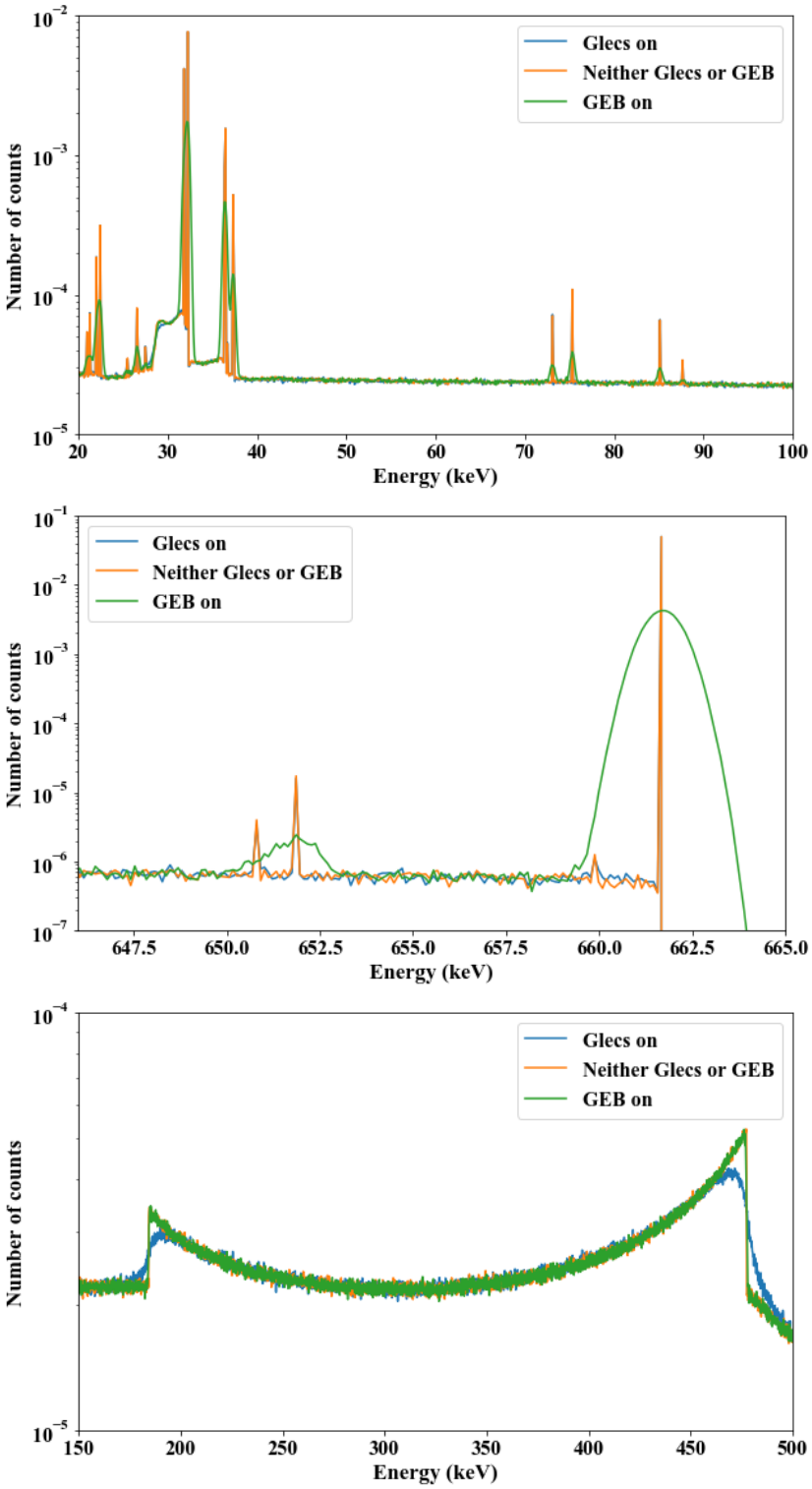


FIGURE C.3 – *The spectral signature of* ^{137}Cs *with varying entry parameters, zoom on the low energies, the full absorption peak and the compton continuum features.*

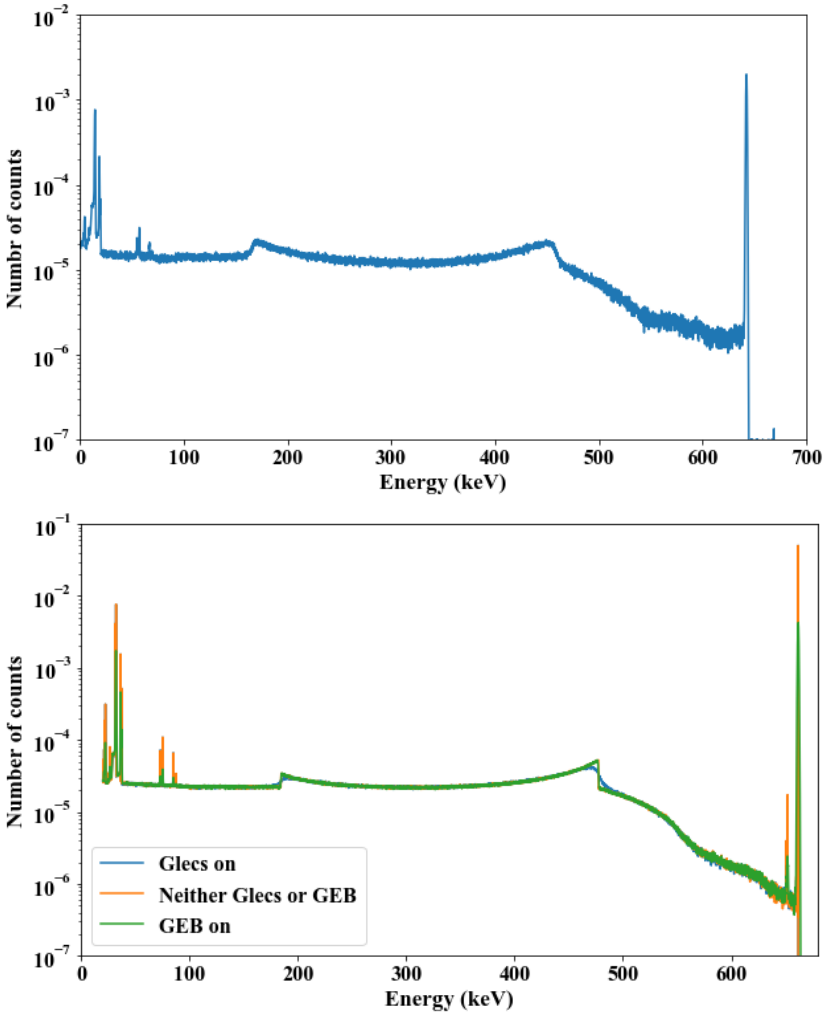


FIGURE C.4 – *On top, the measurement of a standard source of* ^{137}Cs *compared to the simulation on the bottom.*

C.3.1 . Detailed comparison between Genie 2000 and Full spectrum analysis

In this thesis, we showed the results of the spectral unmixing alone. In this annex we show a study performed at the beginning of my thesis and published in [20] that compares the results of the peak-based analysis using Genie 2000 [7] software and the results by spectral unmixing. In this study we have the perfect setting to compare the performance of both methods as standard source are available for the 3 radionuclides that are present in the sample we analyse.

In this study, we analyse four synthetic standard sources with known activity concentrations of three radionuclides : ^{210}Pb, ^{137}Cs and ^{54}Mn. To keep the problem somewhat similar to the aerosol filter spectra analysis, ^{210}Pb has an activity level that is customarily observed in aerosol filters (around 100 Bq), while ^{137}Cs and ^{54}Mn have been chosen as substitutes for ^{7}Be and ^{137}Cs with similar energy ranges (662 keV for 477 keV, and 835 keV for 662 keV, respectively) and activity levels. The ^{137}Cs activity level is thus around 1000 Bq. The activity of ^{54}Mn varies in the four sources from 18 mBq to 4.9 Bq. It is important to note that even the lowest activity in our samples is ten times over the activity we usually have in the environment for ^{137}Cs. Unfortunately it was not possible to have sources with certified activities of ^{54}Mn around few mBq, which would correspond to the lowest level of ^{137}Cs in aerosol filters. The spectra of the four mixing sources measured with the studied HPGe are shown in figure **??** with a zoom on the energy range around the characteristic peak of ^{54}Mn. This highlights that this radionuclide is hardly visible for the lowest activity concentration.

We used both Genie 2000 and the spectral unmixing algorithm to analyse the measurement spectra of the four mixing sources. The results are reported in Table C.1 with activities (in Bq) and associated uncertainties (in Bq and for $k = 2$) for the 3 radionuclides in the 4 mixing sources : reference values, values estimated by Genie 2000 and values estimated by the spectral unmixing.

It is important to note that :

— The estimation made with Genie 2000 and spectral unmixing are similar for high activity radionuclides, as testified by the results obtained for ^{137}Cs and ^{210}Pb (Fig.C.5) and the mixtures 3 and 4 (denoted mix 3 and 4 in the figures) that have the highest levels of ^{54}Mn (Fig.C.6).

— As shown on Fig. C.5 the spectral unmixing estimation is more accurate and more precise than Genie 2000 estimation for ^{210}Pb with estimated value more accurate (ratio close to 1) and uncertainties smaller due to

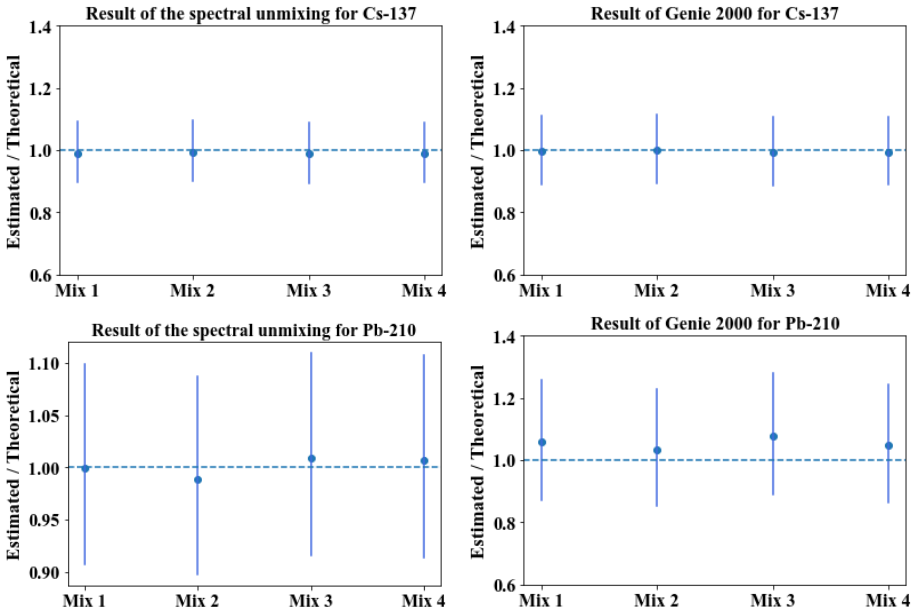


FIGURE C.5 – *Ratios between the estimated activities and the reference valuers with uncertainties (k=2). Results are presented for both analysis methods (spectral unmixing on left and Genie 2000 on right) for both radionuclides present at highest levels :* ^{137}*Cs (top) and and* ^{210}*Pb (bottom) at* 1000 *Bq and* 100 *Bq respectively.*

the fact that the information of the full spectrum is used in activity estimation procedure.

— Activity estimation at low level is more challenging ; however, the spectral unmixing have better performances than Genie 2000 in this case. Indeed ^{54}Mn is detected even at the lowest level (18 mBq) and the estimation for the second lowest level (49 mBq) is better than the one made with Genie 2000 (see Fig.C.6).

— The activity of ^{54}Mn is estimated properly by the spectral unmixing for the first mixture (in which the activity of ^{54}Mn is 17.96 ± 0.76 mBq) while it is not detected by Genie 2000, which classifies this radionuclide as having activity under the detection limit (38 mBq).(see Fig.C.6).

— At low level activity (like the one of ^{54}Mn in mixtures 1 and 2 the information contained in the peaks is not sufficient to estimate the activity of the radionuclides properly. Full spectrum analysis is therefore an appealing solution to this problem in this regime.(see Fig.C.6).

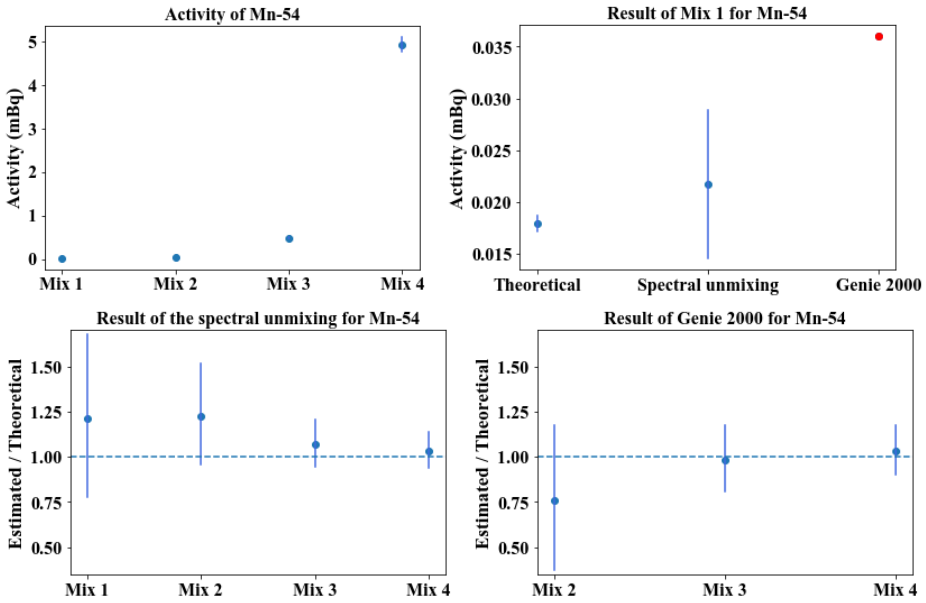
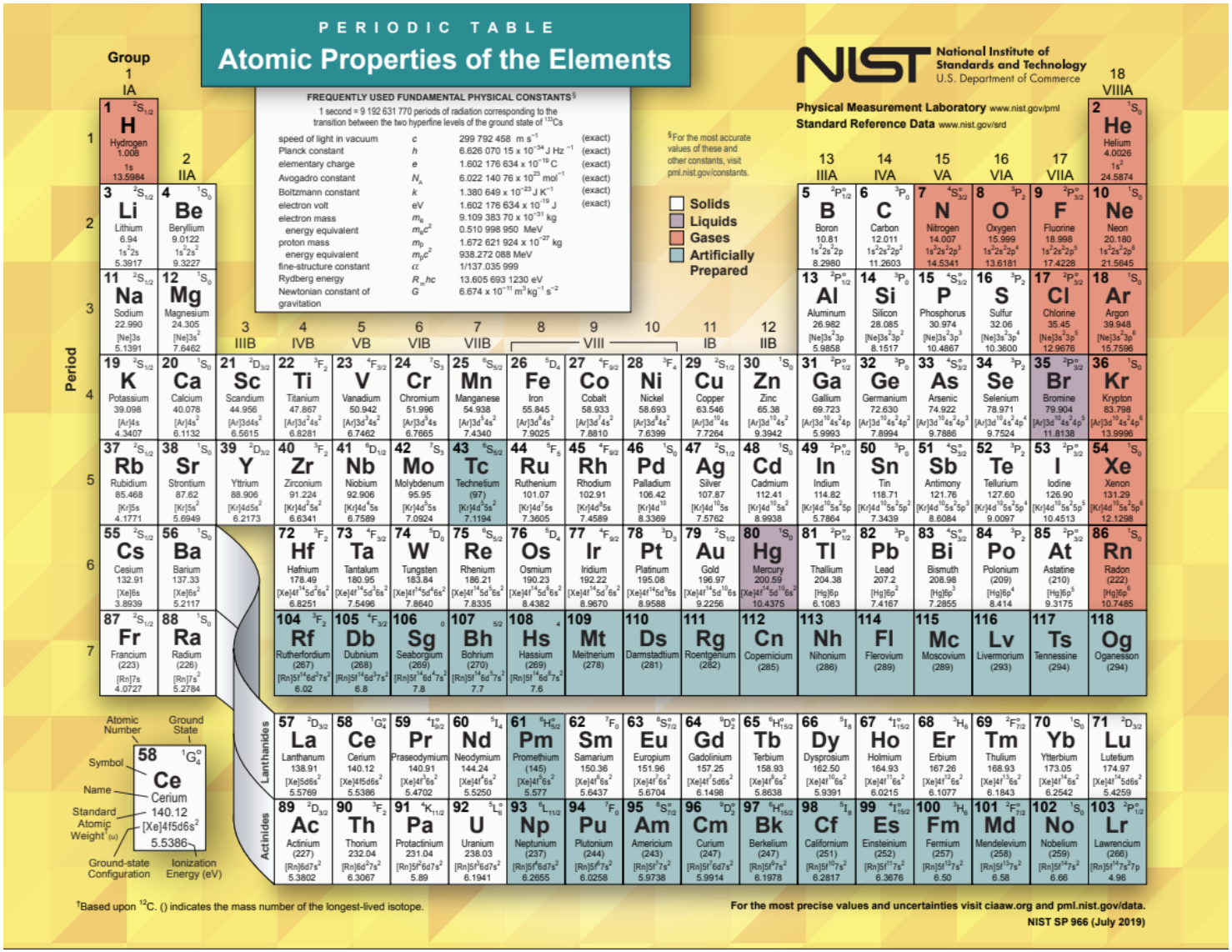


FIGURE C.6 – *Results of spectral unmixing (bottom-left) and Genie 2000 (bottom-right) on the different sources for* ^{54}Mn *with varying activities (top-left). Results for the first source with very low activity (18mBq) are compared on the top-right figure with the reference value, the value estimated by spectral unmixing and the detection limit (38 mBq) calculated by Genie 2000 following the ISO 11929 standard* ($\alpha = \beta = 0.025$)*, in red.*

Mix.	Radionucl.	Reference	Genie 2000	Spectral unmixing
1	Cs-137	1026 ± 39	1022 ± 73	1017 ± 61
	Pb-210	95.9 ± 3.4	101 ± 15	95.8 ± 5.8
	Mn-54	0.01796 ± 0.00076	< 0.038 (DL)	0.02172 ± 0.0071
2	Cs-137	1014 ± 39	1013 ± 72	1008 ± 60
	Pb-210	101.1 ± 3.5	105 ± 15	100.0 ± 6.1
	Mn-54	0.0493 ± 0.0021	0.037 ± 0.018	0.060 ± 0.011
3	Cs-137	993 ± 38	985 ± 70	981 ± 59
	Pb-210	94.4 ± 3.3	102 ± 15	95.3 ± 5.8
	Mn-54	0.485 ± 0.020	0.476 ± 0.069	0.520 ± 0.041
4	Cs-137	1030 ± 39	1024 ± 73	1019 ± 61
	Pb-210	99.8 ± 3.5	105 ± 15	100.5 ± 6.1
	Mn-54	4.92 ± 0.17	5.08 ± 0.51	5.08 ± 0.31

TABLE C.1 – *Activities and associated uncertainties (k=2) in Bq : certified values of the standard sources (Reference), estimated values by peak based analysis (Genie 2000) and spectral unmixing algorithm. DL refers to the detection limit calculated by Genie 2000 following the ISO 11929 standard ($\alpha = \beta = 0.025$).*


PERIODIC TABLE
Atomic Properties of the Elements
NIST National Institute of Standards and Technology U.S. Department of Commerce
Physical Measurement Laboratory www.nist.gov/pml
Standard Reference Data www.nist.gov/srd
FREQUENTLY USED FUNDAMENTAL PHYSICAL CONSTANTS§
1 second = 9 192 631 770 periods of radiation corresponding to the transition between the two hyperfine levels of the ground state of 133Cs
speed of light in vacuum c 299 792 458 m s−1 (exact)
Planck constant h 6.626 070 15 x 10−34 J Hz−1 (exact)
elementary charge e 1.602 176 634 x 10−19 C (exact)
Avogadro constant NA 6.022 140 76 x 1023 mol−1 (exact)
Boltzmann constant k 1.380 649 x 10−23 J K−1 (exact)
electron volt eV 1.602 176 634 x 10−19 J (exact)
electron mass me 9.109 383 70 x 10−31 kg
energy equivalent mec2 0.510 998 950 MeV
proton mass mp 1.672 621 924 x 10−27 kg
energy equivalent mpc2 938.272 088 MeV
fine-structure constant α 1/137.035 999
Rydberg energy R∞hc 13.605 693 1230 eV
Newtonian constant of gravitation G 6.674 x 10−11 m3 kg−1 s−2
§For the most accurate values of these and other constants, visit pml.nist.gov/constants.
Solids
Liquids
Gases
Artificially Prepared
Group
Period
Lanthanides
Actinides
Atomic Number
Symbol
Name
Standard Atomic Weight†(u)
Ground-state Configuration
Ground State
Ionization Energy (eV)
†Based upon 12C. () indicates the mass number of the longest-lived isotope.
For the most precise values and uncertainties visit ciaaw.org and pml.nist.gov/data.
NIST SP 966 (July 2019)

Bibliographie

[1] MR Bhat. « Evaluated nuclear structure data file (ENSDF) ». In : *Nuclear Data for Science and Technology: Proceedings of an International Conference, held at the Forschungszentrum Jülich, Fed. Rep. of Germany, 13–17 May 1991*. Springer. 1992, p. 817-821.

[2] Laboratoire National Henri Becquerel. *Lara web*. URL : http://www.lnhb.fr/Laraweb/.

[3] G. F. Knoll. *Radiation Detection and Measurement*. 4th. John Wiley & Sons, 2010.

[4] A. Magré. « Développement de l'analyse isotopique 135Cs/137Cs adaptée aux sols et aux sédiments faiblement contaminés pour identifier l'origine du radiocésium dans l'environnement ». Theses. Université Paris-Saclay, mars 2023. URL : https://theses.hal.science/tel-04058451.

[5] H. Paradis. « Développement de la mesure par spectrométrie gamma en coïncidence ». Theses. Université Paris Saclay (COmUE), sept. 2016. URL : https://theses.hal.science/tel-01376229.

[6] J. Xu. « Development of analysis tools for gamma-ray spectrometry ». Theses. Université Paris-Saclay, nov. 2020. URL : https://theses.hal.science/tel-03080461.

[7] Mirion-Canberra. *Genie 2000*. Version 3.4.1. Nov. 2016. URL : https://www.mirion.com/products/genie-2000-gamma-analysis-software.

[8] J. M. Kirkpatrick et B. M. Young. « Poisson Statistical Methods for the Analysis of Low-Count Gamma Spectra ». In : *IEEE Transactions on Nuclear Science* 56.3 (juin 2009), p. 1278-1282. ISSN : 0018-9499. DOI : 10.1109/TNS.2009.2020516.

[9] M. Hýža et P. Rulik. « Low-level atmospheric radioactivity measurement using a NaI (Tl) spectrometer during aerosol sampling ». In : *Applied Radiation and Isotopes* 126 (2017), p. 225-227.

[10] U. Stöhlker et al. « Spectro-dosemeter-based gamma dose rate network in Germany ». In : *Applied Radiation and Isotopes* 182 (2022), p. 110077. ISSN : 0969-8043. DOI : https://doi.org/10.1016/j.apradiso.2021.110077. URL : https://www.sciencedirect.com/science/article/pii/S0969804321004681.

[11] P.H.G.M. Hendriks, J. Limburg et R.J. de Meijer. « Full-spectrum analysis of natural γ-ray spectra ». In : *Journal of Environmental*

Radioactivity 53.3 (avr. 2001), p. 365-380. ISSN : 0265931X. DOI : 10.1016/S0265-931X(00)00142-9.

[12] Y. SEPULCRE, T. TRIGANO et Y. RITOV. « Sparse Regression Algorithm for Activity Estimation in γ Spectrometry ». In : *IEEE Transactions on Signal Processing* 61.17 (sept. 2013), p. 4347-4359. ISSN : 1053-587X. DOI : 10.1109/TSP.2013.2264811.

[13] R. ANDRÉ et al. « Metrological approach of gamma-emitting radionuclides identification at low statistics: application of sparse spectral unmixing to scintillation detectors ». In : *Metrologia* (2020). URL : http://iopscience.iop.org/article/10.1088/1681-7575/abcc06.

[14] J. XU et al. « Sparse spectral unmixing for activity estimation in γ-RAY spectrometry applied to environmental measurements ». In : *Applied Radiation and Isotopes* 156 (2020), p. 108903.

[15] J. XU et al. « Analysis of gamma-ray spectra with spectral unmixing, Part 1: Determination of the characteristic limits (decision threshold and statistical uncertainty) for measurements of environmental aerosol filters ». In : *Applied Radiation and Isotopes* 182 (2022), p. 110109.

[16] P. MALFRAIT, J. BOBIN et A. DE VISMES OTT. « Spectral unmixing of multi-temporal data in gamma-ray spectrometry ». In : *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1045 (2023), p. 167547.

[17] Turner A.N. et al. « Convolutional Neural Networks for Challenges in Automated Nuclide Identification ». In : *Sensors (Basel)* 21 (2021), p. 5238. DOI : https://doi.org/10.3390/s21155238.

[18] Z. CHAOUAI et al. « Application of adversarial learning for identification of radionuclides in gamma-ray spectra ». In : *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1033 (2022), p. 166670. ISSN : 0168-9002. DOI : https://doi.org/10.1016/j.nima.2022.166670. URL : https://www.sciencedirect.com/science/article/pii/S0168900222002248.

[19] F. M.F. DE OLIVEIRA, G. DANIEL et O. LIMOUSIN. « Artificial gamma ray spectra simulation using Generative Adversarial Networks (GANs) and Supervised Generative Networks (SGNs) ». In : *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1047 (2023), p. 167795. ISSN : 0168-9002. DOI : https://doi.org/10.1016/j.nima.2022.167795. URL : https://www.sciencedirect.com/science/article/pii/S0168900222010877.

[20] J. Xu et al. « Analysis of gamma-ray spectra with spectral unmixing, Part 2: Recalibration for the quantitative analysis of HPGe measurements ». In : *Applied Radiation and Isotopes* 182 (2022), p. 110082.

[21] H. Bateman. « Solution of a System of Differential Equations Occurring in the Theory of Radio-active Transformations ». In : *Proc. Cambridge Philos. Soc* 15 (1910), p. 423-427.

[22] A. Chambolle et T. Pock. « A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging ». In : *Journal of Mathematical Imaging and Vision* 40.1 (mai 2011), p. 120-145. ISSN : 0924-9907. DOI : 10.1007/s10851-010-0251-1.

[23] D. D. Lee et H. S. Seung. « Learning the parts of objects by non-negative matrix factorization ». In : *Nature* 401.6755 (1999), p. 788-791.

[24] A. Berlizov. *A Correlated Particle Source xtension of a General Purpose Monte Carlo N-particle Transport Code MCNP-CP Upgrade Patch Version 3.2.* 2012, p. 183-194.

[25] H. Paradis et al. « Leda: A gamma-gamma coincidence spectrometer for the measurement of environment samples ». In : *Applied Radiation and Isotopes* 126 (2017). Proceedings of the 7th International Conference on Radionuclide Metrology – Low-Level Radioactivity Measurement Techniques, p. 179-184. ISSN : 0969-8043. DOI : https://doi.org/10.1016/j.apradiso.2016.12.049. URL : https://www.sciencedirect.com/science/article/pii/S0969804316304377.

[26] G. Gilmore. *Practical Gamma-ray Spectrometry.* Second. Wiley-Blackwell, 2001.

[27] Y Ménesguen et M-C Lépy. « COLEGRAM, a flexible user-friendly software for processing of ionizing radiation spectra ». In : *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1003 (2021), p. 165341.

[28] W. C. Evans. « Quantitative methods for continuous particulate air monitoring ». In : *IEEE Transactions on Nuclear Science* 48.5 (2001), p. 1639-1657.

[29] A. Quarteroni, R. Sacco et F. Saleri. *Numerical mathematics.* T. 37. Springer Science & Business Media, 2010.

[30] Détermination des limites caractéristiques (seuil de décision, limite de détection et limites de l'intervalle élargi) pour le mesurage des rayonnements ionisants — Principes fondamentaux et applications. 2020.

[31] R. A. Fisher. *Statistical methods and scientific inference*. Oliver et Boyd, 1956.