PRÉDICTIONS DE MESURES NUCLÉAIRES DÉGRADÉES PAR DES AÉROSOLS AVEC L'INTELLIGENCE ARTIFICIELLE

G. Dougniaux[1*], J. Baccou[2]

[1]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSN-RES, SCA, Gif-sur-Yvette, 91192, France
[2]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSN-RES, SEMIA, Saint-Paul-Lez-Durance, 13115, France
*gregoire.dougniaux@irsn.fr

TITLE

Prediction of nuclear measurement degraded with coarse aerosol by artificial intelligence

RESUME

Dans les installations nucléaires, la surveillance obligatoire de la contamination aéroportée est opérée par des instruments dédiés (CAM – *Continuous Air Monitor*) qui collectent les aérosols sur un filtre, mesurent la radioactivité déposée et déclenchent une alarme lorsqu'un seuil prédéterminé en activité est dépassé. Cette mesure et donc les alarmes sont très influencées par les variations en taille et concentration des aérosols. Afin de les prendre en compte, nous utilisons des méthodes d'apprentissage automatique à partir d'un jeu de données semi-synthétique. Nous présenterons dans cette communication la construction de ce jeu de données et les résultats des apprentissages : la présence d'une contamination a été correctement évaluée par un réseau de neurones.

ABSTRACT

In nuclear facilities, the mandatory airborne contamination surveillance is operated by dedicated Continuous Air Monitors (CAM) that collect all airborne aerosol on a filter, measure the radioactivity and trig an alarm when a predetermined activity concentration is exceeded. This measurement and therefore the alarms are very integrated by the variations in size and concentration of the aerosols. To consider it, we use machine learning methods from a semi-synthetic dataset. We present in this communication the construction of this dataset and the learning results: the presence of a contamination was correctly provided by a neural network.

MOTS-CLÉS : Spectre alpha, CAM, IA, aérosol / **KEYWORDS:** Alpha spectrum, CAM, AI, aerosol

1. INTRODUCTION

Sur les sites nucléaires, et en particulier les chantiers de démantèlement, des moniteurs de la radioactivité aéroportée sont déployés afin de s'assurer que les valeurs cibles de radioprotection des travailleurs sont respectées. Pour ce faire, ce type d'instrument prélève les aérosols de l'air ambiant en continu sur un filtre et mesure en temps réel la radioactivité accumulée sur ce filtre. Il doit alerter les personnels en cas de détection d'une contamination atmosphérique.

Pour la détection des émetteurs alpha artificiels (par exemple le ^{239}Pu), il doit estimer en continu et en temps réel la contribution des descendants du radon qui interfère dans la voie de mesure alpha artificielle (Hayes, 2003; Hoover & Newton, 1998; Justus, 2021; Li et al., 2013). En temps normal, cette estimation est très correctement faite mais, dans le cas particulier des chantiers de démantèlement, les variations soudaines de l'ambiance aérosols, tant en termes de granulométrie qu'en concentration, entraînent une modification de la mesure nucléaire, donc une estimation du bruit de fond incorrecte et *in fine* une augmentation fausse de l'activité alpha artificielle. Cela peut conduire à une fausse alarme et à un arrêt de chantier (Dougniaux et al., 2016). Afin d'améliorer l'estimation du bruit de fond, il a été montré que la prise en compte de la distribution granulométrique de l'aérosol était nécessaire (Hoarau, 2020; Hoarau et al., 2022).

Dans ce travail, on s'intéresse donc au développement d'un algorithme permettant de détecter de façon automatique la présence ou l'absence d'émetteurs alpha artificiels à partir de la connaissance de la mesure et de la taille des aérosols déposés sur le filtre. L'utilisation d'un tel algorithme pourrait être une solution prometteuse pour apprendre une nouvelle stratégie de mesure de la contamination. Cette communication décrit une première étude qui a été réalisée pour évaluer l'apport des techniques de machine-learning (Azencott, 2019) pour la construction de cet algorithme. Les phases d'apprentissage et de test de l'algorithme exploitent la base de données acquise en laboratoire (Hoarau, 2020) sur le banc d'essais ICARE de l'IRSN (Ammerich, 1989).

DOI : 10.25576/ASFERA-CFA2023-32933 Cet article est publié sous la seule responsabilité de ses auteurs

2. MATÉRIEL ET MÉTHODES

2.1. Description de la base de données expérimentales

La base de données acquise en laboratoire consiste en 63 mesures. Une mesure est l'enregistrement du spectre en énergie mesuré par le moniteur pendant les deux heures qui suivent l'événement de remise en suspension bref (< 2 min) d'une poudre d'alumine connue. Dans le même temps, le moniteur va prélever un aérosol de CsCl marqué des descendants du ^{222}Rn. Le tableau 1 présente ce dernier (noté S) ainsi que les quatre granulométries (notées de A à D) utilisées. De plus, chaque événement de remise en suspension conduit à une masse d'aérosols déposés sur le filtre différente : on mesure des masses de 0,5 à 15 mg. Finalement, la base de données est composée de 63 spectres de référence.

Tableau 1 : Caractéristiques des aérosols pendant les essais

Identifiant	S	A	B	C	D
Matière	CsCl	SPM105 – Al_2O_3	SPM102 – Al_2O_3	SPM95 – Al_2O_3	SPM84 – Al_2O_3
DAMM (µm)	0,2	9	13	18	57
σ_g	1,8	1,9	1,8	1,3	1,4
χ	1,08	1,6	1,5	1,6	1,2

DAMM : diamètre aérodynamique médian en masse
σ_g : écart-type géométrique
χ : facteur de forme dynamique

La figure 1 présente deux spectres en énergie tels que mesurés par le moniteur. Ce dernier est un ABPM203M (Mirion), que l'on désigne fréquemment sous l'acronyme anglais de CAM (*Continuous Air Monitor*). On note aisément sur cet exemple les différences entre spectre noir, mesure de référence sans aérosol grossier, et le spectre orange, notablement affecté par l'aérosol A.

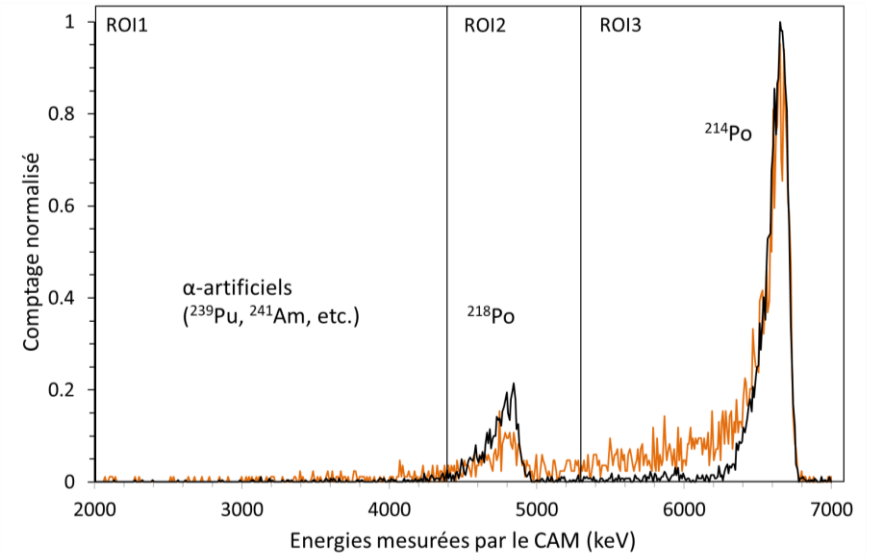

Figure 1 : Exemple de résultats de mesure. En noir le spectre de référence non dégradé, lié à l'aérosol S, et en orange un spectre dégradé lié à l'aérosol A. Le CAM (*Continuous Air Monitor*) utilisé ici est un ABPM203M (Mirion).

2.2. Description de la base de données semi-synthétique

Afin de pouvoir utiliser la base de données pour l'apprentissage et le test d'un modèle de machine-learning, il est nécessaire d'augmenter très significativement le nombre d'entrées. Ainsi, nous allons en particulier générer aléatoirement des spectres de différentes durées. Chaque essai consiste en 2 h de mesure, qui vont donc être étendues de 3 s à 2 h.
Pour ce faire, pour un spectre, on tire aléatoirement un aérosol et une masse (par exemple D et 2,9 mg). Il en résulte un spectre dans la base de données expérimentale (Figure 2, gauche). On tire alors un temps (par exemple 400 s), et un nouveau spectre est tiré selon la distribution de référence (Figure 2, droite).

DOI : 10.25576/ASFERA-CFA2023-32933 Cet article est publié sous la seule responsabilité de ses auteurs

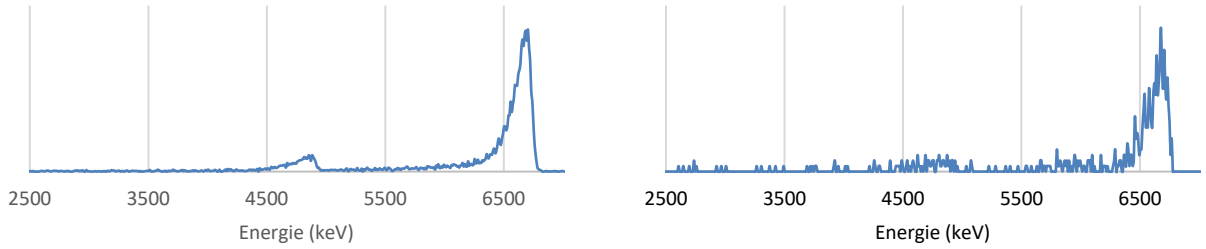

Figure 2 : Exemple de donnée synthétique. Le spectre de référence (à gauche) est rééchantillonné sur 400 s (à droite).

De cette façon, une nouvelle base de données, dite semi-synthétique, peut être générée de la dimension nécessaire à l'apprentissage supervisé.

2.3. Description de la construction de l'algorithme de détection

La détection de la présence ou de l'absence d'émetteurs alpha artificiels est un problème de classification. Les entrées de l'algorithme que l'on cherche à construire sont ici le spectre enregistré par l'appareil de mesure, le temps d'acquisition du spectre et la granulométrie de l'aérosol déposé sur le filtre. La sortie de l'algorithme est une grandeur binaire (0 ou 1) appelée label, 0 correspondant à l'absence d'émetteurs alpha artificiels et 1 à leur présence.
Une fois que la base de données a été générée, la construction de l'algorithme de détection est faite en 2 étapes :

<u>Étape 1 – Transformation de la base de données pour réduire la dimension des entrées</u>
Le spectre est enregistré sur 1024 canaux d'énergie. En prenant en compte le temps d'acquisition et la granulométrie, la dimension des entrées est donc de 1026. On procède donc d'abord à une réduction du nombre de canaux d'énergie en se focalisant sur les énergies comprises entre 2500 et 5500 keV, puis en effectuant une Analyse en Composantes principales (Lebart et al., 1997) ou une transformée en ondelettes (Daubechies, 1992).

<u>Étape 2 - Apprentissage et test d'un modèle de machine-learning à partir de la base transformée</u>
Plusieurs modèles de machine-learning (régression logistique, SVM, forêts aléatoires, réseau de neurones de type perceptron multicouche) sont appris et leurs performances sont évaluées. Classiquement, les performances sont évaluées par une matrice de confusion qui prend la forme décrite dans le tableau 2.

Tableau 2. Définition des paramètres nécessaires au tests de performance

		Label prédit	
		0	**1**
Label exact	**0**	$M_{0,0}$	$M_{0,1}$
	1	$M_{1,0}$	$M_{1,1}$

L'entrée (i, j) de cette matrice correspond au nombre de données ayant comme label exact i pour lesquelles le label j a été prédit par le modèle de machine-learning.

Plusieurs indicateurs de performance sont alors calculés à partir des coefficients de cette matrice :

- la justesse (*accuracy*) : ratio de bonnes prédictions ; elle s'écrit $\frac{M_{0,0}+M_{1,1}}{M_{0,0}+M_{1,1}+M_{0,1}+M_{1,0}}$,
- la précision (*precision*) : elle se focalise sur le nombre de données de label 1 bien prédites dans l'ensemble des données pour lesquelles le label 1 a été prédit ; elle s'écrit $\frac{M_{1,1}}{M_{1,1}+M_{0,1}}$,
- le rappel (*recall*) : il représente la proportion de données de label 1 bien prédites par rapport au nombre total de données ayant un label exact égal à 1 ; il s'écrit $\frac{M_{1,1}}{M_{1,1}+M_{1,0}}$,
- la mesure $f1$ ($f1$-*score*) : combinaison entre la précision et le rappel ; elle s'écrit $f1_score = 2 \times \frac{précision \times rappel}{précision + rappel}$.

3. RÉSULTATS

La base de données comporte 100 000 données (spectres, temps d'acquisition, granulométrie) avec un label associé (0 ou 1). 80% de la base sont utilisés pour l'apprentissage et 20% pour le test. Le tableau 3 fournit les valeurs des indicateurs de performance de quelques modèles pour une transformation de la base de données par ACP. On se limite ici à un modèle de régression logistique et de réseau de neurones (perceptron multicouche) qui ont fourni la moins bonne et la meilleure performance parmi l'ensemble des modèles choisis.

Tableau 3. Résultats de l'apprentissage pour les deux algorithmes

Réduction de dimension par ACP en conservant les 10 premiers axes principaux
Dimension des entrées : 12

Modèle	**Ensemble d'apprentissage**				**Ensemble test**			
	Accuracy	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score
Régression logistique	0,66	0,70	0,56	0,62	0,65	0,69	0,55	0,61
Réseau de neurones, 2 couches cachées avec 70 neurones chacune	0,97	0,98	0,97	0,97	0,97	0,97	0,96	0,96

Il ressort de cette première étude que la modélisation par réseau de neurones permet d'atteindre des performances très satisfaisantes (i.e. scores proches de 1) pour la détection automatique d'émetteurs artificiels. Il est également important de noter que ces résultats ont été obtenus à partir de la base transformée dont les entrées sont de dimension 12 au lieu de 1026 dans la base initiale.

4. CONCLUSION

Cette première étude a montré que les techniques de machine-learning étaient des outils efficaces pour construire un algorithme de détection automatique des émetteurs alpha artificiels à partir du spectre enregistré par l'appareil de mesure, du temps d'acquisition du spectre et de la granulométrie de l'aérosol déposé sur le filtre.

Les travaux futurs concernent d'abord la consolidation des résultats de cette étude qui intègre une phase d'optimisation des hyperparamètres des modèles de machine-learning. Afin de renforcer la confiance dans le modèle retenu, il s'agira ensuite d'identifier les caractéristiques des données mal prédites puis de progresser sur l'analyse de l'explicabilité du modèle. Enfin, il est envisagé de construire la même démarche pour quantifier précisément les rayonnements artificiels et identifier les caractéristiques des aérosols.

RÉFÉRENCES

Ammerich, M. (1989). *Réalisation d'une installation d'étalonnade de moniteurs de contamination atmosphérique à l'aide d'aérosols radioactifs calibrés (ICARE)*.

Azencott, C.-A. (2019). *Introduction au machine learning* (Dunod (ed.)).

Daubechies, I. (1992). Ten Lectures on Wavelets. *Ten Lectures on Wavelets*. https://doi.org/10.1137/1.9781611970104

Dougniaux, G., Monsanglant-Louvet, C., Teppe, A.-L., Marcillaud, B., Dieux Lestaevel, B., Gensdarmes, F., Michielsen, N., Bondiguel, S., Boussetta, B., & Quentel, G. (2016). Results from a measurement campaign in dismantling nuclear sites: a study of the false alarms emitted by CAM. In *European Aerosol Conference*.

Hayes, R. B. (2003). False CAM alarms from Radon fluctuations. *The Radiation Safety Journal*, *85*(2), S81–S84.

Hoarau, G. (2020). *Étude de la limite de détection et des fausses alarmes émises par les moniteurs de mesure de la contamination radioactive atmosphérique dans les chantiers de démantèlement*. Université Paris-Saclay.

Hoarau, G., Dougniaux, G., Gensdarmes, F., Ranchoux, G., & Cassette, P. (2022). Impact of the coarse indoor non-radioactive aerosols on the background radon progenies compensation of continuous air monitor. *Health Physics*, *122*(5), 563–574. https://doi.org/10.1097/HP.0000000000001530

Hoover, M. D., & Newton, G. J. (1998). Performance testing of continuous air monitors for alpha-emitting radionuclides. *Radiation Protection Dosimetry*, *79*(1–4), 499–504.

Justus, A. (2021). Technical Details of the Sigma Factor Alarm Method within Alpha CAMs. *Health Physics*, *120*(4), 442–453. https://doi.org/10.1097/HP.0000000000001341

Lebart, L., Morineau, A., & Piron, M. (1997). *Statistique exploratoire multidimensionnelle* (Dunod (ed.)).

Li, H., Jia, M., & Wang, K. (2013). Critical level setting of continuous air monitor. *Radiation Protection Dosimetry*, *154*(3), 391–395. https://doi.org/10.1093/rpd/ncs250