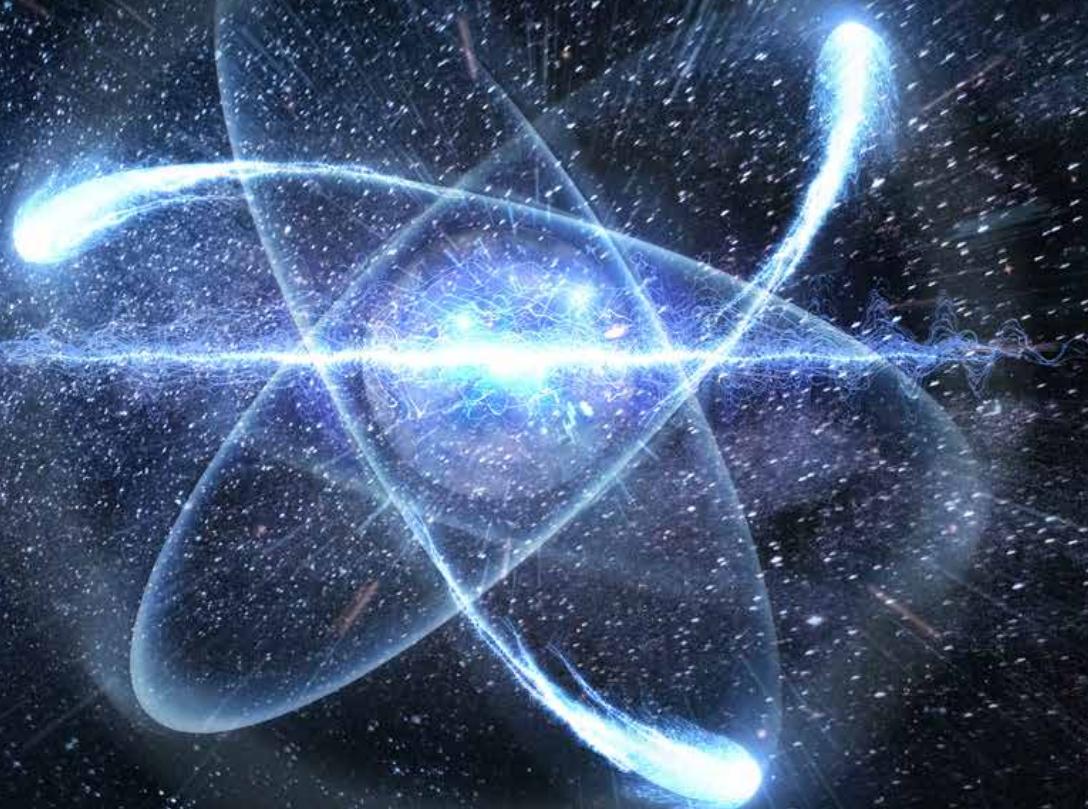




# ESARDA BULLETIN

The International Journal of Nuclear Safeguards  
and Non-Proliferation



ISSN 1977-5296  
KJ-BB-23-001-EN-N

Volume 65  
December 2023

**Editor**  
Elena Stringa

**Assistant Editor**  
Andrea De Luca

European Commission, Joint Research Centre,  
Directorate G.II - Nuclear Science, Safeguards  
and Security  
Unit G.II.7 – Digital Systems for Safeguards and  
Non-proliferation,  
TP421, I-21027 Ispra (VA), Italy  
Tel. +39 0332-786182  
EC-ESARDA-BULLETIN@ec.europa.eu

ESARDA is an association formed to advance and harmonize research and development for safeguards. More information can be found at the following address:

<https://esarda.jrc.ec.europa.eu/>

#### Editorial Board

K. Axell (SSM, Sweden)  
K. Aymanns (FZJ, Germany)  
S. Cagno (EC, JRC, J.1, Italy)  
A. De Luca (consultant at EC, JRC, G.II.7, Italy)  
S. Grape (UU, Sweden)  
T. Krieger (FZJ, Germany)  
O. Okko (STUK, Finland)  
I. Popovici (CNCAN, Romania)  
G. Renda (EC, JRC, G.II.7, Italy)  
A. Reznicek (Uba GmbH, Germany)  
R. Rossa (SCK-CEN, Belgium)  
J. Rutkowski (SNL, USA)  
Z. Stefánka (HAEA, Hungary)  
E. Stringa (EC, JRC, G.II.7, Italy)  
A. Tomanin (DG ENER, Luxembourg)

Papers submitted for publication are reviewed by independent authors including members of the Editorial Board.

Manuscripts have to be sent to the Editor (EC-ESARDA-BULLETIN@ec.europa.eu) following the paper guidelines available in the ESARDA Bulletin section of the ESARDA website (<https://esarda.jrc.ec.europa.eu/>) where the bulletins can also be viewed and downloaded.

Accepted manuscripts are published free of charge.

N.B. Articles and other material in the ESARDA Bulletin do not necessarily present the views or policies of either ESARDA nor the European Commission.

ESARDA Bulletin is published jointly by ESARDA and the Joint Research Centre of the European Commission.

The publication is authorised by ESARDA.

© Copyright is reserved, but part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopy, recording, or otherwise, provided that the source is properly acknowledged.

Cover design and layout by Christopher Craig Havenga, (consultant at EC, JRC, G.II.7, Italy).

# ESARDA BULLETIN

The International Journal of Nuclear Safeguards  
and Non-Proliferation

## Contents: Volume 65

### Editorial

Elena Stringa ..... 1

### Peer Reviewed Articles

**Improving the Confidence Associated with Passive Total Neutron Counting in the Nuclear Weapon Disarmament Verification Process** ..... 2  
Ville Bogdanoff, Kari Peräjärvi

**List Mode Inference Using Linear Classifiers for Nuclear Arms Control Verification** ..... 10  
Eduardo Padilla, Heidi Komkov, Christopher Siefert, Adam Hecht, Ryan Kamm, Kyle Weinurther, Jesus Valencia

**Dynamic Network Analysis of Nuclear Science Literature for Research Influence Assessment** ..... 19  
Samrat Chatterjee, Dennis Thomas, Daniel Fortin, Karl Pazdernik, Benjamin Wilson, and Lisa Newburn

**Simulating submarine reactor fuel in light of the AUKUS deal** ..... 34  
Sophie Grape, Erik Branger, Cecilia Gustavsson, Robert Kelley, Vitaly Fedchenko

**Data Validation Experiments with a Computer-Generated Imagery Dataset for International Nuclear Safeguards** ..... 44  
Zoe N. Gastelum, Timothy M. Shead, Matthew Marshall

**Benefits and Drawbacks of Non-Destructive Assay (NDA) for Nuclear Safeguards – An Overview** ..... 63  
Matteo Bolzonella

**Editorial**

Elena Stringa

Dear Readers,

It is with great pleasure that we present volume 65 of the 'ESARDA Bulletin - The International Journal of Nuclear Safeguards and Non-proliferation'.

This volume contains high quality and very interesting contributions on the following topics: arms control and disarmament verification, data analytics, non-proliferation and NDA techniques.

Please allow me to use this editorial to announce the ESARDA course on Nuclear Safeguards and Non-Proliferation that is co-organised by the JRC's Nuclear Science, Safeguards and Security directorate and the ESARDA Training and Knowledge Management working group. The course will be held online on April 15-19, 2024 and the deadline for registration is March 11th, 2024. The students providing the best essays that fulfill the ESARDA Bulletin publication policies will be invited to submit their contributions to the journal. The last article published in this volume is indeed a paper of an ESARDA course student that successfully passed the review process. There are also two additional articles from the ESARDA course students currently under review: if you wish to attend the ESARDA course, 2024 edition, please take into account this opportunity.

Please let me remind you that the ESARDA Bulletin is now a rolling publication. Articles can be submitted on a continuous basis and will be published online with their DOI as soon as they are ready, usually between 4 and 8 weeks after the submission.

If you wish to publish your work in the ESARDA Bulletin, send your article at any time together with the paper submission form duly filled and signed to EC-ESARDA-BULLETIN@ec.europa.eu. If accepted, the article will be published as soon as the review process will be completed. Before submitting your work, please ensure that your paper fits the Bulletin scope and that its content presents some novelties: we do not accept work that has already been published in other journals or conference proceedings, unless new aspects of the work are introduced (e.g. new results and related discussion). You can find the publication policies in the ESARDA Bulletin website under documents and forms.

Before concluding, I would like to thank the authors for their interest in publishing their findings in our journal, and the reviewers for their hard work, that lead to the publication of high quality articles. Many thanks also to Andrea De Luca (assistant editor) for all the work done to improve our journal and all the work he's doing in registering DOIs for the single articles and for the ESARDA Bulletin website. Thank you also to Guido Renda and Simone Cagno for their valuable advice: their suggestions have greatly improved the quality of the journal. Finally, thank you to Christopher Havenga, our graphic designer, who has designed the Bulletin cover and edits all the articles to fit the publication layout.

Enjoy the reading,

**Dr. Elena Stringa, PhD**

Editor of the ESARDA Bulletin - The International Journal of Nuclear Safeguards and Non-Proliferation  
 EC-ESARDA-BULLETIN@ec.europa.eu  
[https://esarda.jrc.ec.europa.eu/publications-0/esarda-bulletin\\_en](https://esarda.jrc.ec.europa.eu/publications-0/esarda-bulletin_en)

# Improving the Confidence Associated with Passive Total Neutron Counting in the Nuclear Weapon Disarmament Verification Process

Ville Bogdanoff

University of Jyväskylä, Department of Physics, P.O. Box 35, FI-40014 Jyväskylä, Finland, E-mail: ville.p.bogdanoff@jyu.fi

Kari Peräjärvi

Radiation and Nuclear Safety Authority (STUK), Jokiniemenkuja 1, FI-01370 Vantaa, Finland, E-mail: kari.perajarvi@stuk.fi

## Abstract:

*Passive total neutron counting is an important tool in the nuclear weapon disarmament monitoring and verification process proposed by the International Partnership for Nuclear Disarmament Verification (IPNDV). In the process, neutron measurements of given treaty accountable items (TAIs) are repeated multiple times in different locations and environments, and the measured neutron count rates are expected to remain unchanged. However, neutrons are heavily scattered in the environment, and the change in location or geometry of the environment can produce varying results in neutron measurements which can deteriorate the confidence of passive total neutron counting. In this paper, we have studied different kinds of neutron detection instruments and methods in various environments to determine the effects of the environment on passive total neutron counting and to develop recommendations and procedures to minimize and take these environmental factors into account. As a result, the moderated  $^3\text{He}$  proportional counter was the most promising type of instrument in terms of how the change in an environment impacted the measured neutron count rates. However, even with the  $^3\text{He}$  counter, the environmental influence increased rapidly with the source-to-detector distance (SDD). For example, with one-meter, two-meter, and three-meter SDDs, the maximum difference in count rates between outdoor and indoor measurements was 10.06(7) percent, 18.6(3) percent, and 28.1(5) percent, respectively. To mitigate the impact of the measurement environment, we propose to use a bare  $^{252}\text{Cf}$  reference source measured in the same geometries as the nuclear TAIs to estimate the influence of the environment on the measured neutron count rates. Using this technique in the same conditions as above produces indoor predictions that differ by 2.55(3) percent with a one-meter SDD, 0.334(5) percent with two-meter SDD and 1.93(4) percent with a three-meter SDD from the actual indoor measurements.*

**Keywords:** nuclear disarmament; IPNDV; neutron counting; neutron scattering

## 1. Introduction

Since 2014, International Partnership for Nuclear Disarmament Verification (IPNDV) has been identifying challenges associated with nuclear disarmament verification and developing potential procedures and technologies to address them [1]. Among other things, IPNDV has developed a 14-step nuclear weapon dismantlement framework [2]. Monitoring and verification technologies associated to different steps have also been studied in more detail [3]. Declarations, inspections, the chain of custody, and various passive radiation measurements and information barriers play a key role throughout the disarmament process [4], [5]. Information barriers are employed to prevent the release of classified information while allowing meaningful conclusions [6]. Note that IPNDV considers disarmament verification activities from the viewpoints of both the host (the owner of the nuclear explosive devices (NED)) and the inspecting party.

This article primarily concentrates on passive total neutron counting, which is used to confirm the presence of neutron-emitting special nuclear material (SNM) in the treaty accountable items (TAIs). In this article TAI refers both to complete NED and/or dismantled SNM from the NED. In passive total neutron counting, the neutron count rate is measured at a certain distance using a neutron detector, and the neutron rate is at some certain level proportional to the total mass of the SNM in the TAI. During the disarmament verification process, the TAI may be measured multiple times in different locations and environments, and the neutron count rate should remain unchanged throughout the process. A dramatic drop in the count rate could indicate reduction of the mass of the SNM during the process, which can deteriorate confidence in the disarmament verification process.

Neutron counting is prone to environmental influence, as neutrons are easily scattered and reflected by the environment, including floor, walls, air, etc. [7], [8]. The neutron count rate acquired in passive total neutron counting is determined by neutrons not only coming directly from the source but also through scattering processes. Consequently, changes in the environment can change the portion of scattered neutrons. From the confidence point of

view, it is important to comprehend how much the environment can influence the measured neutron count rate.

In this paper, we have studied how the environment influences the neutron count rates of different kinds of neutron detection instruments and methods, including moderated  $^3\text{He}$  proportional counter,  $^6\text{LiF}/\text{ZnS}(\text{Ag})$  scintillator-based detector, as well as  $\text{LaBr}_3$  and  $\text{NaI}(\text{Tl})$  scintillators that detect neutron induced high-energy gamma-rays. The moderated  $^3\text{He}$  proportional counter can be considered the gold standard for neutron detection. However, due to the global shortage of  $^3\text{He}$  isotope [9], it was important to study how the alternative neutron detection technologies perform in passive total neutron counting. In this study, we have also developed and tested an additional monitoring procedure to mitigate the impact of environmental factors on the confidence associated with neutron counting and the overall monitoring and verification process.

## 2. Materials and Methods

Experimental work was conducted at the Radiation Metrology Laboratory (RML) at the Finnish Radiation and Nuclear Safety Authority (STUK). RML is responsible for maintaining national measurement standards of ionizing radiation in Finland and provides calibration services for various companies and institutions.

### 2.1 Experimental Setup

Due to the limited measurement time available in the RML, different types of commercially available neutron detectors were irradiated simultaneously with an industrial neutron source, and the total neutron count rates were measured as a function of source-to-detector distance (SDD) in three different environments: 1) outdoors, 2) the neutron source

placed approximately in the middle of the RML's calibration hall (subsequently referred to as Indoor I) and 3) the neutron source placed one meter away from the concrete wall at the back of the calibration hall (subsequently referred to as Indoor II). The outdoor measurements were conducted in the asphalt parking plot belonging to STUK. The calibration hall is 16 meters in length, 5.5 meters wide, and 5 meters in height and has approximately one-meter-thick concrete walls. The measurement setups are depicted in Figures 1, 2 and 3.

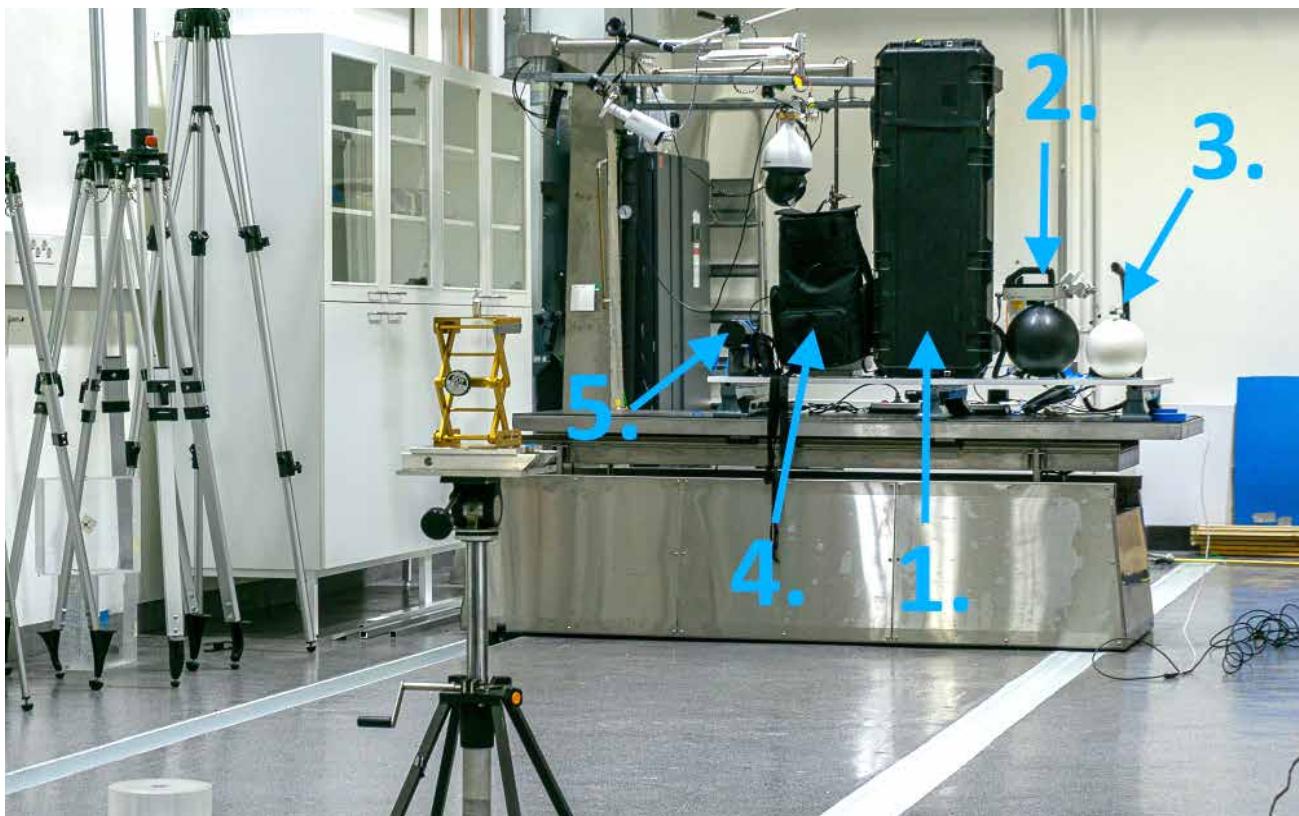


**Figure 1:** The setup depicting indoor measurements (Indoor II setup) with a moderated  $^{252}\text{Cf}$  source placed on the tripod one meter away from the concrete wall at the back of the calibration hall.

In the indoor measurements, detectors were placed side by side on the electronically movable table (height of one meter), and the neutron source was placed approximately 1.1 meters above the ground on the tripod. In the outdoor measurements, detectors were placed on the stationary table (height of 0.8 meters), and the neutron source (placed



**Figure 2:** The setup depicting outdoor measurements. Moderated  $^{252}\text{Cf}$  source was placed on the table trolley and the source-to-detector distance was measured with a laser range finder.



**Figure 3:** The setup depicting indoor measurements (Indoor I setup) with bare  $^{252}\text{Cf}$  source placed on the tripod in the middle of the calibration hall. Detectors numbered in the figure are: 1) Environics RanidPort, 2) Berthold LB 6411, 3) Mirion SN-D-2, 4) Environics RanidPro200  $^6\text{LiF}/\text{ZnS}(\text{Ag})$  neutron plate, 5) Environics RanidPro200  $\text{LaBr}_3$ .

approximately 0.9 meters above the ground) was moved with a table trolley. Neutron count rates with bare and moderated source were measured in the SDD range of 1 to 7 meters with one-meter intervals. Based on the IPNDV verification process, the neutron detector cannot be placed in contact with a bare or containerized NED due to safety reasons. Therefore, a minimum distance of one meter, also discussed by the experts representing the nuclear-weapon states in the IPNDV meetings, was selected for the present study. The maximum distance of seven meters was due to calibration hall dimensions preventing us from operating the movable table beyond seven meters from the source (Indoor I measurements).

The SDDs were measured from the centre of the neutron source to the front face of the central detector ( $\text{NaI}(\text{Tl})$ ), and the distances to other detectors were calculated with the Pythagorean theorem by also knowing the distances between the central detector and the other detectors. With each SDD, data was collected for several minutes to ensure enough data was gathered to limit statistical error. Background measurements without the neutron source were performed before and after each of the three series of measurements, and the background count rate of the latter measurement was subtracted from the measured neutron count rates.

A neutron-emitting  $^{252}\text{Cf}$  isotope was used as the source of neutrons in this study. The neutron energy spectrum of

$^{252}\text{Cf}$ -spontaneous fission [10] with an average neutron energy of 2 MeV is very similar to special nuclear materials (SNM), such as  $^{239}\text{Pu}$  [11] and  $^{235}\text{U}$  [11] used in nuclear weapons. The  $^{252}\text{Cf}$  used in this study had a calculated activity of 179.5 MBq (nominal activity of 500 MBq with a reference date of Nov-15-2016) emitting  $2.15 \times 10^7$  neutrons per second. The source (made by QSA Global Inc.) was a cylindrical sealed capsule with a single stainless-steel encapsulation surrounded by a custom aluminium shield.

The measurements were performed with a bare and moderated  $^{252}\text{Cf}$  source. Moderation reduces the speed of fast neutrons and increases the portion of slower (thermal) neutrons, making them more susceptible to neutron capture. In the nuclear disarmament verification process, the exact geometry of the TAI is not known, e.g., the nuclear material can be shielded, moderated, or bare. Therefore, it was important to study whether moderation enhances or diminishes the influence of the environment on the results of passive total neutron counting. We used a partially hollow cylinder of polymethyl methacrylate (PMMA) to moderate the  $^{252}\text{Cf}$  source. The moderation was 9 cm thick from the sides, 8 cm thick from the bottom, and 6.8 cm thick from the top.

## 2.2 Neutron Detectors

The detectors used in this study are depicted in Figure 3. Detector in the centre (numbered 1 in the figure) is the

RanidPort Mobile, a radiation portal monitor made by Envionics (a subsidiary of Bertin Technologies). RanidPort features a four-liter volume 4"x4"x16" NaI(Tl) scintillator and is capable of indirectly detecting neutrons through high-energy gamma radiation using the energy gate of 3.5 to 8.5 MeV [12]. The natural gamma-ray background above 3.5 MeV is reasonably low (see Section 3). In the present experimental conditions, additional high-energy gamma-ray signals are caused by neutron capture reactions occurring in the detector and the environment, and there are also prompt fission gamma rays being emitted directly from the source. The biggest advantage of using an indirect neutron detection method based on high-energy gamma-ray spectroscopy can be performed simultaneously with the same detector. In the disarmament verification process, conventional gamma-ray spectroscopy performed behind the information barrier using a medium energy resolution detector can provide useful information from the TAI [13].

The second detector was the Berthold LB 6411, which is used as the reference neutron ambient equivalent dose meter in the RML. Berthold LB 6411 is a <sup>3</sup>He and methane gas-filled proportional counter with wide neutron energy measuring range from thermal to 20 MeV. The gas-filled counter tube inside the detector is surrounded by a 250 mm polyethylene moderator sphere moderating fast neutrons to lower energies.

The third detector was the Mirion SN-D-2 neutron dose probe (also moderated <sup>3</sup>He gas-filled proportional counter) connected to the Canberra Colibri TTC survey meter. The data from the SN-D-2 probe was not utilized in this study but was used for internal purposes at STUK to compare the neutron count rates between Berthold LB 6411 and SN-D-2.

The fourth detector was the <sup>6</sup>LiF/ZnS(Ag) neutron detector plate made by Symetrica and is part of the RanidPro200 backpack-operated radionuclide identification device made by Envionics. The plate is composed of two <sup>6</sup>LiF/ZnS(Ag) neutron screens coupled to a wavelength-shifting plastic (polyvinyl toluene) and not containing any moderation other than the external plastic shell and the nylon backpack fabric surrounding the plate.

The fifth detector was the 2"x2" LaBr<sub>3</sub> scintillator which is also part of the Envionics RanidPro200 backpack. The LaBr<sub>3</sub> detector is capable of indirect detection of neutrons through high-energy gamma radiation (energy gate of 3.5 MeV to 5.7 MeV) as the NaI(Tl) detector. In this study, the LaBr<sub>3</sub> detector was taken out from the backpack and placed on the measuring tables.

### 3. Results

Figure 4 shows the background subtracted data sets including estimations for uncertainties caused by the statistics and distance, measured in Indoor I, Indoor II, and Outdoor measurement setups. The uncertainties for the distance were calculated using the standard method of error propagation. The background count rates in Outdoor measurements were 0.007(3) cps for the <sup>3</sup>He proportional counter, 0.14(2) cps for the LaBr<sub>3</sub> scintillator, 0.23(2) cps for <sup>6</sup>LiF/ZnS(Ag) scintillator and 13.4(2) for the NaI(Tl) scintillator. Inside the calibration hall, the background rates were somewhat higher. In the Figure 4, the neutron count rate has been plotted as a function of source-to-detector distance (SDD) for each detector using moderated and bare <sup>252</sup>Cf source. For each data set in Figure 4, we have fitted an inverse power function (equation (1)) using the least-squares fitting weighted with the absolute statistical uncertainty of each data point. The fitting was performed using the curve\_fit function of Python's SciPy-library [14].

$$R = \frac{A}{r^x} + B, \quad (1)$$

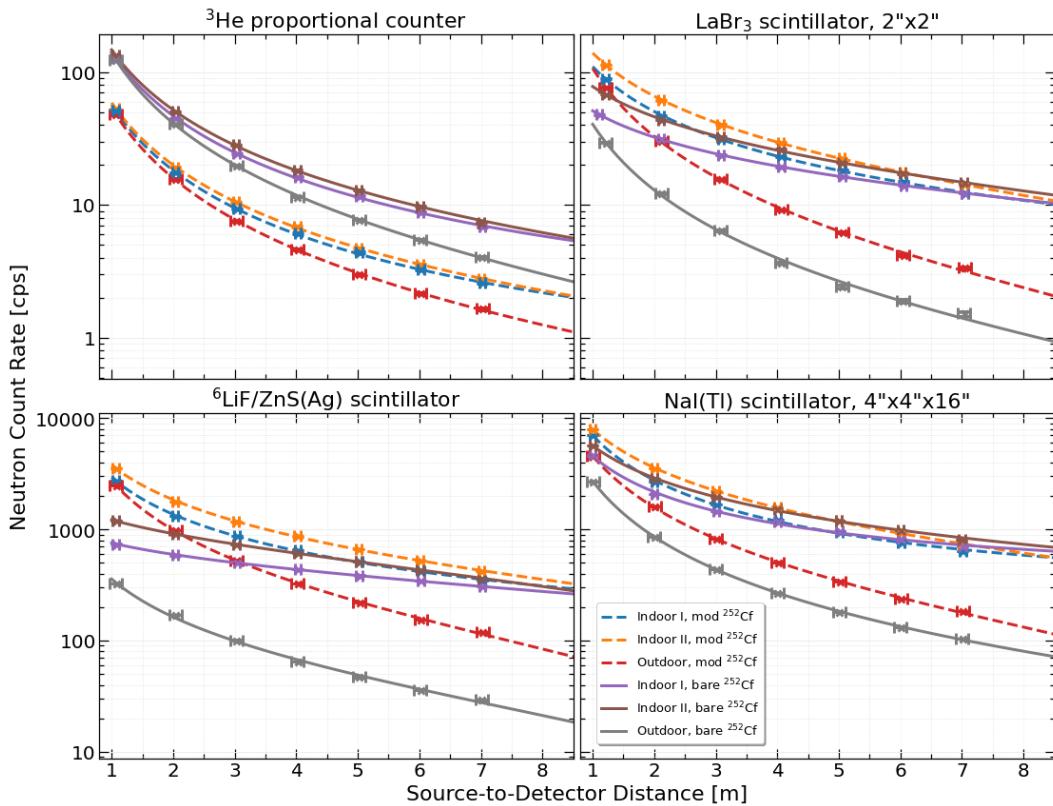
where  $r$  is the SDD and  $A$ ,  $B$ , and  $x$  are parameters of the inverse power function. Similar function with parameter  $x$  fixed to two was used in the ref [15]. Note that the parameter  $A + B$  equals the count rate at 1-meter SDD. Parameter  $B$  is needed to improve the performance of the fitting function. Without scattering, the intensity of radiation is inversely proportional to the square of the SDD, and the parameter  $B$  would be equal to zero. However, in the case of neutron radiation, the neutron scattering will cause the intensity to diverge from the inverse square law and result in a lower exponent than two and non-zero values of  $B$ .

The parameters gained by fitting equation (1) to data sets are shown in Table 1. As depicted, none of the data sets comply with the inverse square law, but all result in a lower exponent than two. The relative percentage differences (RPD) of  $A$ ,  $B$ , and  $x$  parameters when comparing the indoor measurements to the outdoor measurements are also listed in Table 1. For the <sup>3</sup>He proportional counter, the parameters  $A$ ,  $B$ , and  $x$  change the least between the measurement setups. Thus, from the confidence point of view, results obtained by the <sup>3</sup>He proportional counter are the least influenced by the environment and can be considered the most promising type of neutron detector compared to the other detectors tested in this study. Note from Figure 4 that the neutron count rates measured with the <sup>3</sup>He proportional counter in the Indoor I configuration at longer SDDs ( $> 5$  m) are comparable to the rates in the Indoor II configuration with both the bare and the moderated source. This is probably caused by the neutrons that are backscattered from the back wall to the detectors. At seven meters SDD, the back wall was approximately two meters from the detectors, and thus with longer SDDs, the

Configuration	$^3\text{He}$ counter			$^6\text{LiF}/\text{ZnS}(\text{Ag})$			$\text{LaBr}_3$			NaI(Tl)		
	A	x	B	A	x	B	A	x	B	A	x	B
	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>	% <sup>a</sup>
Indoor I, bare $^{252}\text{Cf}$	140.1 (6)	1.61 (2)	0.9 (2)	793.5 (12)	0.465 (1)	b	58.6 (12)	0.56 (3)	-7.6 (15)	4263 (4)	1.156 (3)	300 (3)
	0.448 (3)	7.02 (6)	220 (50)	107.9 (3)	55.0 (4)	b	44 (1)	66 (4)	2000 (800)	56.81 (8)	28.02 (8)	3800 (400)
Indoor II, bare $^{252}\text{Cf}$	147.5 (4)	1.50 (1)	-0.4 (2)	1307.2 (9)	0.5848 (6)	b	88.2 (9)	0.64 (2)	-10.5 (11)	5722 (4)	0.947 (2)	-65 (4)
	5.73 (3)	13.6 (1)	50 (21)	242.5 (6)	43.4 (3)	b	116 (2)	60 (2)	3000 (1100)	110.5 (2)	42.5 (1)	710 (80)
Outdoor, bare $^{252}\text{Cf}$	139.5 (6)	1.73 (1)	-0.8 (2)	381.7 (9)	1.033 (6)	-23.3 (7)	40.8 (4)	1.62 (3)	-0.33 (11)	2718 (3)	1.647 (3)	-8.1 (7)
	c	c	c	c	c	c	c	c	c	c	c	c
Indoor I, mod $^{252}\text{Cf}$	56.3 (3)	1.66 (2)	0.4 (1)	2909 (2)	1.120 (2)	27 (2)	109.5 (4)	1.13 (2)	0.3 (4)	6776 (4)	1.418 (2)	232 (2)
	4.44 (4)	4.72 (5)	320 (120)	3.741 (4)	20.47 (5)	141 (8)	2.17 (2)	32.5 (5)	130 (140)	45.68 (5)	5.217 (9)	410 (7)
Indoor II, mod $^{252}\text{Cf}$	58.7 (2)	1.56 (2)	-0.1 (1)	3948 (3)	0.963 (2)	-178 (3)	144.8 (5)	1.02 (2)	-5.6 (6)	8137 (5)	1.095 (2)	-223 (4)
	9.03 (7)	10.7 (2)	80 (200)	40.78 (5)	31.59 (8)	170 (3)	35.0 (3)	39.0 (6)	500 (100)	74.95 (7)	26.86 (5)	199 (5)
Outdoor, mod $^{252}\text{Cf}$	53.9 (4)	1.74 (2)	-0.2 (1)	2804 (3)	1.408 (3)	-66 (1)	107.2 (7)	1.67 (2)	-0.9 (2)	4651 (4)	1.50 (1)	-74.7 (11)
	c	c	c	c	c	c	c	c	c	c	c	c

<sup>a</sup> Relative percentage difference (RPD). <sup>b</sup> The fitting was done by setting the parameter *B* fixed and equal to zero. <sup>c</sup> Reference environment for RPD.

**Table 1:** Parameters *A*, *B* and *x* obtained by fitting equation (1) to the data sets. The relative percentage difference (RPD) in fit parameters *A*, *B* and *x* are shown for each indoor data set when the outdoor measurement setup was used as the reference environment. Uncertainties of the fitted parameters and RPD values are given in parenthesis.



**Figure 4:** Neutron count rate vs. source-to-detector distance for various neutron detectors using the bare and moderated  $^{252}\text{Cf}$  source in three different measurement setups. Equation (1) was fitted for the data points and the resulting parameters are given in Table 1. The uncertainties for neutron count rates and SDDs are marked for each data point.

Indoor I setup was somewhat similar in geometry to the Indoor II setup.

For the other detectors evaluated, the influence of the environment on neutron count rates was more dramatic. The worst performer seemed to be the  ${}^6\text{LiF}/\text{ZnS}(\text{Ag})$  scintillator-based neutron detector, as it had the largest difference in parameters  $A$ ,  $B$ , and  $x$  values between configurations. Also, the indirect high-energy gamma-ray-based neutron detectors did not perform as consistently as the  ${}^3\text{He}$  proportional counter, where a significant environmental sensitivity even at 1-meter SDD is visible ( $A + B$ ). In all measurement setups, the highest count rates were associated to Nal(Tl) and  ${}^6\text{LiF}/\text{ZnS}(\text{Ag})$  detectors.

As can be seen from Figure 4, the moderation had a substantial impact on the neutron count rates measured with all detector types. For instance, with the  ${}^3\text{He}$  proportional counter, the measured count rates were reduced by almost half when the source was moderated. For the other detectors tested, the moderation impacted oppositely: the count rates were higher when the source was moderated. Consequently, it is relevant for the inspectors monitoring the nuclear disarmament to know if the shielding or moderation i.e., the surrounding material of the TAI has changed in between neutron measurements.

#### 4. Discussion

Although the moderated  ${}^3\text{He}$  proportional counter was the most reliable neutron detector tested, the measured neutron count rates of the  ${}^3\text{He}$  proportional counter still varied significantly in various environments. Figure 5 shows the relative percentage differences in the neutron count rates

vs. the source-to-detector distance when comparing outdoor measurements to indoor measurements (circles). As depicted, the RPD increases rapidly when the distance increases. In Outdoor/Indoor II data set, the difference is 18.6(3) percent with distance of two meters but increases to 28.1(5) percent with three meters. Therefore, the neutron detector should be placed as close as possible to the object which is to be verified. Also, for simplicity, a fixed distance and preferably a similar type of detection instrument should be used in every neutron measurement performed during the disarmament verification process.

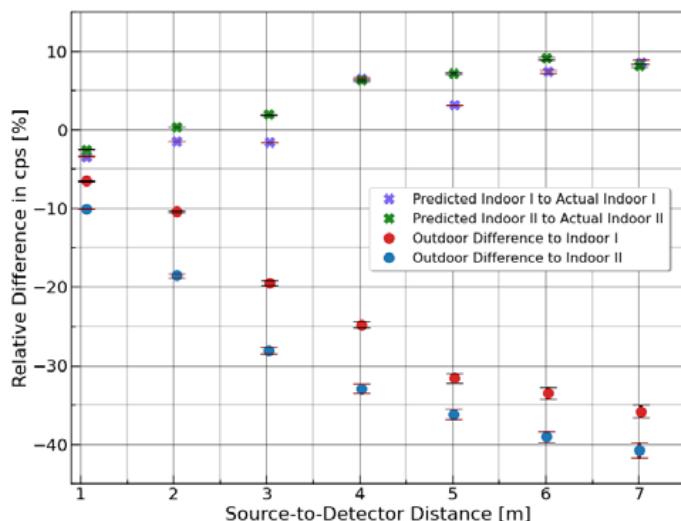
Notice that even with the SDD of 1.07 meters, there would still be a 10.06(7) percent difference in the neutron count rates between the Outdoor and Indoor II measurement configurations. Assuming that the neutron count rate correlates with the SNM mass of the TAI, then the 10.06(7) percent reduction in the Outdoor count rate would indicate that a significant amount of SNM has gone missing between the Indoor II and Outdoor measurements. As shown earlier, if the SDDs get longer the differences in count rates become larger, which can further deteriorate the confidence in the disarmament verification process.

To mitigate the environmental impact on the confidence of neutron measurements in the nuclear disarmament verification process, we propose adding a  ${}^{252}\text{Cf}$  reference source measurements to it. In the modified process, the reference source is always measured in the same geometries as the TAIs. This approach is based on the assumption that the TAI and  ${}^{252}\text{Cf}$  count rates have similar behaviour as a function of measuring environment. If the assumption is valid then the TAI and  ${}^{252}\text{Cf}$  measurement data can be used to predict the TAI count rates ( $R$ ) in new environments, see equation (2).

$$\frac{R_{252\text{Cf}, \text{location 2}}}{R_{252\text{Cf}, \text{location 1}}} = \frac{R_{\text{predicted TAI}, \text{location 2}}}{R_{\text{TAI}, \text{location 1}}}. \quad (2)$$

The performance of the method is examined in Figure 5, where we assume that the moderated  ${}^{252}\text{Cf}$  source is a sensitive TAI. In the figure, crosses show how much the predicted (using equation (2)) indoor neutron count rates differ from the actual measured values. As an example, using one-meter (1.07 m exactly), two-meter (2.03 m exactly) and three-meter (3.03 m exactly) SDDs, the predicted and measured values of Indoor II differ by 2.55(3) percent, 0.334(5) percent, and 1.93(4) percent, respectively. These values demonstrate the potential associated with the proposed technique.

The introduced  ${}^{252}\text{Cf}$  reference source can also be applied for other purposes. For example, it can be employed to test and calibrate the performance of the host-provided neutron



**Figure 5:** Relative percentage difference in neutron count rate of  ${}^3\text{He}$  proportional counter vs. source-to-detector distance. In the figure, circles represent how much the measured neutron count rates of the outdoor measurements differ from the indoor measurements, and crosses show how much the predicted neutron count rates differ from the actual measured indoor values. Uncertainties are included for each data point.

detection instruments. By comparing data from new and older measurements performed in similar conditions as well after calculating the required decay corrections one can make conclusions about the device performance and make the necessary corrections if needed. By applying the lessons learned from this study such instrument testing/calibration should be performed in as an open environment as possible using a short (preferably constant) SDD. Between inspections, the host could store  $^{252}\text{Cf}$  reference source in a sealed container.

## 5. Conclusions

In this paper, we show that all the tested neutron detection instruments and methods are sensitive to source shielding/moderation. Therefore, changes on those should be communicated to the inspecting party. The selection of an optimal neutron detector for verification use is important. Neutron monitoring simplifies if only one type of instruments is used. In the neutron measurements as short measurement distances as possible should be employed to minimize the influence of the environment on the measurement results (using a constant measurement distance would facilitate the analysis work even further). Testing a host-provided instrument with a well-calibrated  $^{252}\text{Cf}$  reference source in an open environment and with a short measurement distance adds confidence to the overall process (instrument performance check). Between inspections, the host could store such reference sources in a sealed container.

The study also shows that the confidence provided by the neutron measurements can be significantly enhanced by performing measurements with the bare  $^{252}\text{Cf}$  reference source in the same environments and geometries as the actual measurements of the TAIs. In the future, one could continue the studies of this approach by using more realistic TAI surrogates. Note that the neutron count rates recorded from the TAIs in different environments are probably classified information, but their comparisons may not be. Therefore, part of the data analysis may have to be conducted behind the information barrier.

## 6. Acknowledgements

This work was financially supported by the Ministry for Foreign Affairs of Finland. The authors would like to thank Dr. Jussi Huikari from the Radiation Metrology Laboratory of STUK for assisting with the measurements.

## 7. References

- [1] Hinderstein C; International partnership for nuclear disarmament verification: Laying a foundation for future arms reductions; *Bulletin of the Atomic Scientists*; vol. 74; 2018.
- [2] Dunn LA; Walkthrough Exercise Summary; Nuclear Threat Initiative; 2017; p 5; <http://www.jstor.org/stable/resrep14254> (accessed December 19, 2022).
- [3] IPNDV; Results of the SCK CEN Exercise for Disarmament Verification Technologies; IPNDV Technology Track; 2021; <https://www.ipndv.org/reports-analysis/results-of-the-sck-cen-exercise-for-disarmament-verification-technologies/> (accessed January 16, 2023).
- [4] Götsche M, Kirchner G; Measurement Techniques for Warhead Authentication with Attributes: Advantages and Limitations; *Science & Global Security* vol. 22; no. 2 (2014): p 83-110.
- [5] Runkle RC, Bernstein A, and Vanier PE; Securing special nuclear material: Recent advances in neutron detection and their role in non-proliferation; *Journal of Applied Physics*; vol. 108; 111101; 2010; <https://doi.org/10.1063/1.3503495>.
- [6] Close D, Macarthur D, Nicholas N; Information Barriers - a Historical Perspective, Office of Scientific & Technical Information Technical Reports; Los Alamos National Lab., NM (United States); 2001; <https://digital.library.unt.edu/ark:/67531/metadc724908/> (accessed December 19, 2022).
- [7] Sonder S, Hebel S, Prünke C, Kirchner G; Impact of Concrete Building Structures on Neutron Radiation and its Mitigation; *ESARDA Bulletin - The International Journal of Nuclear Safeguards and Non-proliferation*; vol. 64(2); 2022; p 2-9; <https://doi.org/10.3011/ESARDA.IJNSNP.2022.8>.
- [8] Gregor J, Baron M, Kesten J, Kroeger EA; Comparison of the response of handheld neutron detectors in differing deployment environments: Measurements, calculations and practical implications, *Radiation Measurements*; vol. 143; 106571; 2021; <https://doi.org/10.1016/j.radmeas.2021.106571>.
- [9] Shea DA, Morgan D; The helium-3 shortage: Supply, demand, and options for congress; *LIBRARY OF CONGRESS WASHINGTON DC*; 2011; <https://sgp.fas.org/crs/misc/R41419.pdf> (accessed January 17, 2023).
- [10] Plompen AJM et al; The joint evaluated fission and fusion nuclear data library, JEFF-3.3; *The European Physical Journal A*; vol 56; 2020; <https://doi.org/10.1140/epja/s10050-020-00141-9>.

- [11] Chadwick MB et al; ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data; Nuclear Data Sheets; vol.148; 2018; p 1-142; <https://doi.org/10.1016/j.nds.2018.02.001>.
- [12] Holm P, Peräjärvi K, Sihvonen A-P, Siiskonen T, Toivonen H; Neutron detection with a NaI spectrometer using high-energy photons; Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment; vol. 697; 2013; p 59-63; <https://doi.org/10.1016/j.nima.2012.09.010>.
- [13] Seager K et al; Trusted radiation identification system; Proceedings of the 42nd Annual INMM Meeting; 2001; <https://media.nti.org/pdfs/SNL-12 SAND2001-1866C.pdf> (accessed January 17, 2023).
- [14] Virtanen P et al; SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python; Nature Methods; vol. 17; 2020; p 261-272; <https://doi.org/10.1038/s41592-019-0686-2>.
- [15] Kemp A et al; A study of neutron room scattering at RPCF; Fermi National Accelerator Lab; 1996.

# List Mode Inference Using Linear Classifiers for Nuclear Arms Control Verification

Eduardo Padilla <sup>1,2</sup>, Heidi Komkov <sup>2</sup>, Christopher Siefert <sup>2</sup>, Adam Hecht <sup>1</sup>, Ryan Kamm <sup>2</sup>,  
Kyle Weinfurther <sup>2</sup>, Jesus Valencia <sup>2</sup>

<sup>1</sup> University of New Mexico, Department of Nuclear Engineering, Albuquerque, NM, USA, 87131

<sup>2</sup> Sandia National Laboratories, PO Box 5800, Albuquerque, NM, USA, 87185

## Abstract:

*In potential future nuclear arms control treaties, methods to confirm the presence or absence of a nuclear warhead or nuclear components are likely to be a central function of a verification regime. Higher confidence in verification methods can be achieved through more rigorous, thus potentially sensitive, analysis of radiation signatures from treaty accountable items. Therefore, methods that protect sensitive information while allowing for rigorous analysis are a critical component of any potential nuclear treaty verification system; these methods are referred to as information barriers. In this paper, we describe the development of a novel radiation analysis method for list-mode (time-stamped pulse heights, pulse-by-pulse) inference using linear classifiers trained on a large set of synthetically generated high resolution gamma spectra. In practice, each detector pulse would be fed into a linear classifier with the applied weight incrementing or decrementing counters for each class. After a set number of pulses, the highest output score determines the classification of the source of radiation. As such, this method serves as both a verification algorithm and information barrier combined. This new method achieves reliable discrimination (83% accuracy) of notional nuclear weapons grade treaty accountable item radiation signatures from those of a diverse, largely unconstrained, set of nuclear, medical and industrial radioisotope combinations. Importantly, this is shown to be achievable without the collection or processing of a potentially sensitive gamma radiation spectrum. This study serves as a proof of concept for the development of an intrinsic information barrier for attribute identification supporting nuclear arms control treaty verification.*

**Keywords:** Information Barrier, Warhead Verification, Machine Learning.

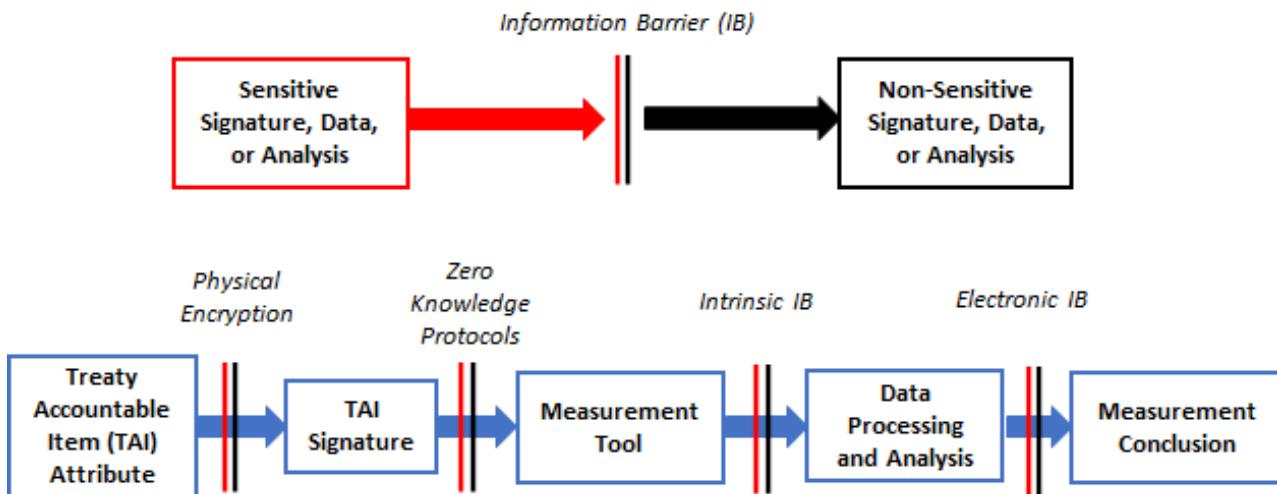
## 1. Introduction

Potential future nuclear arms control treaties are likely to require much more rigorous and intrusive measures for verification as nuclear weapons states move beyond current absence verification methods such as those employed in New START (Evans, 2021). As such, methods more advanced than neutron detection above a threshold to confirm the presence or absence of a nuclear warhead or component are likely to be a central function of a verification regime. While standard methods for nuclear assay such as gamma ray spectroscopy and/or neutron measurements are largely capable of performing this function, the amount of information revealed during the analysis is likely too high for a nuclear arms control regime; nuclear warhead design information can be inferred from these measurements, thus potentially disclosing sensitive strategic information to a treaty partner. Hence, a critical need for nuclear arms control verification is a method that produces high-confidence assessments without revealing sensitive information such as can be inferred from gamma ray spectra or detailed neutron signatures.

### 1.1 Background

There are numerous approaches to protecting sensitive information, also referred to as information barriers (IBs). Figure 1 illustrates several types of information barriers which can be used within a verification process to act as information reducers and prevent the passage of sensitive information (red side) while allowing a reduced or transformed subset of non-sensitive information to proceed (black side). The approach described in this paper is a form of an intrinsic information barrier, wherein the pulses from a gamma and/or neutron detector are analyzed immediately and individually, without the creation or storage of accumulated spectra, dose rates, or other potentially sensitive information. This type of information barrier is complementary to other approaches, such as physical encryption, zero knowledge protocols (ZKP) and electronic information barriers.

Yan and Glaser (Yan & Glaser, 2015) provide a comprehensive review of past warhead verification systems incorporating several types of information barriers. Additional systems (Hamel, 2018) (White, 2012) (Wolford & White, 2000) have been included for background consideration in this



**Figure 1:** Information Barriers within a Verification Process

paper and a subset are summarized in Table 1, categorizing them according to the proposed conventions in Figure 1.

Table 1 categorizes historically-developed arms control verification concepts and systems by three primary design aspects: 1 – template or attribute verification approach, 2 – active or passive measurement and 3 – the type of information barrier employed. Each of these design aspects have associated strengths and weaknesses and cannot be truly evaluated independent of a well-defined treaty regime. For example, plutonium absence verification can reasonably be expected to be achievable using a much simpler method (gross neutron counting (Harahan, 1993)

than neutron tomographic imaging combined with ultra-high-resolution gamma spectroscopy; the simplest proposed method of performing a specific treaty verification task has a higher likelihood of negotiated implementation.

Compared to attribute verification systems, template verification systems are often considered easier to implement, since these can be designed to be performed behind an information barrier, sometimes requiring little to no a priori knowledge about the treaty accountable item (TAI). All that matters is that a TAI matches a measurement to a reference TAI or “golden copy”. The crux of template-based verification systems (TRIS, NMIS, CIVET, CONFIDANTE,

System	Description (Template/Attribute)	Active Interrogation	Information Barrier
TRIS (Seager, et al., 2001)	Low resolution gamma spectrum <b>template</b>	Passive	Electronic IB
TRADS (Mitchell & Tolk, 2000)	HPGe-based Pu <b>attribute</b> measurement (minimum mass and enrichment)	Passive	Electronic IB
(F)NMIS (Hamel, 2018)	Fast neutron <b>template</b> imaging	Active	N/A
AVNG (Langner, et al., 2002)	Neutron Multiplicity and HPGe-based <b>attribute</b> measurement	Passive	Electronic IB
3G-AMS (Dale, et al., 2009)	HPGe and Neutron slab detector based <b>attribute</b> measurement	Passive	Electronic IB
UKNI (Chambers, et al., 2010)	HPGe-based plutonium <b>attribute</b> measurement	Passive	Electronic IB
INPC (Hamel, 2018)	HPGe-based <b>attribute</b> measurement	Passive	Electronic IB
CIVET (Vanier, et al., 2001)	HPGe-based gamma spectrum <b>template</b>	Passive	Electronic IB
CONFIDANTE (Marleau & Krentz-Wee, 2020)	Fast neutron coded aperture <b>template</b>	Passive	ZKP
Princeton ZKP (Glaser, Barak, & Goldston, 2014)	Neutron radiography <b>template</b>	Active	ZKP
Princeton/MIT (Hecla & Danagoulian, 2018) (Engel & Danagoulian, 2019)	Nuclear resonance <b>template</b>	Active	Physical Encryption/ZKP

**Table 1:** Summary of Previous Arms Control Verification Systems

Princeton ZKP, Princeton/MIT), then, is the authenticity of the golden copy, and how the inspecting party can attain confidence in the item presented as a golden copy. Owing to the difficulty of certifying a golden copy, this huge consideration is often deferred as part of future work.

In an attempt to address the golden copy obstacle, Hecla and Danagoulian (Hecla & Danagoulian, 2018) propose a method by which a golden copy warhead is selected at random and with minimal notice from a fielded system. Even this approach has many potential pitfalls, as this approach could only work for ground-based ICBMs subject to overhead imagery and persistent monitoring; submarine/ship launched warheads, as well as the myriad bombs, cruise missiles and tactical nuclear munitions are more easily moved and not subject to persistent monitoring by design. During the Intermediate Range Nuclear Forces Treaty (INF), inspection notices gave up to six hours of time to the host country to allow an inspection (Harahan, 1993), ample time for golden copy spoofs to be emplaced. Further, the method proposed by Hecla and Danagoulian was to scan only the pit of a warhead, due to the possibility of neutron and x-ray/gamma ray shielding materials being present in a fully assembled nuclear weapon. The disassembly of a nuclear weapon is a highly sensitive operation and would need to be performed in private, thus allowing the host country to modify (e.g., smash, shield, or otherwise obfuscate the true form and signature) the pit before placing in a black box for subsequent golden copy template generation. Combined with undisclosed and host-controlled anti-mask templates, a flattened and shielded pit used as a golden copy could then allow for simple spoofing of warhead dismantlement.

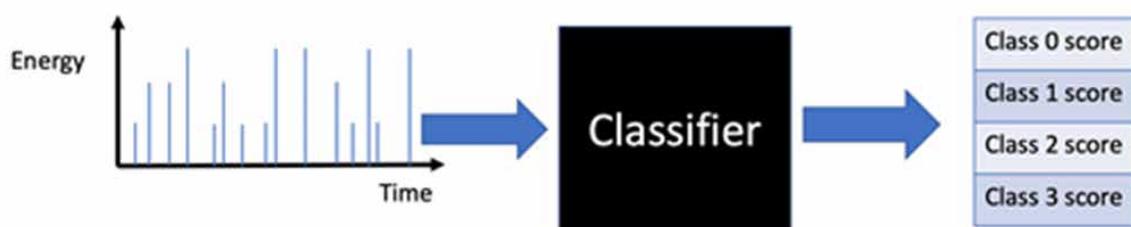
This inherent difficulty in golden copy certification demonstrates the value of attribute verification systems. Instead of blindly comparing two items, these systems seek to verify one or multiple signatures consistent with various characteristics (attributes) of a warhead, such as the presence of weapons grade nuclear material, certain isotopic ratios, geometric extent of intrinsically radioactive material, minimum mass of fissile material, etc. To achieve this, attribute verification systems generally require the measurement and analysis of more sensitive information, such as gamma spectra, neutron multiplicity and/or radiographic imaging.

Therefore, this analysis is generally performed behind an electronic information barrier to protect against the release of sensitive measurement data to an inspector.

While template verification systems can more easily limit the generation of sensitive information, they are completely reliant on the veracity of the golden copy template, creating a single point failure. On the other hand, attribute verification systems can be designed to confirm the veracity of TAIs (or even golden copies themselves), while relying on electronic information barriers and more rigorous authentication and certification needs. Depending on the specific treaty regime and agreed upon implementation protocols, either a template, attribute or combined approach may be most effective.

When considering passive measurements versus active interrogation, the simplest proposed solution to address the needs of the verification regime is more likely to result in successful implementation negotiations, as seen in INF negotiations (Harahan, 1993). As nuclear arms control reduction treaties progress from New START-like treaty regimes (absence verification), more intrusive inspection approaches are likely to be necessary. If the nature of nuclear arms control treaties follows a progressive track towards complete global nuclear disarmament, solutions spanning multiple levels of intrusiveness and complexity will be required. It follows that the complexity of system hardware is directly proportional to the level of intrusiveness of the inspection technology, and also to the difficulty of performing authentication and certification on inspection equipment. Active interrogation systems will need to be authenticated by the inspection team and certified by the hosts as whole, meaning additional effort for developing trust in imaging sources (linacs, nuclear reactors, x-ray generators, etc.) will have additive effort and the potential for reduced trust as they introduce more attack vectors (each piece of hardware must be authenticated and certified down to individual electronic components. (Greenberg, 2019))

When designing information barriers, having the IB further to the left (Figure 1) lowers the number of potential vectors for sensitive host information exfiltration. Once the sensitive information is stripped out it cannot be regenerated. Thus, from a host perspective, pushing the IB as far to the left as



**Figure 2:** List-Mode Linear Classifier Architecture

the verification process allows is desirable, and incorporating redundant IB's of independent design will add trust.

In contrast, an inspector may gain higher confidence in a measurement by performing rigorous analysis on the raw signatures of TAI's with an IB as far to the right as possible, depending on the verification technologies involved. These competing design constraints result in the development of vastly different approaches to the challenge of nuclear warhead verification.

Instead of using an electronic information barrier to separate sensitive data from an output display (far right in Figure 1), the method we propose is inherently limited in the amount of information it collects. Individual gamma ray detection pulses from a detector are input into the linear classifier individually, and only four floating point values are saved (Figure 2). The gamma ray spectrum, which is sensitive information, is never collected.

Figure 2 illustrates our linear classifier system architecture, which will ingest a pre-defined number of pulses in list-mode, storing only running scores for a small number of classes.

## 1.2 Scope

The concept of operation for this method in a treaty verification scenario is that a spectroscopic gamma detector system would be developed to run exclusively in list-mode operation and set to process a pre-defined number of pulse events sequentially. Ingesting a set number of pulses is a key normalization function allowing for source strength information to be largely obviated and relevant radiation signatures appropriately weighted. However, administrative controls for minimum and maximum count rates would be necessary to guard against highly shielded sources or detector saturation, respectively. During a verification process, the detector system would be set up to measure the treaty accountable item, and at the end of collection the

highest class score would be used to determine the type of item being measured (Figure 2).

This paper does not directly address authentication and certification concerns, as that will be done in future work. The primary goal of this paper is to present a novel information barrier and algorithmic approach to warhead verification. As discussed in the previous section, there has never been a complete, end-to-end verification technology solution to the many problems posed by nuclear arms reduction treaties; many systems have been developed to address specific issues at various points in a more broadly comprehensive nuclear arms control treaty. This system is envisioned as a flexible option capable of tailored attribute measurement.

## 2. Approach

The necessarily transparent nature of nuclear arms control verification research and development often requires the use of publicly available and non-sensitive datasets. While more constrained (and thus potentially more sensitive) datasets might yield better algorithm performance, the ability to co-develop and share methods and approaches is highly prioritized in the arms control verification research community. For this initial proof of concept, our team used an algorithmic approach to generate synthetic spectra, which were fed into a linear classifier described in the following sections.

### 2.1 Data Generation

GADRAS, a software suite developed to perform detector response modeling, is used to generate realistic gamma-ray spectra for a multitude of potential detectors to nearly any radiological source of interest (Thoreson, et al., 2019). With ongoing development for over three decades, the built-in library of radioisotopic sources is robust, and rapid radiation transport modeling allows users to generate

Material	Very Highly Enriched Uranium	Highly Enriched Uranium (20-85%)	Weapons Grade Pu	Reactor Grade Pu 33 MWd/kg	Reactor Grade Pu 65 MWd/kg	$^{233}\text{U}$	Am	Np
<b>Composition (weight %)</b>	$^{234}\text{U}$ , 0.70 $^{235}\text{U}$ , 85-92 $^{236}\text{U}$ , 0.3 $^{238}\text{U}$ , rest	$^{234}\text{U}$ , 0.70 $^{235}\text{U}$ , 20-85 $^{236}\text{U}$ , 0.3 $^{238}\text{U}$ , rest	$^{236}\text{Pu}$ , 5e-9 $^{238}\text{Pu}$ , 0.015 $^{239}\text{Pu}$ , 93.63 $^{240}\text{Pu}$ , 6.0 $^{241}\text{Pu}$ , 0.355	$^{236}\text{Pu}$ , 3e-8 $^{238}\text{Pu}$ , 1.2 $^{239}\text{Pu}$ , 59.0 $^{240}\text{Pu}$ , 24.0 $^{241}\text{Pu}$ , 11.8 $^{242}\text{Pu}$ , 4.0	$^{236}\text{Pu}$ , 4e-8 $^{238}\text{Pu}$ , 4.6 $^{239}\text{Pu}$ , 49.36 $^{240}\text{Pu}$ , 23.92 $^{241}\text{Pu}$ , 12.49 $^{242}\text{Pu}$ , 9.63	$^{232}\text{U}$ , 3e-4 $^{233}\text{U}$ , rest	Am	Np
<b>Age (y)</b>	0 - 65	0 - 65	0 - 20	0 - 20	0 - 20	0 - 5	0 - 20	0 - 20
<b>Mass (kg)</b>	1 - *	1 - *	0.5 - 10	1 - 13	1 - 13	1 - 16	1 - 60	1 - 60
<b>Density (g/cc)</b>	18.95	18.95	15.75	15.75	15.75	18.95	12.0	20.45

**Table 2:** Fissile materials and their associated parameters (Nelson & Sokkappa, 2008)

simulated spectra for fairly complex sources. Users can specify radiation emitting materials as well as shielding material layers in arbitrary configurations. The catalogue of training data used in this study comprises two principal classes of simulated sources: nuclear material and nuisance sources.

A method for generating random sources containing various forms of nuclear material is described by Nelson and Sokkappa (Nelson & Sokkappa, 2008). Following the algorithm for generating nuclear threat objects in this “Spanning Set” paper, tens of thousands of randomly generated fissile and fissionable material objects were created as GADRAS 1D models and transported to produce simulated gamma ray spectra. Material age and isotopic ratios were sampled as prescribed by the algorithm and outlined in Table 2. Some targeted model generation was performed to allow for class balanced training, e.g., the branching ratio specified for models containing two layers of fissile material was 10%, and of these many were supercritical and therefore not usable. Table 2 – Fissile materials and their associated parameters (Nelson & Sokkappa, 2008)

Nuisance sources encompass 184 radionuclides contained in GADRAS’s built-in library; most commonly-known medical, industrial, and natural radioisotopes are available for simulating detector responses. These radionuclides were randomly selected and grouped up to three at a time, in varying activities from 10  $\mu\text{Ci}$  to 1 mCi. Modeled as point sources, these mixed isotope “cocktails” were then placed inside randomly generated layers of shielding as prescribed in the “Spanning Set” paper (Nelson & Sokkappa, 2008). The product of this data generation process is a continuously growing library (over 90,000 spectra at the time of writing this paper) of highly-realistic gamma ray spectra representing a very diverse set of medical, industrial and nuclear radiological sources of varying strength and shielding configurations.

## 2.2 Linear Classifier

The requirement to not store a full spectrum, even temporarily, necessitates processing each pulse as it arrives to the classifier. Instead of constructing a spectrum – summing the data in energy bins before it enters the classifier – we instead apply classifier weights to each pulse, keeping a running sum of the classifier’s output. Notably, this is incompatible with typical classification algorithms such as neural networks with nonlinear activations, because for a nonlinear function  $\sigma$ , the function of a sum is not necessarily equal to the sum of the function for scalar inputs  $x_i : \sigma\left(\sum_i x_i\right) \neq \sum_i \sigma(x_i) \forall x_i$  for scalar inputs  $x_i$ . List-mode processing can be done with models that have no nonlinear elements, such as linear classifiers.

## 3. Theory

A linear classifier is a linear mapping of inputs  $x$  to output scores  $y$ , which can be described in terms of vectors representing sets of data and outputs as:

$$\vec{y} = \mathbf{W}\vec{x} + \vec{b} \quad (1)$$

where  $\vec{x}$  is a vector of inputs,  $\mathbf{W}$  is a matrix of weights,  $\vec{b}$  is a vector of biases, and  $\vec{y}$  is a vector of output scores. The weights and biases are tunable parameters, which are trained using an optimization algorithm such as stochastic gradient descent. During inference, the predicted output class is determined by the index of the maximum value in the vector of output scores,  $\vec{y}$ . The desired output ( $\hat{y}$ ), also called the ground truth, is represented by a vector of zeros with 1 in the index of the true class.

Linear classifiers have the advantage of being highly interpretable, which is useful in an arms verification context. The input-output mapping is plainly shown by the weights. There are several complementary interpretations of the weights: the first is that they define templates onto which inputs are projected. A dot product of an input onto a template that is similar to it results in large output magnitude. The second interpretation is that the weights and biases are slopes and intercepts of decision planes in feature space, in which every input is a point. The planes make binary separations of the points into classes.

To train machine learning models, data is separated into a training set with which the model’s weights and biases are adjusted, a validation set used to monitor the model’s performance during training, and a test set used to measure the model’s final accuracy. The weights and biases are randomly initialized. The output of each training example is computed, and a loss function quantifies the error between the computed output scores and the ground truth (the desired output). A common choice of cost function in a classification task is categorical cross-entropy loss. First, the softmax function,  $s$ , (also called the normalized exponential function) is computed over the scores,  $y$ , converting them to normalized probabilities:

$$s_i = \frac{e^{y_i}}{\sum_{j=1}^C e^{y_j}} \quad (2)$$

where the sum is taken over C output classes. Then the cross-entropy (CE) between the softmax output and the ground truth is computed:

$$CE = -\sum_{i=1}^C \hat{y}_i \log(s_i). \quad (3)$$

Because  $\hat{y}$  is a vector in which there is a single nonzero entry of value 1, the sum can effectively be removed from the equation, thereby making equation (4) the full categorical cross-entropy loss function, where  $y_{true}$  is the output in the index of the true class as given by  $\hat{y}$ .

$$L = -\log \left( \frac{e^{y_{true}}}{\sum_{j=1}^C e^{y_j}} \right) \quad (4)$$

In training, the weights and biases are adjusted to minimize the loss function by taking steps in the direction of the downward gradient of the loss function with respect to each tunable parameter. Despite the simplicity of our model, optimization is difficult because the inputs to the linear classifier are poorly-conditioned: gamma-ray spectra have significant differences in the orders of magnitude of their input features, and due to constraints of our algorithm, no nonlinear pre-processing transformations are permissible. The Adam optimizer (a form of gradient descent optimization) (Kingma & Ba, 2014) was selected for this application because of its individual adaptive learning rates for every parameter, with an initial learning rate of 0.1, for 100,000 epochs.

### 3.1 Equivalence of Linear Classifier Inference on Binned and List Mode Data

Assume that at some time  $t > 0$  we have recorded  $p$  pulses. Let  $v_i \in [1, N]$  be the bin associated with pulse  $i$ , for  $i = 1, \dots, P$ . Let  $e_{V_i}$  be a vector of length  $N$  that is zero except for the  $v_i$ -th entry, which is one. Let our energy spectrum  $x$ , be defined as  $x = \sum_{i=1}^P e_{v_i}$  which is the count of the pulses in each bin. Finally, let our linear model be defined as  $y = Wx + b$ . Then,

$$y = Wx + b, \quad (5)$$

$$= \sum_{i=1}^P e_{v_i} + b, \quad (6)$$

$$= \sum_{i=1}^P We_{v_i} + b, \quad (7)$$

which means we can apply the  $W$  portion of the linear model to each individual pulse, rather than the whole accumulated  $x$  vector and get the same answer.

Therefore, a linear classifier trained on spectra may perform inference on a spectrum, or inference on list-mode data while keeping a running sum of the outputs, and the results will be the same.

## 4. Experiment

For this initial study, the standard detector response function for an ORTEC Detective EX-100 HPGe was used, with all spectra including default Albuquerque, NM natural background radiation. All spectra generated for model training were ideal, without Poisson noise; the impacts of varying the background and the counting statistics were not considered as part of this study. In general, this effect can be mitigated through administrative controls requiring the object of interest count rate to exceed a minimum threshold value based on background count rates ( $3\sigma$  is a commonly used multiplier).

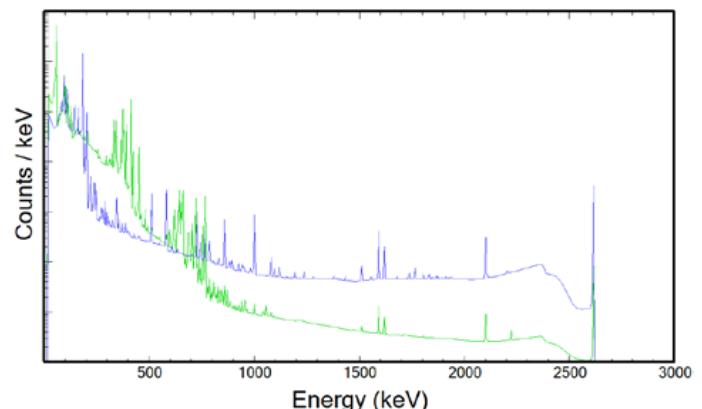
To test our approach, we compared models containing weapons-grade material, defined for this study as 94% Pu-239 and greater than or equal to 90% enriched U-235 (weight percentages) to models containing reactor-grade material to determine if we could discriminate the different material types. Weapons-grade-containing models were discriminated from models containing reactor-grade material, highly enriched uranium (HEU) material just below the arbitrary threshold of weapons-grade used in this study, and standard radiological sources such as industrial and medical isotopes. Class 1 contains the 90%+ U-235 samples, class 2 contains the 94% Pu-239 samples, and class 3 contains samples with a combination of uranium and plutonium layers, where at least one shell layer is weapons grade. All other samples, whether sub-threshold or containing only industrial and/or medical isotopes are defined as class 0.

Our training data consists of 41,595 samples, of which 10% are used for validation, and our testing set consists of 5,136 samples. Every spectrum in the dataset is normalized so that the features (in this case counts in each channel bin) sum to 1. There are 8127 features, spanning energies from 20keV – 3.27 MeV (the default bin structure of the Detective EX-100 is 8192 channels, though 65 were below the lower-level discriminator and thus excluded from our optimization and training processes.

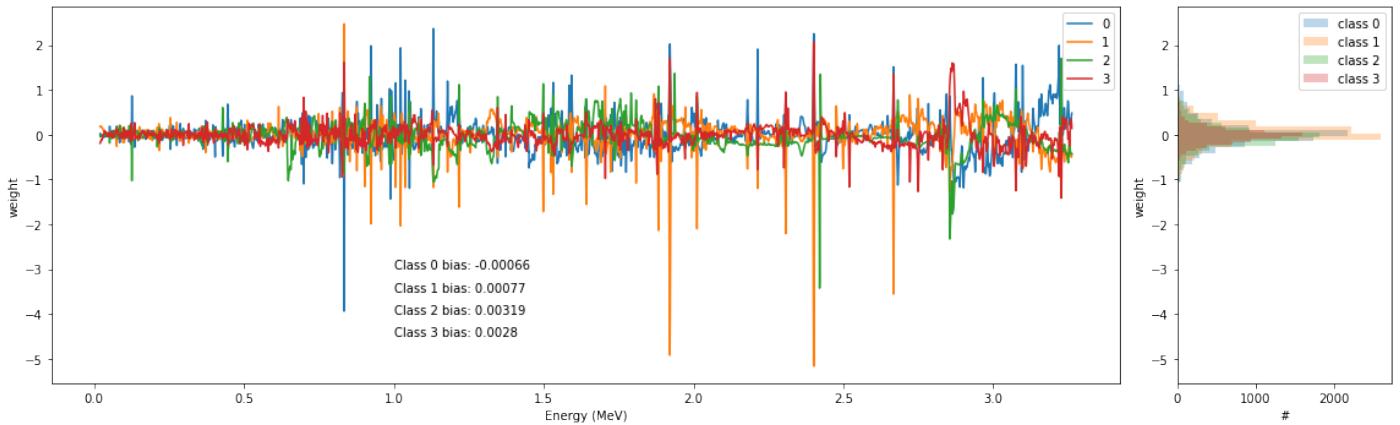
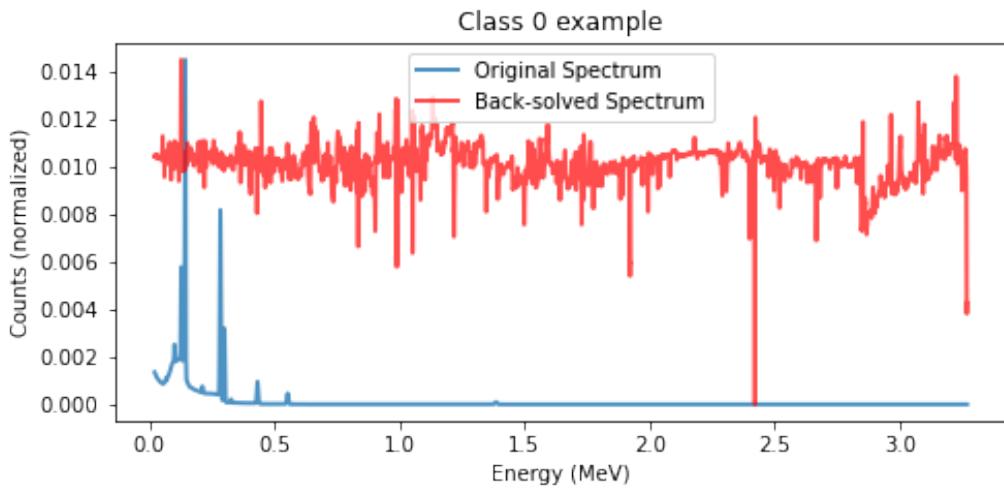
After the training and loss curves fully converged, the validation set accuracy was compared to other commonly available machine learning models included in MATLAB's classification learner (The MathWorks, Inc., 2022).

## 5. Results

Primary emissions from HEU (blue) and Pu-239 (green) are illustrated in Figure 3, which are representative, ideal, plots from GADRAS of both weapons grade plutonium (WGPU) and HEU sources as would be measured with standard Albuquerque, NM USA terrestrial and cosmic background.



**Figure 3:** Example Gamma Spectra from HEU (Blue) and WGPU (Green)

**Figure 4:** Weights and Biases of Linear Classifiers**Figure 5:** Example Reconstructed Gamma Energy Spectrum

The methods described in the previous section were used to optimize the weights and biases for this dataset, shown in Figure 4. Weights are plotted against energy bins for each class on the left, with the distribution of weights plotted on the right.

Due to the large variation in magnitudes of the features and the inability to apply any nonlinear preprocessing techniques (incompatible with list-mode data), convergence of the model was extremely slow; 100,000 epochs were run to achieve the results presented in this paper.

An important feature of an information barrier is the concept of irreversibility, wherein the sensitive input signatures cannot be reconstructed given information to which an inspector may have access. In this proposed system, that would include the classifier's outputs and linear classifier weights. Figure 5 shows an example spectrum that was reconstructed by multiplying the classifier's output by the weights:

$$\vec{x} = \mathbf{W}^+ (\vec{y} - \vec{b}) \quad (8)$$

Where  $\mathbf{W}^+$  is the Moore-Penrose pseudo-inverse of the weight matrix. The back-solved spectrum is then rescaled to match the maximum and minimum of the original spectrum for convenient visual comparison. The results shown in Figure 5 are representative of all examples visualized; the input spectrum or pulse train is unrecognizable from the backwards reconstruction.

Accuracy results on the test set are summarized in Table 3, with our linear classifier confusion matrix shown in Figure 6. We show class-weighted accuracy measures as well as two measures specific to our dataset. The first study-specific accuracy, red/green, measures the binary classification accuracy of natural, industrial, medical, and sub-threshold special nuclear material (SNM) sources of radiation (class 0) versus all weapons grade nuclear material as defined in this study (classes 1, 2 and 3). The second study-specific accuracy, Class 3f, is a 4-class classification which "forgives" any misclassification of class 3 material (containing layers of both uranium and plutonium with at least one of them weapons grade) as class 1 or class 2. The logic here is that a sample which contains WGPu nested outside of HEU may preferentially self-shield the

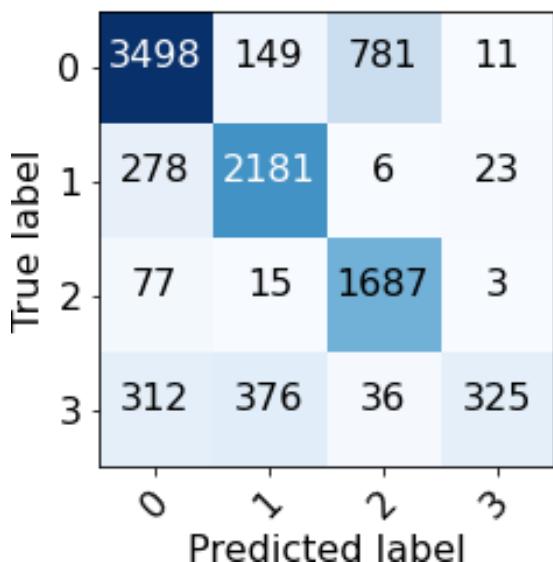
emissions from uranium and therefore appear to contain only plutonium. The class 3f accuracy measure considers such a classification as correct instead of erroneous. These caveated accuracy results are relevant to a notional treaty verification regime in which the inspector may only care whether an object contains weapons grade nuclear material or not.

The Tree and k-nearest neighbors (KNN) models are computed using MATLAB's classification learner app (The MathWorks, Inc., 2022) and consists of all of the "quick to train" models available in the app. None of the MATLAB models are compatible with list mode data, but we have included them for the sake of comparison to illustrate relative performance of our linear classifier to existing mature classifiers without the additional self-imposed limitations of this application.

Accuracy:	Class-weighted	Red-Green	Class 3f
fineTree	70.40	83.07	86.87
medTree	65.28	78.62	84.82
coarseTree	54.66	74.15	76.92
fineKNN	65.94	79.87	77.22
medKNN	67.16	81.63	82.46
coarseKNN	66.07	80.51	84.70
cosineKNN	67.14	81.82	82.15
cubicKNN	67.21	82.04	82.65
weightedKNN	68.97	82.09	83.11
linear classifier	73.03	83.13	82.85

**Table 3:** Accuracy Results for Associated Classifiers

The linear classifier presented in this paper achieved the highest class-weighted and Red-Green accuracy scores, while achieving slightly above average for the class 3 forgiving score.



**Figure 6:** Linear Classifier Confusion Matrix

From the confusion matrix (Figure 6), it is apparent that the largest single source of error in our linear classifier comes from misclassifying "other" sources (class 0) as 90%+ highly enriched uranium (class 2). This is somewhat expected, in that HEU is a relatively low-intensity source with most gamma emissions in the sub-200 keV energy range; with a minimal amount of shielding, HEU can be very hard to detect and therefore identify reliably. Further, due to the prescribed structure of the Spanning Set data generation algorithm, HEU enrichment was varied linearly from 20-92%, meaning much of the class 0 data is sub-threshold HEU (89% or less) with nearly identical signatures to 90-92% enriched HEU.

## 6. Conclusions and Future Work

We have shown an inherently information-limited method to classify radioactive sources using a linear classifier performing inference on list-mode gamma ray data. A sensitive spectrum is never collected, and the input cannot be reconstructed from the values that are stored.

Our results show 83% classification accuracy in distinguishing weapons grade nuclear material (as defined here) from nuisance sources, which include special nuclear material and thousands of combinations of medical and industrial isotopes. This initial result is a promising indicator that our algorithm will perform well with further refinement. Particularly interesting would be a closed-loop data generation method to maximize generation of spectra on the decision boundaries, therefore generating data that maximally improves the model.

Beyond additional data generation, there are opportunities to add complexity to the linear model to enable a more complicated decision surface defining class boundaries. One option is to include additional output classes, possibly by an unsupervised clustering of existing data into similar groups. More sophisticated non-invertible list-mode-compatible architectures also hold promise, such as an autoencoder with list-mode encoder and nonlinear decoder, storing intermediate values between them.

From a signature verification perspective, increased performance is expected when adding other radiation detection modalities such as neutron counting or multiplicity, potentially in addition to further constraining the class definitions to include attributes such as minimum mass of weapons grade material or the presence of high explosives. Substituting gamma-only scintillators with lithium loaded neutron-sensitive inorganic detectors such as  $\text{Cs}_2\text{LiYCl}_6:\text{Ce}$  (CLYC) or  $\text{Cs}_2\text{LiLaBr}_{6-x}\text{Cl}_x:\text{Ce}$  (CLLBC) is also of interest. Parametric studies investigating background variation, statistical sampling in list mode (accuracy vs. total counts), minimum mass of SNM and detector resolution are also being pursued.

## 7. Acknowledgements

This work was supported by Laboratory Directed Research and Development funding at Sandia National Laboratories.

## 8. References

1. Chambers, D. M., & al, e. (2010). UK-Norway Initiative: Research into Information Barriers to all warhead attribute verification without release of sensitive or proliferative information. 51st Annual Meeting of the Institute of Nuclear Materials Management. Baltimore, MD, USA.
2. Dale, C., & al, e. (2009). Third-Generation Attribute Measurement System Conceptual Design Report. Los Alamos, NM, USA: Los Alamos National Laboratory.
3. Engel, E. M., & Danagoulian, A. (2019). A physically cryptographic warhead verification system using neutron induced nuclear resonances. *Nature Communications*(10).
4. Evans, D. (2021). Strategic Arms Control Beyond New START - Lessons from Prior Treaties and Recent Developments. Johns Hopkins University Applied Physics Laboratory LLC.
5. Glaser, A., Barak, B., & Goldston, R. J. (2014). A zero-knowledge protocol for nuclear warhead verification. *Nature*, 510, 497-502.
6. Greenberg, A. (2019, October 10). WIRED. (Condé Nast) Retrieved January 23, 2023, from <https://www.wired.com/story/plant-spy-chips-hardware-supermicro-cheap-proof-of-concept/>
7. Hamel, M. (2018). Next-Generation Arms-Control Agreements Based on Emerging Radiation Detection Technologies. Institute of Nuclear Materials Management 59th Annual Meeting. Baltimore, MD, USA.
8. Harahan, J. P. (1993). On-Site Inspections Under the INF Treaty, A History of the On-Site Inspection Agency and Treaty Implementation, 1988-1991. Washington, DC USA: Library of Congress Cataloging-in-Publishing Data.
9. Hecla, J. J., & Danagoulian, A. (2018). Nuclear disarmament verification via resonant phenomena. *Nature Communications*, 9.
10. Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations. San Diego.
11. Langner, D., & al, e. (2002). Attribute Measurement Equipment for the Verification of Plutonium in Classified Forms for the Trilateral Initiative. Vienna, Austria: International Atomic Energy Agency.
12. Marleau, P., & Krentz-Wee, R. (2020). CONFIDANTE Demonstration Prototype Report. Albuquerque, NM, USA: Sandia National Laboratories.
13. Mitchell, D. J., & Tolk, K. M. (2000). Trusted Radiation Attribute Demonstration System. INMM 41st Annual Meeting. New Orleans, LA, USA.
14. Nelson, K., & Sokkappa, P. (2008). LLNL-TR-408407 A Statistical Model for Generating a Population of Unclassified Objects and Radiation Signatures Spanning Nuclear Threats. Livermore, CA: Lawrence Livermore National Laboratory.
15. Seager, K. D., & al, e. (2001). Trusted Radiation Identification System. Albuquerque, NM, USA: Sandia National Laboratories.
16. The MathWorks, Inc. (2022). Classification Learner. Retrieved April 4, 2022, from <https://www.mathworks.com/help/stats/classificationlearner-app.html>
17. Thoreson, G., Horne, S., Theisen, L., Mitchell, D., Harding, L., & Amai, W. (2019). SAND2019-14305 GADRAS Version 18 User's Manual. Albuquerque, NM: Sandia National Laboratories.
18. Vanier, P. E., & al, e. (2001). Study of the CIVET Design of a Trusted Processor for Non-intrusive Measurements. Annual Meeting of the Institute of Nuclear Materials Management. Indian Wells, CA, USA.
19. White, G. (2012). Review of Prior U.S. Attribute Measurement Systems. Livermore, CA, USA: Lawrence Livermore National Laboratory.
20. Wolford, J. K., & White, G. K. (2000). Progress in Gamma Ray Measurement Information Barriers for Nuclear Material Transparency Monitoring. Institute of Nuclear Materials Management 41st Annual Meeting. New Orleans, LA, USA.
21. Yan, J., & Glaser, A. (2015). Nuclear Warhead Verification: A Review of Attribute and Template Systems. *Science & Global Security*, 23, 157-170.

# Dynamic Network Analysis of Nuclear Science Literature for Research Influence Assessment

Samrat Chatterjee, Dennis Thomas, Daniel Fortin, Karl Pazdernik, Benjamin Wilson, and Lisa Newburn

Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99354 USA,

E-mail: samrat.chatterjee@pnnl.gov, dennis.thomas@pnnl.gov

## Abstract:

Analyzing nuclear science literature via data-driven methods is key for assessing research influence and technology advancements. Indicators of scholarly activities may be buried in publications and collaboration networks over time. Mining for relevant scholarly influence trends in large volumes of text can be computationally challenging; however, open-source information over time can offer opportunities to extract meaningful insights. While network centrality analysis of scholarly research provides topology-based insights, additional emphasis on dynamics associated with information diffusion through these networks is important. This paper represents a step in that direction through the development of a novel dynamic network analysis framework and computational engine to identify key entities and capabilities over time within global scholarly nuclear science collaboration networks. Network theoretic, stochastic simulation, and optimization methods are used to analyze variability in scholarly interactions, influence propagation, and collaboration patterns. A topic-aware influence maximization algorithm is developed to identify key influential authors over time, along with an efficient parallelized implementation to reduce computational costs. A case study using open-source Scopus data with 33,517 published nuclear research papers from 2000-2019 is presented and representative analytic insights are generated. Broad implications of these insights are discussed, and future research directions are also identified.

**Keywords:** nuclear; collaboration; network; topic; influence

## 1. Introduction

Activities such as publishing nuclear research and expanding scholarly networks include indicators of research influence – defined as the ability to have the greatest reach (or spread) in a scholarly network – and technology advancements. These types of indicators, often buried in large volumes of technical publication data, may reveal spatiotemporal patterns associated with a dynamic research collaboration landscape. Recent advances in data science methods and computational platforms, combined with nuclear domain knowledge, might provide the appropriate mix of analytic tools that can generate key research influence insights. While data-driven learning technologies are promising for analyzing patterns in large volumes of text, applying these methods to a research influence assessment problem over time can be computationally challenging. Readily available open-source information on research collaborations, such as journal papers and technical reports, can offer insights into an evolving nuclear research and technology domain. However, there is a need for network-theoretic techniques to better exploit time-varying metadata from publications, including authorship, collaboration, and topics of interest.

In this paper, a dynamic network analysis framework is presented for addressing the challenge of identifying key entities and capabilities in nuclear research networks. An entity may be broadly defined as an author, organization, or state. The focus here is on author-level collaboration networks based on open-source publication metadata over a 20-year time period from 2000-2019. Research goals comprise of: (1) identifying key network influencers based on nuclear research topics, (2) comparing network topology-based measures with information diffusion-based outcomes over time, and (3) characterizing influential author collaboration dynamics, including persistence and emergence of connections. Data-driven approaches implemented to meet research goals include network construction, topic modeling, centrality analysis, information diffusion, influence maximization, and temporal dynamics analysis. A case study application is also presented.

The remainder of this paper is organized as follows. We first present a brief background on network analysis of nuclear science literature. Thereafter, we describe the

influential entity identification problem. Next, a modular data-driven dynamic network analysis framework is presented including network construction, topic modeling, topic-aware influence maximization, and temporal dynamics analysis. This is followed by a case study application. Finally, concluding remarks and steps for future research are included.

## 2. Network analysis of nuclear science literature

Nuclear technology capability development and transfer can include physical items of trade as well as knowledge shared in scientific networks by researchers (Molas-Gallart, 1997). Detecting early signs of proliferation activities, such as based on analysis of text-based data from scientific publications, may provide additional intervention options further “left of boom” (Sheffield, 2020). Analyzing scientific literature for evaluating a State’s nuclear activities also leads to diversification of information sources for comparison to generate safeguards conclusions (Feldman et al., 2013). Further, the International Atomic Energy Agency (IAEA) Physical Model represents a consolidated framework for data fusion and analysis that also includes areas of nuclear research and development (Liu & Morsy, 2007). As a result, analysis of research networks may be used to discover collaboration communities and influencers engaged in nuclear fuel cycle related research and development (Iancu, Wilson, Calle, & Gagne, 2018).

Network analysis and text mining techniques have been applied before to analyze scientific networks for nuclear capabilities assessment (Kas et al., 2012; Stewart et al., 2018; Diab et al., 2018; Iancu et al., 2018; Goldblum et al., 2019). Kas et al. (2012) developed a text-mining tool to construct a terminology thesaurus based on nuclear physics research with over 20,000 articles. This tool uses citation networks to identify key entities – individuals, organizations, and nation states – engaged in nuclear research and topics of interest by mapping key terms to capabilities. Weighted citation networks have also been studied in the biomedical domain under the assumption that all cited papers may not have equal influence on the publication of interest (Kim et al., 2018). Weights are typically based on topic similarity and relative importance of a paper in terms of content.

Stewart et al. (2018) describe a data collection, fusion, storage, analysis, and visualization architecture using open-source information for nonproliferation applications. Algorithmic methods that were developed include natural language processing, knowledge and ontology-graph based approaches, link analysis, and geoparsing. Diab et al. (2018) describe a natural language processing approach for identifying key terms that distinguish uranium from other mining processes. Iancu et al. (2018) model publication co-author relations, patent ownership, and organizational affiliation with a directed multigraph. More recently, Goldblum et al. (2019) describe a multiplex network science

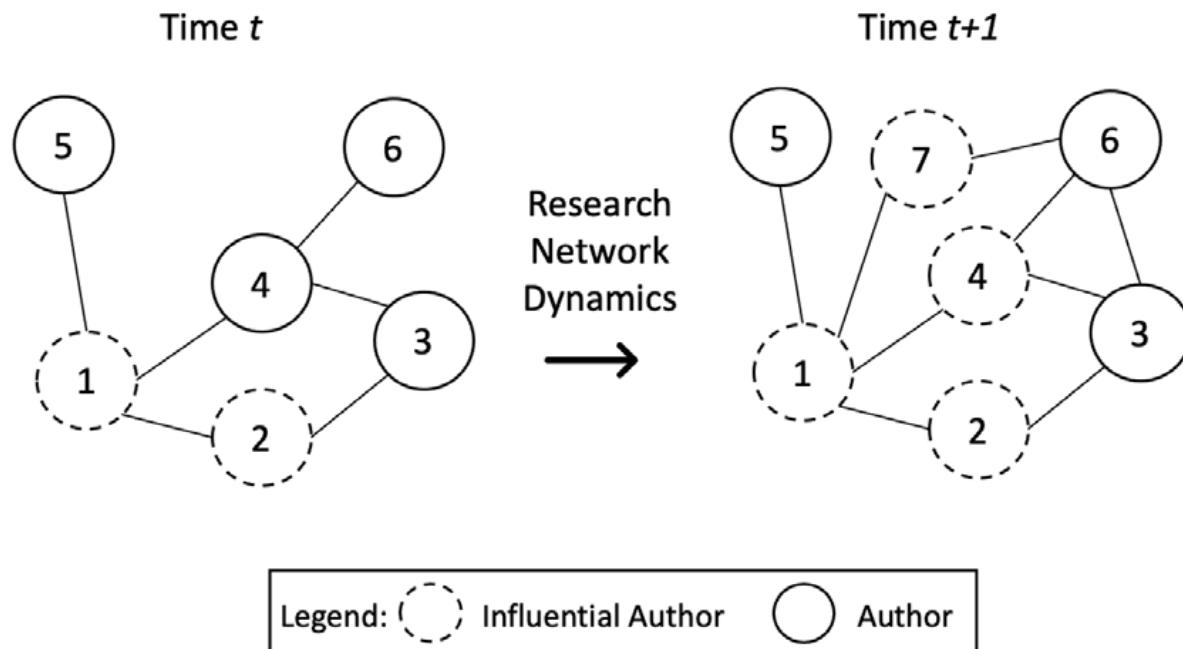
framework for modeling state proliferation decisions using trade, conflict, alliance, and cooperative agreement networks. A policy-oriented aggregated proliferation metric over time was defined based on measures of centrality and correlation across network layers.

Network analysis methods above typically rely on topology-based measures of centrality over time to identify anomalies and key researchers. However, research influence assessment requires additional emphasis on the role of dynamics associated with these networks. Specifically, research network dynamics in terms of context or topic-based information diffusion and evolving collaboration behavior. This paper represents a step in that direction and presents network-based algorithms and insights from a case study application with focus on dynamic research network analysis.

## 3. Influential entity identification problem

Identifying influential entities of interest (i.e., author, organization, or state) as well as the evolution of their capabilities, in a computationally efficient manner, within a dynamic scientific research network setting is a challenging problem. Temporal patterns in collaboration networks can contain significant information on sequencing of events and evolution of technology advances. Figure 1 presents a conceptual illustration of research network dynamics and the evolution of influential authors. At a time  $t$ , nuclear technology capability of an entity of interest may be reflected via the prominence of authors within a nuclear research collaboration network. On the other hand, the interest in pursuing research and development on a particular topic most likely is motivated by the desire for advancing technology capability. As illustrated here at time  $t+7$ , with an evolving research landscape, the technology capability of an entity might continue to grow as might be evident from increased number of prominent authors within a research network.

The focus of this study is on developing data-driven algorithms and computational pipelines that may be useful for identifying potential influential entities and their capabilities over time. Using a global nuclear research collaboration network over a 20-year time period, three research questions were defined: (1) how to identify key network influencers based on evolving research topics over time? (2) how does network topology-based measures of centrality compare against information diffusion dynamics-based outcomes? and (3) how to characterize influential author collaboration dynamics (in terms of persistence and emergence of connections)? A modular dynamic network analysis framework, developed to address these research questions, is described next.



**Figure 1:** Conceptual Research Network Dynamics Illustration. Research network elements may evolve – grow, shrink, or remain consistent – over time. Nodes in a research network represent authors and edges represent collaboration among authors on a research paper. A node with dotted lines here represents an influential author. In this illustration, the number of influential authors in the network increases from 2 at time  $t$  to 4 at time  $t+1$ .

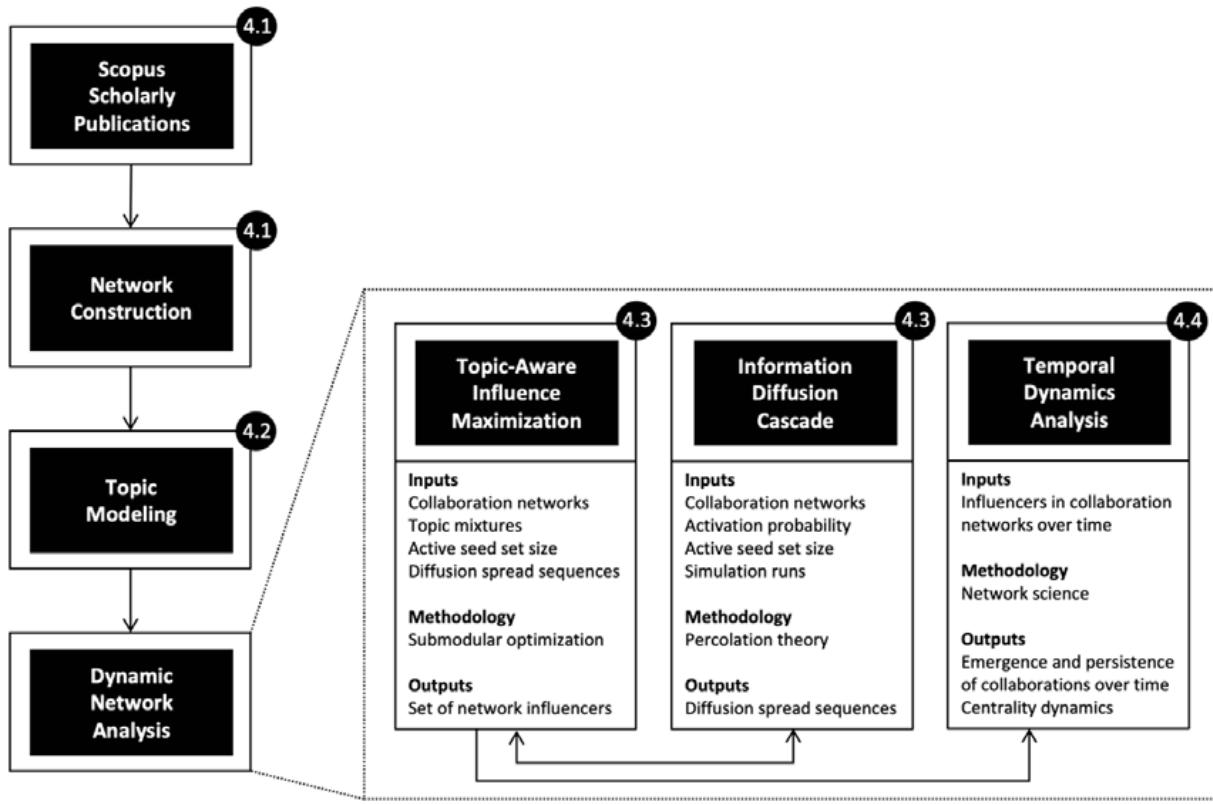
#### 4. Dynamic network analysis framework

Dynamic network analysis is a scientific area of study that fuses concepts from network science, graph theory, network optimization, and stochastic simulation to characterize topology and dynamics associated with networked systems (Newman, 2018; Barabási, 2016; Carley, 2003). Mathematically, a network or graph may be defined as  $G = (V, E)$ , where  $V$  is a set of vertices or nodes and  $E$  is a set of edges or links. A network with order  $n$  (i.e., number of nodes) is specified by the adjacency matrix,  $A$ , an  $n \times n$  square matrix where  $A_{ij}$  indicates a link connecting node  $i$  and node  $j$ . Dimensions of temporal effects on networks may be broadly categorized through: (1) node addition/removal, (2) link addition/removal, (3) dynamic flows across networks, and (4) node/link state transitions. Typically, graph analytic methods for influence and capability assessment model individuals, organizations, or events as nodes, and the relations between nodes as different types of links. For example, collaboration networks represent authors as nodes and joint authorship on a manuscript as a link between two nodes. In this study, analytics from collaboration networks over time were generated to address the research questions above associated with identifying key network influencers, comparing topology-based measures of centrality with diffusion dynamics-based outcomes, and influential author collaboration dynamics. Pairing of methods from network science, simulation, and optimization within a flexible computational environment was accomplished to generate insights about key entities and

capabilities over time. Computational costs associated with various algorithms were also taken into account and parallel computing paradigms were implemented to accelerate simulation runs.

Figure 2 presents our modular dynamic network analysis framework and computational engine. On the left, a step-wise workflow begins from the top with data collection from Scopus scholarly publications on nuclear research followed by author collaboration network construction. This is followed by topic analysis using metadata information such as title and abstract. The resulting network structures were thereafter subject to dynamic network analysis algorithms that led to characterization of key entities and capabilities over time. On the right, an overview of the computational engine for dynamic network analysis is described in greater detail with three connected modules: (1) topic-aware influence maximization, (2) information diffusion cascade, and (3) temporal dynamics analysis.

Methodological details under each of the framework elements in figure 2 including compute modules within dynamic network analysis are described in sections 4.1 to 4.4. Section 4.1 describes network construction steps including data collection and author collaboration network representation. Section 4.2 focuses on topic modeling based on the Non-negative Matrix Factorization (NMF) algorithm. Next, in section 4.3, topic-aware influence maximization algorithm is discussed which is based on submodular optimization and information diffusion cascade simulation. Finally, section 4.4 includes temporal dynamic



**Figure 2:** Modular dynamic network analysis framework. Inputs, methodology, and outputs are linked through analytic information flow pipelines within and across computational modules described in sections 4.1 to 4.4.

analysis based on network centrality measures. The overarching goal of this dynamic network analysis framework and computational engine was to generate insights on key entities and capabilities over time using network-theoretic, stochastic simulation, and optimization methods that address variability in scholarly interactions, influence propagation, and author collaboration patterns.

#### 4.1 Network construction

Steps involved in constructing global collaboration research networks are briefly described below.

##### 4.1.1 Data collection

The data for constructing the global collaboration research networks was obtained from Scopus (Elsevier, 2021). We used the PyScopus Python package (Zuo, 2023) to search and obtain the Scopus identifiers (IDs) of nuclear related articles, published in the years from 2000 to 2019, and then used the Scopus Application Programming Interface (API) to directly search Scopus by the Scopus IDs to obtain full metadata information, particularly, author affiliations and abstracts. The list from pyscopus returned the Scopus ID, title, publication name, ISSN, volume, page range, date, doi, citation counts, publication type, author affiliations (name, city, country), author IDs, and full-text links (incomplete) of each article. The articles were identified by searching for terms such as, *nuclear fuel*, *nuclear energy*, and

*nuclear reactor* in the title, abstract and keyword list of each article record in Scopus. For example, the query used for finding the articles published in the year 2000 was specified as: "TITLE-ABS-KEY('nuclear PRE/0 fuel') OR TITLE-ABS-KEY('nuclear PRE/0 energy') OR TITLE-ABS-KEY('nuclear PRE/0 reactor') AND PUBYEAR IS 2000". The same query format is used for other years. The list, however, did not show one-to-one correspondence between each author and their affiliation; this information was obtained from the full metadata information obtained using the Scopus API. Records from journals that published less than 10 nuclear-related articles in a year were not included in the final dataset. The final dataset contained a total of 33,517 records (articles) in .json format.

Within Scopus, each article has a unique numeric ID for each author. For the publications from the years 2000-2019 used in this study, there were 64,312 authors from around the world. We observed that authors over time may change organizational affiliations or use different names (e.g., initials or full name) in their publications. As a result, an author may get assigned multiple numeric IDs. This may lead to additional nodes (representing author IDs) in the collaboration networks over time. We wrote Python scripts to check for cases where an author had multiple IDs but maintained a single affiliation--there were 393 (or 0.61% of 64,312) such instances among all authors over the 20-year time period in

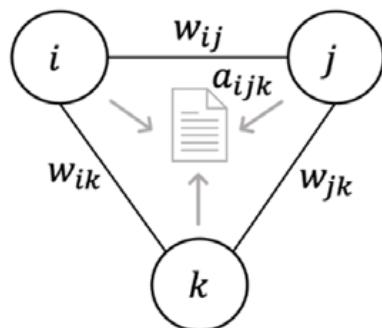
this study. In order to merge or split nodes, additional information is needed that indicates whether an author used different forms of their name or whether two authors may have the exact same name along with affiliation. While additional information related to possible author name resolution can refine our analysis further, it is a non-trivial problem to address for a global-scale network and was outside the scope of this study. Moreover, given the relatively small proportion of instances (0.61%) in this study where an author had multiple IDs but maintained a single affiliation, we did not artificially merge, split, or discard any author information.

#### 4.1.2 Author collaboration network

The author collaboration network is a co-authorship based influence network, where a node represents an author (identified by author id) and an edge represents co-authorship between two authors if they had co-authored at least one article. Figure 3 presents author collaboration network elements and an illustration of network construction when multiple authors collaborate on the same article. The network was constructed using the *NetworkX* Python package (Schult and Swart, 2008). From each record (.json file), we extracted the author names and ids; as well as their institution name, city, and country from their respective affiliations. To quantify the strength or closeness of each pair of co-authors, we calculated Newman-Fowler (NF) weights (Fowler, 2006; Perianes-Rodriguez et al., 2016) using the formula:

$$f_{ij} = f_{ji} = \sum_p \frac{a_{pq}}{n_q - 1},$$

where  $a_{pq} = 1$ , if  $i$  and  $j$  are co-authors of the same publication  $p$ , and 0 otherwise;  $n_q$  is the number of authors of publication  $q$ . The  $-1$  in the denominator  $n_q - 1$  is used to ignore self-links. The NF weights were inverted and assigned as edge weights in the graph, so that edge weights can be interpreted as cost or distance between two co-authors for centrality analysis.



**Figure 3:** Author collaboration network elements. In this illustration, all authors  $i, j$ , and  $k$  (nodes) collaborate on the same article  $a_{ijk}$ , and inverted Newman-Fowler weights,  $w_*$  represent collaboration strength along edges.

Since we are interested in finding authors with maximum influence, we only consider the Giant Connected Component (GCC) of the network for the influence analysis. The GCC is defined as a sub-network that is the largest connected collection of nodes from the original network (Kitsak et al., 2018), and was determined using the *connected\_components* function from *NetworkX*. Although it may be possible that influential authors are present outside of a network's GCC, that possibility was not considered for this analysis. Before creating the GCC, we removed densely and weakly connected components of the network by removing edges with weights above a threshold value of 20, which corresponds to an edge representing one paper co-authored by 21 authors. Papers co-authored by many authors will appear as densely connected networks in the author collaboration graph. These networks may represent relatively weak collaborations and can end up being selected in the GCC. Papers with more than 100 co-authors are also present in our dataset. Therefore, we applied the criterion to remove such densely and weakly connected components by removing edges with weights above a threshold value.

#### 4.2 Topic modeling

Topic modeling was performed using state-of-the-art NMF algorithm (Kuang et al., 2015). NMF is a linear algebra based dimension reduction algorithm where the input is a normalized term frequency-inverse document frequency (TF-IDF) matrix and the outputs are two non-negative matrices representing words by topics and topics by documents. We used the NMF outcomes for subsequent influence maximization analysis.

We implemented the NMF algorithm using the *Scikit-learn* Python package (Pedregosa et al., 2011). The title and abstract of each record were combined and used as the text for the topic modeling. The NMF model was fitted with a maximum of 1,000 features extracted from the text of 33,517 records. Common English stop words and corpus-specific words occurring in the text from only one or two records or from at least 95% of the 33,517 records were removed during feature extraction. The features for the NMF model were extracted using *TfidfVectorizer* function in the Scikit-learn package. A total of ten topic categories (or topics) were obtained. The number of topics (categories) was set to ten, which we found to meaningfully classify the records with minimal overlap. The weights of the ten topics discovered by the NMF model were fitted using *sklearn.decomposition*. The NMF module with *Frobenius* norm minimization and regularization, where the L1 to L2 ratio was set at 0.5, and *alpha*, the constant multiplying the regularization terms, was set at 0.1, to avoid overfitting.

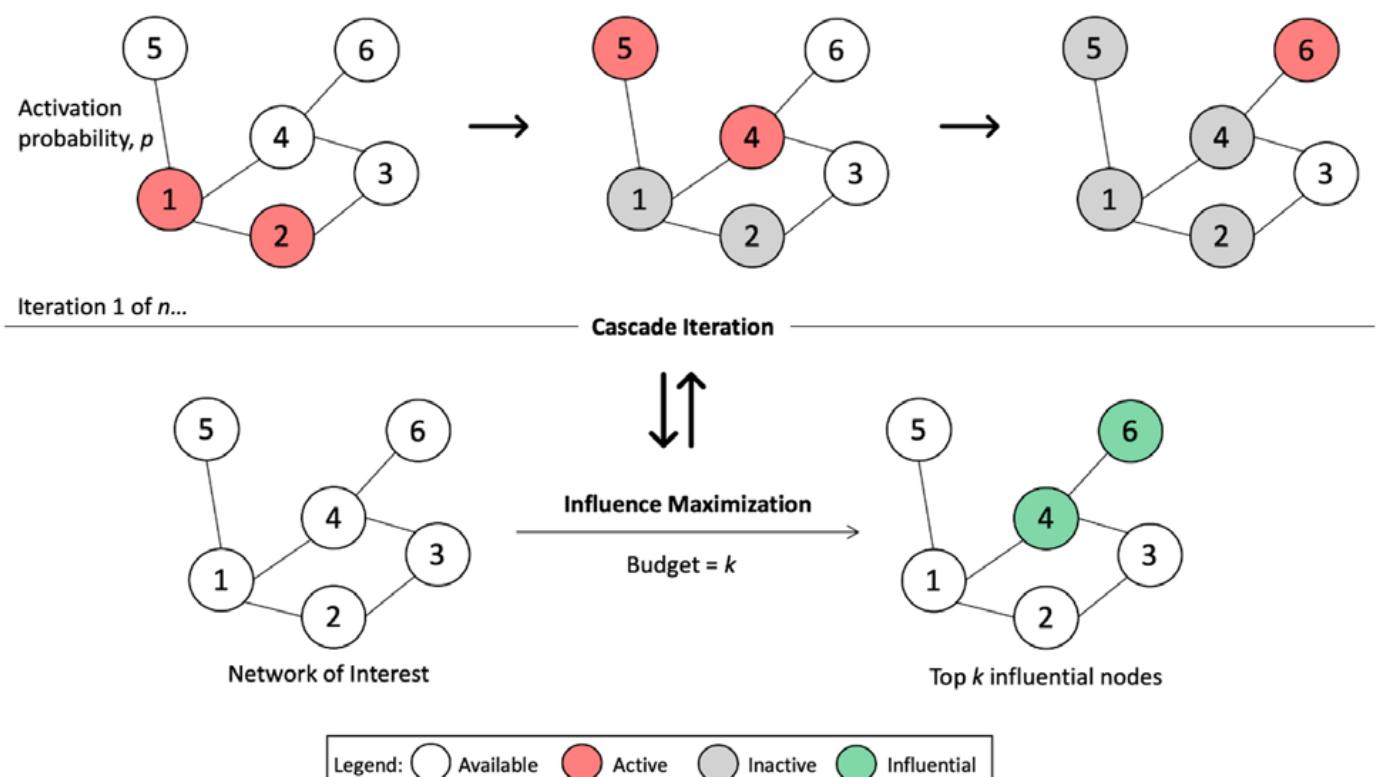
Each topic is a group of keywords (features), where each keyword contributes a certain weight to the topic. The top 10 words in each topic were used to analyze the meaning

of the topic. The model outputs were the weight distributions of the ten topics for each record. These weights were normalized using the formula,  $\nu_i / \sum_{i=1}^{10} \nu_i$ , where  $\nu_i$  is weight of topic  $i$  for a record. The topic with the highest normalized weight was treated as the dominant topic for that record.

#### 4.3 Topic-aware influence maximization

The topic-aware influence maximization method (Chen et al., 2015) was applied to identify the top 5 authors who can influence the information diffusion of a topic mixture in a network. The topic mixture is defined as a vector,  $\lambda = \{\lambda_i | i \in [1, 10]\}$ , where  $\lambda_i$  indicates whether topic  $i$  is to be included ( $\lambda_i = 1$ ) or excluded ( $\lambda_i = 0$ ) in the activation probability calculations. For example,  $\lambda = \langle 1, 0, 0, 0, 1, 0, 1, 0, 1, 0 \rangle$  represents a mixture of topics 1, 5, 7, and 9. Figure 5 illustrates the steps in our research topic-based estimation of activation probabilities for the influence analysis. The activation probability along each edge in the network was computed by taking the dot product of the topic mixture vector and NMF topic weight vector obtained by averaging the topic weights of the co-authored papers along the edge followed by normalization.

The influence maximization method (Kempe et al., 2003) uses an information diffusion model to identify the key authors that have the potential to cause maximal spread of information on a network based on activation probability at an edge. We consider the Independent Cascade (IC) approach (Kempe et al., 2003) as the information diffusion model in this study. Figure 4 illustrates this process of information diffusion cascade and the feedback with influence maximization. Mathematically, the influence maximization problem can be defined as follows. Consider a graph  $G = (V, E)$  that abstracts a complex network, where  $V$  is the set of nodes  $V$  and  $E$  is the set of edges  $\{(u, v) | u, v \in V\}$ . There are three types of nodes: (1) active - refers to an author who is influenced in the current step in an iteration path, (2) inactive - refers to an author who was active before and cannot influence others in subsequent time steps in an iteration path, and (3) available - refers to an author who can be influenced in the next step in an iteration path. The edge  $(u, v)$  implies that  $u$  can influence  $v$ . Additional simulation conditions are posed by the choice of diffusion model. For instance, in the IC-based diffusion model, an activated node  $u$  has a single chance to activate its available neighboring node  $v$  with an activation probability of  $p_{uv}$ . Given the possibility of initially activating  $k$  nodes, the



**Figure 4:** Information diffusion cascade and influence maximization process. In this illustration, a cascade iteration represents a simulation instantiation where an initial set of active nodes result in network impact. Simulation outcomes inform the influence maximization leading to identification of influential nodes, and candidate seed set samples are sent back to the simulation engine via a feedback loop.

```

Input:  $G, \lambda$            ▷ initial author collaboration network,
                           topic mixture vector
Output:  $H$              ▷ network with activation probabilities
1: procedure ACTIVATIONPROBABILITY( $G$ )
2:    $H \leftarrow CreateCopy(G)$ 
3:   for each edge  $e \in Edges(G)$  do
4:      $a_u, a_v \leftarrow GetCoAuthorIDs(e)$ 
5:      $scopusIDList \leftarrow GetCoAuthoredPaperScopusIDs(a_u, a_v)$ 
6:      $W \leftarrow GetTopicWeightMatrix(scopusIDList)$ 
7:      $m \leftarrow RowSize(W)$            ▷ number of papers
8:      $n \leftarrow ColumnSize(W)$          ▷ number of topics
9:      $t \leftarrow CreateArray(size = n)$ 
10:    for  $j = 1 : n$  do ▷ average weights of all papers for each topic
11:       $t_j \leftarrow (\sum_{i=1}^m W_{ij}) / m$ 
12:    end for
13:    for  $j = 1 : n$  do ▷ normalize the averaged topic weights
14:       $t_j \leftarrow t_j / \sum_{j=1}^n t_j$ 
15:    end for
16:     $p \leftarrow DotProduct(\lambda, t)$ 
17:    if  $p = 0$  then
18:      RemoveEdgeFromGraph( $H, e$ )
19:    else
20:       $H \leftarrow AddEdgeAttribute( $e, p$ )$ 
21:    end if
22:  end for
23:  return( $H$ )
24: end procedure

```

**Figure 5:** Algorithm for activation probability calculation in author collaboration network.

influence maximization problem aims to find a set of  $k$  seed nodes called the seed set  $S$ , that when activated result in maximal activations on the network among all possible such sets of  $k$  nodes. The seed size ( $k$ ) for all the simulation runs was set at 5 (i.e., to identify top 5 influencers), and the number of iterations ( $n$ ) for each IC simulation was set at 10,000. The IC model is a stochastic simulation where the node activation is a random process and the expected cascade size is a random variable. Multiple iterations of the IC are required to obtain the largest expected cascade size. In our simulations, we determine the authors with the largest expected cascade sizes as the top influencers. The IC simulation-based optimization converges resulting in a stable list of top  $k$  authors.

The influence maximization computation was performed using a Message Passing Interface (MPI) Python implementation with parallel computing of the ordinary greedy algorithm (Kempe et al., 2003) using the *mpi4py* Python package (Dalcin et al., 2005). A Monte Carlo loop was implemented for 10,000 iterations to compute the expected spread in activations. A key input in the information cascade and influence maximization process described above is the characterization of activation probability ( $p$ ) that models the likelihood of information flow across networked entities. Figure 5 illustrates the steps in our research topic-based estimation of activation probabilities for use in influence analysis. The activation probability for a given topic mixture of an author to activate another author was computed using the weight distribution of the NMF topics

obtained from the topic modeling results. Starting with collection of co-authored or cited papers corresponding to edges in the network, topic modeling led to a distribution of topic weights per paper. These topic weights were averaged and normalized over the collection, and aggregated topic mixture weights were computed to serve as activation probabilities in the scientific network.

#### 4.4 Temporal dynamics analysis

Analyzing temporal dynamics over scientific networks involves exploring multiple dimensions of information exchange. Below are brief descriptions of the methods implemented to characterize such temporal behaviors.

##### 4.4.1 Collaboration dynamics

The collaboration dynamics of the influencers can be characterized by analyzing their ability to form new collaborations (i.e., emergence) and to maintain old collaborations (i.e., persistence) in time. Mathematically, the number of new collaborations in a given year was estimated by computing the difference in the set of collaborators in that year and the aggregated set of collaborators in prior years. The number of old collaborations in a given year was estimated by computing the intersection of the set of collaborators in that year and the aggregated set of collaborators in prior years.

##### 4.4.2 Centrality dynamics

Centrality dynamics of the influencers was characterized by first computing the topology-based degree and betweenness measures of centrality of influencers in the collaboration network of each year (Barabási, 2012). The change in the centrality measures over time indicates the evolution of key influencers based on topology. The centrality measures (*degree and betweenness*) were calculated using the Python library *NetworkX* (Schult and Swart, 2008). Degree centrality indicates who is well-connected (popular) based on the number of connecting edges. Betweenness centrality indicates who controls information flow (or acts as a bridge) between two authors based on how often a node appears on the shortest paths between all other nodes in the network. The edge weights (as defined in the network construction section above) were used in the betweenness centrality calculations.

## 5. Case study

The case study application addresses the overarching research goals of identifying key influencers, comparing network measures of centrality with information diffusion-based outcomes, and characterizing collaboration dynamics. The results and discussion support the dynamic network analysis framework and computational engine described in Figure 2.

Year	Whole network				GCC network	
	Nodes	Edges	Edges (NF wt $\geq 0.05$ )	Number of components	Nodes	Edges
2000	3022	67361	5303	1136	42	116
2001	3462	26920	8443	1247	139	606
2002	3522	17219	7553	1091	110	438
2003	3505	27948	8636	1159	56	167
2004	4654	14195	12152	1209	159	573
2005	4866	24636	12921	1383	282	1457
2006	6487	431823	12668	3002	99	504
2007	5731	34945	15789	1644	617	3488
2008	6021	24681	14394	1566	131	525
2009	5890	25055	15423	1417	86	291
2010	6415	31206	16608	1666	111	538
2011	5540	32185	14635	1429	150	836
2012	5171	14128	13597	1105	223	914
2013	6296	28237	17586	1441	177	824
2014	6727	36303	19687	1426	952	4753
2015	7472	62235	20690	1789	276	1151
2016	8285	93836	22426	1998	419	1728
2017	8032	74851	23255	2047	351	1294
2018	10353	784580	26165	2747	934	3526
2019	8289	30427	23218	1523	891	3629

**Table 1:** Number of nodes and edges in the author collaboration networks.

## 5.1 Author collaboration data

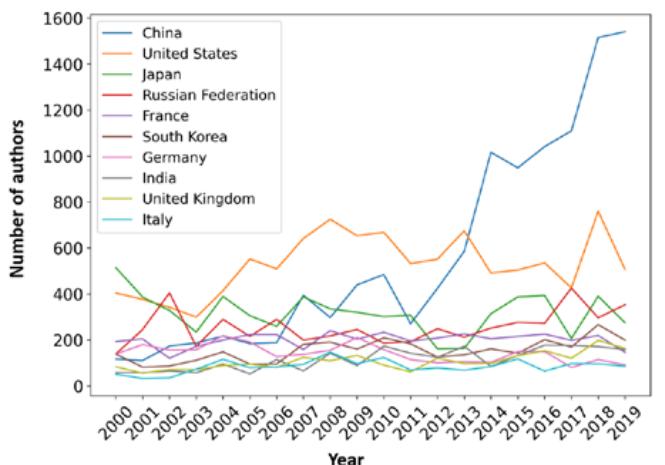
The author collaboration networks were created based on 33,517 records. Table 1 presents the number of nodes and edges from the overall author collaboration network over time before and after removing edges based on the NF weight criterion. The year 2018 had the largest network, with 10,353 nodes and 784,580 edges. After removing edges with NF weights below 0.05, the number of edges reduced from 784,580 to 26,165. By counting the number of authors affiliated with each country in a year, and averaging the count over the 20-year period, we can find that

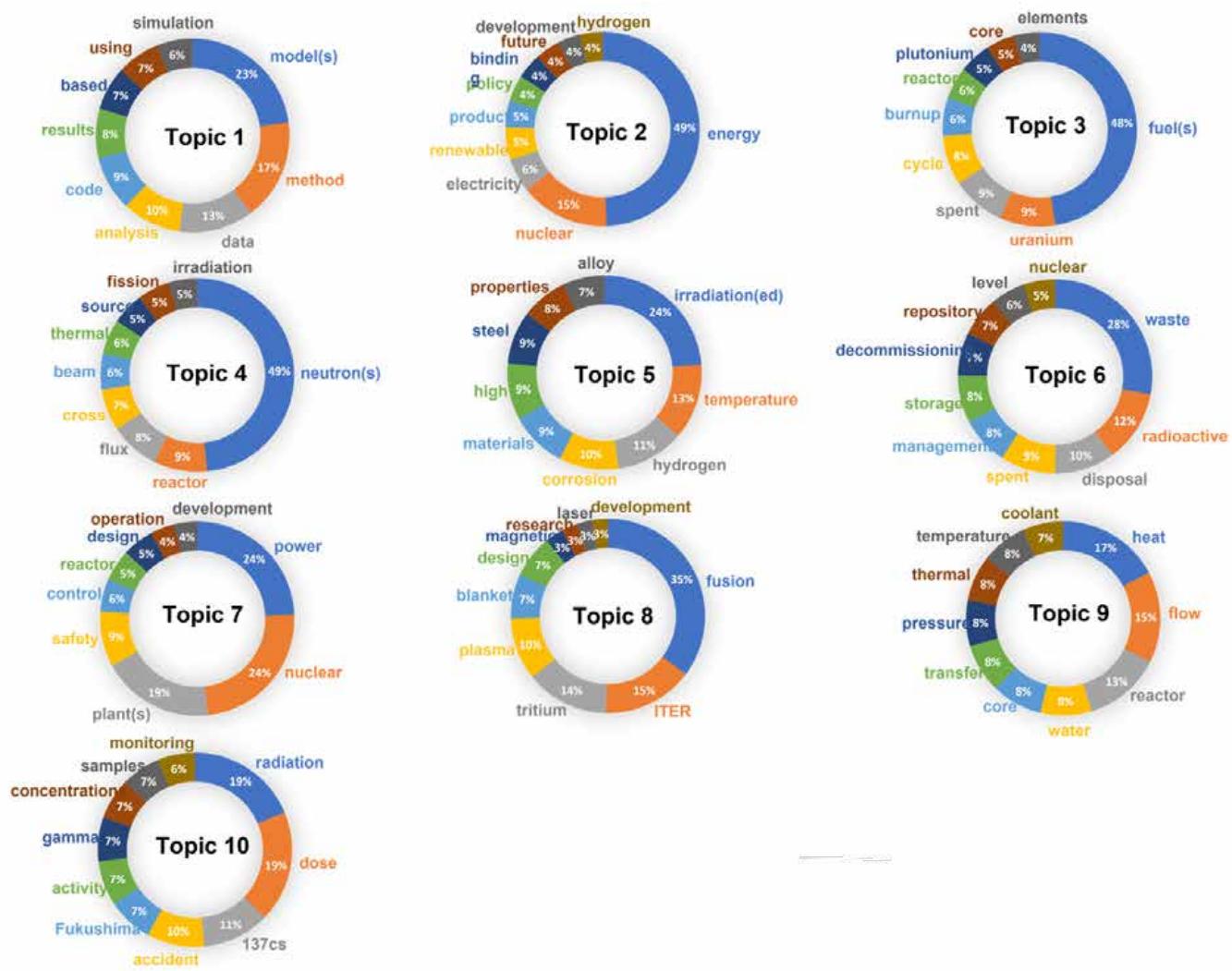
the highest average number of authors were from China, followed by United States, Japan, and Russia. Figure 6 depicts how the number of authors changed over a period of 20 years for the top ten countries that had the highest average number of authors. The figure indicates that there has been a drastic increase in the number of authors from China after 2011.

The GCC networks from the author collaboration networks were computed after removing edges with zero activation probabilities. We observe that in several cases the size of GCC author collaboration networks is roughly an order of magnitude less than the original networks, thereby contributing to higher computational efficiency while capturing significant network connectivity.

## 5.2 Topic modeling

Figure 7 presents ten topics identified by the NMF model with distinct keywords for each topic along with normalized weight contributions of the top ten keywords in each topic. Each topic is labeled by a topic index number from 1 to 10. In each topic, the top two to three keywords are highly weighted compared to the other keywords; which, suggests that the model is able to identify distinct keywords that capture the overall meaning of each topic. Specifically, the authors' interpretations of the topics based on the keywords were:

**Figure 6:** Top ten countries with the highest number of authors from 2000 to 2019.



**Figure 7:** Normalized weight distribution of the top 10 keywords of each NMF topic.

Topic 1: Modeling and data analysis;

Topic 2: Nuclear energy production;

Topic 3: Spent uranium in the nuclear fuel cycle;

Topic 4: Neutron flux in nuclear reactors;

Topic 5: Material irradiation at high temperatures;

Topic 6: Radioactive waste management;

Topic 7: Reactor control in nuclear power plants;

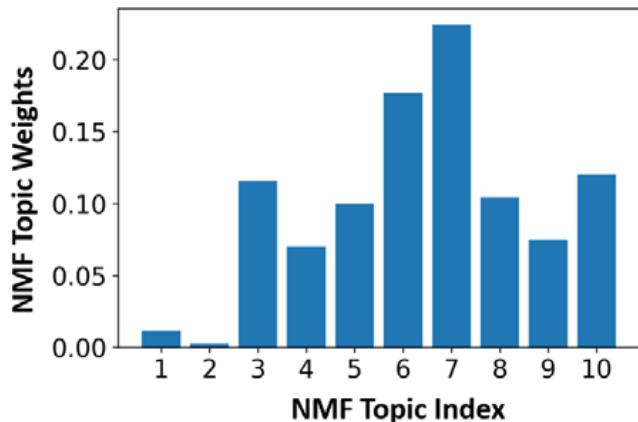
Topic 8: Nuclear fusion and ITER (International Thermonuclear Experimental Reactor);

Topic 9: Heat flow in nuclear reactors; and

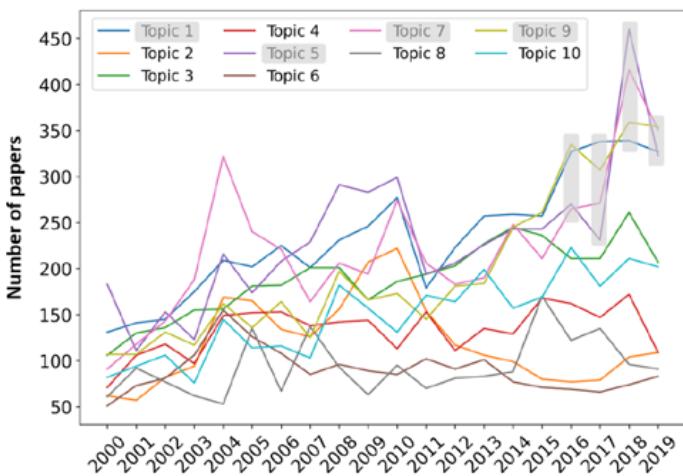
Topic 10: Radiation dose from nuclear accidents.

The word “nuclear” appears in topics 2, 6, and 7, indicating that it was not ignored during the feature extraction and more than 5% of the records do not have the word “nuclear” in their titles and abstracts. These are the records where one or more of the query phrases (*nuclear fuel*,

*nuclear energy*, or *nuclear reactor*) occurred only in the keyword list and not in the title and the abstract used for the topic analysis. Although the NMF topics have unique meanings, it is typical for a record to belong to multiple topics with varying weights. Figure 8 shows an example where all the topics contribute a non-zero weight to a record, with the highest-weighted topic being topic 6. In this study, the topic with the highest weight is considered the dominant topic for a particular record. Per figure 9, topics 1, 5, 7, and 9 were the top four *dominant topics* among all records in the most recent years from 2016-2019. In addition, these were also the top four dominant topics based on cumulative number of records from 2000-2019 (see Table 2). In this study, these top four prevalent dominant topics were selected as the topic mixture for information diffusion as part of topic-aware influence maximization analysis.



**Figure 8:** An example showing the non-zero normalized weights of each NMF topic of a record.



**Figure 9:** Number of records published every year on each dominant NMF topic. Highlighted topics were the top four dominant topics in the most recent years.

Dominant NMF Topic	Cumulative Number of Records from 2000-2019
Topic 1	4689
Topic 2	2399
Topic 3	3795
Topic 4	2669
Topic 5	4666
Topic 6	1789
Topic 7	4502
Topic 8	1871
Topic 9	3956
Topic 10	2983

**Table 2:** Cumulative number of records for each dominant NMF topic from 2000-2019. Top four dominant topics and corresponding cumulative number of records are highlighted.

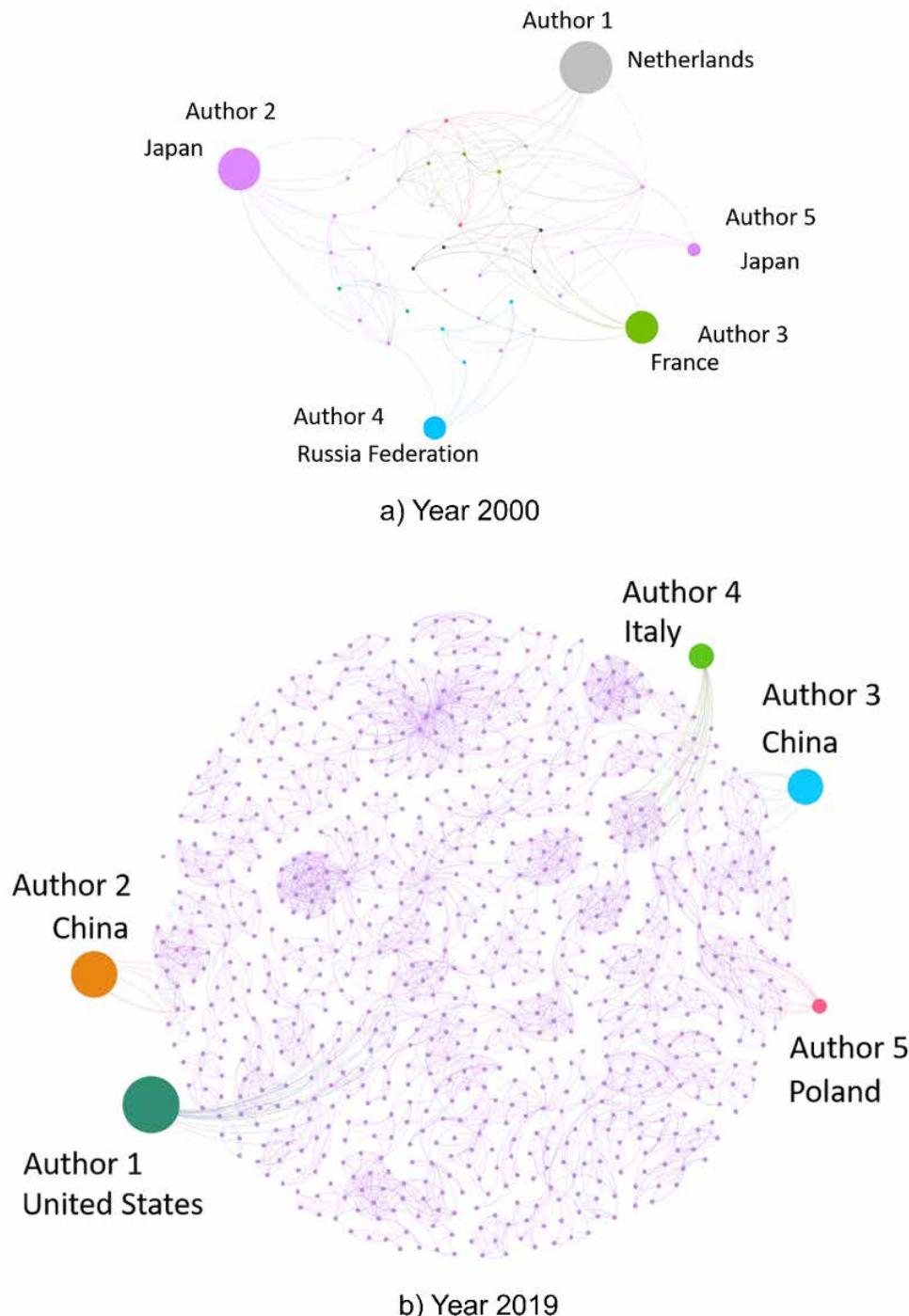
### 5.3 Influence analysis with collaboration network

As described in the topic-aware influence maximization section, activation probabilities for influence analysis may be computed using a mixture of NMF topic weights yielding different results from the topic-aware influence maximization (TAIM). In this case study, we present TAIM results to identify key influencers over time using an NMF topic mixture. Specifically, we use TAIM to identify the top 5 authors who can influence the diffusion of information pertaining to a mixture of the top four dominant NMF topics (see figure 9 and table 2). These topics are modeling and data analysis (topic 1), material irradiation at high temperatures (topic 5), reactor control in nuclear power plants (topic 7), and heat flow in nuclear reactors (topic 9).

#### 5.3.1 TAIM analysis on author collaboration networks

The TAIM analysis was applied to identify the top 5 influencers in the collaboration network from each year. The computations were performed using a parallelized (MPI) version of the TAIM algorithm. For example, a single TAIM simulation run with GCC of a network with 952 nodes and 4,753 edges using our optimized algorithm converged in about 6 hours using 32 processors; without MPI, the processing time for this run was about 30 times slower or about 7 days. The computational acceleration allowed us to run the TAIM algorithm on large graphs (e.g., 5,000+ nodes and 28,000+ edges) and complete all simulation runs in the order of a few days. Figure 10 shows the top 5 influencers for the years 2000 and 2019. The top influencer in 2000 was from Netherlands, followed by Japan, France, Russia Federation, and Japan again. The top influencer in 2019 was from United States, followed by two from China, and one each from Poland and Italy. Similar analysis was applied on collaboration graphs for years 2001 to 2018 (results not shown here). The change in the top influencers from year to year is indicative of the network dynamics.

The primary advantage of the TAIM analysis is that it helps to identify the top influencers who can diffuse information about the selected topic mixture through their high-spread influence network (collaborators), more efficiently than others. Whether or not they leveraged their positions as top influencers is subject for further investigation. Particularly, it is beneficial to discover if any of the influencers have used their high influence spread in a network to gain new collaborations and prominence in a research area. If they have, then it would be important to analyze the collaboration and publication dynamics to determine whether their positions as the top influencers in a particular year affect or was affected by their collaborations, publication track records, and research impact in past and future years. At the same time, we observe that the influencers may not spread their influence if they or members of their high-spread influence

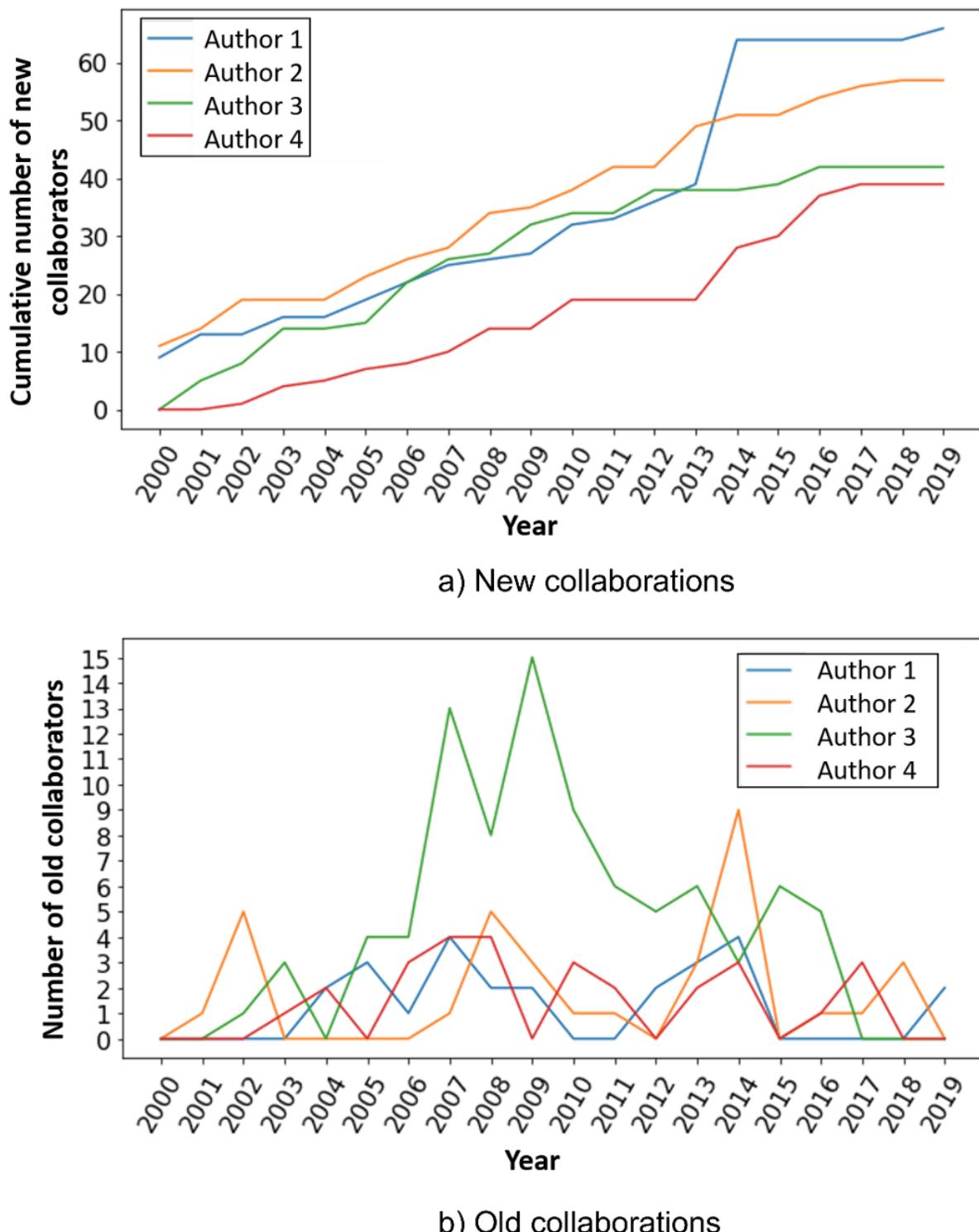


**Figure 10:** GCC of the author collaboration network, showing the top 5 influencers from years 2000 and 2019 for a mixture of dominant NMF topics 1, 5, 7, and 9. Author IDs have been masked for privacy, node size indicates influence, and node color indicates authors.

network are not actively collaborating or publishing for a long time period.

Among the 99 unique influencers that were identified from 2000 to 2019 (one author was a two-time top 5 influencer), 54 actively published papers for only less than 5 years, 26 were active for 5 to 9 years, 15 were active for 10 to 14 years, and the remaining 4 were active for at least 15 years. Thus, not all influencers from a given year actively

published in subsequent years. This may be due to possible factors such as nature and duration of the research, transition of collaborators, or evolving research interests. Influence analysis provides a way to track the collaboration and publication dynamics of influencers pertaining to any topic mixture over a period of time. Furthermore, the ability to generate and monitor scholarly influence dynamics may possibly contribute to identifying technology advances and readiness.

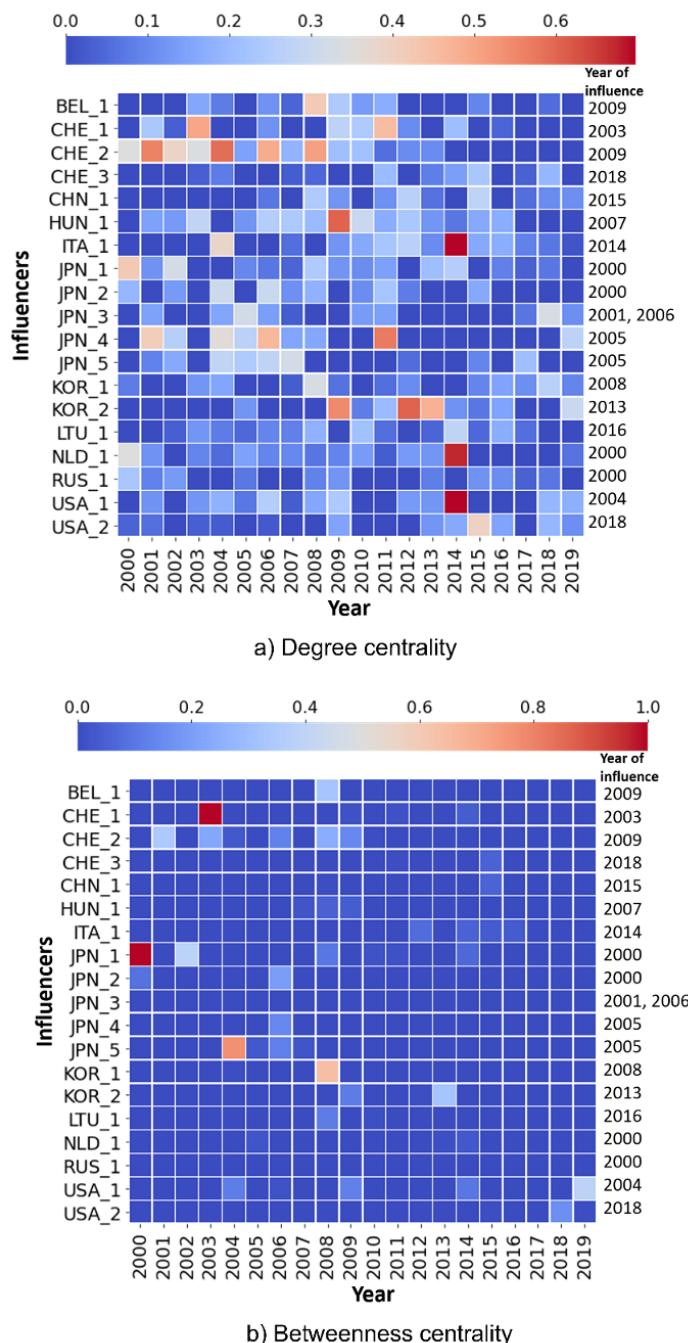


**Figure 11:** New and old collaborations of influential authors who published papers in at least 15 years between 2000 and 2019. Author IDs are masked for privacy.

### 5.3.2 Collaboration dynamics of influential authors

The collaboration dynamics of the influencers from 2000 to 2019 were analyzed based on their persistence to maintain old collaborations and the emergence of connections through their ability to form new collaborations. The new collaborators in a given year were those the influencer did not co-author a paper within the years prior to that year.

Figure 11 shows the collaboration dynamics of the four influencers who published papers for at least 15 years. Two of them were among the top 5 influencers in year 2000, one was from 2007, and the other from 2016. Figure 11a presents the rate at which each influencer formed new collaborations over time. While all the influencers have been actively making new collaborations over time, we observe that they exhibit different collaboration signatures over time.



**Figure 12:** Heat map showing the centrality values of influencers who published at least for 10 years from 2000 to 2019. Centrality values are normalized by the maximum value in a given year. Author IDs are masked by their 3-letter country codes and count index.

From the results presented, it can be understood that it is important for an influencer to work with new collaborators to maintain or increase research productivity and influence. For example, the top influencer from 2000 (author 1 in figure 11) gained 25 new collaborations in the year 2014. Such an influencer could have suddenly gained new collaborations due to contributions on a project with large number of new team members or based on new and multiple projects, or possibly even new leadership positions over multiple projects. Regardless of the factors that might have

contributed to the sudden gain (or even decline) in the number of new collaborations, the analytic insights can be used to characterize the growth rate of new collaborations for each author. The old collaboration dynamics in figure 11b indicate that the same influencers continued to maintain about two to five of their old collaborations. The influencer from 2007 (author 3 in figure 11b) in fact shows a strong collaborative behavior with their old collaborators. This might suggest continuing collaborations with known researchers from established areas of expertise and prior history of teaming together.

### 5.3.3 Centrality analysis of author collaboration networks

Network centrality measures (e.g., degree and betweenness) are commonly used topology-based outcomes to identify authors who are well connected and may control the flow of information in a network. It is interesting to analyze if the top influencers from TAIM analysis (including information diffusion-based outcomes) also have high degree or betweenness centrality values over time. Figure 12 presents the centrality values over time for the 19 influencers who published for at least 10 years. Three influencers, each from 2000, 2004, and 2014, have high degree centralities in the year 2014 (see figure 12a), which are indicative of their high number of collaborations. Two influencers, one from 2000 and another from 2003, also ranked among the top for the highest betweenness centrality (see figure 12b). Since high centrality values over time for influencers are not consistently observed from the heatmap, authors with high-influence spread are not necessarily also the ones with high centrality values. These results signify that a direct correspondence between rankings based on centrality measures and influence spread may not occur since the centrality analysis does not take into account topic awareness and information diffusion. Thus, the TAIM analysis provides novel information that centrality analyses do not provide, which can be valuable for understanding technology advances.

## 5.4 Discussion

The range of representative analytic outcomes presented in the case study provide insights from global nuclear research collaboration networks that could support influence and capability assessment. These insights are based on modeling and simulation assumptions as well as conditions described under section on dynamic network analysis framework. In light of the overarching challenge problem of identifying key entities and their capabilities over time for influence and capability assessment, four main takeaways from the case study are summarized below:

- Author collaboration networks represent different forms of influence that may lead to varied scholarly research publication patterns in support of technology advancements over time. Collaboration is more direct and indicate evolution of connections in the form of disaggregat-

ed networks over time. The GCC of these networks may serve as a computationally efficient reduced order representation while preserving significant network connectivity properties.

- Within the dynamic network analysis framework and computational engine, additional mixed-topic research areas can also be defined by subject matter experts. NMF based topic modeling algorithm may still serve as useful prior information.
- TAIM accounts for complex topologies and information cascade dynamics simultaneously to identify and rank key influencers by country over time, using topic mixture weights as inputs for computing activation probabilities of influence among authors. This represents an advance in the state-of-the-art of network analysis with nuclear science literature. Parallelized implementation of the TAIM algorithm led to up to 30 times faster compute times in some cases. These key influencers, identified via stochastic simulation-based network optimization, may possess the ability to diffuse information over a network more efficiently than others, and might include authors who are not necessarily those with high values of topological measures of centrality. This is important because using just topology-based measures may miss other influential authors and their collaborators.
- Even among influential authors there is variability in the way they form collaborations over time. Some influencers may choose to continue partnerships with their old collaborators exhibiting persistence of connections, while others may choose to continually seek new collaborators to pursue research goals exhibiting emergence. These collaboration signatures may reveal patterns of scholarly behavior that might help in a more robust assessment of technology advancements and capabilities.

## 6. Conclusion

The novel data-driven dynamic network analysis framework and computational engine developed in this paper is comprised of three connected computational modules: (1) topic-aware influence maximization, (2) information diffusion cascade, and (3) temporal dynamics analysis. Network theoretic, stochastic simulation, and optimization methods were leveraged to identify key entities and capabilities over time within global scholarly nuclear science research collaboration networks. The analytic insights associated with variability in scholarly interactions, influence propagation, and collaboration patterns over time via network connections can be useful for assessing technology advancements and capabilities. The main element of the dynamic network analysis engine is a topic-aware influence maximization algorithm that enables identification and ranking of key authors who have the potential to influence the spread of information in networks over time. A critical insight from our case study is that influential authors may have unique collaboration behaviors and may or may not exhibit high

values of topological measures of centrality. As a result, using just topology-based measures may not lead to a comprehensive assessment of the nuclear research landscape and technology advancements.

The results described using author collaboration networks represent analytic examples to illustrate the value of our dynamic network analysis engine for assessing research influence and technology advancements. Further work may include analysis of authors who collaborate with influencers and their evolution as potential influencers in the future. Future research may further involve expansion of information sources to include other data types such as corporate, trade, patent, and professional affiliation network activities over time to yield even more comprehensive understanding of key entities and capabilities over time. Further research may also include multi-layer network representations with corresponding topology and dynamics to capture importance and influence across network layers; as well as transformer-based topic analysis along with the use of graph representation learning for characterizing uncertainty (due to missing or unobserved information) in research connections. In summary, identifying influential authors can enable estimation of research trajectories in a country, and possibly in collaborating countries over time. Such information can be vital for detecting early signs of proliferation activities and generating safeguards conclusions.

## 7. Acknowledgments

This study was supported by the U.S. Department of Energy's National Nuclear Security Administration (NNSA). Pacific Northwest National Laboratory (PNNL) is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

## 8. References

1. Barabási A-L; Network Science; Cambridge University Press; 2016
2. Carley K M; Dynamic Network Analysis; 2003
3. Chen S, Fan J, Li G, Feng J, Tan K-L, Tang J; Online Topic-Aware Influence Maximization; Proceedings of the VLDB Endowment; 8; 2015; p 666-677
4. Dalcin L, Paz R, and Storti M; MPI for Python; Journal of Parallel and Distributed Computing; 65(9); 2005; p 1108-1115
5. Diab J, Burr P, and Stohr R; Using Machine Learning and Natural Language Processing to Enhance Uranium Mining and Milling Safeguards; IAEA Symposium on International Safeguards; 51; IAEA; 2018; p 1-7
6. Elsevier; Scopus; <https://www.elsevier.com/solutions/scopus>; Online; accessed 30 March 2021

7. Feldman Y, Barletta M, Ferguson M, and Norman C; Scientific and Technical Information as a Source for State Evaluation; INMM 54th Annual Meeting; 2013; Atlanta, GA
8. Fowler J H; Legislative Cosponsorship Networks in the US House and Senate; *Social Networks*; 28; 2006; p 454-465
9. Goldblum B L, Reddie A W, Hickey T C, Bevins J E, Laderman S, Mahowald N, Wright A P, Katzenzon E, and Mubarak, Y; The Nuclear Network: Multiplex Network Analysis for Interconnected Systems; *Applied Network Science*; 4; 2019; p 1-17
10. Iancu I, Wilson B, Calle D, and Gagne D; Detection of Undeclared Nuclear Material and Activities using the Collaborative Analysis Platform; Proceedings of the {IAEA} Symposium on International Safeguards; 2018
11. Kas M, Khadka A G, Frankenstein W, Abdulla A Y, Kunkel F, Carley L R, and Carley K M; Analyzing Scientific Networks for Nuclear Capabilities Assessment; *Journal of the American Society for Information Science and Technology*; 63; 2012; p 1294-1312
12. Kempe D, Kleinberg J, and Tardos É; Maximizing the Spread of Influence Through a Social Network; Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2003; p 137-146
13. Kim M, Baek I, and Song M; Topic Diffusion Analysis of a Weighted Citation Network in Biomedical Literature; *Journal of the Association for Information Science and Technology*; 69; 2018; p 329-342
14. Kitsak M, Ganin A A, Eisenberg D A, Krapivsky P L, Krioukov D, Alderson D L, and Linkov I; Stability of a Giant Connected Component in a Complex Network; *Physical Review E*; 97; 2018; p 012309
15. Kuang D, Choo J, and Park H; Nonnegative Matrix Factorization for Interactive Topic Modeling and Document Clustering; In *Partitional Clustering Algorithms*; Springer; 2015; p 215-243
16. Liu Z, and Morsy S; Development of the Physical Model; IAEA Safeguards Symposium; 29; 2007; p 1-7
17. Molas-Gallart J; Which Way To Go? Defence Technology and the Diversity of 'Dual-Use' Technology Transfer; *Research Policy*; 26 (3); 1997; p 367-385
18. Newman M; *Networks*; Oxford University Press; 2018
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, and Duchesnay E; Scikit-learn: Machine Learning in Python; *Journal of Machine Learning Research*; 12; 2011; p 2825-2830
20. Perianes-Rodriguez A, Waltman L, and Van Eck, NJ; Constructing bibliometric networks: A comparison between full and fractional counting; *Journal of Informetrics*; 10.4; 2016; p 1178-1195
21. Schult D A and Swart P; Exploring Network Structure, Dynamics, and Function using NetworkX; Proceedings of the 7th Python in Science Conferences (SciPy 2008); 2008.
22. Sheffield A; Developing the Next-Generation of AI Systems to Push the Detection of Foreign Nuclear Proliferation Further "Left of Boom"; *Countering WMD Journal*; 21; 2020; 100-102
23. Stewart I J, Lee A, and ElGebaly A; Automated Processing of Open Source Information for Nonproliferation Purposes; *Journal of Nuclear Materials*; 46; 2018; p 21-36
24. Zuo S; *pyscopus* 1.0.3; <https://pypi.org/project/pyscopus/>; Online; accessed 9 May 2023

# Simulating submarine reactor fuel in light of the AUKUS deal

S. Grape<sup>a</sup>, E. Branger<sup>a</sup>, C. Gustavsson<sup>a</sup>, R. Kelley<sup>b</sup>, V. Fedchenko<sup>b</sup>

<sup>a</sup>Uppsala University, Department of Physics and Astronomy, Box 516, 751 20 Uppsala, Sweden

<sup>b</sup> Stockholm International Peace Research Institute, Signalistgatan 9, 169 72 Solna Sweden

## Abstract:

*This work investigates fuel properties of submarine reactor fuel from a non-proliferation and safeguards perspective in light of the deal involving Australia, the United Kingdom and United States known as AUKUS. This study investigates the isotopic composition of the spent fuel at the end of intended reactor life. The fuel in the proposed AUKUS submarine is modelled after a Virginia Class fast attack submarine, discussed as an option for Australia.*

*The vast majority of civil experience with plutonium production is with fuel starting at low enrichments for shorter burnups. The AUKUS fuel at the start of irradiation campaign is assumed to contain uranium enriched to between 93% and 97.3%. It is burned at high power for about 33 years before retirement. Because the fuel is mostly uranium-235 initially, there are very few thermal captures leading to production of plutonium-239. In the submarine, the majority of non-fissile captures lead to the production of uranium-236 with other capture chains that do not lead to the production of weapons grade plutonium or weapons usable uranium.*

*This study concludes that the final isotopic composition of the AUKUS spent fuel is no longer VHEU, but a low grade of HEU diluted largely by uranium-236 instead of uranium-238. Several kilograms of plutonium are produced but it is composed of several different plutonium isotopes with a large fraction of plutonium-238. There is little likelihood that spent AUKUS fuel will be reprocessed by any of the countries involved. But if it were reprocessed, the resulting uranium and plutonium will have very unusual isotope compositions. Resulting materials would be subject to safeguards but would not, in fact, be well-suited as fissile material for weapons purposes.*

**Keywords:** AUKUS, non-proliferation, safeguards, submarine, reactor fuel.

## 1. Introduction and objective

Nuclear-powered submarines are operated by six countries in the world: the US, Russia, Britain, France, China and India. The first five countries are NWSs under the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), while India has never signed the NPT but is in possession of nuclear weapons. Nuclear-powered submarines have also to varying extent been considered by a number of non-nuclear weapon state parties to the NPT. These countries are Australia, Brazil, Canada, Iran and South Korea [1], as well as, potentially, Argentina and Japan [2]. A comprehensive overview of submarine nuclear reactors used for propulsion can be found in [3]. This work focuses solely on Australia in light of the recent AUKUS deal, but the implications are valuable in the context of other countries as well.

The United States and the United Kingdom have proposed supplying Australia with nuclear powered submarines under the so-called AUKUS security pact. These submarines are not designed to carry nuclear weapons. However, there are two nuclear safeguards issues associated with them. First, very highly enriched uranium (VHEU) in the form of reactor fuel would be in the custody of a non-nuclear weapon state (NNWS) and the fuel would probably be provided to Australia by a nuclear weapon state (NWS). Second, normal IAEA safeguards would not be in force on VHEU fuel in the submarine reactor core. These two facts have been interpreted in the literature as proliferation risks (or even loopholes in the legal frameworks) associated with the AUKUS deal. This paper's objective is to investigate those potential risks by assessing proliferation attractiveness of nuclear material in the AUKUS reactor fuel during its lifetime.

## 2. The AUKUS security pact and non-proliferation concerns

AUKUS is a trilateral security pact between Australia, the UK and the US. All three countries are parties to the NPT, a treaty aimed to prevent the spread of nuclear weapons by regulating, among other issues, how fissile materials can be used by states. Non-nuclear weapon states (NNWS) such as Australia, are explicitly forbidden to develop or acquire nuclear weapons, and must place their nuclear material intended for peaceful nuclear activities under nuclear safeguards. In 2021, it was announced that the pact would

cooperate on a range of topics, including nuclear-powered submarines as mentioned in references [4] and [5]. This announcement raised multiple questions, some of which were practical in nature and concerned how such a cooperation could be arranged and facilitated. It also raised a number of non-proliferation concerns [4], [5], [6] and it is argued in reference [7] that Australia “will have to become the first non-nuclear-weapon state to exercise a loophole that allows it to remove nuclear material from the inspection system of the International Atomic Energy Agency (IAEA)”. Indeed, paragraph 14 of the safeguards agreement INFCIRC/153/Corr. allows for the “Non-Application of Safeguards to Nuclear Material to be used in Non-Peaceful Activities”, which is often interpreted to include nuclear-powered ships and submarines, military space vehicles, nuclear reactors and radio-thermal generators (RTGs) for military bases or isolated radar stations [8]. A comprehensive overview of the paragraph 14 provisions can be found, describing the challenges for the IAEA safeguards system associated with states showing interest in nuclear-powered submarines and ships and in exempting nuclear material for safeguards [2]. A number of non-proliferation concerns are raised reference [7], and reference [9] states that the above-mentioned paragraph about non-application of safeguards could be considered a loophole in the NPT where removing nuclear material from nuclear safeguards could become a precedent for future proliferators to use naval reactor programs as a cover to develop nuclear explosive devices (NEDs). Already, Iran has been pointed out as one country that may be benefiting from the AUKUS deal, by being able to use the arguments put forward by Australia to support their own expanded nuclear ambitions [10], [11].

A number of diversion scenarios associated with the naval fuel programs have been identified in reference [9]. Diversion scenarios specifically associated with the AUKUS submarine deal were identified to include:

- diversion of low enriched uranium (LEU) and HEU from the enrichment facility;
- diversion of enriched stockpiled product intended for fuel fabrication;
- diversion of nuclear material from a fuel fabrication plant;
- the establishment of undeclared enrichment plant; and
- diversion of spent fuel from storage, followed by reprocessing.

In this work, we are interested in investigating non-proliferation concerns of the AUKUS deal from a technical perspective, focusing specifically on the last one of the potential diversion scenarios listed above. We will, using openly available information, model submarine reactor fuel and its irradiation to estimate how the isotopic composition changes as a function of irradiation. The objective of the work is to, from a technical perspective, analyse the usability of the

fuel material (beginning-of-life, middle and end-of-life) for use in a NED. Of specific interest is the spent fuel at the end-of-life, as indicated in the list above, but as a complement to that we will also investigate fresh and partially irradiated fuel material as such results will become available as well. We will also make assessments of technical non-proliferation concerns related to the AUKUS deal that could have an impact on the nuclear safeguards community. Note that this work does not consider other parts of the fuel cycle except the operation of the submarines, as we assume that Australia will only operate the reactors and then return them to the host state which is a NWS in control of all military fuel cycle activities leading up to the production of nuclear weapons. However, if Australia were to use the AUKUS deal to motivate a need to control the front-and back-end fuel cycle including uranium enrichment facilities, fuel fabrication facilities, additional stockpiles or reprocessing facilities, the associated proliferation concerns would look very different.

### 3. Submarine reactor cores

The UK and US submarines under consideration in AUKUS are the British Astute-class submarines and the American Virginia-class submarines [9]. The intent is for Australia to operate the submarine reactors throughout their lifetime of 33 years, meaning that Australia will not produce the reactor fuel and that refuelling is not needed [12]. After having reached their end-of-life, the reactors will be returned to the supplier, meaning the US or UK [13].

All nuclear-powered submarines in the world, except the Russian ones, are equipped with one nuclear reactor, with the British and American nuclear-powered submarines are using pressurized water reactor (PWR) cores, fuelled with HEU having a starting enrichment level of at least 93% uranium-235 (Ma, 2008). The core is typically very compact and designed to give a high heat transfer area per fuel volume [12]. The power of the reactor core depends on the design. The power of the Virginia class submarines is believed to be slightly higher than that of the Los Angeles submarine, known to be 130 MW<sub>th</sub>, because the Virginia reactor core is slightly larger [14]. This also agrees with other estimates stating that the power is 150 MW<sub>th</sub> [15].

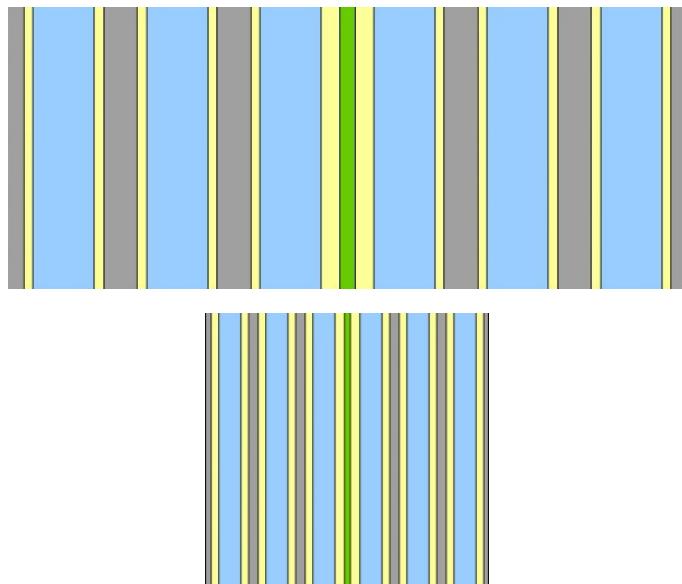
#### 3.1 Earlier works on submarine reactor models

The design information about a submarine reactor core is in general difficult to obtain because it is sensitive, but some information is openly available. A Virginia class attack submarine can be assumed to have a core of about 0.4 tons of weapons-grade uranium [14]. More details on submarine reactor cores are provided on the French Rubis-class submarines [16]. While details about the actual reactor operation are unavailable, existing models assume that the HEU fuelled submarines spend 240 days per year at sea and that they operate at 25% of maximum capacity,

corresponding to 60 full power days (FPD) per year [16]. The total reactor lifetime is up to 20 years, corresponding to 1200 full power days in total. This may not be a completely accurate information, but with limited data available on the operation of HEU-fuelled submarines, it is a reasonable estimate, and we will assume a similar history to allow for benchmarking. In the same work, a number of different fuel materials and submarine cores of 50 MW were modelled with the EPRI-Cell code, and depletion calculations were performed with the CINDER code.  $\text{UO}_2$  fuel with  $\text{Gd}_2\text{O}_3$  burnable absorbers, configured in a fuel plate geometry to ensure efficient cooling of the fuel, was considered. A number of different fuel geometries were modelled and three different HEU fuel designs were implemented. It was shown that the material composition of the core was dependent on the initial enrichment of the fuel as well as on the core geometry. Research was also published on the possibility of replacing HEU cores in submarines with LEU cores using the modern Monte Carlo code Serpent 2 [15]. In that work, a more realistic operation of the submarine reactor is employed, together with an HEU fuel plate geometry and burnable poison in the form of  $\text{Gd}_2\text{O}_3$ .

### 3.2 Submarine reactor models used in this work

In this work, we have chosen to study multiple reactor cores. There are several reasons for this. Obviously, the main interest is in studying a Virginia-like reactor core, but there is no openly available information on it. The most detailed information about submarine reactor cores available to the authors was found in reference [16]. One core design from that work is here referred to as Model 1 and used as a benchmark, although another design (here known as Model 2) was more Virginia-like. Results on the isotopic



**Figure 1:** The Model 1 (upper) and Model 2 (lower) geometries implemented Serpent2. The  $\text{Gd}_2\text{O}_3$  plate is shown centrally in green, the zirconium cladding in yellow, the fuel (or fissile) material in grey and water in blue. Dimensions are provided in table 1.

composition of the spent nuclear fuel were however obtained using a completely different simulation framework available 30 years ago, which did not facilitate as comprehensive simulations that can be made today. We thus decided to make simulations of both models to more accurately study the composition of the end-of-life core, complementing the results published earlier.

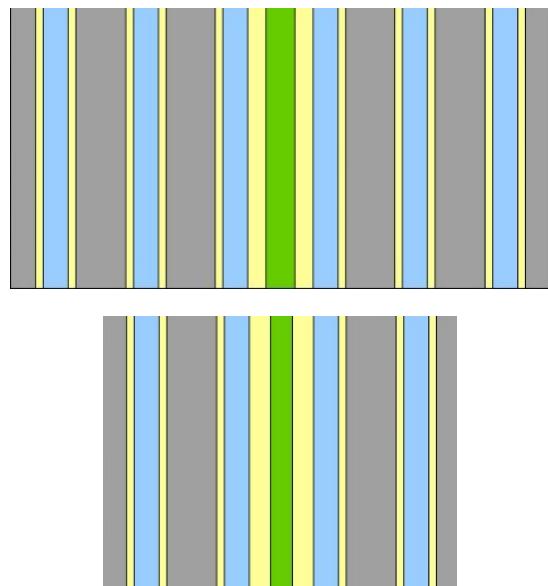
More recent work did study operational and safety aspects of a Virginia-like core with higher power and slightly lower initial enrichment compared to research in [16], using modern Monte Carlo codes [15]. The isotopic composition of the spent nuclear fuel was however not investigated nor reported in that work. For this reason, it was decided to update the core designs with information about the core geometry and irradiation based on published data [15], here known as Model 3 and 4, and make new simulations focusing on non-proliferation aspects of the fuel.

## 4. Methodology

In this work, a number of different reactor cores are studied and different designs are implemented in Serpent 2 [18]. The operation of the reactor cores is modelled, and the depletion calculations show how the nuclide inventory changes over time. After having reached its end-of-life, the isotopic composition of the fuel is evaluated, with special attention paid to the uranium and plutonium vectors.

### 4.1 The submarine core model and its irradiation

All cores modelled in this work assume that the fuel is made of HEU that is dispersed in the form of  $\text{UO}_2$  in a metal alloy containing both uranium and zirconium as



**Figure 2:** The Model 3 (upper) and Model 4 (lower) unit cell design implemented in Serpent 2. Both unit cells are infinitely reflected in two dimensions.  $\text{Gd}_2\text{O}_3$  is shown centrally in green, the zirconium cladding in yellow, the fuel material in grey and water in blue. Dimensions are provided in table 1.

described in references [15] and [16]. The fuel material is located in fuel plates, with water channels surrounding the plates on both sides. This design is chosen because of the extremely high burnups (compared to light-water reactor fuel where the typical discharge burnup is around one-tenth of the studied submarine cores) and the need to ensure sufficient cooling of the fuel during operation. Gadolinium is also present to control the reactivity throughout the reactor lifetime.

In this work, the objective is to study the fuel composition evolution over time. For such analyses, it is not necessary to model support structures or the core in its full geometry. We have thus implemented a number of geometries consisting of a unit cell comprising fuel plates, gadolinium

plates and water channels. The details of the geometries are provided in table 1. The implementation of Model 1 and 2 are also shown in figure 1, and Model 3 and 4 in figure 2. Numerical values relating to the submarine reactor cores either come from references [15] and [16], or result from scaling the Model 1 and 2 parameters to the size of the Model 3 and 4 core.

Both Model 1 and 2 consists of a unit cell centred around a  $\text{Gd}_2\text{O}_3$  plate surrounded by 2.5 fuel plates on each side, infinitely reflected. Model 2 has reduced plate thickness compared to Model 1 and also reduced water channels, found to result in a higher inventory of uranium-235 at the end-of-life (and thus a potentially more attractive material for use in a NED) [16]. The water density was unrealistically

	Model 1	Model 2	Model 3 & 4
<b>Power [MW]</b>	50	50	150
<b><math>\text{UO}_2</math> enrichment [%]</b>	97.3	97.3	93
<b><math>\text{UO}_2</math> volume% in fuel [%]</b>	20	33	24.1
<b>Zr volume% in fuel</b>	80	67	75.9
<b>Start-up uranium composition [%]</b>	U235: 97 U238: 3.0	U235: 97 U238: 3.0	U234: 0.739 U235: 93.0 U238: 6.27
<b>Beginning-of-life composition [kg]</b>	U235: 106	U235: 152.5 U238: 4.3	U234: 5.1 U235: 638.7 U238: 43.1
<b>Temperatures</b>			
<b>Fuel temperature [<math>^{\circ}\text{C}</math>]</b>	927 (1200 K)	927	400
<b>Cladding temperature [<math>^{\circ}\text{C}</math>]</b>	927	927	400
<b>Water temperature [<math>^{\circ}\text{C}</math>]</b>	327	327	327
<b><math>\text{Gd}_2\text{O}_3</math> temperature [<math>^{\circ}\text{C}</math>]</b>	327	327	327
<b>Densities</b>			
<b><math>\text{UO}_2</math> density [g/cm<sup>3</sup>]</b>	10.3	10.3	10.97
<b>Zr density [g/cm<sup>3</sup>]</b>	6.56	6.56	6.56
<b><math>\text{UO}_2\text{Zr}</math> density [g/cm<sup>3</sup>]</b>	6.56	6.56	6.56
<b><math>\text{Gd}_2\text{O}_3</math> density [g/cm<sup>3</sup>]</b>	7.64	7.64	7.64
<b><math>\text{H}_2\text{O}</math> density [g/cm<sup>3</sup>]</b>	7.64	7.64	7.64
<b>Dimensions</b>			
<b>Fuel material thickness [mm]</b>	1.45	0.5	2.5
<b>Zr cladding thickness [mm]</b>	0.385	0.385	0.4
<b>Fuel plate thickness [mm]</b>	2.22	1.27	3.3
<b><math>\text{Gd}_2\text{O}_3</math> thickness [mm]</b>	0.0665	0.0285	0.1425
<b>Water channel [mm]</b>	2.63	1.12	1.25
<b>Cell thickness [mm]</b>	29.1	14.32	27.27
<b>Irradiation conditions</b>			
<b>Irradiation scenario</b>	1200 full-power days	1200 full-power days	Duty cycle of 6 months/year (183 days). Operation at 25% of full power during that time.
<b>Power density [kW/g]</b>	0.457142	0.31892	0.05525
<b>Discharge burnup [MWd/kgU]</b>	548.57	382.70	333.65

**Table 1:** Properties of the submarine reactor cores modelled here.

high at 1 g/cm<sup>3</sup>, and results on core volume and reactivity was later corrected for this in order to ensure more realistic results. In this work, the same (unrealistic) water density was used to allow benchmarking with previously published results [16].

Two other models describing a more Virginia-like core were also implemented; they are referred to as Model 3 and 4. The two designs were largely taken from published research [15]. One exception is the gadolinium plates, which were not included in the previous work (the atomic density of the fuel was simply increased by adding gadolinium atoms in varying amounts throughout the core, noted as being an unphysical manner) [15]. In this work, two different Virginia-like designs are investigated. In Model 3, a gadolinium plate was again surrounded by 2.5 fuel plates on each side, and the ratio of gadolinium volume over fuel volume in the unit cell was identical to that in Model 2. In Model 4, this ratio was changed as 1.5 fuel plates surrounded the gadolinium plate on each side. The unit cells of Model 3 and 4 are shown in figure 2. The UO<sub>2</sub> density in this work was chosen as the theoretical maximum density to allow for comparison of results with previous works [15].

## 5. Analysis and discussion

### 5.1 Results of Model 1 and 2

The results of the simulations show that the inventory of uranium-235 and uranium-238 are, as expected, reduced over time, while uranium-236 and various plutonium isotopes build up. The results from Model 1 and 2 are on a general level in accordance with previous results [16], with a few exceptions. Earlier works [16] did not include any other heavy isotopes than uranium-235, uranium-236, uranium-238 and plutonium-239 to plutonium-242, and the results here show that including additional isotopes is essential for a proper evaluation of the end-of-life core properties. The inclusion of plutonium-238 is found to be especially important, as it is dominating the plutonium vector for the

end-of-life core in Model 1 and 2. Excluding this isotope from the analysis shows that the results of this work are in good agreement with already published results [16], on a relative scale. As Model 1 was intended as a benchmark of this work against such earlier results, this finding supports the conclusion that the modelling done here is sufficiently good.

The evolution of the plutonium in Model 1 and 2 can be seen in figure 3. It is seen that the production of plutonium-239 increases sharply in the beginning, while production of plutonium-238 and plutonium-240 to plutonium-242 sets in after 400 full power days (FPDs). Of the total amount of plutonium at the end-of-life of the core, about 64% is plutonium-238 in Model 1, and 51% in Model 2. In all cases the relative content of plutonium-238 is below 80%, a value defined in the nuclear safeguards community, at which point the plutonium is considered to be of such quality that it is exempted from safeguards. This means that if the nuclear material used for naval propulsion purposes were to be returned to the civil fuel cycle, it would have to be placed under safeguards. It can be noted that the relative content of plutonium-240 in Model 1 and 2 is around 5%, which is lower than the “impurity” level of 7% limit, defined by the US Department of Energy (U.S. Department of Energy, 2009), below which plutonium is classified as weapon-grade. This classification is however a result of how plutonium has been produced historically, and where plutonium-239 would make up 93% of the plutonium isotopes. The amount of plutonium-240 found in the plutonium produced in this work should thus not be interpreted to indicate that the material is suitable for weapons manufacture, as other impurities (such as plutonium-238) are also present and the fraction of plutonium-239 is far below 93%.

The results also show that the end-of-life cores in Model 1 and 2 have a uranium-235 enrichment of 61.3% and 76.8%, respectively, which is slightly below those published earlier in [16]. Similarly, the results show a somewhat higher absolute content of uranium-238. The explanation for the lower

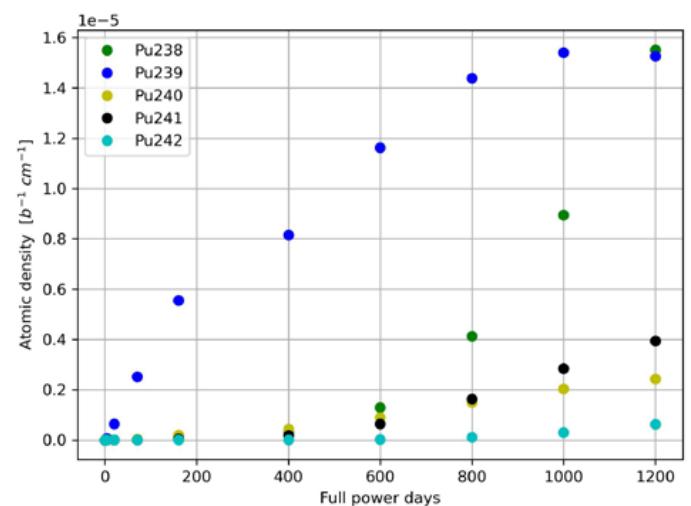
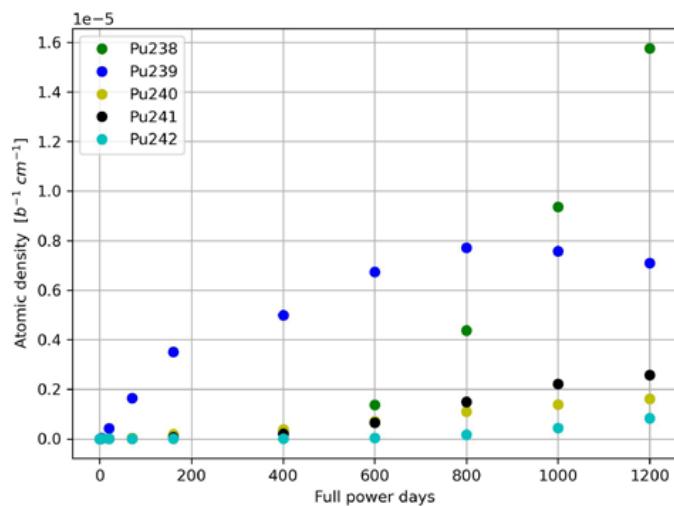


Figure 3: The plutonium composition in Model 1 (left) and 2 (right) as a function of operation time of the submarine.

enrichment is the fact that the gadolinium in the absorber plates did not last throughout the entire lifetime of the reactor (it is depleted after about 1000 FPDs), despite the importance of balancing the fuel volume against the need for  $\text{Gd}_2\text{O}_3$  to ensure reactivity control [16]. After full depletion of the gadolinium, the neutron spectrum will be more thermalized, leading to relatively less fast fission and neutron resonance capture in uranium-238, consistent with the findings here of more uranium-238. It will also lead to less plutonium-239 than in previous works [16]. The findings can also be explained by the inclusion of the complete inventory of heavy nuclides in the simulations, which in turn results in less neutron absorption in uranium-238 and more absorption in other nuclei. The depletion calculations also reveal that the end-of-life core in Model 1 consists of 1.38 kg of neptunium (1.54 kg in Model 2), which can be compared with the total plutonium mass of 906 g. The vast majority of that neptunium, over 99%, is neptunium-237.

Compared to the results of Model 1, Model 2 has a slightly higher relative content of uranium-235 and a plutonium vector slightly less dominated by plutonium-238. More notably, the relative amount of plutonium-239 is almost 60% higher in Model 2 than Model 1. All the results are however essentially an effect of the lower discharge burnup in combination of a different start-up fuel, as the volume fraction of  $\text{UO}_2$  is 33% in Model 2 instead of 20% as in Model 1, rather than an effect of the modified geometry. Table 2 and 3 show the relative and absolute content of major uranium and plutonium isotopes in the end-of-life core for Model 1 and 2.

Model	Relative content [%]			
	1	1 [16]	2	2 [16]
<b>U234</b>	0.0252	-	0.0167	
<b>U235</b>	61.3	68.1	76.8	81.8
<b>U236</b>	33.8	28.4	19.6	-
<b>U237</b>	0.0655	-	0.0303	-
<b>U238</b>	4.83	3.55	3.55	-
<b>Pu238</b>	64.4	-	50.8	-
<b>Pu239</b>	19.6	52.3	31.2	-
<b>Pu240</b>	5.04	14.0	5.72	-
<b>Pu241</b>	7.30	24.4	10.1	-
<b>Pu242</b>	3.70	9.30	2.25	-

**Table 2:** Results from the simulations of Model 1 and 2 corresponding results published in [16]. Isotopes with a relative content below 0.01% have been excluded.

Model	Absolute content [g]			
	1	1 [16]	2	2 [16]
<b>U234</b>	11.2	-	15.3	
<b>U235</b>	27 400	33600	70500	74900

<b>U236</b>	15200	14000	18100	1670
<b>U237</b>	29.6	-	28.1	-
<b>U238</b>	2190	1750	3 300	3 300
<b>Tot U</b>	44 900	49 400	91 900	79 870
<b>Pu238</b>	582	-	497	-
<b>Pu239</b>	177	135	306	294
<b>Pu240</b>	45.9	36	56.4	56
<b>Pu241</b>	66.8	63	99.6	92
<b>Pu242</b>	34.0	24	22.4	16
<b>Tot Pu</b>	900	258	981	458

**Table 3:** Results from the simulations of Model 1 and 2 corresponding results from [16]. Isotopes with a relative content below 0.01% have been excluded.

It can be noted that the results of this work show a difference in total uranium content in the end-of-life cores when compared to previously published results [16]. Some of it can be explained by the increased plutonium production (of which the majority is plutonium-238), and some of it by the production of neptunium, not considered earlier. Remaining differences could be due to that the results here are obtained using more modern depletion code and fuel libraries.

## 5.2 Results of Model 3 and 4

Results from the simulations of the more Virginia-like cores Model 3 and 4 are shown in table 4. The first thing to note is that although the two models have a different frequency of absorber plates, the composition of the spent nuclear

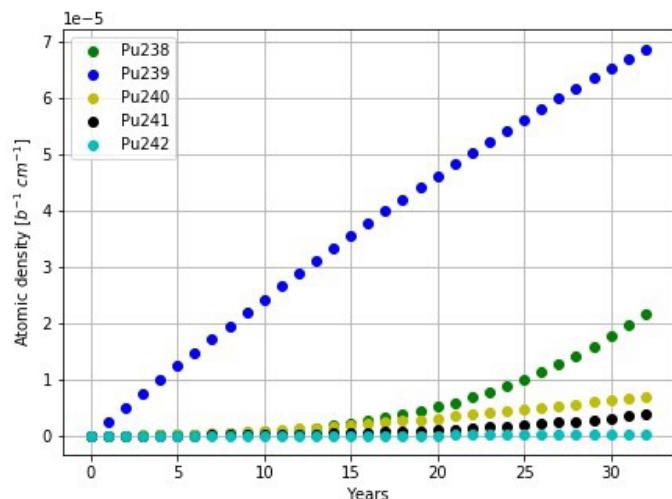
	Model 3		Model 4	
	Relative cont. [%]	Absolute	Relative	Absolute
<b>U234</b>	0.764	cont. [g]	cont. [%]	cont. [g]
<b>U235</b>	73.3	3310	0.75	3260
<b>U236</b>	17.9	319000	73.2	318000
<b>U238</b>	7.98	78300	18.1	79200
<b>Tot U</b>		35200	7.94	35000
		436000		436000
<b>Pu238</b>	22.1			
<b>Pu239</b>	66.8	1520	22.6	1610
<b>Pu240</b>	7.16	4620	66.4	4750
<b>Pu241</b>	3.66	497	6.91	497
<b>Pu242</b>	0.289	255	3.85	278
<b>Tot Pu</b>		20.3	0.31	22.8
<b>Pu241</b>	3.66	6900		7200
<b>Pu242</b>	0.289	20.3	0.31	22.8
<b>Tot Pu</b>		6900		7200

**Table 4:** Results from the simulations of Model 3 and 4. Isotopes with a relative content below 0.01% have been excluded.

fuel is almost identical in the two designs because in both cases, the gadolinium amount is sufficient to last throughout the lifetime of the reactor.

A key goal of this study was to calculate the end-state enrichment of the uranium fuel. The high burn-up long-life submarine core is very different from most familiar power reactor fuel cycles. The breeding of uranium-236 is a significant feature of this fuel cycle. Uranium-236 is a non-fissile diluent in HEU. We argue that it behaves much like uranium-238 and increases the critical mass, and hence diminishes utility for weapons. The uranium enrichment of the end-of-life cores is around 73%, which is below weapons-grade (above 90%) but still a very high enrichment. On the absolute scale, the uranium content is considerably larger than in Model 1 and 2, because the Virginia-like reactors are larger.

It is not our intent to speculate on the utility of 73% enriched uranium for NED. There is little information about intermediate levels of uranium-235 enrichment, such as 70% in the literature, presumably because it is an unattractive composition. The “Little Boy” device had “most” of its HEU enriched to 89% uranium-235, “for an average enrichment of only about 80%” [19]. It can also be noted that the critical mass of 70% enriched uranium is about 60% higher than the critical mass of 93% enriched uranium-235 diluted with uranium-238 [20]. This larger critical mass would require a larger NED and the effect may be nonlinear considering high explosives and the necessary implosion energy required to reach explosive supercriticality. Nevertheless, it cannot be excluded that a crude NED could be manufactured using reprocessed uranium from the end of core life. This should however not be interpreted to indicate that a crude NED is something that can be easily manufactured. Also a crude weapon would require substantial efforts in a large number of weapon design and manufacturing areas to do anything but fizzle. If something more advanced than a crude NED is desired, it may also be attractive to try to



**Figure 4:** The plutonium composition in Model 3 as a function of operation time of the submarine.

enrich the 73% HEU to above 90%. However, the enriching will be more expensive compared to enrichment of natural uranium (where there is a mass difference of three atomic mass units between the isotopes in the feed), because the reprocessed uranium contains additional uranium isotopes making the mass difference between uranium-235 and other uranium isotopes (such as uranium-236 and heavier isotopes) as low as one atomic mass unit.

The evolution of the plutonium composition in Model 3 can be seen in figure 4. The main difference between the Virginia-like designs and Model 1 and 2 with respect to material composition of the irradiated fuel, is found in the plutonium vector. In Model 3 and 4, the relative content of plutonium-238 is less than half of that in Model 2 and roughly a third of that in Model 1. At the same time, the plutonium-239 content is roughly a factor of two higher than for Model 2 and a factor of three higher than for Model 1. The relative content of plutonium-240 is also higher for Model 3 and 4 than for Model 1 and 2, and is around 7%. Depending on how plutonium is classified, this could mean that the material may reach a fuel-grade classification rather than weapon-grade, assuming that the classification considers only the fractions of plutonium-239 and plutonium-240. Should a more comprehensive classification of plutonium be available that takes into account also other plutonium isotopes, the classification could however be very different. It can also be seen in figure 4 that the relative contribution of plutonium-238 and plutonium-240 to plutonium-242 is considerably lower in the first 15 years of operation. In fact, during the first 14 years, the plutonium consists of over 90% plutonium-239 and less than 5% plutonium-238. For shorter irradiation times, the relative contribution of plutonium-239 increases, and that of plutonium-238 decreases. The amount of neptunium in the end-of-life core is 5795 g in Model 3 (5997 g in Model 4), which could possibly be another non-proliferation concern, should the irradiated core be reprocessed.

It has been investigated what the reason is for the dramatic change in plutonium composition for Model 3 and 4 compared to Model 1 and 2. It was found that the lower fuel temperature in Model 3 and 4 had a minimal effect on the uranium and plutonium composition, as did the actual fuel irradiation history. The slightly lower enrichment of the Virginia core (93% as opposed to 97.3%) had only a minor impact, as did the slightly different volume fractions of the fuel. In fact, the change in plutonium composition was found to be a result of the relatively thick plates of fissile material in combination of the thinner water channels and that there is enough gadolinium in the absorber plates to last through the entire lifetime of the reactor. Thicker plates of fissile material lead to more neutron captures in uranium-238, which produces more plutonium-239. At the same time, more  $\text{Gd}_2\text{O}_3$  in the core means that more plutonium-239 (which would otherwise fission at thermal energies) remains at the end-of-life. An important production path of

plutonium-238 starts from uranium-235 (of which there is plenty in this fuel), which captures neutrons to form uranium-236 and uranium-237 before beta-decaying to neptunium-237. Neptunium-237 then absorbs another neutron and beta decays to plutonium-238. The Virginia-like core has less water and more fuel than the Model 1 and 2 cores, and neutron energies are thus higher. A higher neutron energy increases the probability that uranium-236, uranium-237 and neptunium-237 undergo fission, and decreases the probability that neptunium-237 absorb a neutron and decays to plutonium-238. All these effects lead to a lower plutonium-238 production.

Thus, the results here suggest that the details of the design may have a considerable effect on the composition of the plutonium that is produced, although it should be pointed out that on an absolute scale the plutonium amounts are relatively small (around 7 kg); just below the 8 kg defined by the International Atomic Energy Agency as a significant quantity (SQ). We have chosen to relate these values with the well-known SQs, despite knowing that it does not reflect the material needed in a NED since devices can be made with considerably less material. The reason for referring to SQs here is simply that it offers a way to relate results and numbers to a quantity that is commonly known in the safeguards community.

## 6. Conclusions and outlook

This work has been a study of non-proliferation aspects of the AUKUS deal, through studies of the submarine spent reactor fuel. Four main conclusions can be drawn from this work.

### 6.1 Technical non-proliferation concerns

Upon irradiation, the enrichment level decreases, and plutonium starts to be produced. The end-of-life core is still highly enriched to a level of around 70%, where the uranium could potentially be useful in certain weapon designs. [20] In total, the end-of-life core from each Virginia-like core contains about 436 kg HEU at an enrichment of 73%, which corresponds to more than 17 SQs according to the IAEA. The fresh fuel, however, contains almost 640 kg HEU enriched to 93% (more than 25 SQs).

The first conclusion is thus that the fresh, unirradiated HEU should be considered as a bigger proliferation threat than the irradiated HEU. However, given that this fuel is produced and supplied by a NWS already in possession of weapons material, we do not assess that the fabrication of the fuel and placement of it in a submarine increases the proliferation risk. However, if the core material was to be removed from the submarine in Australia, or if the operation of the submarine would be interrupted early on and require unplanned maintenance in Australia, this could be a source of concern.

On a higher level, proliferation risks related to additional states using the AUKUS deal as a precedent or inspiration for their own pursuit of nuclear-powered submarines would be more concerning if such activities were to be used to motivate the establishment of new facilities in the state. The state could then either choose to remove nuclear material originally part of the civil fuel cycle and placed under safeguards for naval propulsion purposes, or investigate ways to increase the amount of nuclear material available. Both alternatives would pose a challenge for the non-proliferation regime. In the former case, amounts of nuclear material outside international control would increase. On the latter case, additional nuclear facilities could be established to openly or covertly produce the fissile material needed for the reactor cores, or to manage the disposal or reprocessing/recycling of the fissile material after irradiation. The establishment would be associated with a non-zero risk of diversion of nuclear material from e.g., fuel fabrication plants or enriched stockpiled product intended for fuel fabrication (see Section 2 of this work). The nuclear facilities producing or processing the nuclear material would however not be exempted from safeguarded in signatory states to the NPT, which means that safeguards verification would come in at that point.

### 6.2 Fuel composition and NED usability

The results show that it is crucial to include heavy isotopes beyond uranium-235, uranium-236, uranium-238 and plutonium-239 to plutonium-242 in the analysis. Especially plutonium-238 is found to play a major role as it is produced in relatively large amounts, and significantly degrades the plutonium quality beyond what would be acceptable in a NED (although studies of the implications of varying fractions of plutonium-238 in a NED design can be found [21]). The quality of the plutonium produced depends on the irradiation conditions, and varies from being dominated by plutonium-238 to being dominated by plutonium-239 but with a significant contribution from plutonium-238. This contribution is however considerably lower in the first years of operation of the submarine, which in fact produces very small quantities of plutonium suitable for NEDs. In absolute terms, about 7 kg plutonium, just short of the 8 kg corresponding to an SQ, is produced at the end-of-life in the Virginia-like reactor cores studied here. Note however, that in order to obtain the plutonium, reprocessing activities are needed to separate it from the uranium and waste products. Typically, weapons-grade plutonium is used in NEDs, where the contribution from plutonium-238 is less than 0.05%. (Nuclear Weapons Archive, 1999) Practically all plutonium compositions with a plutonium-238 contribution < 80% (i.e. not only weapons-grade plutonium) can be used in a NED. [23] An evaluation of what the physics behind the 80% level is, is beyond the scope of this work, but it could be worth studying further whether this is truly a lower limit. How well the NEDs perform will vary depending on the plutonium composition, where those using a composition with

more plutonium-238, plutonium-240 and americium-241 are more likely to perform poorer than those with less.

The second conclusion is thus that plutonium suitable for use in a NED is produced at short irradiation times, but in small quantities. The plutonium produced at the end-of-life has significant contributions from (above all) plutonium-238 but also plutonium-240, which makes the material ill-suited for, but not impossible to, use in a NEDs. The reasons to why plutonium-238 is very undesirable in NEDs are that i) it emits so much thermal energy that it is very difficult to handle; that ii) plutonium-238 has a significant spontaneous neutron emission which may cause a nuclear explosive device to pre-initiate; and that iii) plutonium-238 has a very high alpha particle emission rate (short half-life) which contributes to alpha-neutron reactions with light impurities, again resulting in copious neutrons that may cause pre-initiation. Accordingly, we draw the conclusion that the use of this fuel material in submarines operated by a NNWS, does not automatically mean that material attractive for use in a NED will become available. However, large quantities of a nuclear material that could at least theoretically be used in a NED - neptunium - will be produced. Neptunium is also produced in civil reactors, but not in these quantities.

### 6.3 Material classification

The results show that the classification of plutonium in the nuclear safeguards community as weapons-grade, fuel grade or reactor grade based only on the plutonium-240 content appears inadequate here. In fact, one could see this study as an illustration of a situation where using definitions designed for nuclear safeguards thresholds are inappropriate for determining weapons applications. The universal safeguards definition of plutonium with a plutonium-238 content <80% was not intended to deal with discharged material from very high enriched fuel with ultra-high burnups, where specifications for minor isotopes and plutonium-238 content (responsible for producing heat from alpha decays) cannot be ignored from a NED-usability point of view.

A third conclusion from this work is thus that a more suitable definition of weapons-grade plutonium should be considered in the light of this work, stating limits on for instance plutonium-238 and plutonium-240, alternatively on plutonium-238, plutonium-239 and plutonium-240. This would better reflect the usability and proliferation concerns associated with plutonium, and also potentially direct safeguards resources to where they are best needed.

### 6.4 Civil reprocessing incentives

There are no indications that the irradiated nuclear fuel from the submarine reactor cores will be reprocessed. Nuclear fuelled submarines out-of-service are in many cases still awaiting dismantlement. In other cases the

decommissioned nuclear submarine reactor cores are stored. The simulations in this work have however shown a relatively large production of plutonium-238, which could potentially make the irradiated reactor material attractive for applications in civil use, such as heat sources for spacecraft and motivate reprocessing of the fuel. This would require separation of both uranium and plutonium, and possibly neptunium. Such reprocessing activities potentially motivated by the recovery of plutonium-238, would constitute a considerable challenge for the non-proliferation regime.

A fourth conclusion is thus that a pressing challenge for the non-proliferation regime could be a future interest in reprocessing the irradiated submarine reactor fuel. Should this happen, the safeguards community and the IAEA must also be prepared to monitor and verify elements such as neptunium in a regular fashion, similarly to uranium and plutonium, and not just on a voluntary basis.

## 7. Acknowledgments

We would like to acknowledge the Alva Myrdal Centre for Nuclear Disarmament at Uppsala University for supporting this work.

## 8. References

- [1] Von Hippel, F. (April 2019). Mitigating the Threat of Nuclear-Weapon Proliferation via Nuclear-Submarine Programs, Journal for Peace and Nuclear Disarmament, vol. 2, issue 1. <https://doi.org/10.1080/25751654.2019.1625504>
- [2] Rauf, T. (2022). The challenge of the naval nuclear fuel cycle to IAEA safeguards. First annual conference of the Alva Myrdal Centre for Nuclear Disarmament, Uppsala, Sweden.
- [3] Moore, G.M, Banuelos, C.A. and Gray, T.T. (2016). The History of HEU in Submarines and Surface Warships, in Replacing Highly Enriched Uranium in Naval Reactors Report (Nuclear Threat Initiative). Pp. 19-37. <http://www.jstor.com/stable/resrep14271.10>
- [4] Clarke, M. (September 21, 2021). The AUKUS Nuclear Submarine Deal: Unanswered Questions for Australia. The Diplomat. <https://thediplomat.com/2021/09/the-aokus-nuclear-submarine-deal-unanswered-questions-for-australia/>
- [5] Philippe, S. (September 17, 2021). The new Australia, UK, and US nuclear submarine announcement: a terrible decision for the nonproliferation regime. Bulletin of the atomic scientist, <https://thebulletin.org/2021/09/the-new-australia-uk-and-us-nuclear-submarine-an>

- nouncement-a-terrible-decision-for-the-nonproliferation-regime/
- [6] Ogilvie-White, T and Gower, J. (October 2021). A Deeper Dive into AUKUS: Risks and Benefits for the Asia-Pacific. Special report, Asia-Pacific Leadership Network for Nuclear Non-Proliferation and Disarmament, <https://cms.apln.network/wp-content/uploads/2021/10/A-Deeper-Dive-into-AUKUS-1.pdf>
- [7] Acton, J.M. (September 21, 2021 ). Why the AUKUS Submarine Deal Is Bad for Nonproliferation—And What to Do About It. Carnegie Endowment for International Peace, Commentary, <https://carnegieendowment.org/2021/09/21/why-aukus-submarine-deal-is-bad-for-nonproliferation-and-what-to-do-about-it-pub-85399>
- [8] IAEA, (1972). The Structure and Content of Agreements between the Agency and States Required in Connection with the Treaty on the Non-Proliferation of Nuclear Weapons. INFCIRC/153 (Corrected), <https://www.iaea.org/sites/default/files/publications/documents/infcircs/1972/infcirc153.pdf>
- [9] Carlson, J. (October 8, 2021). IAEA Safeguards, the Naval “Loophole” and the AUKUS Proposal. Vienna Center for Disarmament and Non-Proliferation, <https://vcdnp.org/wp-content/uploads/2021/10/Safeguards-and-naval-fuel-JC-211008.pdf>
- [10] Castelli, L. (February 23, 2022). Aukus and Iran: A noteworthy entanglement, European leadership network. European Leadership Network, Commentary, <https://www.europeanleadershipnetwork.org/commentary/aukus-and-iran-a-noteworthy-entanglement/>
- [11] Kamat, D. (September 23, 2021). AUKUS deal could strengthen Iran’s nuclear ambitions, The Arab Weekly, <https://theearabweekly.com/aukus-deal-could-strengthen-irans-nuclear-ambitions>
- [12] Naval Nuclear Propulsion. (1995). Report on Use of Low Enriched Uranium in Naval Nuclear Propulsion. Prepared by: Director, Naval Nuclear Propulsion, <https://fissilematerials.org/library/onnp95.pdf>
- [13] Carlson, J. (September 17, 2021). AUKUS Nuclear-Powered Submarine Deal – Non-proliferation Aspects. Commentary, Asia-Pacific Leadership Network, <https://www.apln.network/analysis/commentaries/aukus-nuclear-powered-submarine-deal-non-proliferation-aspects>
- [14] Ma, C and von Hippel, F. (Feb 2008). Ending the Production of Highly Enriched Uranium for Naval Reactors. The Nonproliferation Review, vol. 8, issue 1, pp. 86-101, <https://doi.org/10.1080/10736700108436841>
- [15] Hanson, B.P. (2013). Validation of the use of low enriched uranium as a replacement for highly enriched uranium in US submarine reactors. Thesis (M.S.), Massachusetts Institute of Technology, Dept. of Nuclear Engineering. <https://dspace.mit.edu/bitstream/handle/1721.1/103656/953286533-MIT.pdf>
- [16] Ippolito, T. (1990). Effects of variation of uranium enrichment on nuclear submarine reactor design. M.Sc. thesis, Massachusetts Institute of Technology, Dept. of Nuclear Engineering, <https://dspace.mit.edu/handle/1721.1/14094?show=full>
- [17] U.S. Department of Energy. (2009). Management of nuclear materials, National Nuclear Security Administration, Office of Nuclear Materials Integration, DOE O 410.2. Approved 8-17-09
- [18] Leppänen, J., Pusa, M., Viitanen, T., Valtavirta, V., and Kaltiaisenaho, T. (2015). The Serpent Monte Carlo code: Status, development and applications in 2013. Ann. Nucl. Energy, vol. 82, pp. 142-150, <https://doi.org/10.1016/j.anucene.2014.08.024>
- [19] Hansen, C. (2007). Swords of Armageddon, vol. 5, Chukelea Publications: Sunnyvale, CA, p. 112.
- [20] Glaser, A. (2006). On the Proliferation Potential of Uranium Fuel for Research Reactors at Various Enrichment Levels, Science and Global Security, vol. 14, pp. 1–24, <https://doi.org/10.1080/08929880600620542>
- [21] Kessler, G. (2017). Plutonium Denaturing by 238Pu. Nuclear Science and Engineering, vol. 155, issue 1, pp. 53-73, <https://doi.org/10.13182/NSE07-A2644>
- [22] Nuclear weapon archive. (1999). Nuclear weapons frequently asked questions. 6.2.2.9 Weapon Grade Plutonium, Nuclear weapon archive, Version 2.18: 20 February 1999, <http://nuclearweaponarchive.org/Nwfaq/Nfaq6.html>
- [23] U.S. Department of Energy. (1997). Nonproliferation and Arms Control Assessment of Weapons-Usable Fissile Material Storage and Excess Plutonium Disposition Alternatives. Technical Report, DOE0NN-0007, pp. 37-38. <https://doi.org/10.2172/425259>

# Data Validation Experiments with a Computer-Generated Imagery Dataset for International Nuclear Safeguards

Zoe N. Gastelum <sup>a</sup>, Timothy M. Shead <sup>b</sup>, Matthew Marshall <sup>c</sup>

<sup>a</sup> Sandia National Laboratories, International Safeguards & Engagements Department, 1515 Eubank SE, Albuquerque, NM, USA, 87123

<sup>b</sup> Sandia National Laboratories, Machine Intelligence and Visualization Department, 1515 Eubank SE, Albuquerque, NM, USA, 87123

<sup>c</sup> Sandia National Laboratories, Nuclear Verification Department, 1515 Eubank SE, Albuquerque, NM, USA, 87123

## Abstract:

Computer vision models have great potential as tools for international nuclear safeguards verification activities, but off-the-shelf models require fine-tuning through transfer learning to detect relevant objects. Because open-source examples of safeguards-relevant objects are rare, and to evaluate the potential of synthetic training data for computer vision, we present the Limbo dataset. Limbo includes both real and computer-generated images of uranium hexafluoride containers for training computer vision models. We generated these images iteratively based on results from data validation experiments that are detailed here. The findings from these experiments are applicable both for the safeguards community and the broader community of computer vision research using synthetic data.

**Keywords:** Computer vision, synthetic data, international nuclear safeguards, uranium hexafluoride.

## 1. Introduction

The International Atomic Energy Agency (IAEA) operates under the United Nations and is responsible for verifying that nuclear materials and facilities across the globe are limited to peaceful use. They do so by implementing and monitoring international nuclear safeguards: measures to account for nuclear materials and verify the design and operation of nuclear facilities. Increasing interest in nuclear energy technologies, growing inventories of nuclear material, and limited IAEA safeguards resources are compelling the IAEA to be more efficient in safeguards monitoring.

Computer vision models could increase IAEA safeguards efficiency, by augmenting visual tasks conducted as part of the IAEA's safeguards mission. Examples of visual tasks for which computer vision research and development is currently underway throughout the safeguards community include:

- Object and change detection for nuclear-relevant sites via satellite imagery analysis (Rutkowski, Carty, & Nielsen, 2018).
- Collection, triage, and information recall for open-source images (Feldman, Arno, Carrano, Ng, & Chen, 2018) (Gastelum & Shead, 2018) (Arno, 2018).

- Reviewing surveillance camera data for specific objects or patterns of life (Smith, Hamel, Hannasch, Thomas, & Gaiten-Cardenas, 2021) (Thomas, et al., 2021) (Wolfart, Casado Coscolla, & Sequeira, 2022).
- Supporting inspector indoor localization at complex nuclear facilities (Wolfart, Sanchez-Belenguer, & Sequeira, Deep Learning for Nuclear Safeguards, 2021).
- Supporting inspectors with digital assistants for visual tasks (Smartt, Gastelum, Rutkowski, Peter-Stein, & Shoman, 2021).

Despite this surge in research, access to sufficiently large, relevant datasets remains a challenge. Relevant data for international safeguards research and development are rare for multiple reasons. First, real international safeguards data are sensitive and held in confidence by the IAEA and are therefore inaccessible for most research. Second, safeguards-relevant data may be either commercially sensitive or have national security sensitivities for states. Third, relevant data may be lost to history due to obsolete file formats, data corruption, or lack of digitization. Finally, relevant data might not exist; for example, images of technologies that are physically feasible but not widely adopted may be of interest to detect future proliferation activities, but images of these technologies are non-existent.

In response to the rarity of available safeguards data, we have created a large, open-source, safeguards-relevant imagery dataset called Limbo. Limbo contains one million synthetic (computer-generated) images intended for computer vision research and development. The images include detailed, automatically-generated segmentation mask, contour, and bounding box annotations, see Figures 8 – 10 for examples. We also provide a small collection of annotated real-world images for validation that include well-documented copyright information to simplify publication. Our goal for the Limbo data, and for synthetic data more broadly, is to develop computer vision models trained solely on synthetic data that can achieve state-of-the-art performance when evaluated on real-world data.

We applied several criteria in selecting a subject matter for our synthetic data. We wanted the subject to be:

- Unclassified, for easier development and dissemination of the data.
- Visually distinct, to facilitate labeling of real-world validation data.
- Relatively common, to ensure that we would have sufficient real-world data to support our validation activities.
- Prevalent within the nuclear fuel cycle, so the generated data could support the broadest possible research and development, without being tied to a single process or type of facility.<sup>1</sup>

Based on these criteria, we opted to generate images of containers used to store and transport uranium hexafluoride (UF6) throughout the commercial nuclear fuel cycle. We specifically focused on two general models of UF6 containers: 30B and 48-type containers.

30B containers are 30-inch cylinders used to transport uranium-235 enriched up to 5%. These containers are primarily found at uranium enrichment facilities (as the product output) and fuel fabrication plants (as the product input). See Figure 1 for a real-world example.

48-type containers refer to a class of 48-inch containers used to store and transport natural and depleted UF6. We included three common designs of 48s: 48X and 48Y containers are used for storage and transportation, while 48G containers are characterized by the lack of an apron and are used exclusively for storage. 48-type containers can be found at uranium conversion plants (as the product output), uranium enrichment plants (as the input, and to store depleted tails), and fuel fabrication facilities (as input for natural uranium fuel). See Figure 2 for a real-world example of 48Y containers.

In addition to relevant containers, the Limbo data includes examples (both real and synthetic) of distractor objects including propane tanks, gas canisters, beer kegs, 55-gallon drums, and more. Synthetic distractors have the full metadata suite, while real-world distractor metadata includes only the class “distractor”.

In the remainder of this paper, we describe the data generation process (Section 2), validation workflow (Section 3), data validation experiments and results (Section 4), and discussion and implications for future research (Section 5). We also provide information on how to access and use the Limbo data, and descriptions of the Limbo dataset contents (Section 6).



**Figure 1:** 30B uranium hexafluoride container at the IAEA Low Enriched Uranium Bank in Kazakhstan. Credit: IAEA, 2019.



**Figure 2:** 48Y containers at Urenco, Netherlands. Credit: IAEA, 2015.

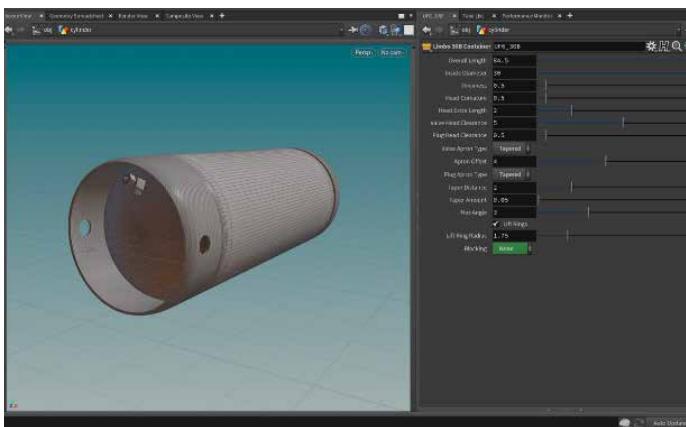
## 2. Data Generation

In this section, we describe our workflow to generate synthetic images. This process includes the creation of three-dimensional (3D) models of UF6 containers, random sampling of 3D model parameters, and placement in real or virtual environments, followed by rendering to produce 2D images and metadata.

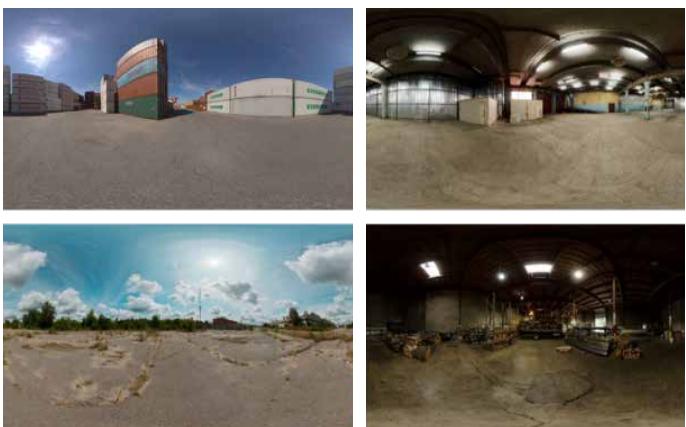
### 2.1 Developing 3D Models

We developed 3D models of our UF6 containers using SideFX Houdini (<https://www.sidefx.com/products/houdini/>), a procedural 3D modeling and animation tool widely used in films, television, and game design. A screenshot of the Houdini workspace with a parameterized 30B container model is provided in Figure 3. The 3D models were informed by technical standards and specifications published in open sources by industry partners and professional societies, with some subjective adjustments to better match the containers in real-world images. Sources that were especially useful for our model development included:

<sup>1</sup> Through a collaboration with researchers at Lawrence Livermore National Laboratory, we had access to a set of images collected from open sources that provided an indication of overall prevalence in open sources and served as a seed for additional data collection.



**Figure 3:** 3D CAD model of a 30B UF6 container in the Houdini software.



**Figure 4:** Sample indoor and outdoor HDR environments.

- Uranium Hexafluoride: A Manual of Good Handling Practices (United States Enrichment Corporation, 1995);
- American National Standard for Nuclear Materials - Uranium Hexafluoride – Packaging for Transport (American Nuclear Standards Institute, 2001); and
- Uranium Hexafluoride: Handling Procedures and Container Descriptions (Oak Ridge Operations, 1987).

Once the cylinder models were created in Houdini, we used the Allegorical (now Adobe) Substance 3D paint software to generate multiple sets of “paint job” textures for the cylinders in varying styles and levels of wear.

In addition to the cylinders, Limbo also includes a variety of distractor objects. Unlike the UF6 containers, the distractors are common objects not specific to the nuclear fuel cycle (such as propane tanks, welding gas cylinders, wine barrels, etc.) that are widely available commercially. Therefore, for the distractor objects we procured 3D models from an online 3D model marketplace (<https://turbosquid.com>) with appropriate permissions for use and distribution.

## 2.2 Model Placement and Environment

As backgrounds for our 3D container models, we provided two major classes of environment: real-world and synthetic.

Real-world environments were created using panoramic High Dynamic Range (HDR) photographs, which capture a 360-degree image of a scene, and—unlike normal photographs—use special techniques to record the full range of light intensity for each pixel. In this way, an HDR image samples light intensity from all directions simultaneously. This makes it possible for an HDR image to provide the photographic backdrop for a scene while also supplying realistic, nuanced lighting. We used HDR images collected from open sources with appropriate permissions, including indoor and outdoor scenes. A majority of the images were



**Figure 5:** Synthetic 3D oil refinery environment.

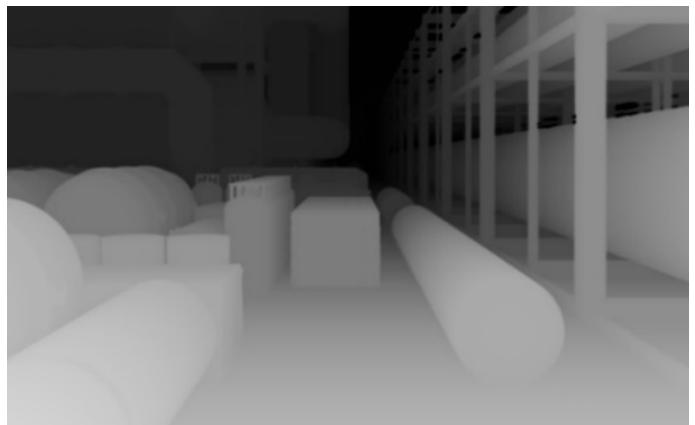
industrial scenes similar to environments where real 30B and 48 containers would be found, but we also included several studio and other scenes for variety. Several examples of our HDR backgrounds are shown in Figure 4 (the images are warped by the panoramic perspective but display normally when projected into the final 2D rendered images).

A limitation of using real images as backgrounds is that the scale and perspective of the background may not match the 3D objects in the foreground. This can lead to cylinders that seem unusually large or small, relative to their surroundings, or appear to be floating in air instead of properly grounded. Although ultra-realistic synthetic images are not necessarily required for robust model training (Tremblay, et al., 2018), we addressed this by providing a fully synthetic 3D environment in later Limbo images, based on an outdoor scene of an oil refinery. The 3D oil refinery provided a large and diverse setting for our containers, had industrial features similar to a nuclear fuel cycle facility, and guaranteed that 3D foreground objects perfectly matched the background in proportions and perspective.

We inserted skies from the real-world HDR images into the synthetic oil refinery for additional control and manipulation of lighting. A scene from the refinery is shown in Figure 5.



**Figure 6:** Examples of synthetic 30B containers in a variety of real-world HDR environments.



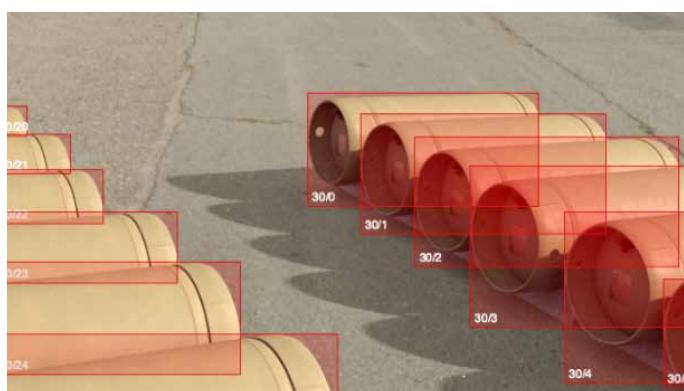
**Figure 7:** Depth map image suitable for use as LiDAR ground truth.



**Figure 8:** Segmentation masks derived from per-instance occupancy data.



**Figure 9:** Object contours for scene objects.



**Figure 10:** Bounding boxes derived from per-instance masks.

The refinery scene was procured from the same online 3D model marketplace as our distractor objects.

For both HDR and synthetic 3D environments, we used Houdini to assemble complete scenes via random sampling of parameters such as the number, type, organization (scattered versus rows) and placement of containers within the environment; camera location, orientation, and lens; environmental lighting conditions; container material appearance; container condition (new, scratched, rusty, etc.); and type of cradles (wood, concrete) used to support the containers.

### 2.3 Rendering and Metadata

Once the individual 3D scenes were assembled, we used Redshift 3D - a GPU-accelerated, biased render engine implementing a physically based rendering (PBR) lighting model – to render 2D images. Importantly, each of our 2D images comprises several layers and multiple files created explicitly with the needs of computer vision research in mind. Each of our images includes the following:

1. A 720 by 720-pixel HDR visible spectrum image (Figure 6).
2. A corresponding depth map image, where the value of each pixel is its distance from the camera. This data can be used by researchers interested in training models on light detection and ranging (LiDAR) information (Figure 7).
3. Sub-pixel occupancy data for every object in the image (Figure 8). Storing the per-pixel areas occupied by multiple objects allows us to generate a variety of perfect ground truth information, including per-instance and per-class segmentation masks, contours (Figure 9), bounding boxes (Figure 10), and tags for image classification. Sub-pixel occupancy data is stored in compressed form using the efficient and elegant Cryptomatte file format (Friedman & Jones, 2021).

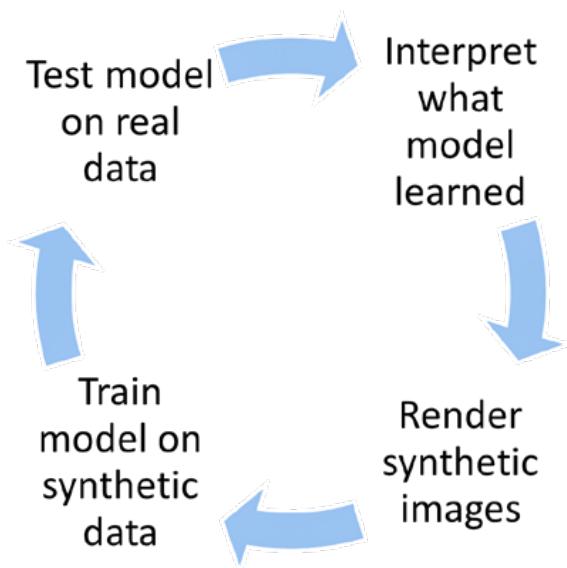
4. Metadata including the contents of a scene, background, lighting, and camera parameters.

The images were rendered as a series of thematic “campaigns”, which are used to describe the image sets in experiments in Section 4, and in describing the data in Section 6. Importantly, the synthetic images are perfectly labelled because the labels themselves are generated at the same time as the images, using the same 3D scene information.

In addition to the data proper, we also developed an application programming interface (API) to simplify accessing the full data and metadata for each image. Information on the API is available at: <https://limbo-ml.readthedocs.io/>.

### 3. Data validation procedure

We developed a data validation workflow to ensure that computer vision models could be trained using our synthetic images. In this section, we describe the data validation workflow, the findings from our validation activities, and how they informed later iterations of the Limbo data. This was a crucial step in the data generation process since the Limbo data is intended for computer vision research and development. Our workflow was iterative, including four steps: rendering synthetic images, training models on synthetic data, testing models on real data, and interpreting what the models learned. Then, we incorporated those lessons when rendering new synthetic data. The workflow is depicted in Figure 11, and each step will be described in additional detail below.



**Figure 11:** Synthetic data workflow.



**Figure 12:** Sample Limbo reference image labelled by our research team (right).

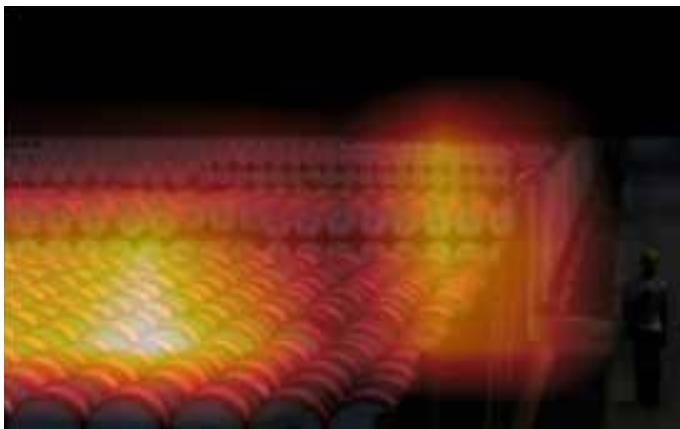
#### 3.1 Training Models on Synthetic Data

Our primary goal was to validate that popular computer vision models could learn from our synthetic images. We selected two main types of computer vision models for validation: image classifiers and object detectors. For each model type, we fine-tuned multiple pre-trained architectures. For image classification we used ResNet-50 (He, Xiangyu, Shaoqing, & Sun, 2016) and Inception (Szegedy, et al., 2015). For object detection, we used YOLO-v5s (Jocher, et al., 2022) (which is built upon Yolov3 (Redmon & Farhadi, 2018)), SSS (Liu, et al., 2016) and Faster R-CNN (Ren, He, Girshick, & Sun, 2015).

We conducted a series of experiments that trained the models using subsets of the synthetic Limbo data, to validate that the models could learn from the data and to identify any issues with the data. The results of those experiments are discussed in Sections 4.2 and 4.3.

#### 3.2 Testing Models on Real Data

After training models using synthetic data, each model was tested on the curated collection of real-world data—which we refer to as reference data—that is included with Limbo. The reference data contains images of both types of relevant UF6 containers and numerous distractors. Each real-world image in the reference dataset is accompanied by metadata that includes copyright information and ground truth bounding boxes manually labelled by members of our project team (Figure 12). As one can see in Figure 12, the manually drawn bounding boxes are not as perfect as those generated automatically for our synthetic data. However, we followed a consistent protocol for bounding box labeling, which was subject to inter-rater quality checks within our team. We think this protocol resulted in higher quality labels than many of the open-sourced labels used in the large benchmark datasets, which have documented errors and quality issues (Northcutt, Athalye, & Mueller, 2021).



**Figure 13:** Example explanation from a false positive image classification, using GradCAM to visualize salient pixels.

### 3.3 Interpreting What the Models Learned

We interpreted the results of our models on reference data to identify potential issues with how the computer vision models were learning from the Limbo data. The mechanisms we used to interpret model learning differed by model type.

For the image classification models, we used machine learning explainability methods to visualize the pixels of an image that were most influential in each prediction. We reviewed the false positive and true positive predictions to interpret the features that were informing positive classification results.

Due to the variation in responses from machine learning explainability techniques, we simultaneously viewed the explanations from three explainability models: GradCAM (Selvaraju, et al., 2017), Guided GradCAM, and Gradient SHAP. An example from an early classification model's false positive explanation is in Figure 13. From these explanations, we interpreted what features of the synthetic Limbo data were more relevant during model training and inference on the real-world data. Figure 13 shows an early example of indications that we needed to add distractor cylindrical objects into the dataset. Additional details of the classification model validation results are below in Section 4.2.

For the object detection models, we opted to use the placement of the bounding boxes to interpret the most relevant areas of an image used to make an inference. For example, in Figure 14 the model incorrectly detected 48-type containers around a human and a 30B container. Similar to how we interpreted the image classification results, we reviewed the object detection true positives and false positive detections and anecdotally devised potential implications of our Limbo data based on what we observed the object detection models were learning. The example in Figure 14 is one of dozens of false positive detections that prompted us to integrate synthetic people into our Limbo data.



**Figure 14:** Example false positive detection results, in which the YOLO-v5s model identified a human and a 30B container as 48-type containers.

We also evaluated images of false negative classifications and detections. We attempted to observe common features of the images that may have impacted the failure to correctly classify or detect the object of interest in the images. The process we used for reviewing misclassifications and what we learned from the process is described in (Gastelum, Shead, & Marshall, 2022).

### 3.4 Rendering New Synthetic Data

From our analysis of image classification and object detection results, we made multiple additions to our Limbo dataset, including the addition of cylindrical distractor objects and arranging containers into rows. After making updates to the Limbo data, we re-trained and re-tested our models. Selected results and findings from those activities, including experimentation with subsets of the Limbo data, are described in the following sections.

## 4. Data validation experiments

The purpose of the data validation experiments was to confirm that computer vision models could be successfully trained on our synthetic data and tested on real data. Though we imagined that researchers or model developers could have access to a small amount of real data, we intended to prepare this data under the assumption that it would not necessarily be augmented by real data. Prior research on the use of synthetic data for training models typically includes large quantities of real-world data such as (Ekbatani, Pujol, & Segui, 2017) (Gaidon, Wang, Cabon, & Vig, 2016) and (Movshovitz-Attias, Kanade, & Sheikh, 2016), and achieves good model performance. However, the sizes of the real datasets in these papers (tens of thousands of real images) are still beyond the reach of our intended application spaces. We have previously examined the impact of augmenting synthetic data with small numbers of real images, with resulting model performance being approximately the same (Gastelum & Shead, 2020). Therefore, our

validation experiments focused on training models exclusively with synthetic data and testing them with real data.

Descriptions of our validation experiments in which we train models on synthetic data and test them on real data are detailed in Sections 4.2 and 4.3. These experiments utilize image classification models and object detection models. While there are other relevant computer vision model types available such as image segmentation models, we think that these two types provide sufficient evidence for our validation tests. Additional experiments with segmentation masks or other model types could prove to be interesting future research.

Model performance in these experiments was measured in two ways. First, for image classification models, accuracy measures are dependent on the class ratios present in the test data, so we evaluated model performance using two common computer vision performance metrics: precision and recall. Precision is the percentage of items predicted to be members of a class that actually are members of that class (true positives divided by the sum of true positives and false positives), while recall is the percentage of class members that are predicted to be members of that class (true positives divided by the sum of true positives and false negatives).

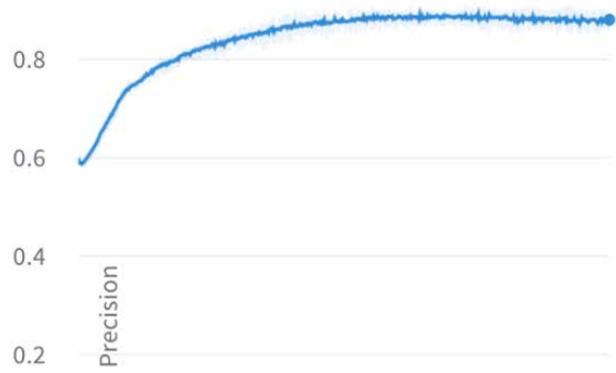
Second, for object detection models, we used a hybrid scoring approach. We first evaluated object detection models with the industry standard measure of performance mean Average Precision (mAP), which considers model performance on multiple object types, evaluation of positive and negative identifications, and evaluation of the predicted bounding box compared to the ground truth bounding box (for a useful tutorial, see (Tan, 2019)). For our evaluations, we set the intersection over union (IOU) threshold of 0.25. We used this lower-than-typical IOU standard based on our deployment assumption that the detection of a relevant object, even with an imperfectly aligned bounding-box, could still support analysts in finding indications of nuclear activity.

It is important to note that it was not the intent of these experiments to spend significant resources in fine-tuning hyper-parameters for best model performance. Rather, we used these validation experiments to suggest improvements for our synthetic data and to obtain a rough estimate of model performance when using it for training.

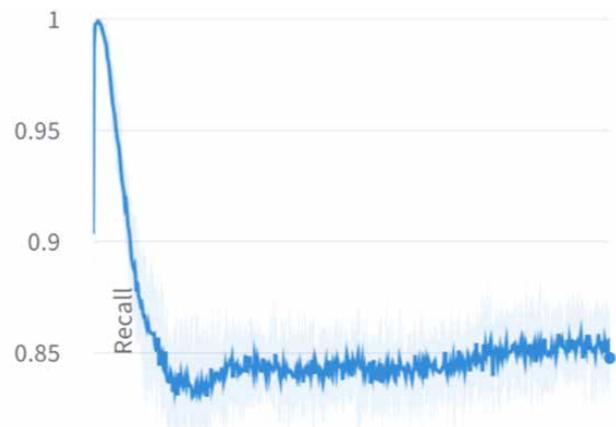
#### 4.1 Confirmation of Model Implementation

Although our focus for eventual deployment is on the train-synthetic, test-real use-case described above, all of our experiments are tested on synthetic data during training too - this allows us to validate that the code is working properly and the models are successfully training. As one extant example, the following figures show train-synthetic, test-synthetic results for one set of experiments where we

trained ten ResNet-50 models for 500 epochs using 5000 synthetic images of type 48 containers from campaign 17, and tested using 1000 additional synthetic images from the same campaign. As can be seen, we achieve excellent precision (>87%) (see Figure 15) and recall (>85%) (see Figure 16) on the type-48 identification task (metrics are the results averaged from evaluating all ten models).



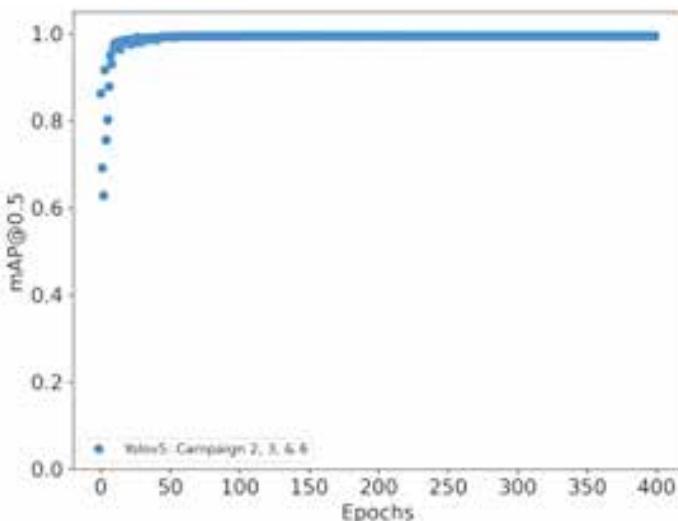
**Figure 15:** Precision results for image classification implementation test.



**Figure 16:** Recall results for image classification implementation test.

For our object detection models, we fine-tuned the pre-trained YOLO-v5 model with synthetic images from the Limbo dataset. We used 8,000 synthetic images for training, and 2,000 synthetic images for testing. The dataset was comprised of images of single 30B or 48-type containers from Campaigns 2 and 3, respectively, and background (no containers) images from Campaign 6. We balanced the dataset with equal number of negative (background) and positive (either a 30B or 48-type container present) examples. For the positive examples, we had the same number of 30B and 48-type containers. The YOLO-v5 model was trained for 500 epochs.

Like the image classification models, we expected the performance of our train synthetic-test synthetic object



**Figure 17:** mAP scores using 0.5 detection threshold implementation test.

detection models to be high. Using a threshold of 0.5 for mAP, performance of the object detection models was near-ceiling as shown in Figure 17.

#### 4.2 Image Classification Validation

Our first set of computer vision validation experiments were focused on image classification. For these experiments, we fine-tuned pre-trained ResNet-50 models using our synthetic Limbo data. The models were trained as one-class classifiers, with a sigmoid output between zero and one, where larger numbers indicated stronger predictions of the container class, and lower numbers indicated lower prediction of the container class. We elected 0.5 as the threshold for container classification, so that images with scores higher than 0.5 were considered a container class and images with scores lower than 0.5 considered a non-container class for the purposes of our evaluation.

We trained 10 models for each experimental run, using randomized initiation points for each model to ensure that training results were not serendipitous. We used this approach instead of cross-validation in order to train the models exclusively on synthetic data in each run and test them exclusively on real data (where cross-validation techniques would shuffle these training and test data sets). And we tested their performance on our full set of real images and recorded the average of the models' performance.

In Figure 18 and Figure 19, we show the results for all ten of the models but describe overall performance in relation to the mean of the ten models. Our classification experiment focused on single 30B container classifications and the experimental manipulation of the content of the negative training examples—either plain backgrounds, or synthetic distractors. For each trained model, we used an equal split of positive and negative examples.

In the first set of models (yellow/green tones along the bottom of Figure 18 and the top of Figure 19), the ResNet-50 model was trained on synthetic images of single 30B containers, with negative examples from background images without any containers. For these initial models, precision scores centered around 0.28 (lower cluster in Figure 18), and recall scores around 0.5 (higher cluster in Figure 19). The second set of models (in red/orange tones along the top of Figure 18 and the bottom of Figure 19), the ResNet-50 models were trained on the same relevant containers, but with distractor containers as negative examples instead of backgrounds only. The precision scores for these models increased significantly, to around 0.58 (higher cluster in Figure 18), while the recall scores were around 0.35 (lower cluster in Figure 19).

The large increase in precision between the first and second set of models indicates that the models trained with synthetic distractors were better at selecting images with relevant containers and not selecting images without relevant containers. The decrease in recall scores between the first and second set of models indicates that the models became less likely to classify relevant containers than before.

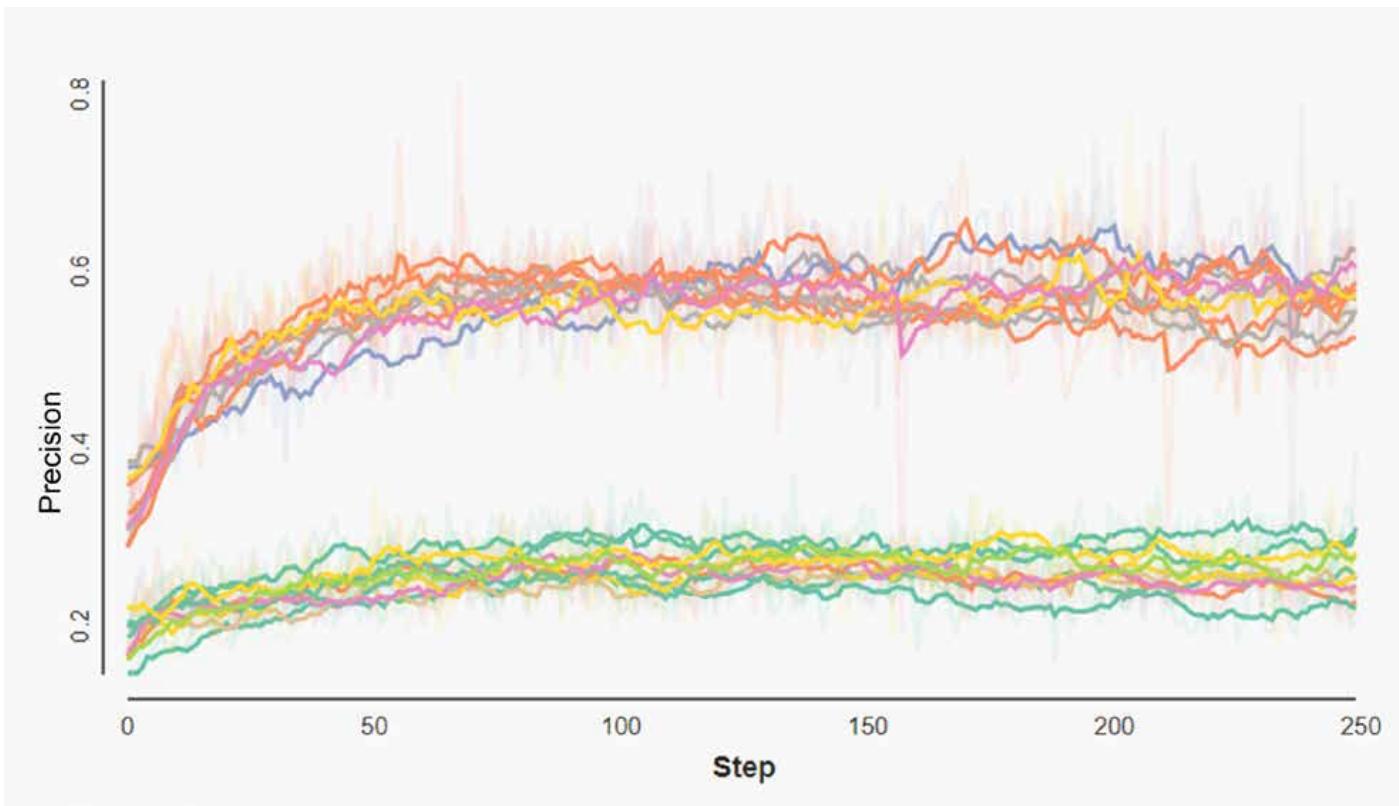
We observed that as we made changes to the content type of the synthetic data, the models reacted in predictable ways—specifically, learning to be more discriminating with cylindrical objects before classifying them as relevant containers.

#### 4.3 Object Detection

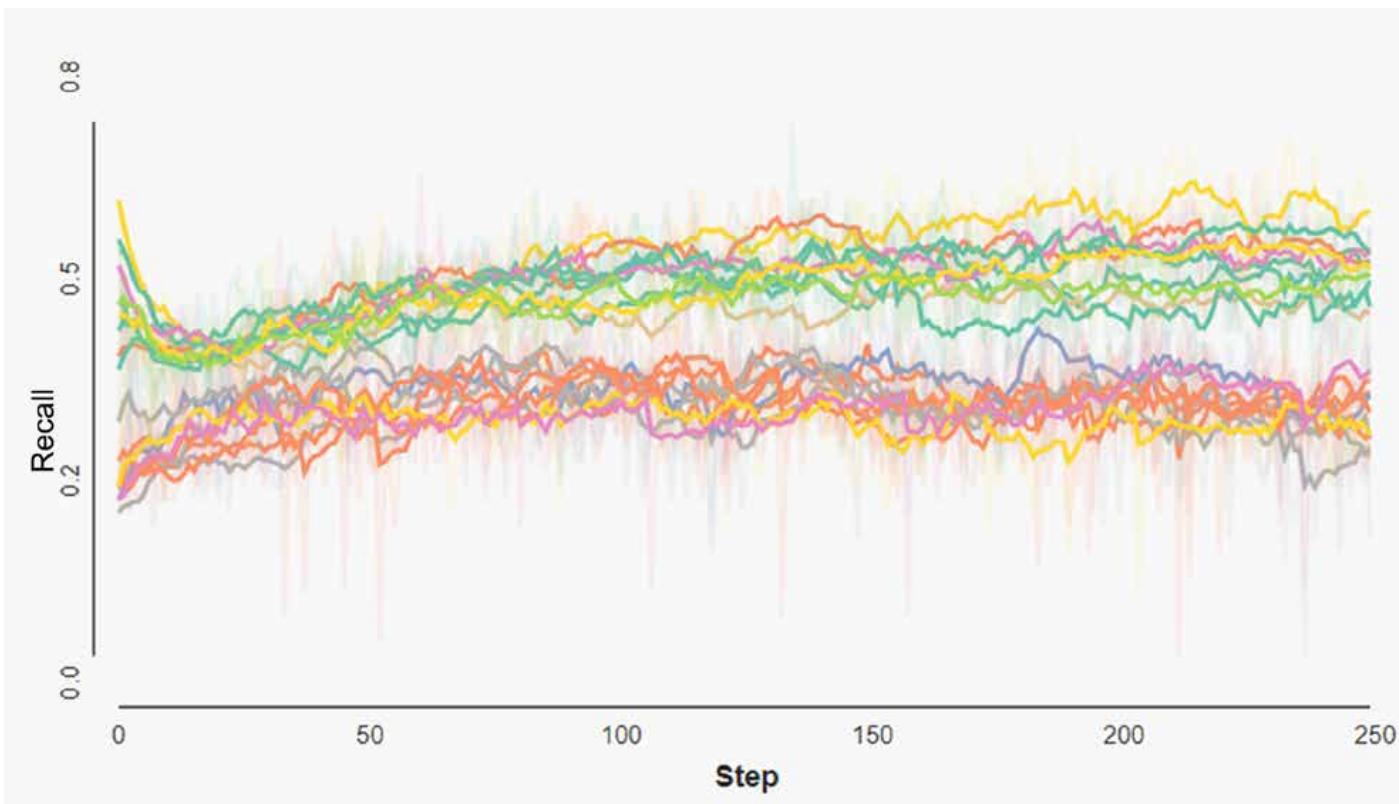
Our second set of validation experiments focused on object detection models. The object detection experiments evaluated models trained using subsets of Limbo to see how those subsets impacted model performance. In these experiments, we used an equal number of positive and negative examples to train the model. We considered a positive example to include one or more relevant containers of interest, and a negative example to contain no objects of interest (only background images or distractor containers). In these experiments, we used the YOLO-v5s object detection model, with an intersection over union (IOU) threshold of 0.25. Additionally, we calculated mean Average Precision (mAP) scores only for the 30B and 48-type containers.

As a baseline for performance, we trained models with 10,300 images containing individual containers (30B and 48-type containers). We compared performance of the baseline models to two alternatives: first, we trained models with images containing single containers (30B and 48) and single rows of containers (30B and 48). Second, we trained models with the same images, plus images containing distractors and individual containers.

Like we did for image classification, for each experiment, we trained 10 models with randomized initiation



**Figure 18:** Precision scores from our image classification experiment show that image classification models trained with distractor objects (top cluster of red/orange lines) had higher precision than models trained without distractors using only background scenes as negative examples (bottom cluster of yellow/green lines).



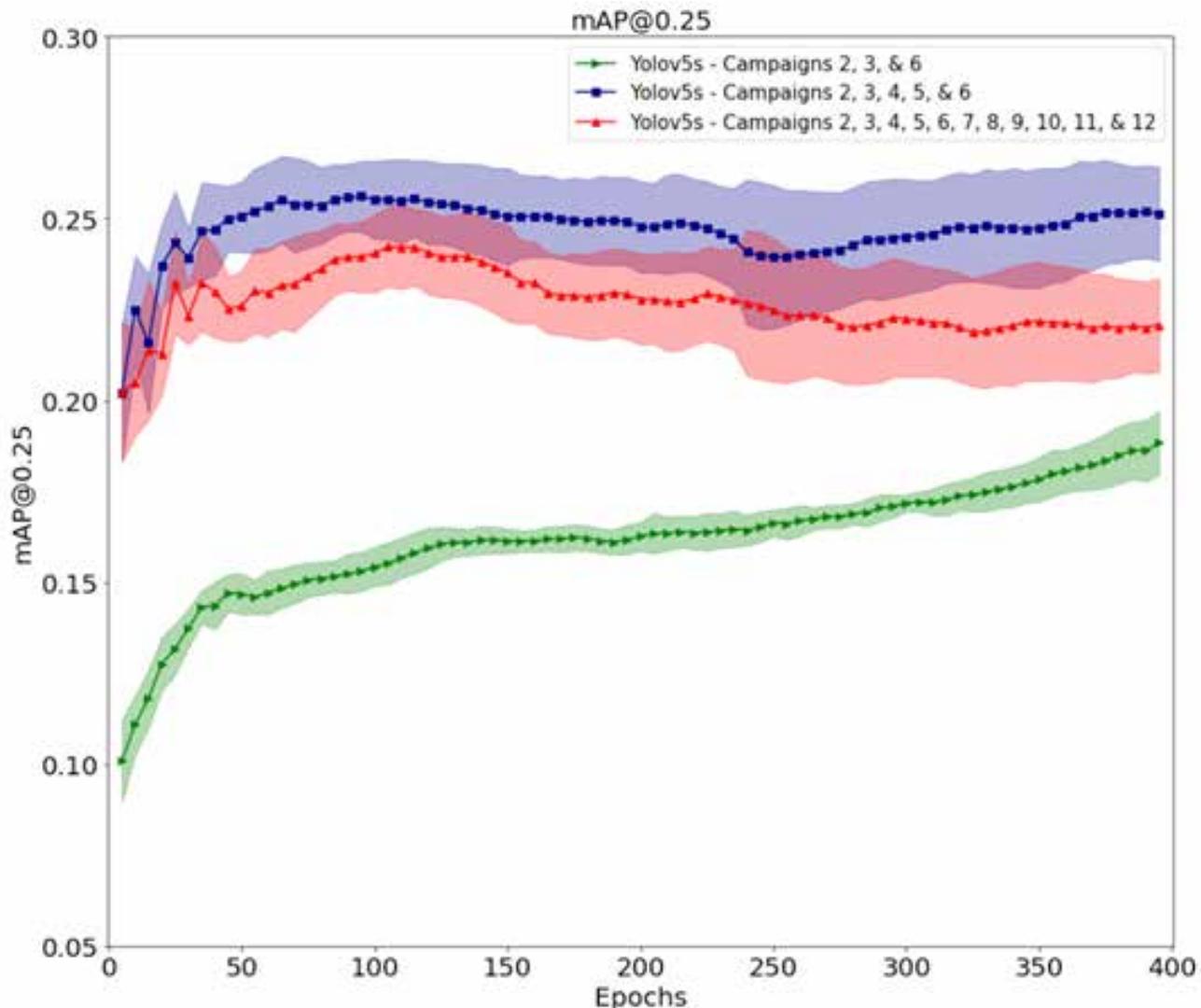
**Figure 19:** Recall scores for an image classification experiment show that image classification models trained with distractor objects as negative examples (bottom cluster of red/orange lines) had lower recall than models trained with background scenes as negative examples (top cluster of yellow/green lines).

parameters and took the mean of their results to ensure that test results were not the product of an especially high- or low-performing model. We found that by including images with rows of containers along with others showing individual containers during training, the mAP score improved compared to the baseline model where only individual containers were present, as shown in Figure 20. The real-world images contain scenarios where the relevant containers are in rows, and through inspection of the object detection results, we noticed models trained without examples of containers in rows, i.e., only using individual containers, struggled to identify examples when presented with a row of containers. By providing the model with examples of containers in rows in the training data, the model was able to learn that more containers were present and detect them.

The model trained using both containers and distractors increased the mAP score relative to the baseline model but did not improve performance relative to individual

containers and rows. The two scenarios are within statistical deviation of each other, but the mean mAP score for models trained on individual containers and rows is higher. In this case, we hypothesize that by training the model with examples of distractors, especially distractors occluding UF6 containers, the model learned features of the occluding object and incorrectly associated it with the 30B or 48 containers, lowering the mAP score. Furthermore, by including distractors in the categories for the model to learn from in the training set, the object detection problem becomes harder because the model has more options to choose from, and we observed that the model confused 30B or 48-type containers for distractors in some instances, which also lowered the mAP score.

To better compare the impact of different synthetic images on our image classification and object detection models, we conducted an analysis in which we judged both models using common metrics. We adapted the signal detection



**Figure 20:** Mean Average Precision (mAP) Scores for Object Detection Experiments. As variety and complexity of training data increased, so did model performance. There was a minor difference in performance between models that were trained with individual and rows of relevant containers and models that also included distractor objects. The green, blue, and red lines indicate training runs with single containers (30B and 48) and background; single containers, background, and single rows of containers; and single containers, background, distractors plus single containers, and single rows of containers, respectively.

performance metrics used in image classification (true positive, true negative, false positive, and false negative) for object detection as follows:

- For any image that had an object of interest (as defined by our team's labeling), an object detection-generated bounding box for that type of object in the image was considered a true positive regardless of its location within the image.
- For any image that did not have an object of interest, the absence of an object detection-generated bounding box of that type in the image was considered a true negative.
- For any image that had an object of interest, but the object detection model did not place a bounding box of that type anywhere in the image, it was considered a false negative.
- For any image that did not have an object of interest, but the object detection model placed a bounding box of that type anywhere in the image, it was considered a false positive.

We present the performance of our object detection models when they were evaluated as classifiers in Figure 21. We provide a summary of observations from assessing our object detection models as classifiers, and model-to-model performance comparisons, below.

First, we observed that increases in performance from including more diverse images in training (as shown in Figure 20) was not as great for the object detection models when they were used as classifiers. This is likely due to an increase in the baseline model performance coming from the lower bar for true positives than for correct object detection.

Second, we had not previously tested differences in computer vision model performance between 30B and 48-type containers. Our early image classification testing focused mostly on 30B containers, and we did not differentiate container types in performance reporting in earlier object detection models. With this new testing, which included both types of containers and easily differentiated results based on how classification results are reported, we found that the object detector—when measured like a classifier—has a higher precision and recall with 48-type containers compared to the 30B containers. This may be due to the more visually distinct features of the 48-type containers compared to the 30B containers.

Third, we found stable patterns in the trade-off between precision and recall as we increased the variety in the training data. In Figure 18 and

Figure 19, we show that as we increased the variety, precision scores increased (i.e., a higher proportion of data

retrieved was relevant) and recall scores decreased (i.e., fewer of the total relevant items were recalled), we see the same pattern in the object detection models when they are evaluated as classifiers.

We had anticipated this result, that as the models learn more about other types of containers that exist, they become more discriminating in their classifications and therefore may also miss more relevant items.

Finally, when we compared the performance of our image classifiers and object detectors on the same metrics of precision and recall, we found that both model types had similar precision scores, but recall scores were significantly higher for the object detection models. The underlying models are different, so it is difficult to make broad generalizations about what this could mean for computer vision generally, but it could indicate that models trained as object detectors are better able to identify—based on the more specific nature of the training data—relevant features that could increase recall, thereby decreasing the potential number of relevant items missed by these models.

#### 4.4 Interpreting Model Results

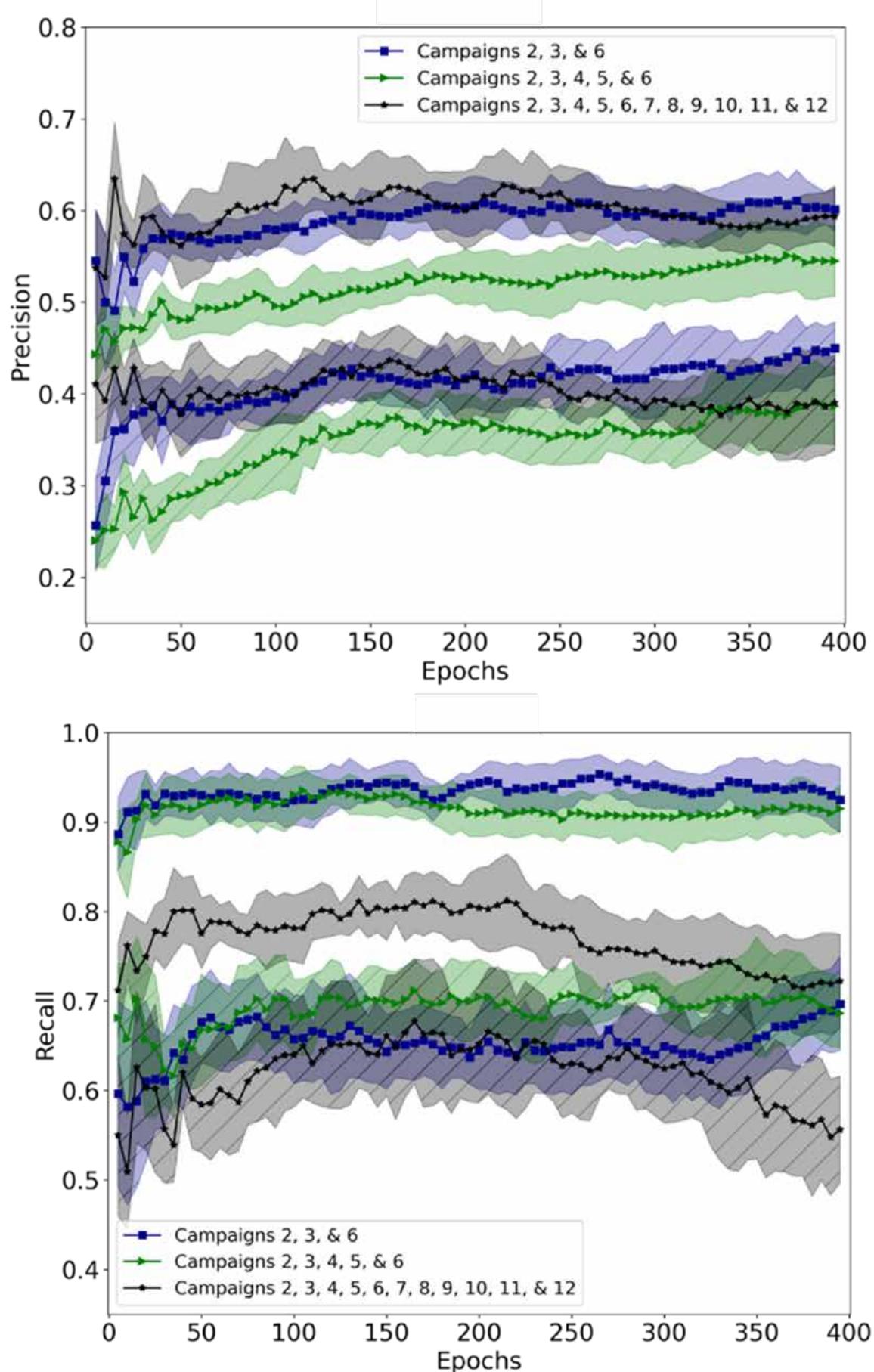
As described in Section 4, we interpreted the computer vision results using explainability techniques for the image classification models and visualization of the bounding boxes for the object detection models. Our most notable observations, and their subsequent impacts on the Limbo data, are described here.

**Relevant containers in rows.** One of our first observations from the image classification explainability activities was that when relevant containers were pictured in rows, a model that was trained on single containers only appeared to be focusing primarily on the first one or two containers. In response, we began generating rows of relevant containers such as might be seen in a shipping or storage area. These changes can be observed in campaigns 4 and 5.

**Synthetic distractors.** We also observed in our image classification explainability tests that the models were recognizing many real-world cylindrical objects as 30B or 48-type containers. We think this was caused by negative examples in early trials, which consisted of backgrounds without any additional synthetic content, such as synthetic cylindrical distractors. In response, we introduced synthetic distractors—primarily cylindrical, round, industrial objects. These changes can be observed starting in Campaign 7.

**Synthetic distractors in groups.** We thought it would be informative to render our distractor objects in groups or clusters, instead of the well-aligned rows of campaigns 4 and 5. This change can be seen starting in campaigns 8 and 9.

**Partially occluded containers.** As a follow-up to the changes made in point 1, we also wanted to occlude containers with distractor objects rather than just relevant containers.



**Figure 21:** Applying signal detection performance assessments for determining precision (top) and recall (bottom) to the object detection results. The hatched lines represent performance on 30B containers, and the solid (no hatch marks) represent performance on the 48-type containers. See Table 1 for a description of the campaign details.

Combinations of distractors with relevant containers appear starting with campaigns 8 and 9.

Synthetic people. As seen in Figure 14, our object detection models frequently mis-labelled people as containers. In response, we introduced synthetic 3D people in campaigns 18 and 19.

Animated walk-through. During its development, we collaborated with partners who wanted to use the Limbo data for their own R&D. One project—the 3D Computer Vision for Safeguards project—is developing container counting capabilities intended for use by a safeguards inspector walking through a facility. In anticipation of their needs, Limbo campaign 20 provides an extensive animated walk-through of the synthetic environment that could be used for frame-by-frame tracking and counting of objects.

## 5. Discussion and future work

During our iterative image validation process, we made several general observations about training computer vision models with synthetic data, which we briefly summarize here along with thoughts on additional research.

First, negative examples are more effective when they include distractors. This observation came directly from our validation activities and is described in Section 4.2 and 4.3, as well as our discussion of updates to the data as an outcome of the validation process in Section 4.4.

Second, object configuration and positioning had a larger influence on detection rates than expected. This was also addressed in Section 4.2 and 4.3, and included an update in our synthetic data described in Section 4.4.

Third, training computer vision models to be more discriminating through the inclusion of distractor objects in training data can lead to a classic performance trade-off of improved precision, but lower recall.

Fourth, computer vision models are generally learning the wrong lessons from training data. Anecdotally, there are many synthetic images in the Limbo dataset that our human colleagues found difficult to distinguish from real-world data. The problem of domain shift between datasets has been well-documented in computer vision research, and (Movshovitz-Attias, Kanade, & Sheikh, 2016) describes its relevance to synthetic as well as real datasets. However, we note that even when human observers can tell which images are real and which are synthetic, they still have no difficulty correctly recognizing the (real or synthetic) cylinders. Yet computer vision models display significant differences in performance when evaluating real and synthetic images. This implies not only that there are differences between the synthetic and real feature distributions, but that the models are making decisions based on image features that humans somehow ignore as irrelevant.

We acknowledge that the premise of computer vision models learning the wrong features may be controversial. However, it is our aspiration that computer vision models respond more like human observers and recognize the intended item across many varied environments. We think the ability of computer vision models to learn the defining visual characteristics of relevant objects is especially crucial for high consequence domains such as nuclear nonproliferation, where learning irrelevant features could have serious security consequences.

At this point, we think attention must be placed back on feature engineering and the models themselves: what are the features they are learning, and can we force them to learn only the features we deem important? Can we identify and prune features that are irrelevant? We believe the next step in computer vision research and development—especially for high-consequence domains where real-world data is limited and synthetic data will likely play a significant role—will require new ideas and new architectures that allow model trainers to explicitly specify the relevance of data.

## 6. Accessing and using the data

The images, metadata, reference data, and documentation for the Limbo dataset are available to the public as unclassified, unlimited release data. While Sandia does not own the reference data, we have checked copyright information to the best of our ability and have included only data that we believe is shareable. The full Limbo dataset, including one million synthetic images, hundreds of real-world reference images, and all associated metadata is hosted in the Lawrence Berkely National Laboratory’s Berkeley Data Cloud (BDC). The data is open source and available to anyone with a free BDC account. Before accessing BDC, however, we recommend reading the documentation, terms of use, and API information detailed at:

<https://limbo-ml.readthedocs.io/>

The Limbo data is organized into a series of topical campaigns that provide a manageable file structure of roughly 50000 images each and reflect the lessons and observations from our data validation experiments (see Section 4). The rendering campaigns are described in Table 1, and in the documentation provided at our website.

Campaign No.	Campaign Description and Associated Figure.
2	30B containers viewed individually, in the relative center of the frame of real-world 3D HDR backgrounds. 50,000 of the images depict 30B containers, and 5,000 images show only the backgrounds without containers for use as negative examples. See Figure 1.
3	48-type containers (X, Y, and G designs intermixed) viewed individually, in the relative center of the frame of real-world 3D HDR backgrounds. 50,000 of the images depict 48 containers, and 5,000 images show only the backgrounds without containers for use as negative examples. See Figure 22.
4	48-type containers (X, Y, and G designs intermixed) arranged in rows in real-world 3D HDR backgrounds. 50,000 of the images depict 48 containers, and 5,000 images show only the backgrounds without containers for use as negative examples. See Figure 23.
5	30B containers arranged in rows framed in real-world 3D HDR backgrounds. 50,000 of the images depict 48 containers, and 5,000 images show only the backgrounds without containers for use as negative examples. See Figure 24.
6	No containers. This campaign contains images from our 3D HDR backgrounds as negative examples. See Figure 25.
7	Single synthetic distractor objects arranged in our real-world 3D HDR backgrounds. See Figure 26.
8	Single 30B containers pictured with a single distractor, in the real-world 3D HDR background. Depending on camera placement and container size, one of the containers might not be visible in some images. See Figure 27.
9	Single 48 containers pictured with a single distractor, in the real-world 3D HDR background. Depending on camera placement and container size, one of the objects might not be visible in some images. See Figure 28.
10	Clusters of distractor objects, including up to three distractor types, in real-world 3D HDR backgrounds. See Figure 29.
11	Single 30B container with up to three types of distractor objects clustered around the container, in real-world 3D HDR backgrounds. This campaign offers more views of occluded containers than previously demonstrated. See Figure 30.
12	Single 48 container with up to three distractor objects clustered around the container, in real-world 3D HDR backgrounds. This campaign offers more views of occluded containers than previously demonstrated. See Figure 31.
13	Highly complex environment with a single 48 container and many distractors of up to 10 different types filling the frame, in real-world 3D HDR backgrounds. These images are intended to test the limits of computer vision applications. See Figure 32.
14	Highly complex environment with a single 48 container and many distractors of up to 10 different types filling the frame, in real-world 3D HDR backgrounds. These images are intended to test the limits of computer vision applications. See Figure 33.
15	Each individual UF6 container type developed for this project, with every possible surface type, viewed from many angles. Backgrounds are real-world 3D-HDR backgrounds. See Figure 34.
16	Between 0 – 50 30B containers with multiple distractors placed in synthetic 3D oil refinery background. See Figure 35
17	Between 0 – 50 48 containers with multiple distractors placed in synthetic 3D oil refinery background. See Figure 36.
18	Single 30B containers with multiple distractors and with the addition of people placed in synthetic 3D oil refinery background. See Figure 37.
19	Single 48 containers with multiple distractors and with the addition of people placed in synthetic 3D oil refinery background. See Figure 38.
20	30B and 48 containers pictured together, with distractor objects, in an animated walkthrough of the synthetic oil refinery background. This campaign is intended for use in computer vision research involving video data. See Figure 39.

**Table 1:** Limbo campaign descriptions.



**Figure 22:** Example from campaign 3, with a single 48-type container.



**Figure 26:** Example from campaign 7, with a single distractor object.



**Figure 23:** Example from campaign 4, showing rows of 48-type containers.



**Figure 27:** Example from campaign 8, showing one distractor and one 30B container.



**Figure 24:** Example from campaign 5, with rows of 30B containers.



**Figure 28:** Example from campaign 9, with a single 48-type container and one distractor.



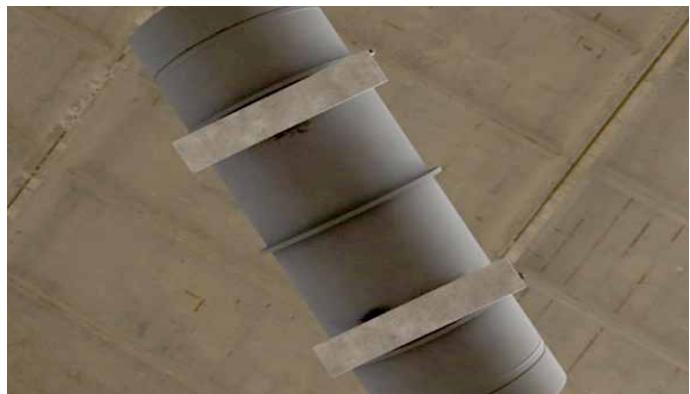
**Figure 25:** Example from campaign 6, showing a background image with no containers.



**Figure 29:** Example from campaign 10, with groups of distractors.



**Figure 30:** Example from campaign 11, showing groups of distractors with one 30B container.



**Figure 34:** Example from campaign 15, in which a 30B container is pictured from below.



**Figure 31:** Example from campaign 12, with groups of distractors with one 48-type container.



**Figure 35:** Example from campaign 16, with several 30B containers and distractors in a synthetic background.



**Figure 32:** Example from campaign 13, with many distractors and one 48-type container.



**Figure 36:** Example from campaign 17, with several 48-type containers and distractors in a synthetic background.



**Figure 33:** Example from campaign 14, with many distractors with one 30B container.



**Figure 37:** Example from campaign 18, with 30B containers, distractors, and people in a synthetic background.



**Figure 38:** Example from campaign 19, with 48-type containers, distractors, and people in a synthetic background.



**Figure 39:** Example from campaign 20, an animated walk-through of 30B and 48-type containers with distractors in a synthetic background.

## 7. Acknowledgments

This work was funded by the U.S. Department of Energy's National Nuclear Security Administration, Defense Nuclear Nonproliferation Research & Development program. Additional thanks go to Lawrence Berkeley National Laboratory for hosting the data developed by this project on the Berkeley Data Cloud, and to Lawrence Livermore National Laboratory for providing seed data for our real-world image collection. Thank you to Sydney Dorawa, Arielle Mattes, and Jason Bolles for their support in collecting and labeling real-world images.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security

Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan. Sandia Tracking Number: 1724688.

## 8. References

- [1] American Nuclear Standards Institute. (2001). American Nuclear Standard for Nuclear Materials - Uranium Hexafluoride - Packaging and Transport. Retrieved from <https://law.resource.org/pub/us/cfr/lbr/002/ansi.n14.1.2001.pdf>
- [2] Arno, M. (2018, October). Streamlining Open-Source Proliferation-Relevant Video Identification, Collection, and Processing. Presented at the Emerging Information Analysis Concepts for Nuclear Nonproliferation and Security Workshop. Ann Arbor, MI.
- [3] Carlini, N., & Wagner, N. (2017). Toward Evaluating the Robustness of Neural Networks. IEEE Symposium on Security and Privacy.
- [4] Ekbatani, H., Pujol, O., & Segui, S. (2017). Data Generation for Deep learning in Counting Pedestrians. ICPRAM, (pp. 318-323). doi: 10.5220/0006119203180323
- [5] Feldman, Y., Arno, M., Carrano, C., Ng, B., & Chen, B. (2018). Toward a Multi-Modal Deep Learning Retrieval System for Monitoring Nuclear Proliferation Activities. Journal of Nuclear Materials Management, XLVI(3), 68-80.
- [6] Friedman, J., & Jones, A. (2021, May 02). Cryptomatte. Retrieved December 05, 2023, from Github: <https://github.com/Psyop/Cryptomatte>
- [7] Gaidon, A., Wang, Q., Cabon, Y., & Vig, E. (2016). Virtual Worlds as Proxy for Multi-Object Tracking Analysis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 4340-4349).
- [8] Gastelum, Z. N., & Shead, T. M. (2018). Inferring the Operational Status of Nuclear Facilities with Convolutional Neural Networks to Support International Safeguards Verification. Journal of Nuclear Materials Management, XLVI(3), 37-47.

- [9] Gastelum, Z. N., & Shead, T. M. (2020). How Low Can You Go? Using Synthetic 3D Imagery to Drastically Reduce Real-World Training Data for Object Detection. Sandia National Laboratories. Retrieved from <https://www.osti.gov/servlets/purl/1670874>
- [10] Gastelum, Z. N., Shead, T. M., & Marshall, M. R. (2022). But it Looks so Real! Challenges in Training Models with Synthetic Data for International Safeguards. Proceedings of the Institute of Nuclear Materials Management Annual Meeting.
- [11] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. arXiv preprint. Retrieved from <https://arxiv.org/pdf/1412.6572.pdf>
- [12] He, K., Xiangyu, Z., Shaoqing, R., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 770-778). Retrieved from [https://openaccess.thecvf.com/content\\_cvpr\\_2016/papers/He\\_Deep\\_Residual\\_Learning\\_CVPR\\_2016\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf)
- [13] Jocher, G., Stoken, A., Chaurasia, A., Borovcik, J., NanoCode012, Kwon, Y., . . . Jain, M. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (v7.0). Zenodo. doi:<https://doi.org/10.5281/zenodo.7347926>
- [14] Joshi, T., Cooper, R., Okamura, A., von Suderth, A., Roberts, B., & Valentine, J. (2021). Multi-Sensor Fusion for Nuclear Material Container Counting and Assay: 3D Computer Vision for Nuclear Material Accountancy in Large Complex Environments. (internal) NSARD program review. Virtual.
- [15] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., . . . Dollar, P. (2015). Microsoft COCO: Common Objects in Context. arXiv. Retrieved from <https://arxiv.org/pdf/1405.0312v3.pdf>
- [16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single Shot Multi-box Detector. European Conference on Computer Vision (pp. 21-37). Springer.
- [17] Movshovitz-Attias, Y., Kanade, T., & Sheikh, Y. (2016). How Useful is Photo-Realistic Rendering for Visual Learning. Computer Vision-ECCV 2016 Workshops. Amsterdam: Springer International Publishing. Retrieved from [https://link.springer.com/chapter/10.1007/978-3-319-49409-8\\_18](https://link.springer.com/chapter/10.1007/978-3-319-49409-8_18)
- [18] Northcutt, C. G., Athalye, A., & Mueller, J. (2021). Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Retrieved from <https://arxiv.org/abs/2103.14749>
- [19] Oak Ridge Operations. (1987). Uranium Hexafluoride: Handling Procedures and Container Descriptions. Retrieved from <https://www.osti.gov/servlets/purl/6304596>
- [20] Papernot, N., McDaniel, P., Wu, X., Somesh, J., & Swami, A. (2016). Distillation as a defense to adversarial perturbations against deep neural networks. IEEE Symposium on security and privacy , (pp. 582-597).
- [21] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv: 1804.02767.
- [22] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Toward Real-Time Object Detection with Regional Proposal Networks. Advanced in Neural Information Processing Systems, 28.
- [23] Rutkowski, J., Canty, M. J., & Nielsen, A. A. (2018). Site Monitoring with Sentinel-1 Dual Polarization SAR Imagery Using Google Earth Engine. Journal of Nuclear Materials Management, XLVI(3), 48-59.
- [24] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Gran-cam: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision. .
- [25] Smartt, H. A., Gastelum, Z. N., Rutkowski, J. E., Peter-Stein, N., & Shoman, N. (2021). Hey Inspecta! Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting.
- [26] Smith, M. R., Hamel, M., Hannasch, D., Thomas, M., & Gaiten-Cardenas, C. (2021). A Deep Learning Workflow for Spatio-Temporal Anomaly Detection in NGSS Camera Data. Proceedings of the Institute of Nuclear Materials Management and European Safeguards Research & Development Association Joint Annual Meeting. Virtual.
- [27] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 1-9).
- [28] Tan, R. J. (2019, March 24). Breaking Down Mean Average Precision (mAP). Retrieved November 27, 2023, from Medium: <https://towardsdatascience.com/breaking-down-mean-average-precision-map-aef462f623a52>
- [29] Thomas, M., Passerini, S., Cui, Y., Rutkowski, J., Yoo, S., Lin, Y., . . . Moeslinger, M. (2021). Deep Learning Techniques to Increase Productivity of Safeguards

Surveillance Review. Proceedings of the Institute of Nuclear Materials Management and European Safeguards Research & Development Association Joint Annual Meeting. Virtual.

[30] Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., . . . Birchfield, S. (2018). Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 969-977).

[31] United States Enrichment Corporation. (1995). Uranium Hexafluoride: A Manual of Good Handling Practices. Retrieved from <https://www.osti.gov/servlets/purl/205924>

[32] Wolfart, E., Casado Coscolla, A., & Sequeira, V. (2022). Deep Learning for Video Surveillance Review. Proceedings of the Institute of Nuclear Materials Management Annual Meeting.

[33] Wolfart, E., Sanchez-Belenguer, C., & Sequeira, V. (2021). Deep Learning for Nuclear Safeguards. Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting, (pp. 1-10).

# Benefits and Drawbacks of Non-Destructive Assay (NDA) for Nuclear Safeguards - An Overview

Matteo Bolzonella

Politecnico di Milano, Department of Energy, Via Lambruschini 4, Milan, Italy, 20156

## Abstract:

*Nuclear safeguards are technical measures used to detect the diversion of nuclear material and technologies from their peaceful use. A key trait of safeguards is represented by measurements, which allow operators and inspectors to verify the nuclear activities and the amount of nuclear material declared by the States. In particular, an important class of measurements are Non-Destructive Assay (NDA) techniques, which allow one to characterize samples while preserving their physical integrity. NDA represents a rapid and effective method for measurements of nuclear materials in a variety of situations (from process lines to finished products, scrap and waste materials), with the possibility to perform on-site measurements, reducing radioactive waste production and offering wide margins for automatization and remote operation. However, there are situations where the use of NDA is not recommended or possible and one has to rely on Destructive Assay (DA) (i.e., techniques which alter the integrity of the item analysed). This happens, for example, in environmental sampling, when dealing with highly diluted nuclear materials, or when the radiation emitted by the sample is weakly penetrating. In conclusion, the integrated use of different NDA and DA techniques is the best possible solution for measurements in the frame of nuclear safeguards.*

**Keywords:** Nuclear Safeguards, Nuclear Instrumentation, Nuclear Measurements, NDA techniques.

## 1. Introduction

Nuclear materials and technologies offer a powerful tool for the progress of society and industry across the world. However, the same instruments may also be used in a harmful way, like for the development of nuclear weapons. To detect the diversion of nuclear material from plants and facilities, preventing the proliferation of nuclear weapons and the illicit trafficking of nuclear and radiological materials, a series of international technical measures have been developed in the form of legally binding agreements. These measures are known as "nuclear safeguards". Nuclear safeguards originated from the signing of the Non-Proliferation Treaty (NPT) in 1968, when non-nuclear weapon States officially declared that they would not develop nuclear weapons and agreed on establishing measures enabling the verifications of compliance to the NPT [1]. These verifications are performed by safeguards inspectorates: at an international level, the reference authority is represented by the International Atomic Energy Agency (IAEA) [2], while, for example, at the European Union level the reference authority is EURATOM [3], [4]. The European Commission collaborates actively with EURATOM in the frame of safeguards development via the DG ENER (EURATOM Safeguards) [5] and the Joint Research Centre (Nuclear Security Unit) [6], [7]. Another important contribution to European nuclear safeguards is given by the European Safeguards Research and Development Association (ESARDA), an association of European organizations established in 1969 to advance and harmonize R&D in the area of nuclear safeguards [8].

Initially, safeguards were only focused on the verification of activities and amounts of material declared by the States to the inspectors, but after the Iraqi crisis of 1991 [9], through the Additional Protocol (AP) (also known as INFCIRC/540) by IAEA [10], they became aimed at providing a more comprehensive picture of a State's nuclear activities.

Through safeguards, a State can demonstrate to others that its nuclear material and technologies are being used for peaceful purposes. The IAEA safeguards were established about 60 years ago and nowadays it is clear that nuclear cooperation and the exchange of scientific and technological expertise among countries have benefited from these measures. At present, more than 180 States worldwide have put into force IAEA safeguards [11]. The continuous enhancement of safeguards efficiency is crucial to

improve the cost-effectiveness of safeguards mechanisms, to encourage international peaceful nuclear collaboration based on the exchange of nuclear technologies and materials, and to adapt to evolving/emerging technologies (e.g., Gen IV reactors, Small Modular Reactors, nuclear fusion, ...). Furthermore, by enhancing safeguards efficiency, the IAEA can maintain its credibility and effectiveness in implementing safeguards on a global scale.

To detect the diversion of nuclear material from declared activities to undeclared activities, it is necessary to account for the declared material via inspections and measurements. In particular, measurements of mass and/or volume and sample analysis play a key role in the control of nuclear material since they allow the characterization of the material both qualitatively (i.e., identifying its nature/components (isotopic composition)) and quantitatively. Measurement techniques can be divided into two main groups: Non-Destructive Assay (NDA) techniques and Destructive Assay (DA) techniques.

Non-Destructive Assay techniques allows the investigation of a sample preserving its physical integrity, as suggested by the name itself. NDA techniques have matured over the years and gained in precision, and nowadays these techniques allow for efficient and accurate measurement of nuclear material, including special nuclear material (SNM) like plutonium and enriched uranium [12], as well as other nuclear materials such as neptunium ( $^{237}\text{Np}$ ) [13], which in turn allows authorities to draw elaborate safeguards for nuclear material accounting and control (NAMC), process control and perimeter monitoring [14].

In contrast to NDA techniques, DA techniques involve sampling from the material under analysis, affecting its physical integrity. Each sample taken from the original item must be representative of the bulk of the material from which it is taken. Moreover, these samples usually undergo a preparation prior to the measurement, which changes their physical/chemical form. For example, specific solutions or oxides may be required to carry out the measurement. Besides the analysis of samples of nuclear material for the verification of declared amounts of material, DA techniques are crucial for the detection of undeclared nuclear activities through environmental sample analysis. Furthermore, DA techniques are crucial to certify working standards used for the calibration of NDA techniques. The most common DA techniques for nuclear safeguards are based on precipitation/weighting of chemical analytes, analytical electro-/radio-chemistry and mass spectrometry.

Besides measurements, IAEA safeguards also involves containment and surveillance as key elements for drawing conclusions.

This paper describes the NDA techniques used for nuclear safeguards – focusing on the main techniques and providing some examples of advanced and novel methods – with

the aim of pointing out their benefits and drawbacks. Instead, a detailed discussion of DA techniques and of containment and surveillance, which is beyond the scope of the present work, can be found, for example, in [15].

## 2. Overview of NDA techniques

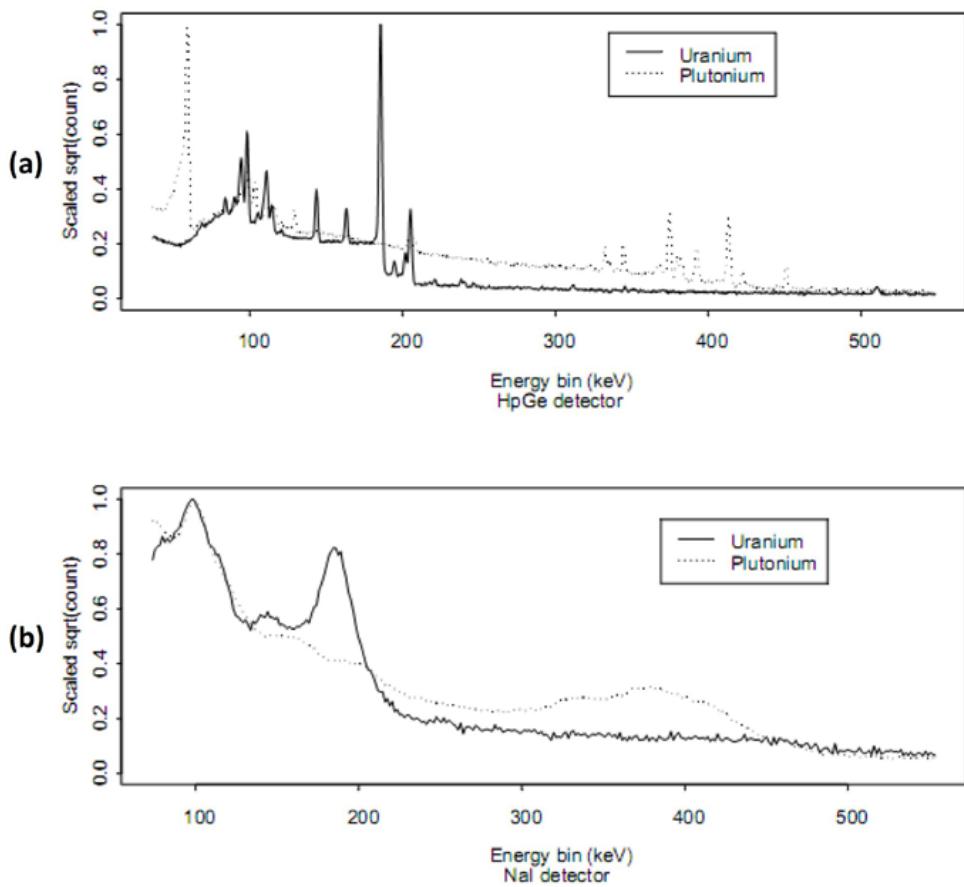
The main NDA techniques used for nuclear safeguards can be classified into three groups:

**a) Gamma-ray spectrometry:** this technique measures the gamma rays (and X-rays) emitted by radioactive nuclides in terms of energy and intensity. Because the photons are a “fingerprint” of the emitting nuclide, gamma-ray spectrometry allows a qualitative analysis of the sample (i.e., it allows one to recognize which nuclides it contains) besides the quantitative analysis linked to the abundance of isotopes in the sample. The reference tools for this NDA technique are gamma-spectrometers, which can include scintillators and semiconductor detectors. The detectors sense radiation based on the excitation (scintillators) and ionization (semiconductors) of the active material by the incoming  $\gamma$  rays [16]. In the context of nuclear safeguards, scintillators (like NaI(Tl) or LaBr<sub>3</sub>), having a good efficiency but a relatively poor energy resolution, are commonly limited to uranium enrichment studies. On the contrary, semiconductors like high-purity germanium (HPGe), offering excellent energy resolution but a limited efficiency, are routinely used to obtain and unfold complex spectra such as the one of Pu isotopes. Figure 1 shows a comparison of spectra of U and Pu acquired with different instruments.

Nevertheless, it should be pointed out that gamma-ray spectrometry is not capable of measuring the isotope  $^{242}\text{Pu}$  due to its very low specific gamma activity. Moreover, the drawback for the excellent resolution of HPGe is that the detector must be maintained at a low temperature (typically at 77 K), which makes the detector setup bulky and expensive. A good compromise between efficiency and resolution is represented by semiconductors like CdTe and CdZnTe, which can be operated at room temperature. In [16] a list of the state-of-the-art scintillators and semiconductor detectors is available, showing all their main features and discussing their performance.

In general, gamma-ray spectrometry is fundamental for the detection of divergences (“defect verification”) between the declared and measured values of spent fuel [18], [19]. As an example, semiconductors are routinely used for the detection of fission products in spent fuel.

Gamma-ray spectrometry is an NDA technique requiring sophisticated instrumentation and skilled operators: deep training [20] is crucial to master both the hardware and the software (including numerical codes) accompanying the spectrometers [21]. The relative error associated with gamma-spectrometry measurements varies with the isotope

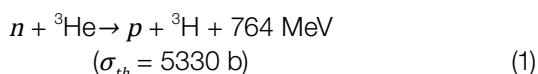


**Figure 1.** Examples of (a) high resolution (HPGe) and (b) low resolution (NaI) spectra for  $^{235}\text{U}$  and  $^{239}\text{Pu}$ . Image adapted from [17].

content (enrichment, fuel burnup, ...) and with the counting time: typical values are in the range of ~0.1%-10% [22]. Generally, high-resolution gamma-spectrometry can detect directly a mass of  $^{235}\text{U}/^{239}\text{Pu}$  greater than about 10 mg [23], [24].

It is worth mentioning that a novel technique based on gamma-ray detection such as partial defect verification: gamma-ray emission tomography [25].

**b) Neutron assay:** this technique is based on the detection of fission neutrons emitted by fissile and special nuclear materials either spontaneously or due to reactions induced by an investigation source. The detectors used for this assay are typically gas-filled detectors, mainly  $^3\text{He}$  proportional counters which can detect thermal neutrons via the reaction:



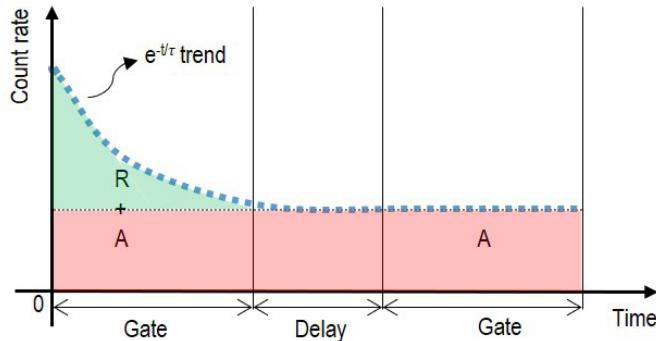
The charged reaction products in Equation (1) interact with the filling gas of the detector (via ionization and excitation events), generating an electrical signal (charge pulse), from which the presence of neutrons is deduced. Since fission events generally release 2-3 neutrons [26], neutrons produced by fission can be discriminated from random

neutrons occurring from background natural/cosmic radiation and from the ones of interrogation sources by using electronics working in coincidence mode. The rationale for discrimination is the following: if two pulses are detected within a short time window ("gate") they are considered as temporally correlated events and are counted as a real fission event ("double"), on the contrary, if only one event is detected it is considered a random individual neutron. This is the basis of the neutron coincidence counting method (NCC), which is the most widespread NDA technique for the determination of the Pu mass in bulk samples. More complex algorithms, such as neutron multiplicity counting methods (NMC) have been developed to account for single, double and triple events detected within the gate, in order to extract quantitative information about the neutron multiplication effect within the sample [27]. If real coincident multiples are present, the distribution in time of events (detector pulses/counts) that follow an arbitrarily chosen starting event (pulse) is given by the Rossi-alpha distribution:

$$S(t) = A + R \cdot e^{-\frac{t}{\tau}} \quad (2)$$

where  $A$  is the accidental coincidence count rate,  $R$  is the real coincidence count rate and  $\tau$  is the die-away time of the detector assembly [23]. The exponential form of the Rossi-alpha distribution comes from the fact that a population of coincident neutrons emitted at a time  $t = 0$  in a

fission reaction will decay in time as  $e^{-t/\tau}$ . Figure 2 shows an example of a Rossi-alpha distribution: a prompt gate starting at  $t = 0$  and ending when the distribution reaches a plateau collects real coincidences ("R", green area) plus accidental events ("A", red area), while a sufficiently delayed gate as wide as the first one collects only accidental events. Counting the events in the two gates and subtracting the events collected in the delayed gate from those collected in the prompt gate gives the real counts. This is done in practice using shift register electronics [15], [23].



**Figure 2.** Rossi-alpha distribution (dashed line) of detection events. Key: R = real coincidence counts (green area), A = accidental coincidences (red area), dotted line: accidental coincidence background,  $\tau$  = neutron die-away time.

The main advantage of passive NMC (i.e., NMC without an investigation source) is that it does not need calibration with a series of representative physical standards to determine the instrument response curve like NCC. Moreover, the additional precision of this technique makes it suitable for the analysis of heterogeneous Pu samples like scrap MOX material.

Since the Pu spontaneous fission yield (due to  $^{238}\text{Pu}$ ,  $^{240}\text{Pu}$  and  $^{242}\text{Pu}$  decays) is higher than that of U, Pu-containing samples can be investigated without the use of an external neutron source inducing fission, while for U measurements an investigation neutron source has generally to be employed. As  $^{240}\text{Pu}$  is usually the major even isotope present in both low- and high-burnup plutonium, the plutonium mass measured with neutron assay is typically expressed as equivalent mass of  $^{240}\text{Pu}$  ( $m^{240\text{eff}}$ ), namely the mass of  $^{240}\text{Pu}$  that would give the same coincidence response as that obtained from all the even isotopes in the actual sample [23]. For measurement of  $^{235}\text{U}$  enrichment in uranium samples, an AmLi ( $\alpha, n$ ) source is typically used to induce fission in the sample because of the low average energy (around 0.5 MeV) of its neutrons, so that fast fission (e.g., on  $^{238}\text{U}$ ) is negligible. Note that an investigation source is also needed to measure  $^{239}\text{Pu}$  mass [23].

Another option for performing neutron assay is by using liquid and plastic scintillators (or gaseous ones [28]), where fast neutrons are detected via the light nuclear recoils (mainly protons) they generate in the scintillator itself. As the

recoil energy is correlated to the initial neutron energy, this technique allows operators to perform a quantitative/spectrometric analysis of the fast neutron field. Moreover, if scintillators are equipped with thermal neutron absorber layers (containing B, Cd, Gd, Li, ...), they are able to detect (count) thermal neutrons too. The use of scintillators is more recent and less widespread than that of gaseous detectors [29].

Finally, note that fission chambers [16] are employed in cases where high neutron fluxes can overcharge the detectors, like in the measurements of spent nuclear fuel.

The most ubiquitous detectors for neutron assay based on  $^3\text{He}$  are the neutron coincidence counter (NCC) – with or without an interrogation source – [23], the active well coincident counter (AWCC) [15], the EURATOM fast collar [30] (routinely used by the European Commission's inspectors) and the underwater coincidence counter (UWCC) [31] (usually dedicated to Pu measurements in MOX fuel assemblies prior to irradiation). Instead, an example of detection system based on (liquid) scintillators is the fast neutron collar (FNC) [32], [33]. The instruments used for neutron counting contain a series of detectors which are typically arranged in compact geometries such as "wells" for relatively small samples or "collars" suitable for the analysis of bigger samples like PWR and BWR fuel pins/assemblies, etc. All these instruments offer compact and easy-to-use solutions for short-time on-site Pu and U measurements, relying on a large set of well-documented and internationally recognized procedures; however, they are bulky instruments often housing radioactive sources, so they are not usually customizable for measurement of "non-standard" samples. For instance, some counters are optimized only for PWR fuel assembly analysis and the presence of a radioactive source rises radiation protection issues. When dealing with massive/self-shielding samples, cadmium liners can be generally inserted in these instruments to improve the penetrability of the interrogating neutron flux in the sample: when cadmium liners are present, the instrument is said to operate in "fast mode" (e.g., the EURATOM fast collar operates in fast mode) [34].

Typically, the IAEA international target values (ITV) for the relative uncertainty of neutron assay techniques are of a few percent, but errors around 1% or less can be usually achieved by properly tuning the measurement time [22], [30], [35]. The minimum detectable mass of  $^{235}\text{U}$  via neutron counting techniques lies commonly in the range 1 mg - 100 mg, while for plutonium the detection limit ranges roughly from 1 mg to 500 mg in terms of  $m^{240\text{eff}}$ , depending on both the method used and the material composition of the item analysed [23], [36], [37], [38].

**c) Calorimetry:** this technique is based on measuring the thermal power produced by the radioactive decay of samples placed inside containers, using calorimeters based on thermocouples and/or thermopiles. The main advantages

of the technique are that the assay is independent of the sample geometry and that it applies to a wide range of material forms. Calorimetry is mainly used to assess the total mass of plutonium inside a sample, but it requires knowledge of the plutonium isotopic mass ratios (usually given by gamma-ray spectrometry) [39], [40]. The Pu measurement via calorimetry is justified by the relatively high thermal power emitted by Pu isotopes and decay products, in particular  $^{238}\text{Pu}$  and  $^{241}\text{Am}$  [39], [40]. Calorimetry is the most accurate NDA technique available for plutonium mass measurements, and it has become a standard in the USA [39]. Nevertheless, it shall be noted that calorimetry measurements typically require a long time for thermal equilibrium to be reached and, moreover, a quite stable room temperature should be kept not to jeopardize the measurement performance, which makes calorimetry not suitable for industrial/workplace environments. The relative error associated with (the sole) thermal power measurement is typically below 1% [41]. It shall be noted that, due to the typical dimensions of the instrumentation used for calorimetry measurements, this technique is often restricted to small samples (< 80 L in volume). However, there exist systems for the calorimetry of samples with a volume of up to about 400 L [42]. The detection limit in terms of Pu mass generally ranges from a few tens of mg to about 10 g of Pu,

depending on the thermal power resolution of the calorimeter [42], [43].

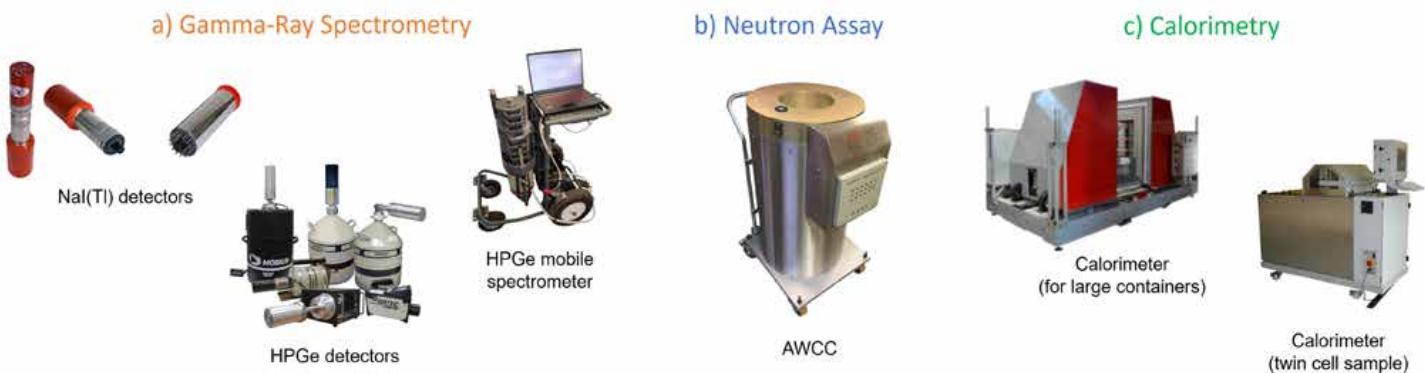
NDA techniques are commonly divided into two categories, active and passive techniques: in the former case, one measures the radiation spontaneously emitted by the sample, in the latter case the radiation measured is induced in the sample employing external sources. Table 1 summarises the main NDA techniques discussed in this Section, while Figure 3 shows some examples of instrumentation typically employed to perform these assays.

The techniques discussed so far represent the most widespread and consolidated NDA methods used for nuclear safeguards (note that the IAEA inspectors use more than 100 different NDA systems [15]). More advanced and/or novel NDA techniques have been developed to increase the effectiveness and efficiency of nuclear safeguards. A few meaningful examples of these techniques are briefly outlined in the following lines, while a more detailed overview of these methods can be found in [15] and [34].

K-Edge Densitometry (KED) consists in irradiating a sample with highly collimated X-rays and measuring the abrupt change of the transmitted X-ray intensity in correspondence of the so-called “absorption edge” of the atomic

NDA	Category	Type	Instrument	Main applications	Advantages	Disadvantages
Gamma-ray spectrometry	Passive	Qualitative + quantitative	Spectrometers (scintillators and semiconductors)	U enrichment determination (LEU and HEU), Pu isotopic composition verification, fresh MOX fuel verification, spent fuel verification.	Possibility to precisely detect/ identify nuclides, a wide range of software and codes for nuclide identification.	Complex instrumentation, need for deep operators' training, possible need to keep detectors at low temperatures, (sometimes) expensive instrumentation.
Neutron assay	Passive or active	Quantitative (and sometimes qualitative)	Gas-filled proportional counters, scintillators, fission chambers	Pu and U mass measurement at practically any enrichment and chemical/physical sample form.	Compact and easy-to-operate instrumentation, well-documented international procedures, high accuracy achievable, relatively short measurement time (< 1 h).	Bulky instrumentation, (possible) presence of a radioactive source (hence, dose to operators), (possible) need for reference field/ sources for calibration, self-shielding effects in the sample can perturb measurements, fixed setup (sometimes).
Calorimetry	Passive	Quantitative	Calorimeters	Total Pu mass measurement.	Applicable to a wide range of material forms, very accurate, simple technology, relatively cheap instrumentation, no calibration source required.	Long time needed to reach thermal equilibrium and/or good statistics, need for near-constant ambient temperature, (often) adapt only for small samples, not suitable for U measurements.

**Table 1.** Summary of the key features of the main NDA technique used for nuclear safeguards.



**Figure 3.** Examples of instrumentation used for the main types of NDA techniques (not to scale). Product Image courtesy of Mirion Technologies, Inc. (<https://www.mirion.com/>) for images a)-left. Image a)-middle courtesy of ORTEC® (<https://www.ortec-online.com/>). Images b) and c)-right courtesy of ANTECH Inc ([www.antech-inc.com](http://www.antech-inc.com)). Image c)-left courtesy of Setsafe (<https://setsafesolutions.com/>).

K-shell of a selected element to determine its concentration in the sample [44]. KED is generally applied to uranium or plutonium solutions with concentrations > 50 g/L, so that it is an ideal candidate for the analysis of reprocessing product solutions [22], [45]. The measurement uncertainty can be below 0.5% (depending on the counting time) [22].

Interferences can occur when elements having similar K absorption edges are present, so that corrections must be applied for samples containing both uranium and plutonium. Alternatively, when dealing with mixtures of uranium and plutonium, one can exploit Hybrid K-Edge Densitometry (HKED) [44], [45], which results from the combination of KED with X-ray fluorescence (XRF). HKED employs a single X-ray source for K-edge densitometry and fluorescence excitation. The concentration of the most abundant actinide is determined by KED, while the concentration of the other is derived from the XRF peaks. HKED is applicable to all U, Th, Pu, U-Th and U-Pu specimens containing at least 0.2 g/L of the main actinide, so that it allows the verification of reprocessing plants solutions having typical U concentrations of 150-250 g/L and Pu concentrations of 1-3 g/L [45]. An accuracy of about 0.20% for U and 0.60 % for Pu can be achieved in samples with a U/Pu ratio of about 100 [45]. Needing a fine X-ray spectrometry, both KED and HKED are normally performed with HPGe detectors [22]. Note that the accuracy of KED and HKED is better than that of high-resolution gamma spectrometry, leading to lower detection limits for the mass of nuclear materials. Another advantage is using an external, tunable investigating source instead of relying solely on the radiation emitted by the sample itself, which may be too weak to detect the source.

Moving to neutron measurements, an advanced NDA technique is the Neutron Resonance Transmission Analysis (NRTA) [46]. NRTA is based on the measurement of the intensity of neutrons that have traversed a sample as a function of the neutron energy. By so doing, it is possible to

correlate the reduction in the intensity of neutrons of a given energy to their absorption inside the sample due to the presence of specific isotopes, which are individuated and quantified. NRTA can determine the mass of all the isotopes of Am, Pu and U that are of interest for nuclear safeguards [46]. NRTA requires, upstream the sample to be investigated, a pulsed high-energy particle accelerator producing neutron bursts, and, downstream, a neutron detector (typically a <sup>6</sup>Li-glass scintillator coupled to a TOF (Time Of Flight) electronics for neutron spectrometry [47]). Each neutron burst comprises neutrons with a continuous energy distribution, but only neutrons with an energy in the 0.1-40 eV energy range are of interest for NRTA. NRTA has found a distinctive application in the analysis of spent fuel pin assemblies, which can be measured in parallel. Typical statistical uncertainties for NRTA are in the range 1%-4% [46]. In contrast to gamma spectroscopy, NRTA can be used to interrogate high-Z bulk material [47]. Moreover, differently from neutron counting techniques, which can detect only fissile and or fissionable nuclides, NRTA can characterize any isotope provided that it has strong and definite resonances in the energy region considered.

As a final example, a novel technique based on neutron measurement is the Differential Die-Away Self-Interrogation (DDSI) [15], [48], [49]. DDSI essentially consists in measuring fissile uranium and plutonium isotopes in spent fuel using spontaneous fission neutrons from <sup>244</sup>Cm that are present in the assembly as the interrogation source. The sensitivity of the fissile mass measurement is enhanced by measuring the sample with and without a cadmium liner between the sample and a surrounding moderator. In addition, the fertile mass can be assessed through the multiplicity analysis of the neutrons detected soon after the initial neutron pulse. Neutrons are detected using several <sup>3</sup>He detectors, which enhance the sensitivity of the measurement. Through DDSI, a statistical uncertainty of around 1% for an average spent fuel assembly (burnup around 40 GWd/tU

and 4%  $^{235}\text{U}$ ) can be achieved in few minutes [48]. DDSI is a powerful tool for detecting small quantities of fissile material, even if it is shielded or placed within a non-fissile matrix.

As described, NDA offers a powerful tool for supporting international safeguards; in the following section, the benefits and drawbacks of NDA for nuclear safeguards compared to other types of measurements and measures will be discussed.

As a final remark, it shall be highlighted that the border between NDA and DA becomes less clear when sampling is concerned: if an NDA technique is applied to a sample taken from an item, the same NDA technique can be considered non-destructive for the sample or destructive referring to the original item because its integrity has been compromised by taking the sample. For instance, this “destructive” sampling is needed when an item does not fit the detector dimensions and a portion of it is taken as sample and used for the selected NDA measurement. For example, HKED is classified as a non-destructive technique when it is assumed that it can be directly applied to the whole item under analysis (or when the sampling phase is neglected) [50], [51], while it is considered destructive when sampling is assumed to be required [45], and the same holds true for the other NDA techniques.

### 3. Discussion

In Section 2 the three main classes of NDA techniques were illustrated, discussing the specific advantages and disadvantages of each specific instrumentation type as they relate to nuclear safeguards. This section is dedicated to a discussion of the benefits and drawback of NDA in general.

Hereunder are listed the (main) benefits of NDA:

- 1) Preservation of the physical integrity of the material under analysis: non-destructive techniques offer the advantage of assessing nuclear material content without causing damage to the material itself. Hence, the physical (and chemical) state of the object investigated is not altered. This is crucial when dealing with valuable or sensitive materials, which can be used or further processed after the measurement.
- 2) Reduction of radioactive waste: NDA techniques do not generate of additional radioactive waste, for example with the preparation of a sample for analysis inside laboratories that, after the measurement becomes a waste. On the contrary, destructive techniques generally produce waste that needs to be managed and disposed of, involving radiation protection issues and, sometimes, safety/security issues.
- 3) Use at standoff distances: NDA methods are non-intrusive, namely they do not require physical contact with

the material being measured, mitigating the radiation hazard to personnel. This is particularly useful when dealing with items that are difficult to handle or manipulate directly. Moreover, this paves the way for automation and remote controlling, which further reduces the exposure to operators.

- 4) Possibility to carry out on-site measurements: a lot of NDA techniques can be adapted for field measurements, allowing inspectors to assess the nuclear material content on-site, thus reducing the need for transportation of materials to dedicated measurement facilities, which in turn limits issues with radiation protection and/or safety and security. This way, issues related to the delay of international transfers of nuclear material which may be necessary for independent destructive analysis are avoided. In some cases, such as for the verification of spent nuclear fuel, on-site NDA is the only possible option.
  - 5) Speed and cost: NDA techniques do not require sampling, which reduces the steps needed for a measurement. Not having to deal with the preparation and manipulation of a sample (generally in a dedicated laboratory) simplifies the measurement procedure and allows the user to get quickly to the result. This is particularly important in the context of nuclear safeguards, where it is crucial to timely verify the declared masses [13]. Furthermore, NDA instruments are typically cheaper than other laboratory instruments such as those used for DA [52].
  - 6) Availability of methods and procedures: a great deal of well-documented, internationally recognized analysis procedures exist for NDA; modern software packages guide the user to perform the measurement from calibration, to data acquisition and analysis/interpretation (for example, the *MAESTRO*<sup>®</sup> software for gamma-ray spectrometry by ORTEC<sup>®</sup> [53]).
  - 7) Possibility of cross-checks: as there are various NDA techniques available, one can use them for cross-validation on the very same sample (it is not modified by the non-destructive measurement), increasing the confidence in measurement results.
- Despite the numerous advantages just presented, there are also a few specific drawbacks connected to the use of NDA, which are given below:
- 1) Limited precision: NDA methods often provide less sensitive, accurate and precise measurements (i.e., larger uncertainties) compared to destructive techniques, which can impact the accuracy of material accountancy and verification [23].
  - 2) Difficulties with some sample geometry/setup: heterogeneous samples or large bulk samples can pose limits to

these techniques, for example, due to self-shielding techniques or due to a variable spatial distribution of burnable poisons. Similar problems arise with highly diluted nuclear materials. If NDA cannot guarantee to provide a measurement representative of the sample, other techniques have to be used; for instance, DA techniques are typically used for environmental sampling analysis for the detection of undeclared nuclear activities.

- 3) Sensitivity to environmental factors: NDA measurements can be performed on-site, but this makes them more dependent on environmental factors such as shielding, background radiation, and detector efficiency, which need to be carefully accounted for to achieve accurate measurements.
- 4) Need for specialized equipment and expertise: analytical methods used for nuclear safeguards, including NDA techniques, often require sophisticated instruments (e.g., neutron sources) which may not be readily available or easily operated by non-experts. Additionally, the interpretation of data can be complex and requires skilled analysts to ensure accurate results [54].
- 5) Limited material identification: some NDAs can quantify the amount of nuclear material (quantitative techniques), but are not able to identify the specific isotopes present in the material. For this purpose, destructive techniques like mass spectrometry or radiochemical analysis can be employed. However, there are NDA techniques capable of directly quantifying specific isotopes in an assay, such as the traditional gamma-ray spectrometry already discussed in the previous section or more exotic techniques like nuclear resonance fluorescence (NRF) [55].
- 6) Relatively high detection limits: NDA techniques have detection limits that are generally higher than those of DA, meaning that they might not be suitable for measuring very low quantities of nuclear material or detecting highly diluted material or low-energy radiation. For the aforementioned cases, a destructive analysis is needed (for example, dissolving the sample into a liquid scintillator after a proper chemical treatment). For example, the limit of detection of NDA for fissile nuclides is typically in the range of micrograms to some milligrams (per gram of sample material), while the most advanced DA techniques like ion mass spectrometry can detect down to tens of picograms of the nuclide of interest [14], [34], [43], [56], [57], [58], [59].
- 7) Tampering risk: because NDA measurements are non-intrusive, there is a risk of tampering or attempts to deceive inspectors by altering the exterior of the material without affecting the nuclear content.

Table 2 offers a visual recap of what has been illustrated.

Pros	Cons
<ul style="list-style-type: none"> <li>• Preservation of material integrity;</li> <li>• Reduction of radioactive waste and personnel's exposure;</li> <li>• Wide margins for automatization and remote control;</li> <li>• Possibility of on-site measurements;</li> <li>• Rapid and (relatively) cheap methods;</li> <li>• Wide international expertise available;</li> <li>• Possibility of cross-checks with different NDAs on the same sample.</li> </ul>	<ul style="list-style-type: none"> <li>• Higher detection limits and uncertainty compared to DA;</li> <li>• Not applicable to weak radioactive sources (e.g., highly diluted radionuclides);</li> <li>• Sensitivity to environmental factors;</li> <li>• Need for specialized equipment and trained operators;</li> <li>• Limited material identifications with certain NDA (e.g., calorimetry);</li> <li>• Tampering risk.</li> </ul>

**Table 2.** Summary of the main advantages and disadvantages of NDA.

#### 4. Conclusion

NDA techniques are crucial to assess and verify the nuclear material content in various items, such as nuclear fuel, waste, or other materials, without altering the sample integrity.

NDA offers several advantages, but it also has a series of drawbacks. Despite NDA being generally less precise than DA, the precision of NDA methods has improved significantly over the years owing to advancements in detector technology, calibration techniques, and data analysis methods. Nevertheless, while NDA methods may have inherent limitations in achieving the same level of precision as destructive assays, they offer the main advantage of preserving the physical integrity of the materials investigated and of allowing on-site field measurements, which is crucial in many practical scenarios. Overall, NDA techniques stand out as valuable tools in nuclear safeguards, but their successful implementation depends on proper instrumentation, skilled analysts, rigorous calibration procedures, and careful consideration of potential uncertainties and errors in the measurements.

In conclusion, a combination of NDA and DA techniques, besides the use of complementary safeguard strategies like containment and surveillance, nuclear forensics and satellite imagery, is the best solution for nuclear safeguards purposes, to maximize the benefits and overcome the limitations of each approach allowing for a more comprehensive assessment of nuclear material content and safeguards effectiveness [60]. The fact that an NDA technique can be applied to a sample taken “destructively” from the item to be analysed is just another example of how non-destructive and destructive operations can be combined for the analysis of nuclear materials.

## References

- [1] IAEA (1970). Treaty on the Non-Proliferation of Nuclear Weapons. Information Circular, INF/CIRC/140, IAEA.
- [2] IAEA (2016). IAEA Safeguards – Delivering Effective Nuclear Verification for World Peace. Leaflet 16-17011.
- [3] Howlett, D. A. (1990). EURATOM and Nuclear Safeguards. London, UK: Palgrave Macmillan.
- [4] European Commission (2012). Consolidated version of the Treaty establishing the European Atomic Energy Community. EU Treaty, EUR-Lex 2012/C 327/01.
- [5] European Commission Directorate-General Energy (DG ENER) official website, <https://commission.europa.eu/about-european-commission/departments-and-executive-agencies/energy>
- [6] European Commission – Joint Research Centre (EC JRC) official website, <https://joint-research-centre.ec.europa.eu>
- [7] European Commission, Directorate-General for Research and Innovation (2012). Euratom research and training programme 2021-2025. EU Publications Office.
- [8] European Commission, ESARDA official webpage, <https://esarda.jrc.ec.europa.eu>
- [9] Wilson, R. (1991). Nuclear Proliferation and the Case of Iraq. Journal of Palestine Studies, 20(3), pp. 5-15.
- [10] IAEA (1997). Model Protocol Additional to the Agreement(s) between State(s) and the International Atomic Energy Agency for the Application of Safeguards. Information Circular, INF/CIRC/540, IAEA (Vienna).
- [11] IAEA (2022). Safeguards. IAEA Bulletin Vol 63/3.
- [12] United States Nuclear Regulatory Commission (USNRC) (2023). USNRC webpage about Special Nuclear Material, <https://www.nrc.gov/materials/types/sp-nuc-materials.html>
- [13] M. Hua, Goddard, B., Lloyd, C., Leppink, E. Abraham, S. A., Noey, J., Clarke, S. D., Pozzi, S. A. (2019). Simulation of the Nondestructive Assay of  $^{237}\text{Np}$  Using Active Neutron Multiplicity Counting. Nuclear Science and Engineering, 194(2), pp. 154-162.
- [14] Gavron, A. (2001). Non destructive assay techniques applied to nuclear materials. ICONE 9: Ninth international conference on nuclear engineering, Nice Acropolis (France). France: Société Française d'Energie Nucléaire.
- [15] IAEA (2011). International Nuclear Verification Series No. 1 (Rev. 2) - Safeguards Techniques and Equipment: 2011 edition. Vienna, Austria: IAEA.
- [16] Knoll, G. F. (2012). Radiation Detection and Measurement - 4th edition, USA: John Wiley and Sons.
- [17] Burr T., Hamada, M. (2009). Radio-Isotope Identification Algorithms for NaI  $\gamma$  Spectra. Algorithms, 2, pp. 339-360.
- [18] Mayorov, M., White, T. A., Lebrun, A., Brutscher, J., Keubler, J., Birnbaum, A., Ivanov, V., Honkamaa, T., Peura, P., Dahlberg, J. (2017). Gamma Emission Tomography for the Inspection of Spent Nuclear Fuel. IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Atlanta, Georgia (US). US: IEEE.
- [19] Jawerth, N. (2016). Revealing Facts Through Science for Nuclear Verification. Article on IAEA official website, <https://www.iaea.org/newscenter/news/auditors-lab-coats-revealing-facts-through-science-nuclear-verification>
- [20] European Commission, EU Science Hub – Training programmes webpage, [https://joint-research-centre.ec.europa.eu/tools-and-laboratories/training-programmes\\_en](https://joint-research-centre.ec.europa.eu/tools-and-laboratories/training-programmes_en)
- [21] Vidmar, T., Aubineau-Laniec, I., Anagnostakis, M., Arnold, D., Brettner-Messler, R., Budjas, D., Capogni, M., Dias, M., De Geer, L., Fazio, A., Gasparro, J., Hult, M., Hurtado Bermudez, S., Jurado Vargas, M., Laubenstein, M., Lee, K., Lee, Y., Lepy, M., Maringer, F., Medina Peyres, V., Mille, M., Morales, M., Nour, S., Planteda, R., Rubio Montero, M., Sima, O., Tomei, C., Vidmar, G. (2008). An intercomparison of Monte Carlo codes used in gamma-ray spectrometry. Applied Radiation and Isotopes 66(6-7), pp. 764-768.
- [22] IAEA, (2022). International Target Values for Measurement Uncertainties in Safeguarding Nuclear Materials, STR-368 (Rev 1.1). Vienna, Austria: IAEA.
- [23] Reilly, D., Ensslin, N., Smith, H. Jr., Kreiner, S. (1991). Passive nondestructive assay of nuclear materials. US: U.S. Nuclear Regulatory Commission.
- [24] Choi, H.-D., Kim, J. (2018). Basic characterization of uranium by high-resolution gamma spectroscopy. Nuclear Engineering and Technology, 50(6), pp. 929-936.
- [25] Jansson, P. (2002). Studies of Nuclear Fuel utilizing Nuclear Spectroscopic Methods. Uppsala, Sweden: Acta Universitatis Upsaliensis.
- [26] IAEA (2008). Handbook of Nuclear Data for Safeguards: Database Extensions. Vienna, Austria: IAEA.

- [27] Ensslin, N., Harker, W. C., Krick, M. S., Langner, D. C., Pickrell, M. M., Stewart, J. E. (1998). Application Guide to Neutron Multiplicity Counting. Los Alamos National Laboratory, Report LA-13422-M.
- [28] Lewis, J. M., Kelley, R. P., Murer, D., Jordan, K. A. (2014). Fission signal detection using helium-4 gas fast neutron scintillation detectors: theory and experiment, *Appl. Phys. Lett.*, 105.
- [29] Darby, F., Hua, M., Pakari, O., Clarke, S., Pozzi, S (2023). Multiplicity counting using organic scintillators to distinguish neutron sources: An advanced teaching laboratory. *Am. J. Phys.* 91, pp. 936-945.
- [30] Clapham, M., Farese, N., Jaime, R., Lezzoli, L. Schillebeeckx, P. (1997). Performance comparison of different Active Neutron Interrogation Techniques for Safeguards Applications. Proceedings of the 19<sup>th</sup> ESARDA Symposium on Safeguards and Nuclear Material Management, Montpellier (France). Luxembourg: Publications Office of the EU.
- [31] Eccleston, G. W., Menlove, H. O., Abhold, M., Baker, M., Pecos, J. (1998). The underwater coincidence counter (UWCC) for plutonium measurements in mixed oxide fuels. Los Alamos National Laboratory, Report LA-UR-98-3303.
- [32] Swinhoe, M. T., Menlove, H.O., de Baere, P., Lodi, D., Schwalbach, P., Rael, C. D., Root, M., Tomanin, A., Favalli, A. (2021). A new generation of uranium coincidence fast neutron collars for assay of LWR fresh fuel assemblies. *Nucl. Instr. And Meth. A*, 1009.
- [33] Beaumont, J. S., Lee, T. H., Mayorov, M., Tintori, C., Rogo, F., Angelucci, B., Corbo, M. (2017). A fast-neutron coincidence collar using liquid scintillators for fresh fuel verification. *J. Rad. Nucl. Chem.*, pp.803-812.
- [34] Abbas, K., Peerani, P., et al. (2023). Nuclear Safeguards and Non-Proliferation – ESARDA Course Syllabus. Luxembourg: Publications Office of the EU.
- [35] Tomanin, A., Bourva, L., Beaumont, J., De Baere, P., Terrasi, A., De pooter, V., Arias Arenas, M. (2019). Inter-Comparison Exercise for the Safeguards Verification of PWR Fresh Fuel Assemblies using Fast and Thermal Neutron Coincidence. Proc. 41<sup>st</sup> ESARDA 2019 conference, Stresa (Italy). Luxembourg: Publications Office of the EU.
- [36] Lousteau, A. L. (2017). Determining 235U Enrichment Using a Dual-Energy Approach for Delayed Neutron Measurements. PhD Diss., University of Tennessee (USA).
- [37] Parker, J. L., Beddingfield, D. H., Menlove, H.O. (1999). Low-Activity Solid Waste Measurements at Tokai Works. Los Alamos, New Mexico (US): Los Alamos National Laboratory.
- [38] De Felice, P., Bogucarska, T., Raiola, F., Pedersen, B. (2021). Good Practice Guide for Validation of a Waste Characterisation System for Very Low, Low and Intermediate Level Radioactive Waste. Luxembourg: Publications Office of the EU.
- [39] ASTM International (2009). ASTM-Standard C 1458-09: Standard Test Method for Nondestructive Assay of Plutonium, Tritium and <sup>241</sup>Am by Calorimetric Assay. West Conshohocken, Pennsylvania: ASTM International.
- [40] Bracken, D. S., Biddle, R. S., Carrillo, L. A., Hypes, P. A., Rudy, C. R., Schneider, C. M., Smith, M. K. (2002). Application Guide to Safeguards Calorimetry. Los Alamos National Laboratory, Report LA-13867-M.
- [41] IAEA, (2010). International Target Values 2010 for Measurement Uncertainties in Safeguarding Nuclear Materials, STR-368. Vienna, Austria: IAEA.
- [42] Setsafe – KEP Technologies. The Heat Check Line - Nuclear measurement. Data Sheet, 01/21, <https://setsafesolutions.com/app/uploads/sites/4/2021/03/HEAT-CHECK-LINE-ENG.pdf>
- [43] Tagziria, H., Bagi, J., Pedersen, B., Schillebeeckx, P. (2012). Absolute determination of small samples of Pu and Am by calorimetry. *Nucl. Instr. And Meth. A*, 691, pp. 90-96.
- [44] Ottmar, H., Eberle, H., Matussek, P., Michel-Piper, I. (1986). How to Simplify the Analytics for Input-Output Accountability Measurements in a Reprocessing Plant. Kernforschungszentrum Karlsruhe, Report 4012.
- [45] Balsley, S.D. (2010). Destructive nuclear material analysis for safeguards: Importance and future trends. IAEA symposium on international safeguards, Vienna, 2010. Vienna, Austria: IAEA.
- [46] Sterbentz, J. W., Chichester, D. L. (2010). Neutron Resonance Transmission Analysis (NRTA): A Nondestructive Assay Technique for the Next Generation Safeguards Initiative's Plutonium Assay Challenge. Idaho National Laboratory, Report INL/EXT-10-20620.
- [47] Klein, E. A., Naqvi, F., Bickus, J. E., Lee, H. Y., Dangoulian, A., Goldston, R. J. (2021). Neutron-Resonance Transmission Analysis with a Compact Deuterium-Tritium Neutron Generator. *Physical Review Applied*, 15.

- [48] Schear, M. A., Menlove, H. O., Tobin, S. J., Evans, L. G., Croft, S. (2011). Development of the Differential Die-Away Self-Interrogations Technique for Spent Fuel Characterization. Los Alamos National Laboratory, Report LA-UR-11-04130.
- [49] Trahan, A. C., McMath, G.E., Mendoza, P. M., Trellue, H. R., Backstrom, U., Pöder Balkeståhl, L., Grape, S., Henzl, V., Leyba, D., Root, M. A., Sjoland, A. (2020). Results of the Swedish spent fuel measurement field trials with the Differential Die-Away Self-Interrogation Instrument. *Nucl. Instr. And Meth. A*, 955.
- [50] Mirion Technologies, Inc. (2011). HKED™, Hybrid K-Edge Densitometry Software. Data Sheet C38779 – 04/2011.
- [51] Shizuma, T., Hajima, R., Hayakawa, T., Fujiwara, M., Sonoda, T., Seya, M. (2011). Proposal for an advanced hybrid K-edge/XRF densitometry (HKED) using a monochromatic photon beam from laser Compton scattering. *Nucl. Instr. And Meth. A*, 654(1), pp. 597-603.
- [52] Tsoulfanidis, N. (Editor) (2012). Nuclear Energy - Selected Entries from the Encyclopedia of Sustainability Science and Technology. New York (US): Springer.
- [53] ORTEC® (2012). MAESTRO® - Software User's Manual, Software Version 7.0, 2012.
- [54] Boulyga, S., Konegger-Kappel, S., Richter, S., Sanguely, L. (2015). Mass spectrometric analysis for nuclear safeguards. *Journal of Analytical Atomic Spectrometry*, 30(7), pp. 1469-1489.
- [55] Ludewigt, B. A., Quiter, B. J., Ambers, S. D. (2011). Nuclear resonance fluorescence for safeguards applications. Lawrence Berkeley National Laboratory, Report LBNL-4350E.
- [56] Forbes, T. P., Szakal, C. (2018). Considerations for uranium isotope ratio analysis by atmospheric pressure ionization mass spectrometry. *Analyst*, 144(1), pp. 317-323.
- [57] Ramebäck, H., Vesterlund, A., Hedberg, M., Jonsson, S., Lagerkvist, P., Vidmar, T. (2023). Gamma spectrometric measurement of uranium isotopic composition and mass in sintered UO<sub>2</sub> pellets using the efficiency transfer method. *Applied Radiation and Isotopes*, 192.
- [58] Andrews, M.T., Goorley, J. T., Corcoran, E. C., Kelly, D. G. (2015). MCNP6 simulations of gamma line emissions from fission products and their comparisons to plutonium and uranium measurements. *Progress in Nuclear Energy*, 79, pp. 87-95.
- [59] European Commission - Joint Research Centre, European Safeguards Research and Development Association (2012). ESARDA Bulletin no. 48. Belgium: Publications Office of the EU.
- [60] Welsh, T. L., McRae, L. P., Delegard, C. H., Liebetrau, A. M., Johnson, W. C., Krick, M. S., Stewart, J. E., Theis, W., Lemaire, R. J., Xiao, J. (1995). Comparison of NDA and DA Measurement Techniques for Excess Pu Powders at the Hanford Site: Operator and IAEA Experience. Proceedings of the 36<sup>th</sup> annual meeting of the Institute for Nuclear Materials Management (INMM), Palm Desert, CA (US). US: INMM.





