◢ESARDA

European Safeguards Research & Development Association

ESARDA BULLETIN

The International Journal of Nuclear Safeguards
and Non-Proliferation

ESARDA

European Safeguards Research & Development Association

ESARDA BULLETIN

The International Journal of Nuclear Safeguards and Non-Proliferation

Contents: Volume 64, n.1

Editorial
Elena Stringa

Dear ESARDA Bulletin readers,

It is with pleasure that I present volume 64 of the 'ESARDA Bulletin - The International Journal of Nuclear Safeguards and Non-proliferation'.

This volume features very interesting articles on the following topics:
- Techniques and standards for non-destructive analysis of nuclear material
- Statistical methodologies for Nuclear Safeguards and Non-Proliferation
- Containment and surveillance
- Data analytics for Nuclear Safeguards and Non-Proliferation

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.

With this publication, I have two news to report. The first one being that from now on the Bulletin will only be available online. The printed copies will no longer be distributed and only a few samples will be printed for libraries requesting them. The decision has been taken in order for the publication to be more sustainable for the environment, in line with the European Green Deal.

The second news that I want to share with you is that we have finally obtained the possibility to assign a DOI for each single published article. This means that the ESARDA Bulletin - The international Journal of Nuclear Safeguards and Non-Proliferation – is now a rolling publication. This milestone is very important because it allows us to publish the articles as soon as they are available, without having to wait for June and December for their publication. The June and December volumes will continue to be issued regularly, by collecting the papers published online in the previous semester. We are now registering the articles already published and making them available online on the Bulletin website (https://esarda.jrc.ec.europa.eu/publications-0/esarda-bulletin_en) in order to include in Scopus also past articles that have yet to be indexed. The single articles also include the DOI number in the header of the first page, and the reference guideline in the footer to provide readers a quick and correct way to cite and reference the article.

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 concluding, I would like to thank 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 advices. 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, and to Lenka Hubert for all the support provided.

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

Studies of the impact of beta contributions on Cherenkov light emission by spent nuclear fuel

Erik Branger, Zsolt Elter, Sophie Grape, Markus Preston

Division of Applied Nuclear Physics,
Department of Physics and Astronomy,
Uppsala University, Uppsala, Sweden
E-mail: erik.branger@physics.uu.se

Abstract

The Digital Cherenkov Viewing Device (DCVD) is one of the instruments used by safeguards inspectors to verify spent nuclear fuel in wet storage. The DCVD can be used for partial defect verification, where the inspectors verify that 50% or more of an assembly has not been diverted. The methodology is based on comparing the measured Cherenkov light intensity with a predicted intensity, calculated with operator information.

Recently, IAEA inspectors have encountered fuel assemblies for which systematic deviations between predictions and measurements could be observed, indicating that the prediction model did not take into account all sources of Cherenkov light production. One contribution to the Cherenkov light intensity that is frequently omitted is the contribution from beta decays, where energetic electrons exit the fuel material and enter the water with sufficient energy to directly produce Cherenkov light. The objective with this work was hence to study beta contributions and evaluate whether that could be the cause of discrepancy between predictions and experimental data.

By simulating the beta contribution for fuel assemblies where the discrepancy was experimentally observed, it was determined that beta decays were the cause. The fuel assemblies had fuel rods with relatively small radii, thin cladding, a short cooling time and an irradiation history that resulted in a relatively large beta contribution for assemblies that had a comparatively low burnup. Therefore, the beta contribution was significant, and caused 10-40% of the total Cherenkov light intensity. By including the beta contributions in the predictions, the RMSE of the deviation between prediction and measurement could be reduced from 20.7% to 11.6% for the available measurement data. The results highlight that the beta contribution can be significant and should be taken into account for accurate predictions.

Keywords: Nuclear safeguards, Cherenkov light, DCVD, beta decay, spent fuel verification

1. Introduction

Following international safeguards agreements, international inspectors from the International Atomic Energy Agency (IAEA) or EURATOM are tasked with verifying that nuclear material is not diverted from peaceful use. One form of nuclear material that must be verified is spent nuclear fuel. To help the inspectors to independently verify nuclear fuel assemblies, a multitude of instruments have been developed [1]. One of the deployed instruments is the Digital Cherenkov Viewing Device (DCVD), which measures the Cherenkov light emission by spent nuclear fuel assemblies in wet storage [2]. The DCVD is capable of performing gross defect verification, where the presence and qualitative characteristics of the Cherenkov light emissions are used to verify that an object is a spent nuclear fuel assembly, and not a non-radioactive dummy object. The DCVD is also used to perform partial defect detection, verifying that part of a fuel assembly has not been diverted. Two methodologies are in use for partial defect detection: one that uses image analysis to determine if rods in visible positions have been removed, and one that quantitatively measures the Cherenkov light intensity to verify that it is consistent with the expected intensity, based on operator declarations of the fuel [3].

1.1 Verification of spent nuclear fuel with the DCVD

For the quantitative Cherenkov light intensity verification, the Cherenkov light intensity is predicted based on operator-declared values of burnup (BU), initial enrichment (IE) and cooling time (CT), or the so-called BIC parameters. In general, these three parameters are sufficient to characterize the fuel assembly, though the irradiation history also has some influence in the abundance of fission products [4]. Once the inspector has completed the predictions of the emitted Cherenkov light intensity, the spent fuel assemblies are measured using the DCVD, which is typically mounted on the railing of a fuel-handling machine above the fuel pond.

In the analysis, measurement data is grouped according to fuel type and measurement campaign. Thus, the measurements of the fuel assemblies in each group can be directly compared with each other, since they have the same design and the measurements were taken under the same conditions. For each group, a multiplicative constant is
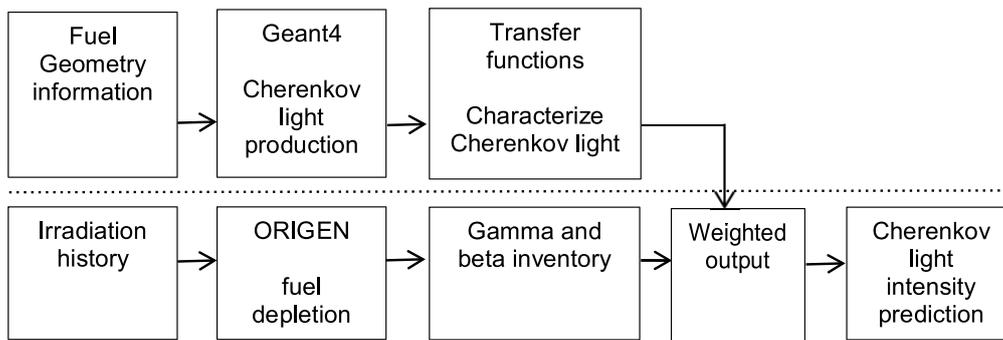
Figure 1: Schematic of the Cherenkov intensity prediction method. The top row contains the Monte-Carlo simulations to parameterize the Cherenkov light productions as a function of the radiation type and energy. The bottom row contains the calculations performed to obtain a prediction.

found to relate the predictions to the measurements using a least-square fitting. This multiplicative constant will compensate for effects that are identical for all assemblies in that group. Such effects include scattering and absorption of Cherenkov light in fuel assembly structures and the surrounding water, the optical efficiency of the detector system, the conversion of light to a measurable charge in the DCVD Charge-Coupled Device (CCD) chip, and the conversion of this charge to an image pixel value. Hence, after this calibration, the measured and predicted Cherenkov light intensities can directly be compared.

1.2 Motivation for this work

Recently, IAEA inspectors have encountered fuel assemblies where a systematic difference between predictions and measurements could be seen. The differences have been observed mainly for fuel assemblies with a CT of 2-5 years, and the magnitude of the difference was found to depend on burnup and fuel design. These systematic deviations indicate that the predictions do not accurately model the Cherenkov light production. The findings motivated this work, where the aim is to investigate whether the discrepancies are caused by the omission of direct beta-decay, and to assess/determine whether the predictions can be improved by including this contribution.

2. Cherenkov light intensity predictions

To predict the Cherenkov light intensity of an assembly, ORIGEN [5] is used to simulate the fuel depletion, either using the operator-provided irradiation history, if available, or using a default irradiation scheme otherwise. ORIGEN will then calculate the gamma-ray emission spectrum of the assembly, which is combined with a transfer function to obtain an estimate of the total Cherenkov light production of the assembly, and the abundance of beta-decaying isotopes. The methodology to predict the Cherenkov light intensity is summarized in Figure 1.

In the case of spent fuel stored in water, Cherenkov light is predominantly caused by gamma decays of fission

products in the fuel [6]. Thus, to predict the Cherenkov light intensity, a model of both the gamma radiation emissions and the subsequent Cherenkov light production is needed. The Cherenkov light prediction model used in the latest version of the DCVD software is based on [7], and this method has been implemented and extended in [8]. As will be detailed later, the prediction model in the DCVD software does not take into account beta decays where the electron passes through the fuel and cladding, and enters the water with sufficient energy to directly produce Cherenkov light. This contribution will be referred to as the "direct beta contribution" in this publication, as opposed to the "indirect beta contribution" which is used to describe beta particles that produce bremsstrahlung in the fuel, which in turn produce high-speed electrons in the water through Compton scattering or photoelectric absorption.

In order to estimate the Cherenkov light intensity that is produced per gamma-quanta of a certain energy, Monte-Carlo simulations that consider the relevant fuel geometry are made. These results are used to create a transfer function that relates the gamma emission energy to Cherenkov light production, as described in [9]. This transfer function is then applied to the gamma emission spectrum of the fuel assembly (as calculated by ORIGEN), to predict the Cherenkov light production by gamma emissions.

In principle, beta-decays can be handled in the same way, but the beta emission spectrum is not calculated by ORIGEN and must be obtained in some other way. Reference [9] suggests simulating beta decays from selected isotopes, to assess their respective direct beta contributions. The results can be used to calculate another transfer function that relates either the isotope activity or isotope mass to a Cherenkov light production, which can then be added to the Cherenkov light prediction due to gamma rays. Note that ORIGEN can calculate the bremsstrahlung emissions due to beta decays being stopped in the fuel, which are included in the gamma emission spectrum. In the DCVD Cherenkov light predictions, the bremsstrahlung contribution is treated as gamma emissions.

Earlier studies showed that the direct beta contribution could approach 5% of the total intensity, in the case of fuel assemblies with long CT, thin rods and thin cladding [9], but that it would typically be closer to 1-2%. However, the experimental data available used to validate those predictions covered fuel assemblies with a CT of at least 5 years, and thus fuel assemblies with shorter CT were not evaluated. Furthermore, the fuel assemblies used to validate the model primarily consisted of fully burnt assemblies; hence, the performance of the prediction methods could not be thoroughly validated at lower BU values. For the fuel assemblies with a CT of at least five years used to validate the model, decays by Sr90/Y90 are the only significant source of beta decays. Due to its modest contribution, the direct beta component was not included in the prediction model available in the DCVD software. For many of the fuel assemblies where the IAEA inspectors have observed a discrepancy between predictions and measurements, the CTs are however shorter than for the measurements used to validate the model. Hence, additional beta-decaying isotopes may still be present, which could potentially contribute more significantly compared to Sr90/Y90.

2.1 Information about the experimental data

To investigate the cause of the systematic difference between predictions and measurements, a set of roughly 300 assemblies with operator provided irradiation history have been measured. The set of fuel assemblies were selected to have a short CT and varying BU, since such fuels have shown the greatest discrepancy between prediction and measurement. Based on the fuel type and irradiation history of the assemblies, a few general remarks can be made:

— The fuel assemblies had a BU in the range of 20-60 MWd/kgU, and CT of 2-5 years. The shortest cooled fuels had the largest range of BU values, while the longer-cooled fuels had typically reached their discharge BU.

— Most of the measured assemblies had smaller rod radii and thinner cladding compared to the fuel assembly measurements used to verify the prediction model [9].

— In general, the fuel assemblies experienced the highest power level during their first few cycles, followed by a varying number of low power cycles, and finally one or more medium power cycles before reaching the discharge BU.

These fuel assemblies thus have fuel parameters such as BU and CT that differ notably from the experimental data used to validate the prediction model [9]. The smaller fuel rod radii and thinner claddings mean that the direct beta contribution could be larger than in the previously studied cases, and the shorter cooling times means that additional

beta-decaying isotopes beyond Sr90/Y90 may need to be taken into account.

The irradiation history of the assemblies in this data set also differed from previously considered irradiation histories where even, high power cycles were assumed for all but the last fuel cycle, which was assumed to be a low-power cycle. The differing power history will affect the abundance of short-lived isotopes present at the short cooling times of 2-5 years in this data set.

3. Methodology

3.1 Depletion calculations

The experimental data consist of DCVD measurements of roughly 300 assemblies of the same type. The assemblies have a wide range of BU and CT, and were selected since the systematic deviations in the predictions were pronounced for this group of assemblies. For each fuel assembly in the data, ORIGEN was run to determine the gamma emission spectrum and beta-decaying isotope contents. The depletion calculations accounted for the the operator provided irradiation history, simulating the correct per-cycle average burnup and the length of all cooling times. The ORIGEN fuel libraries "ge10x10-8" and "atrium10-9" have a similar rod configuration to the measured assemblies, and the Atrium library was chosen since it matched the number of short rods. The ORIGEN gamma spectrum also includes bremsstrahlung.

3.2 Simulating Cherenkov light production

To model the Cherenkov light production in a fuel assembly geometry, a Geant4 [10] based simulation toolkit has previously been created [7]. The results of these simulations were used to set up the transfer functions, relating gamma and beta particle emissions to Cherenkov light production. This code simulates the emissions of gamma and beta particles in the fuel material, their interactions in the fuel, cladding and water, and the production of Cherenkov light in the assembly. This code has been used to study the Cherenkov light production by both gamma and beta decays [6], including both direct beta contribution as well as bremsstrahlung, to identify their respective contributions.

This code package was used to simulate the Cherenkov light production for the fuel assemblies in the experimental data set. The modelled fuel assembly geometry is summarized in Table 1 and was selected to be representative of several different modern BWR fuels in [11]. Although the exact fuel dimensions were not revealed by the operator, the modelled dimensions match the Atrium fuel simulated by ORIGEN. The selected dimensions are also representative of modern BWR designs, matching the experimental data. However, the lack of detailed knowledge does introduce uncertainties in the simulations. As an example, for

Fuel density [g/cm3]	Fuel radius [mm]	Cladding inner radius [mm]	Cladding thickness [mm]	Pitch [mm]	Rod configuration
10.5	4.34	4.42	0.61	13.4	10x10

Table 1: Fuel dimensions used in the simulations.

the modelled cladding thickness, if the thickness were to be changed by 0.1 mm, the direct Cherenkov light production by Y90 may change by a factor of two [6], since the beta electrons are strongly attenuated by the cladding material. Since exact fuel dimensions is typically not available to an inspector, the prediction model should be general enough to be applicable even in the lack of such information, although uncertainties will be introduced by such assumptions when they do not match the assemblies to be measured.

Monoenergetic gamma emissions from the fuel material were simulated for various energies in the range 250 keV to 4 MeV, to determine the Cherenkov light production in the water as a function gamma initial energy. To simplify the simulations, fresh fuel material was assumed, since the difference in gamma attenuation between fresh and spent fuel is less than 3% for the photon energies that can produce Cherenkov light and the BU encountered in this work [12]. The initial gamma particles were distributed uniformly in the radial direction of each rod. Cherenkov photons forming an angle less than 3 degrees to the vertical axis were tallied in the simulations, since [13] notes that the vertical and total Cherenkov light components may behave slightly differently. Due to the measurement setup, the measured intensity will more closely follows the vertical intensity.

To determine the beta contribution, a separate set of simulations were run for each isotope identified to be of relevance for direct Cherenkov light contribution, taking into account the beta energy spectrum of the decay. The beta decay spectra were taken from [14]. Bremsstrahlung was disabled in the simulations, to account for only the direct beta contribution, and because bremsstrahlung is treated as a gamma emission in the prediction model. Uniformly distributed starting locations were used also for the beta simulations, though as discussed in section 3.4, the real distribution is more complicated but is unlikely to be known to an inspector performing a measurement.

3.3 Intensity predictions

Once the ORIGEN depletion calculations were done and the transfer functions were set up, the clip software package [8] was used to extract the gamma emission spectrum and the abundance of the beta-decaying isotopes for each simulated assembly. These were combined with the simulated Cherenkov light production as a function of gamma-ray energy, and the Cherenkov light production per decay

for the beta decaying isotopes, to assess the direct beta Cherenkov light production. The direct beta production was then added to the gamma production, to obtain a total prediction that properly includes both components.

Since the direct beta contribution was handled separately, this allowed for estimating the magnitude of the direct beta contribution relative to the gamma contribution. It also enabled comparisons between measurements and predictions, where predictions either included or excluded the direct beta contribution.

3.4 Limitations of the source distribution assumption

The radiation source distribution in a rod is more complex than a uniform distribution, which does have an impact on the Cherenkov light production, as noted in [6]. Some elements such as Caesium migrate due to heat gradients, and fission product concentrations are higher on the pellet rim due to the high-burnup structure. Especially for beta decays, where only decays on the pellet rim can produce Cherenkov light, effects such as the high-burnup structure at the rim can noticeably enhance the beta contribution [6]. The situation is further complicated by the fact that the fissile material in the pellet rim may be depleted more quickly than the bulk fuel rod material to obtain the high-burnup structure at low rod average burnup. However, if the rim is depleted early in the fuel lifetime, this may relatively suppress the beta contribution at higher rod burnups, when the rim has been depleted for some time. In addition, cracking of the fuel pellets may result in that regions some
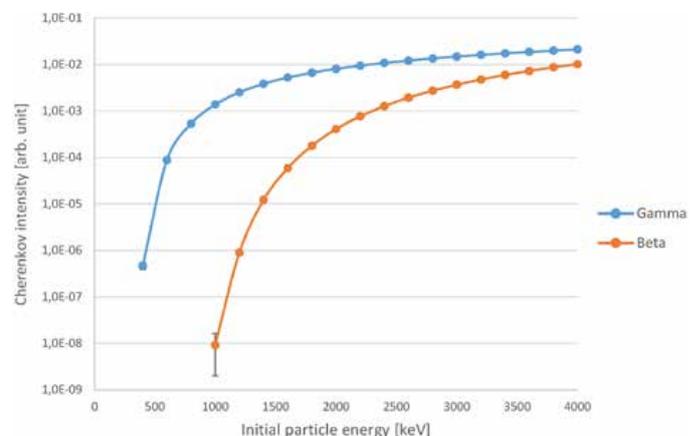
Figure 2: The average Cherenkov light production per gamma and beta particle emission. For the beta emissions, the contribution contains only the Cherenkov light produced directly by the beta decay and neglects any bremsstrahlung, which in turn can result in Cherenkov light production. The error bars refer to the Monte-Carlo statistical uncertainties in the simulations.

millimetres into the pellet have a free path to the cladding, which could allow beta decays from more interior locations of the pellet to contribute more to the Cherenkov light production. The development of a detailed source distribution model that is generally applicable based on the limited information available to an inspector is outside the scope of this work, hence the simplifying assumption of fresh fuel and uniform source distribution is made.

4. Results

4.1 Cherenkov light production by gamma and beta decays

The Cherenkov light production as a function of initial particle energy is shown in Figure 2, for both initial gamma and beta particles. Similar to the results of [6], beta particles require higher energy than gamma particles to produce comparable amounts of Cherenkov light. A kinetic energy of around 250 keV is required for an electron to produce Cherenkov light in water, though as seen in Figure 2, an initial kinetic energy of around 1 MeV is required for the electron to be able to penetrate the fuel and cladding to directly produce Cherenkov light.

4.2 Beta decaying isotopes of interest

Based on the results in Figure 2 and previous considerations, beta-decaying isotopes fulfilling the following characteristics may contribute to the Cherenkov light production at the cooling times seen in the experimental data:

— The fission yield should be high enough that the fission product isotope is abundant in spent nuclear fuel. Isotopes with a cumulative fission yield above 0.1% were included in this work.

— The half-life of the isotope should be short-lived to be active enough to contribute, but sufficiently long-lived to be seen in the experimental data. For this work, isotopes with a half-life between 1 month and 100 years were investigated

— The maximum beta particle energy should be at least 1 MeV according to Figure 2 to contribute at all to the measurable Cherenkov light intensity. The beta emission could come from either the decaying nucleus or a short-lived daughter.

Based on these criteria, three isotopes and their daughters were identified to be of relevance, as summarized in Table 2.

For the three isotopes, the parent nuclei are sufficiently abundant and long-lived to be of interest, and the short-lived daughter emits high-energy beta particles. The simulated Cherenkov light production by these three isotopes is summarized in Table 2. The beta energy spectra used in the simulations were taken from [14].

4.3 Direct beta contribution in the experimental data

Using the Cherenkov light intensity prediction for gamma and beta decays, the fraction of Cherenkov light produced directly by beta decays to the total intensities were calculated. The results are shown in Figure 3. As can be seen in Figure 3, the direct beta contribution to the total Cherenkov light intensity is significant, above 10% in all cases for this data set and up to 40% for the low-BU assemblies. In part, the thinner cladding and smaller rod radii mean that beta decays are more likely to directly produce Cherenkov light compared to previously studied fuel types, which is one reason why their contribution is so significant here. However, part of the explanation is also the irradiation history of these specific fuel assemblies. In Figure 3, the fuel assemblies have been further subdivided into groups according to CT, with group 1 having the shortest CT and group 4 the longest. For group 1, a wide range of BU is present, and the effect of the irradiation history results in a wider spread. The irradiation history for all assemblies typically consisted of initially a few high-power cycles, until a burnup of 20-30 MWd/kgU was achieved. The high-power cycles were followed by several low-power cycles, typically until a burnup of 40-45 MWd/kgU was achieved. Finally, the assembly experienced some medium-power cycles before the assembly reached its discharge BU, of typically 55-60 MWd/kgU. The final irradiation cycle for group 1 fuels may thus be either a low, medium or high-power cycle. For group 2-4, the fuel assemblies had typically reached their discharge BU and experienced a final, medium-power cycle, resulting in a much less pronounced spread in the beta fraction within each group.

The relative fraction of the direct beta contribution by each of the three identified isotopes is shown in Table 3, for three selected groups of fuel assembly parameters. The build-up of Ce144 peaks at a BU of around 20-30 MWd/kgU in the ORIGEN simulations for this data set. Part of the reason for this concentration peak is that these fuel assemblies had just experienced high-power cycles; hence, the production of Ce144 is high. Another cause is the

Isotope	Sr90/Y90	Ru106/Rh106	Ce144/Pr144
Parent half-life	28.9 years	372 days	285 days
Daughter maximum beta energy	2.24 MeV	3.53 MeV	2.99 MeV
Cherenkov photons per decay	2.71E-5	3.87E-4	1.56E-4

Table 2: The three identified isotopes and daughters that contribute to the total Cherenkov light intensity. The Cherenkov light production takes into account the beta emission spectrum of the daughter, and neglects bremsstrahlung.

Figure 3: The fraction of total Cherenkov light production caused by direct beta decays, as a function of the fuel assembly BU. The remainder is caused by either gamma decays or bremsstrahlung due to beta decays. The fuels are grouped according to their CT, with group 1 having the shortest CT, and the CT increases with group number. The vertical spread in group 1 is predominantly caused by the differing irradiation history within the group.

cumulative fission yield of Ce144, which is higher for fission in U235 than in Pu239. Plutonium fissions contribute more to the energy release at higher BU than at lower BU since it builds up with BU, and consequently the Ce144 production is reduced with BU. For Ru106 the cumulative fission yield is significantly higher for fission of Pu239 than U235, hence its importance tends to increase with BU for the assemblies analyzed, although it was occasionally seen to decrease slightly during the low-power cycles. For fully burned assemblies, the importance of Sr90/Y90 increases with CT as it is comparatively long-lived, with the importance of Ce144/Pr144 decreases faster than Ru106/Rh106 due to its slightly shorter half-life.

4.4 Evaluation of prediction performance

A comparison is made between the measurements and the predictions, for predictions with and without the direct beta contribution, shown in Figure 4. Note that one fit is done to relate predictions without direct beta contribution to the measurements, and a second fit is done to relate the predictions including the direct beta contribution to the measurements. After the fitting, the relative deviations between predictions and measurements in the data sets

were calculated and averaged over, to determine the RMSE values of the deviations. As seen earlier in Figure 3, the direct beta contribution varies significantly in this data set. That can be seen as a large spread in the difference between the prediction and measurement when the beta decays are not accounted for, as seen in Figure 4. In total, the predictions without the beta contribution had an RMSE between prediction and measurement of 20.7%. This was lowered to 11.6% when the direct beta contribution was taken into account; a value comparable to other measurement campaigns with long-cooled fuels where beta decay contribution was negligible, such as in [13]. Thus, when the direct beta contribution is included in the predictions, the predicted and measured values are a much better match after the new fit, judging by the RMSE values. This conclusion is valid when assessing the entire group of fuel assemblies in Figure 4, though note that individual predictions for a single fuel assembly may differ (i.e., improve or worsen) as the direct beta contribution is included.

In total, these results show that the direct beta contribution to the Cherenkov light intensity can be significant and should not be neglected for accurate prediction. The

Fuel	Sr90/Y90	Ru106/Rh106	Ce144/Pr144
CT 2 years, BU 20-30 MWd/kgU	5-10%	50-60%	30-40%
CT 2 years, BU 50-60 MWd/kgU	5-10%	60-70%	20-30%
CT 5 years, BU 50-60 MWd/kgU	30-35%	55-60%	10-15%

Table 3: The fraction of the direct beta contribution by each of the three identified isotopes, for three groups of fuel parameters in the experimental data. These groups represent combinations of high and low BU/CT that occur in the data set.
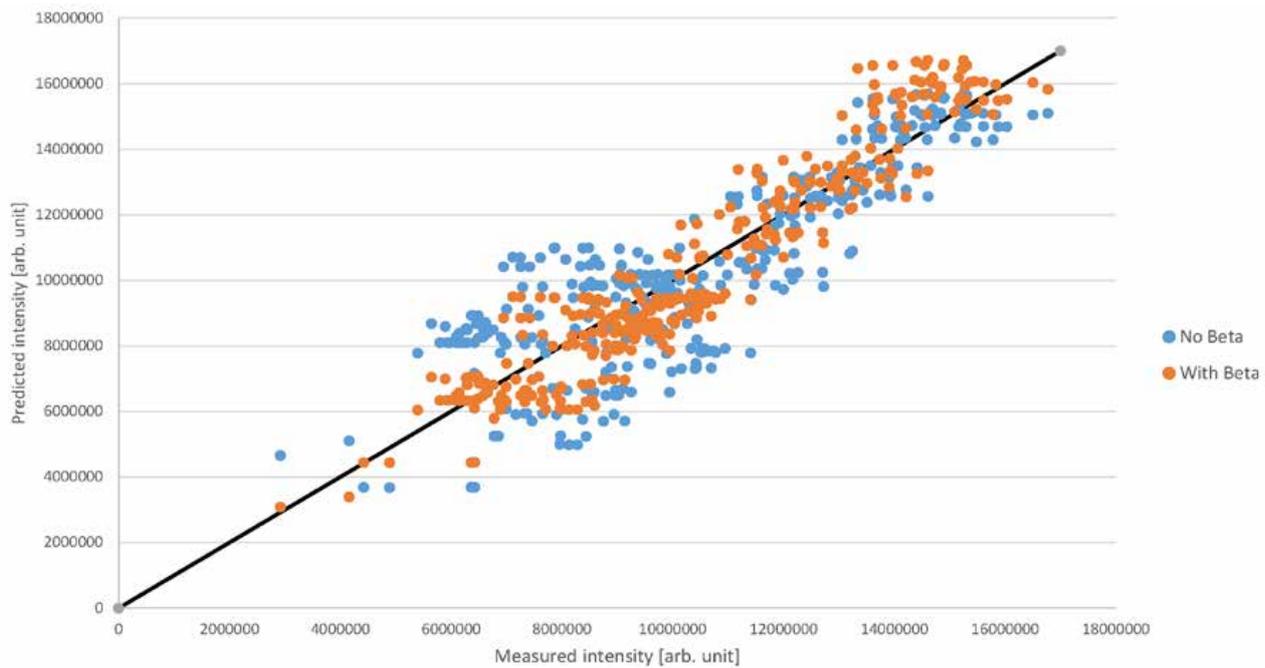
Figure 4: Comparison of the predictions of the two models, with and without the direct beta contribution, with the measured Cherenkov light intensities. The predictions of each model were scaled to match the measured intensity, as detailed in section 1.1. The black line is a guide for the eye, noting where the predictions and measurements match.

results also highlight that the abundance of several of the short-lived beta-decaying isotopes depends on the irradiation scheme used and should be accurately modelled for best results.

5. Conclusion and Discussion

Previous studies have identified that Cherenkov light is produced by beta particles that pass through the fuel material and cladding and directly produce Cherenkov light in the water. However, based on the available experimental data, it has not been clear when this becomes a significant contribution. In experimental measurements of fuel assemblies with cooling times of 2-5 years and burnups from 20-60 MWd/kgU, systematic effects were seen in the comparison between predictions and measurements, which largely can be explained by the direct beta contribution.

The model used to predict Cherenkov light intensities based on the fuel assembly gamma spectrum has previously been extended to also include beta decays explicitly. Since relatively few beta-decaying isotopes that significantly contribute to the Cherenkov light intensity are present, simulations were made for each isotope. This gives information about the intensity of the Cherenkov light produced per decay from each isotope. In turn, this can be combined with the isotope mass abundance in the spent fuel to obtain a prediction of the isotope-wise Cherenkov light production, which can be added to the total prediction. For the experimental data, adding the Cherenkov light contribution caused by these isotopes significantly reduced the systematic effects seen as a function of burnup and cooling time. The deviation between prediction and

measurement was reduced from an RMSE of 20.7% to 11.6%, which is comparable to measurement campaigns with long-cooled fuels where beta decays had a negligible impact on the predictions. In the simulation work performed here, it was found that the relatively thin claddings and the irradiation history resulted in a much larger direct beta contribution than previous results have indicated, with the direct beta contribution varied between 10% and 40% of the total Cherenkov intensity. This shows that there exist fuel assembly populations where the direct beta contribution cannot be neglected in the predictions.

Based on these results, we recommend that the prediction methodology in the DCVD software should be updated to include the direct beta-contribution thereby making accurate predictions available to safeguards inspectors. The methodology itself has already been developed and is ready to be incorporated in the next DCVD software version. Implementing this will require simulations of the Cherenkov light production from identified beta-decaying isotopes for a number of fuel geometries. Should the DCVD be used for even more short-cooled fuel assemblies than those studied in this work, additional isotopes may need to be added to the model.

While this work have obtained improved predictions using a rather simple direct beta modelling, accurately modelling the beta contribution is more challenging, but could potentially further improve the results. Fuel rods feature a high-burnup structure on the pellet rim, as fully moderated neutrons do not penetrate deep in the fuel material. In addition, the high burnup structure may results in pellet cracking, allowing beta decays from slightly deeper within the pellet to have a free path to the cladding. The direct

beta contribution was previously found to be sensitive to the cladding thickness, and effects such as cladding creep, oxygen and hydrogen pickup will likely influence the attenuation of beta particles in the cladding. Hence, for accurate modelling of the direct beta contribution, such effects should be included in the model, and especially how they change with time or burnup. Such models however need to be general enough that they can be applied based on the limited amount of data available to an inspector, in order to be useful in the field.

6. Acknowledgements

7. References

[1] IAEA, Safeguards Techniques and Equipment: 2011 Edition. IAEA, 2011. IAEA/NVS/1/2011 (ISBN:978-92-0-118910-3).

[2] J. Chen, D. Parcey, A. Ferwing, B. Wilcox, R. Kosierb, M. Larsson, K. Axell, J. Dahlberg, B. Lindberg, F.Vinnå, and E. Sundkvist, "Spent fuel verification using a Digital Cherenkov Viewing Device," in Institute of Nuclear Materials Management 46th annual meeting Portland, Oregon, 2009.

[3] J.D. Chen, D.A. Parcey, A.F. Gerwing, P. Carlson, R. Kosierb, M. Larsson, K. Axell, J. Dahlberg, B. Lindberg, S. Jacobsson Svärd, and E. Sundkvist. Partial defect detection in LWR spent fuel using a Digital Cerenkov Viewing Device. In Institute of Nuclear Materials Management 50th annual meeting Tucson, Arizona, pages 12-16, 2009.

[4] A. Bolind, The use of the BIC set in the characterization of used nuclear fuel assemblies by nondestructive assay. Annals of Nuclear Energy, Volume 66, pages 31-50. 2014. https://doi.org/10.1016/j.anucene.2013.11.010.

[5] S. Bowman, L. Leal, O. Hermann, and C. Parks. ORIGEN-ARP, A Fast and Easy-to-Use Source Term Generation Tool. Journal of Nuclear Science and Technology, 37, pages 575-579, 2000.

[6] E. Branger, S. Grape, S. Jacobsson Svärd, P. Jansson, E. Andersson Sundén, On Cherenkov light production by irradiated nuclear fuel rods. Journal of Instrumentation, June 2017. DOI: 10.1088/1748-0221/12/06/T06001

[7] S. Grape, S. Jacobsson Svärd, and B. Lindberg, Verifying nuclear fuel assemblies in wet storage on a partial defect level: A software simulation tool for evaluating the capabilities of the Digital Cherenkov Viewing Device, Nuclear inst. and Meth. A,, vol. Volume 698, pp. Pages 66-71, 11 January 2013.

[8] E. Branger. clip - python package for estimating the Cherenkov light intensity of a spent nuclear fuel assembly for nuclear safeguards purposes. Software available from https://github.com/ebranger/clip/ Retrieved 2022-02-15.

[9] E. Branger, S. Grape, S. Jacobsson Svärd, P. Jansson, E. Andersson Sundén, Comparison of prediction models for Cherenkov light emissions from nuclear fuel assemblies. Journal of Instrumentation, June 2017. DOI: 10.1088/1748-0221/12/06/P06007

[10] S. Agostinelli et. al, Geant4: a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 506(3):250 - 303, 2003.

[11] Nuclear Engineering International. Fuel design data, September 2004.

[12] H. Atak, A. Anastasiadis, P. Jansson, Zs. Elter, E. Andersson Sundén, S. Holcombe, P. Andersson, The degradation of gamma-ray mass attenuation of UOX and MOX fuel with nuclear burnup. Progress in Nuclear Energy, Volume 125, 2020, 103359. ISSN 0149-1970, https://doi.org/10.1016/j.pnucene.2020.103359.

[13] E. Branger, S. Grape, P. Jansson, S. Jacobsson Svärd, On the inclusion of light transport in prediction tools for Cherenkov light intensity assessment of irradiated nuclear fuel assemblies. Journal of instrumentation, January 2019. DOI: 10.1088/1748-0221/14/01/T01010

[14] Live chart of nuclides, IAEA Nuclear Data Section. Retrieved 2022-01-24. Available at https://www-nds.iaea.org/relnsd/vcharthtml/VChartHTML.html

Statistical error model-based and GUM-based analysis of measurement uncertainties in nuclear safeguards – a reconciliation

Oscar Alique[1], Yetunde Aregbe[2], Raffaele Bencardino[3], Robert Binner[4], Thomas Burr[5],
Jeff. A. Chapman[6], Stephen Croft[6], Andy Fellerman[7], Thomas Krieger[8], Klaus Martin[9], Peter Mason[10],
Claude Norman[4]*, Thomas Prohaska[11], Divyesh Trivedi[12], Steve Walsh[13], Dariusz Wegrzynek[14, 15],
Ben Wright[7], Jan Wüster[4]

[1]European Commission - Directorate Translation – Evaluation and Analysis, BECH C3, Rue Alphonse Weicker 5, 2721, Luxembourg, Luxembourg, [2]European Commission – Joint Research Centre (EC-JRC), Directorate G – Nuclear Safety & Security Unit G.2 - Standards for Nuclear Safety, Security & Safeguards, Retieseweg 111, B-2440 Geel, Belgium, [3]European Commission – Directorate General for Energy – EURATOM Safeguards EUFO3260, 10 rue R. Stümper L-2557 Luxembourg, [4]International Atomic Energy Agency - Nuclear Fuel Cycle Information Analysis Section- Division of Information Management - Department of Safeguards, Vienna International Centre, PO Box 100, 1400 Vienna, Austria, [5]United States Department of Energy, Los Alamos National Laboratory, Los Alamos, United States of America, [6]United States Department of Energy, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America, [7]National Nuclear Laboratory (NNL), B170, Sellafield, Seascale, Cumbria CA20 1PG, United Kingdom, [8]Forschungszentrum Jülich, Institute of Energy and Climate Research (IEK), Jülich, Germany, [9]Trier, Germany, [10]United States Department of Energy, NBL Program Office, Oak Ridge, Tennessee, United States of America, [11]Montanuniversität Leoben - Allgemeine und Analytische Chemie, Franz-Josef-Strasse 18, 8700 Leoben, Austria, [12]National Nuclear Laboratory (NNL), 5th Floor, Chadwick House, Birchwood, Warrington, WA3 6AE, United Kingdom, [13]Montana, United States, [14]International Atomic Energy Agency- Coordination and Support Section-Office of Safeguards Analytical Services- Department of Safeguards, Vienna International Centre, PO Box 100, 1400 Vienna, Austria, [15]AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, al.Mickiewicza 30, 30-059 Krakow, Poland

*corresponding author E-mail: c.norman@iaea.org

Abstract

At the occasion of the Consultants Group Meeting held to review the "International Target Values 2010 for Measurement Uncertainties in Safeguarding Nuclear Material" [3], discussions between experts highlighted the need to improve communication between different safeguards measurement communities, e.g. laboratory analysts, non-destructive assay specialists, safeguards data evaluators, and to reconcile their approaches to estimating measurement uncertainties. The purpose of this paper is to contribute to reaching a common understanding of the terminology and methodologies used by different professional groups in the field of uncertainty quantification.

Keywords: Statistical error model, GUM, terminology, uncertainty, measurement, nuclear safeguards

Introduction

Safeguards implementation requires a statistical evaluation of declared and verified amounts of nuclear material quantities to assess whether the differences can be explained by measurement errors or if they warrant further investigation. For this reason, the analysis of measurement error variances in the operator's and inspector's measurement systems and the modelling of their propagation into relevant evaluation statistics was researched within the safeguards community in the 1970s and 1980s. The approaches developed and the associated terminology have been consistently used since and are currently undergoing review and enhancement. Technical progress towards the end of the 20th century pushed the performance of measurement technologies in many disciplines to their practical or even theoretical limits of applicability, while increasing international cooperation, culminating in deep global supply chains, required a commonly understood way to communicate measurement results and associated uncertainties. The metrological community responded to these needs by defining an international standard, the 'Guide to the Expression of Uncertainty in Measurement' (GUM). Since its first publication in 1995, this standardized approach of first-principles ("bottom-up") uncertainty quantification (UQ) and the associated terminology have been adopted by an increasing number of laboratories, including those where nuclear material (NM) samples from the fuel cycle are regularly analyzed and those where instruments for destructive analysis (DA) and non-destructive assay

(NDA) of NM are developed and calibrated. The adoption of the GUM is not complete but will eventually lead to every reputable laboratory measurement result being metrologically traceable and accompanied by a defensible uncertainty statement. On the other hand, the NDA measurement community faces specific UQ challenges, such as incompletely controlled measurement conditions and item-specific biases. Error variance propagation is a key component of UQ using both analytical and Monte Carlo approaches, however there is no general NDA UQ guide analogous to the GUM. The need for more comprehensive bottom-up UQ for NDA, including model-based adjustments of test items to calibration items is recognized. The GUM recommends and, in its capacity as an international guide published by the Joint Committee for Guides in Metrology (JCGM) of the Comité International des Poids et Mesures (CIPM) – i.e. the International Committee for Weights and Measures - also prescribes, a different terminology from the one traditionally employed by some of the authors and specialists in the analysis of measurement errors in safeguards, who typically are educated in the field of statistics. Further, certain sections of the GUM, especially in its earlier versions, explicitly discourage references to concepts that are central to safeguards practice, such as the concept of the true value of a measurand and the concept of an error model that explicitly distinguishes between random and systematic errors.

This dual difference, both terminological and conceptual, complicates communication between professional communities interested in measurement uncertainty, such as the safeguards laboratories and the safeguards statistical evaluation services. Attitudes have ranged from a desire to explain and to convince the other community to a more or less benign mutual neglect, but recently the communities took the opportunity to learn from each other. With this article the authors wish to overcome the undesirable impediments to communication between the relevant professional communities by reconciling the safeguards statistical error model with the GUM-based analysis of measurement uncertainties. The GUM is mostly known for bottom-up UQ used by metrologists, but also includes information regarding top-down UQ often used by statisticians. On the other hand, safeguards evaluators focus on top-down UQ by analysis of paired operator-inspector data, but also make use of bottom-up UQ. Properly interpreted these approaches are complementary rather than contradictory and hold the promise of mutual interdisciplinary cross-fertilization. Motivated by this perspective, the equivalence of top-down paired data analysis as applied by the IAEA with GUM-inspired repeatability and reproducibility analysis has recently been demonstrated. The approaches are not expected to be completely unified, because the underlying objectives are different, but the potential benefits of convergence in areas of overlap are identified and steps towards such convergence are recommended.

Starting in Section 1 with an historical review of UQ in the field of safeguards and the parallel development of the GUM, the paper describes in Section 2 and 3 how UQ methodologies are respectively used by safeguards data evaluators and the DA and NDA laboratories. The purpose of Section 4 is to systematically compare and reconcile the methodologies and terminology used by these communities while Section 5 focuses on the statistical bases of UQ methodologies. The complementarity of their purposes and the mutual benefits of communication and convergence between the professional communities involved are underlined in Sections 6 and 7 which support the paper's conclusion.

1. Historical Developments and current situation

1.1 Safeguards at the IAEA and EURATOM

The International Atomic Energy Agency (IAEA) was established in 1957 as an independent intergovernmental organization in the United Nations system. Article III of the IAEA's statute provides the IAEA with the authority to apply safeguards on nuclear material and other specified items. The IAEA's Department of Safeguards' primary role is to deter the proliferation of nuclear weapons by detecting early any misuse of nuclear material or technology, and by providing credible assurances that States are honouring the obligations stemming from their safeguards agreements.

Also in 1957, the European Atomic Energy Community, or EURATOM, was established and exists next to the European Union as a separate legal entity. Article 77 of the EURATOM Treaty provides the European Commission safeguards system with the authority to ensure that nuclear materials are not diverted from their intended civil uses, while complying with safeguards obligations concluded with third states and international organisations such as the IAEA.

In order to detect diversion of declared nuclear material, nuclear material accountancy (NMA) is used as the basic safeguards measure. NMA is that part of a nuclear material safeguards program that consists of procedures and systems to perform nuclear material measurements, prepare and maintain accounts and records, and perform data analyses. Statistical analysis is an essential element of effective NMA, and over the past 50 years, highly specialized statistical procedures have been developed to address unique problems encountered in NMA and associated verification activities [1]:

- Recognition of multiple error sources in a material balance (e.g., sampling, instrument, analyst, environmental conditions).
- Estimation of variance components associated with each error source.

- Reconciliation of measurement results from different measurement systems and different laboratories, often obtained from independent samples taken at different times.

- Assurance of independent verification of inventories and balances.

The safeguards mandate to independently evaluate material balance differences including the operator-declared *Material Unaccounted For (MUF[1])*, *the Shipper-Receiver Differences (SRD[2])*, and the projected Difference between Operator declaration and Inspector verification D statistic (D), fundamentally depends on the estimates of the measurement error uncertainties associated with all nuclear material quantities that enter the material balance [2]. Because the uncertainties associated with MUF, SRD, and D are obtained by error propagation methodologies applied to estimates of measurement error uncertainties, independent evaluation can only be accomplished if methods are available that allow safeguards evaluators to obtain such estimates.

Hence, early in the history of safeguards a need arose for specialized statistical procedures to estimate measurement uncertainties associated with nuclear material quantities, for both the operators' declared values and for the inspectors' verification results (for verifications by both DA and NDA). These estimates needed to be independent, i.e., not simply declared by the facility operator and accepted by the safeguards authority. Because independent information is gathered by the safeguards authority in the form of verification measurement results on a sample of items, this information needed to be utilized, in conjunction with the corresponding operator's declarations, to obtain estimates of measurement uncertainty for both operator and inspector through the analysis of operator-inspector paired differences (or, in the absence of sufficient independent measurement data, through the use of international target values (ITV) [3], themselves partially derived from historical paired difference analysis and other information such as the evaluation of laboratory measurement capabilities [4].

Methodologies developed for this task as applied to IAEA safeguards were first formulated in 1977 by John Jaech in the *Safeguards Technical Manual (STM)* and further developed in subsequent STM volumes [5], mainly based on his earlier work for the U.S. Atomic Energy Commission [6]. As methodologies were further refined or newly developed, several revisions of the STM were published, with the final

version (revision 5), re-named *Statistical Concepts and Techniques for IAEA Safeguards* [7], published in 1998. More recently, further extensions to the methodologies have been developed and tested (e.g. *Optanova,* a methodology and associated software for determining the optimal top-down estimators of the variances of random and systematic errors for paired and three-laboratory data [8]).

1.2 The Guide to the Expression of Uncertainty in Measurement (GUM)

In 1977, the International Committee for Weights and Measures (CIPM - Comité International des Poids et Mesures) asked the Bureau International des Poids et Mesures (BIPM) to address the problem of a lack of common agreement on expressing measurement uncertainties in order to facilitate comparison of laboratory results. The issue was addressed by the BIPM by convening a working group on the statement of uncertainties in 1980, including members from a number of national metrology institutes from around the world. The chair of the committee specifically stressed that the main goal of the working group was to develop clear and simple rules applicable to the determination of uncertainties, that these should be generally applicable to a large majority of users, and that it would be best to produce guidance that can be used at any level of metrology [9], [10].

While a major motivation was to address the significant issues being faced by the national metrology institutes evaluating measurements, which did not have a transparent or even comparable means of calculating and reporting measurement uncertainties, the principle of broad applicability was stressed from the outset. The result of the working group was recommendation INC-1, which was the progenitor of the modern *GUM*. The recommendations were approved by CIPM in 1981 and reaffirmed in 1986. At that time the CIPM asked the International Organization for Standardization (ISO) to work with a number of other standard setting bodies to develop a detailed guide based on the broad recommendations. The first full expression of the CIPM's recommendations was the GUM, published in 1993 [11] which has been periodically updated and is currently under the auspices of BIPM's Joint Committee on Guides to Metrology Working Group 1. The current GUM, published in 2008 [12], has been widely adopted in the analytical laboratory community.

1.3 Common Ground

At about the same time the GUM was developing, the IAEA established a set of expected measurement uncertainties associated with safeguards at nuclear fuel cycle facilities, but lacked specific details regarding the performance of measurement systems used for the determination of specific safeguarded nuclear materials. The Working Group on DA of ESARDA in 1979 presented a list of 'target values' for the uncertainty components in

---

1 In compliance with safeguards agreements and in application of the related NMA provisions, facility operators have to declare their balance and any MUF at the end of each material balance period (MBP), for each material balance area (MBA) and each nuclear material category. The MUF is defined as the difference between the physical inventory and the book inventory (accountancy ledger).

2 The difference between the quantity of nuclear material in a batch as stated by the shipping MBA and as measured at the receiving MBA.

nuclear material measurements [13]. A number of revisions were published [14,15] in consultation with laboratories and safeguards organizations, and eventually in 1993 the IAEA published a technical report detailing the collaborative effort [16], followed up in 2000 by the International Target Values 2000 (ITV-2000) [17]. The ITV-2000 document listed separately systematic and random components of uncertainty (which is essential for estimating the uncertainties associated with evaluating a material balance) for a number of measurement methods, and specifically stated that the developments related to GUM (referenced as the ISO, NIST and EURACHEM guides that were developing concurrently) involve uncertainty assessments in line with the developing GUM guidance. The 2010 ITV document [3] included for the first time a third column of uncertainty values, labelled "ITV", which are meant to document the total uncertainty associated to the methods listed in the tables.

In the following years, progress was made by both the laboratory and the safeguards communities in moving toward a better understanding of the methodologies used to estimate measurement uncertainties applicable to the respective needs of the two communities. However, misconceptions still exist, which were not entirely resolved during the consultations for establishing and updating the ITVs, and which this article strives to identify and clarify. These collaborative efforts between the safeguards and laboratory communities in establishing reference values for expected uncertainties associated with NMA and safeguards verification activities have increasingly highlighted the need for a common understanding of the statistical basis, terminology, and intended uses of uncertainty estimates applied in the course of performing evaluations of safeguards data. The "bottom-up" [3] approach to uncertainty estimation at the heart of GUM[4] (based on propagation of uncertainties determined for every component identified as influencing the outcome of a measurement) and the "top-down"[3] approach applied to safeguards verification data (based on ANOVA of operator-inspector differences) both arose out of historical need and serve their respective communities well[5].

However, there is much to be gained in 'reconciling', which means increasing the understanding among the safeguards evaluators, laboratories, nuclear facility operators and metrology experts, and in finding a shared language between the two approaches [18,19,20]. It is the intention of this article to bridge 'gaps' and facilitate this common understanding to the benefit of the safeguards measurement and metrology community, by describing how uncertainty quantification methodologies are used by safeguards data evaluators and the DA and NDA laboratories, comparing and reconciling the related methodologies and terminologies and the underlying complementarity of their purposes as well as the mutual benefits of communication and convergence between the professional communities involved.

2. Measurement uncertainty analysis in safeguards

One of the main purposes of safeguards verification activities is to detect in a timely manner and to deter the diversion of nuclear material from declared nuclear fuel cycle facilities. NMA is the basis for the detection of diversion of nuclear material by means of its keystone, material balance evaluation (MBE), which is performed for each material balance area (MBA), each material balance period (MBP) and each nuclear material category (e.g. depleted uranium, natural uranium, enriched uranium, the associated ^{235}U, plutonium, thorium). In bulk-handling facilities, where nuclear material is processed in loose forms such as gas, liquid or powder, and where nuclear material quantities are associated with process losses, where hold-up and waste have to be estimated, and where most accounting records are based on measurement results intrinsically subject to errors, MBE statistics such as MUF, SRD and the difference between the operator's declarations and the inspector's verification measurement results (operator-inspector differences, D) are necessarily non-zero. They have to be statistically tested to determine whether or not they can be explained by the operator's and inspector's measurement uncertainties.

Before statistical tests can be applied, measurement uncertainties must be estimated and propagated from the item level to the level of the MBA and MBP. The estimation of measurement uncertainties is one of the most demanding questions faced by statistical methodologies for safeguards. In some cases uncertainty estimates are documented by facility operators, analytical laboratories and/or instrument developers but to support credible conclusions regarding the absence of diversion, they need to be validated independently and their fitness for purpose needs to be assessed by safeguards analysts.

Estimating measurement error variances can be performed, for example, using a "bottom-up" approach via calibration certificates and the validation of nuclear

---

3   A 'bottom-up' (first principles) approach starts from a measurement equation identifying all variables that influence the measurement results and propagates the corresponding uncertainty components to establish an uncertainty budget. A "top-down" (empirical) approach starts from a statistical measurement model and applies analysis of variance to data comparing measurement results with a reference such as quality control data or other measurement data, e.g., in the context of safeguards, declared data.

4   Note: the GUM also very briefly treats top-down UQ, but it is better known for bottom-up UQ.

5   The deliverable of any laboratory is a measurement result (expressed as a measured quantity value with uncertainty and traceability) of a measurand (analyte in the investigated matrix).The deliverable of the safeguards evaluator community is, inter alia, an assessment of the statistical significances of observed operator-inspector differences and their impact on drawing safeguards conclusions. While the common goal to evaluate measurement data is the same for the laboratory and the evaluator communities, the applied models will generally differ, based on the purpose of their construction and the nature of the deliverables.

operators' data or a "top-down" approach by analysis of variance (ANOVA) of observed operator-inspector differences, i.e. paired-data. Because calibration-based uncertainty quantification (UQ) does not necessarily account for all sources of uncertainty, paired data analysis has traditionally been the method of choice at the IAEA. On the other hand, because paired-data based UQ does not only include the facility operators' measurement uncertainties, the analysis of operators' data has been the method of choice at the European Commission (EC) inspectorate.

In the top-down approach, the observed paired differences reflect the combined effect of the operator's and inspector's measurement errors, and form the basis for the estimation of the relative standard deviations (RSD) associated with these errors, which in turn are needed to obtain uncertainty estimates associated with MUF, SRD and D, to calculate verification sample sizes and to establish rejection limits for individual operator-inspector differences. Measurement uncertainty estimates are quantified by the absolute (relative) standard deviation of measurement errors, respectively denoted σ (δ). The propagation step requires, in addition to the separation of the operator's (σ_o or δ_o) and inspector's (σ_i or δ_i) error standard deviations, a further parsing of both of these into a random e.g. ($\sigma_{O,R}$ or $\delta_{O,R}$) and a (short-term) systematic e.g. ($\sigma_{O,S}$ or $\delta_{O,S}$) component, because the averaging process reduces the effect of random errors in multiple measurements while the effect of systematic errors is not reduced by averaging, and the different mode of propagation of these two error components into material balances makes it essential for safeguards analysts to obtain separate estimates for their respective standard deviations. It must be noted that the separation of error standard deviations into four components is required by the error propagation process regardless of the chosen UQ approach.

One of the main difficulties when applying ANOVA to paired data is to obtain separate estimates of the four different uncertainty components. This task is further complicated by the need to process outliers and to validate various assumptions (e.g. normally distributed random errors) that are necessary for the implementation of certain algorithms. As explained in section 1 above, the methods used by the IAEA to estimate measurement error variances were developed several decades ago (e.g. Grubbs Analysis, 1948 [21]) and are presently being refined.

When a bottom-up approach is applied to estimate the operator's error RSDs, the operator's UQ practice is audited and the resulting uncertainties are confirmed to comply with latest international standards. The uncertainty associated with the operator's declared MUF can then be computed by error propagation in order to perform a statistical test of the hypothesis that it can be explained by measurement errors.

3. GUM in the Laboratory

The chair of the BIPM committee tasked to develop what became the GUM specifically stressed that the main goal of the group was to develop clear and simple rules applicable to the determination of uncertainties, that these should be generally applicable to a large majority of users, and that it would be best to produce guidance that can be used at any level of metrology [9]. The idea of a standardized approach to uncertainty evaluation is to provide a method that is applicable to all types of measurements with results (including uncertainties) that are transparent and easily utilized by a variety of users (i.e. the value and uncertainty should be easily transferrable). While the effort and expertise necessary to produce an uncertainty evaluation is often not a simple task, the basic JCGM 100:2008 (GUM Guide) approach provides a stepwise and relatively easily taught and understood mechanism to uncertainty evaluation that has proved to be of great practical value to measurement practitioners [12]. While the requirements of ISO/IEC 17025:2005 and recently ISO/IEC 17025:2017 (and to some extent ISO/IEC 17043:2010 and ISO Guide 34:2009 and more recently ISO 17034:2016) have driven the utilization of the GUM, in workshops conducted throughout the nuclear measurement community, the use of the GUM approach has engendered much positive discussion and has been typically embraced by laboratory staff and management around the world [22,23,24,25]. In particular, laboratory professionals from bench technicians to measurement experts have indicated that the GUM gives them a usable framework for a better understanding of their measurements, helps them to identify potential problem areas, and provides them with useful guidance on how to report measurement results in a transparent and organized manner.

A simple example demonstrating the practical use and benefit of the GUM arose during an introductory workshop at a US national laboratory when staff assigned to perform assay measurement of a plutonium storage tank provided their measurement method for modelling and evaluation by GUM. The tank was measured for accountability purposes on a semi-annual basis and results were submitted to the material accountancy organization. The procedure specified the use of a random error RSD of 4% for the distance of the detector from the tank (76 cm +/- 3 cm), to account for imprecision in reproducible placement of the detector. Employing the GUM methodology for this procedure was straightforward, with the result that the error RSD ascribed to the distance of the detector from the tank contributed nearly 80% to the overall RSD of the measurement. The entire exercise took about half an hour to perform. The technicians were particularly surprised at the influence of the distance uncertainty on the overall RSD and awareness was created among the technicians, lab manager and statistical staff of the laboratory.

The point of this simple example is to illustrate that the basic principles of the GUM are relatively easily grasped and implemented, that the GUM provides an accessible tool to measurement practitioners of varying expertise, and that these attributes provide a useful means for those performing and/or using measurement results to better understand their measurement processes. Similar positive results have arisen in a number of situations during workshops and discussions among laboratory staff.

More complex examples that arose during GUM uncertainty evaluations included demonstrations emphasizing the use of isotopic ratios rather than abundances in uncertainty determinations, the correlation of mass bias (K-factors) in thermal ionization mass spectrometry (TIMS) leading to unexpectedly small uncertainties for minor isotopes, and identifying unexpected significant contributors to certified reference materials (CRM) production efforts. In the case of the production of CRMs, the creation of GUM uncertainty budgets prior to any analytical effort has been invaluable in deciding the scope and breadth of effort required for the production and certification of a variety of uranium and plutonium CRMs, identifying key contributors to the final product's uncertainty [26]. For example the contributions to the combined standard uncertainty of a measurement result of $n(^{236}U)/n(^{238}U)$ in IRMM-2022 by TIMS are shown in Figure 1 [27,28].

In 2006, the EC-JRC conducted a proficiency testing exercise that included 71 laboratories from 26 countries performing uranium isotopic measurements [29]. A wide range of laboratories participated, active in research, environmental radioactivity measurements, monitoring of nuclear facilities, medical applications and safeguards. Of the 71 laboratories that submitted measurement results, 30% reported they held ISO 17025 accreditation. Nearly half of all laboratories reported that their uncertainties were calculated according to the GUM. While compliance with ISO 17025 requires GUM uncertainty evaluations, many labs in 2006 were already utilizing the GUM without the accreditation. Usage of the GUM has no doubt further increased in the ten years since this exercise was completed, with laboratories finding benefit in utilization of the GUM.

3.1 GUM in Proficiency Testing and Laboratory Self-Evaluation

Laboratories, particularly in safeguards, have to demonstrate their performance over short and long terms by means of conformity assessment and quality control tools [4]. They are required to have measures in place to ensure that the measurement process is stable and in control. The GUM-based uncertainty values in the ITV 2010 document are being utilized by laboratories and proficiency testing providers as benchmarks. The REIMEP-17 inter-laboratory comparison (ILC), reported in 2015, utilized the prescribed ISO 13528:2005 and ISO 17043:2010 statistics for evaluating laboratory results' agreement with the reference values for the distributed materials [30,31]. The laboratories in REIMEP 17 were thus evaluated against the ITV-2010 GUM-based uncertainties to compare their performance to the state of the practice for measurements as determined by the ITV-2010 document. The New Brunswick Laboratory Safeguards Measurement Evaluation Program (SME) utilizes a similar approach, The 2011 NBL SME Report utilized five different fuel-cycle materials, with 23 laboratories participating. The submitted results were compared to the reference values and also the ITV-2010 GUM-based target
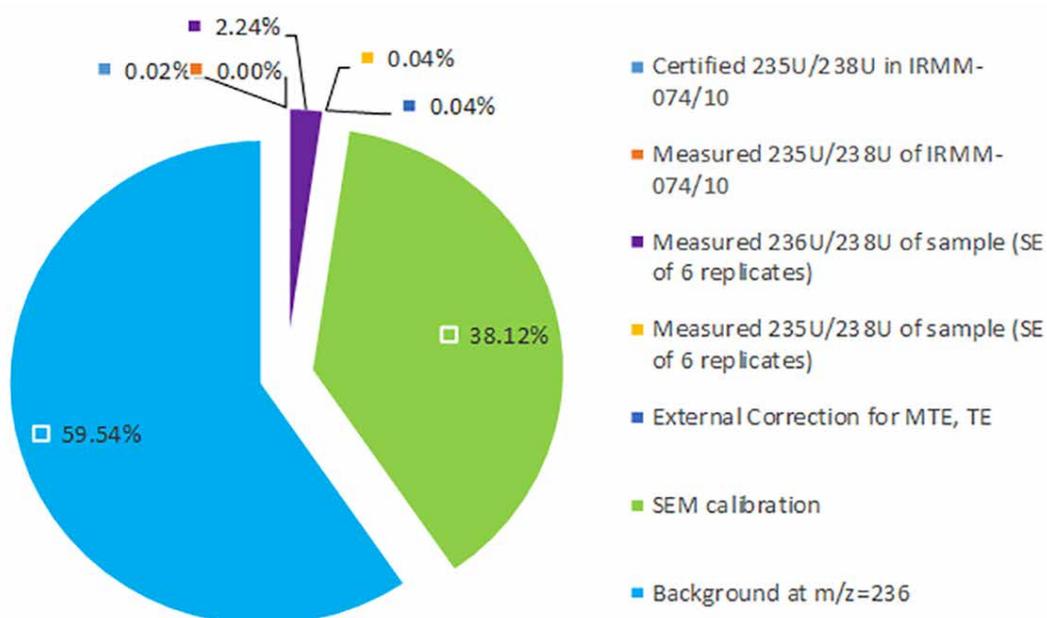
Figure 1: Uncertainty contributions for the measurement of n(236U)/n(238U) in IRMM-2022

values [32]. In most instances, the CRMs used in ILCs are produced by specialized laboratories using state-of-the-art methods and produced painstakingly to yield the smallest possible uncertainties at the highest metrological standards. Laboratories participating in proficiency testing exercises typically perform analyses using standard procedures, at state-of-the-practice levels and in many cases employing analytical methods yielding RSDs two-to-tenfold larger than the certified RSDs of the material. While the ISO 13528 and 17043 evaluation methods are useful for assessing agreement with the certified value in these cases, the additional comparison utilizing the ITV-2010 GUM-based uncertainties is of particular practical value in comparing laboratory and method performance.

Given the increasing use and importance of the ITV-2010 GUM-based RSDs, the refinement of the GUM based values could be addressed in a future revision of the ITV document, more accurately representing GUM-compliant uncertainties as determined in the field. The use of ILC exercise results and individual laboratory reporting is vital to this effort [33,34]. In conjunction with the next revision of the ITVs, the GUM-based laboratory performance values may be issued in a separate document. This would distinguish them from the ITVs, which are derived from historical paired difference analysis and other information, but not exclusively from the evaluation of laboratory measurement capabilities. This will distinguish them from the ITVs, which are partially derived from historical paired difference analysis and other information but not exclusively from the evaluation of laboratory measurement capabilities.

3.2 Fit-for-purpose according to GUM

The purpose of performing a measurement is to provide a result with stated uncertainty and traceability of a measurand to be utilized by one or more users for various purposes. A measurement result lacking a value an uncertainty or traceability is meaningless and not useable. For the purpose of decision making or conformity assessment, the measured value and uncertainty must be transferrable and comparable by the end-users. The result would provide measurement producers and users with the ability to demonstrate fitness for purpose, demonstrate laboratory proficiency and provide assurance of laboratory capability [35]. The methodical and accessible approach to uncertainty evaluation provided for by the GUM has been embraced within the nuclear measurement community and has been established in a large variety of nuclear measurement laboratories worldwide. It is worth noting that, given the specific purpose described in section 2, the IAEA, which is using the top-down paired data approach for UQ, does not need to take reported measurement uncertainties from either operator or inspector laboratories into account.

3.3 Destructive analysis (DA) Laboratory

To achieve traceability, one must link measurand identity and quantity value to a stated reference (preferably via calibration standards and CRMs). To give an example, DA nuclear laboratories are routinely measuring the plutonium amount in a plutonium nitrate solution sample [36]. They need to provide an accurate and traceable measurement result within the respective ITV-2010 GUM-based uncertainties. Often the method of choice is Isotope Dilution -Thermal Ionisation Mass Spectrometry (ID-TIMS). In ID-TIMS, using a ^{242}Pu enriched material as the spike, the ^{239}Pu content in an unknown sample can be determined by isotope dilution, through a measurement of the isotope ratio $R(^{242}Pu/^{239}Pu, B)$ in the blend. Following the GUM's generic measurand equation

$$Y = f(X_1, X_2, ..., X_N) \tag{1}$$

where Y denotes a measurand determined from *N* other quantities $X_1, X_2, ..., X_N$ through the functional relationship *f*.

the plutonium amount content can be calculated as follows [37]:

$$c(Pu, X) = \frac{R(^{242}Pu/^{239}Pu, Y) - R(^{242}Pu/^{239}Pu, B)}{R(^{242}Pu/^{239}Pu, B) - R(^{242}Pu/^{239}Pu, X)} \cdot \frac{\sum_m R(^m Pu/^{239}Pu, X)}{\sum_m R(^m Pu/^{239}Pu, Y)} \cdot \frac{m(Y)}{m(X)} \cdot c(Pu, Y) \tag{2}$$

where:

$R(^m Pu/^{239}Pu, X)$ = amount ratio $^m Pu/^{239}Pu$ in the unknown sample material X

$R(^m Pu/^{239}Pu, Y)$ = amount ratio $^m Pu/^{239}Pu$ in the known spike material Y

$R(^m Pu/^{239}Pu, B)$ = amount ratio $^m Pu/^{239}Pu$ in the measured blend material B

$m(X)$ = mass of the unknown sample used in the measurement

$m(Y)$ = mass of the spike solution used in the measurement

$c(^{239}Pu, X)$ = amount content (moles) of ^{239}Pu / g sample material

$c(^{242}Pu, Y)$ = amount content (moles) of ^{242}Pu / g spike solution

$c(Pu, X)$ = amount content of Pu / kg sample material

$c(Pu, Y)$ = amount content of Pu / kg spike solution

If any of these components (analyte, value, uncertainty, unit) is missing, the measurement is meaningless. Laboratories estimate the combined standard uncertainty of a measurement result by applying either the bottom-up or top-down approach. In the bottom-up approach the uncertainties of each factor in the measurement model are estimated and these individual uncertainties are combined

according to the law of error propagation applied to Eq. (1) [38]

$$u_c(y(x_1,...,x_n)) = \sqrt{\sum_{i=1,n}(\frac{\delta y}{\delta x_i})^2 \times u(x_i)^2} \qquad (3)$$

where y denotes the estimate of Y and x_i denotes the estimate of X_i.

In the top-down approach, combined effects covering several factors - also unknowns - are estimated using uncertainties due to repeatability (u_{rep}), intermediate precision (u_{ip}) and "trueness" (u_t), as established by means of a CRM, combined with uncertainties for calibration (u_{cal}) [39].

$$u_c = \sqrt{u_{rep}^2 + u_{ip}^2 + u_t^2 + u_{cal}^2} \qquad (4)$$

If performed accurately and documented properly, so that it is possible for an external auditor to reproduce how the combined standard uncertainty was estimated, the laboratory is fully compliant with ISO standards independently of the chosen approach. One main advantage of a bottom-up approach is that it yields detailed information for method improvement, whereas such information is not revealed by a top-down approach.

3.4 Non-destructive assay (NDA) Laboratory

NDA of items containing nuclear material uses calibration and modelling to infer item characteristics such as nuclear material mass on the basis of detected radiation such as neutron and gamma emissions. Three specific issues in UQ for NDA are as follows.

NDA is often applied in challenging settings because the detector is brought to the facility where ambient conditions can vary over time, and because the items to be assayed are often heterogeneous in some way. Because of such challenges, dark uncertainty [33] can be large, as is evident whenever bottom-up UQ predicts smaller uncertainty than is observed in empirical (top-down) UQ [34] (by "uncertainty" we mean the reproducibility standard deviation as quantified, for example, in an ILC) [39].

NDA is widely applied in situations where the items subject to measurement differ substantially from the calibration items; therefore, the concept of item-specific bias has long been recognized [40,41].

Currently, there is no general UQ guide for NDA that is analogous to the GUM. But, the GUM is typically followed for the error variance propagation steps in UQ, and each NDA method has a specific and documented implementation of UQ, for example, ASTM C1514 for the enrichment meter principle (EMP) as discussed in full detail elsewhere [42]. However, this NDA example needs to be presented here in a consolidated manner to follow the reasoning

towards reconciliation of complementary approaches as discussed later on in sections 5 and 6

Example: Enrichment Meter Principle (EMP) for gamma spectroscopy

This sub-section provides an example that involves calibration of gamma spectroscopy in order to describe some of the statistical aspects of bottom-up UQ. The amount of ^{235}U in an item can be estimated by using a measured net weight of uranium U in the item and a measured ^{235}U enrichment (the ratio $^{235}U/U$). Enrichment can be measured using the 185.7 keV gamma-rays emitted from ^{235}U by applying the EMP. The EMP aims to infer the fraction (enrichment) of ^{235}U in U by measuring the count rate of the strongest-intensity direct (full-energy) gamma from decay of ^{235}U, which is emitted at 185.7 keV [43,44,45]. The EMP assumes that the detector field of view into each item is identical to that in the calibration items (the "infinite thickness" assumption), that the item must be homogeneous with respect to both the ^{235}U enrichment and chemical composition, and that the container attenuation of gamma-rays is equal or similar to that in the calibration items, so that empirical correction factors have modest impact and are reasonably effective. If these three assumptions are met, the known physics implies that the enrichment of ^{235}U in the U is directly proportional to the count rate of the 185.7 keV gamma-rays emitted from the item. It has been shown empirically that under good measurement conditions, the EMP can have a random error RSD of less than 0.5 % and a long term bias of less than 1 %, depending on the detector resolution, stability, and extent of corrections needed to adjust items to calibration conditions. However, in some EMP applications, the random error RSD can be larger than bottom-up UQ predicts (see next paragraph) and larger than the 0.5% target random RSD. For example, assay of the ^{235}U mass in UO$_2$ drums suggests that there is larger-than-anticipated random RSD in some deployments of the EMP.

To investigate UQ for the EMP, Burr et al. [46] fit the known enrichment in each of several standards to observed counts in a few energy channels near the 185.7 keV energy as the "peak" region and to the counts in a few energy channels somewhere below and above the 185.7 keV energy but outside the peak area to estimate background (two-region EMP method), expressed as

$$Y = \beta_1 X_1 + \beta_2 X_2 + R \qquad (5)$$

where Y is the enrichment, X_1 is the peak count rate near 185.7keV, X_2 is the background count rate in neighbouring energy channels near the 185.7keV peak region, and R is random error. Figure 2 is an example low-resolution (NaI detector) gamma spectrum near the 185.7keV. The two background ROI counts can be combined into one count, resulting in two predictors as in Eq. (5): X_1 is the peak ROI counts and X_2 is the background ROI counts to be used to predict enrichment E in Eq. (5) using least squares
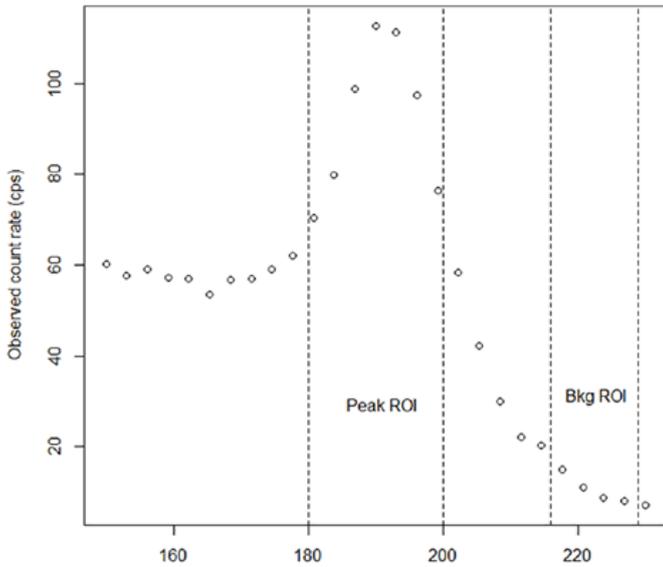
Figure 2: Example low-resolution (NaI detector) gamma spectrum near the 185.7keV peak with two background regions (one region below the 185.7 keV peak and one region above the 185.7 keV peak)

regression. There will be measurement errors in X_1 and X_2 and there will often be correction factors applied to X_1 and X_2, for example, to adjust test item container thickness to calibration item container thickness. Calibration data is used to produce estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ of the two model parameters, β_1 and β_2. The covariance matrix of the random variables $(\hat{\beta}_1, \hat{\beta}_2)$ is not necessarily well approximated by the usual least squares expression because of errors in X_1 and X_2. Therefore, [44,45] suggest that the mean squared error (MSE) in $\hat{Y}$ be estimated using simulation of the calibration procedure, which easily allows for errors in X_1 and X_2 arising from Poisson counting statistics, and also arising from other sources, such as container thickness (with or without an adjustment for the measured container thickness) varying among test items. Errors in X_1 and X_2 due to imperfect adjustment for container thickness can manifest as item-specific bias. The simulation strategy in [44,45] and the summary sub-section below illustrate how item-specific bias can be understood and estimated. The MSE in $\hat{Y}$ is defined as usual, as

$$E((\hat{Y} - Y_{true})^2) = E(\hat{Y} - E(\hat{Y})^2) + (E\hat{Y} - Y_{true})^2 = variance + bias^2.$$

We can express the simple calibration Eq. (5) as in Eq. (1), where we identify X_1 as $\hat{\beta}_1$, X_2 as $\hat{\beta}_2$, X_3 as X_1, and X_4 as X_2, respectively, with cov $(\hat{\beta}_1, \hat{\beta}_2)$ estimated by simulation, so with some effort, GUM's Eq. (1) could be used to estimate var $(\hat{Y}_1)$ and cov $(\hat{Y}_1, \hat{Y}_2)$, although Elster [47] points out that GUM's Eq. (1) is not actually designed to be applied to calibration applications, regardless of whether there are errors in the predictors X_1 and X_2 (which complicates the data analysis). Some of the numerical bottom-up UQ examples in [44,45] have estimated random error RSD ranging from less than its 0.5% target to approximately 1.0% (because of
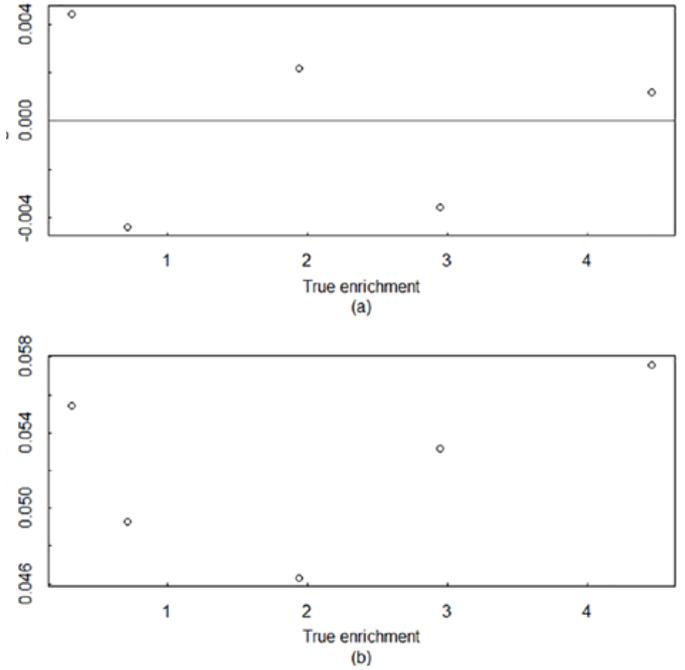
Figure 3: The average residual (a) and the MSE (b) in testing data, using the same values for testing and training. Results are based on 10^5 simulations so simulation error is negligible. The data, collected during 2015 at Oak Ridge National Laboratory are enrichment $Y = \{0.3166, 0.7119, 1.9420, 2.9492, 4.4623\}$, and $X1 = \{5616, 10298, 25093, 37103, 55178\}$, $X2 = \{1803, 1815, 1914, 1984, 2132\}$. The assumed absolute error standard deviations were 0.0035 in Y and 1% of the range of $X1$ in $X1$ and 1% of the range of $X2$ in $X2$.

item-specific biases arising due to container thickness variations and other effects,) but less than the 1.81% reported from empirical (top-down) UQ of the UO_2 drums example by Walsh et al. [48].

Figure 3: plots the average residual versus the true enrichment in fitting Y as a function of X_1 and X_2 (Eq. (5)). Because 10^5 simulations were used, simulation error is negligible [49]. Figure 3: (b) is an example of a simulation-based bottom-up prediction of uncertainty due to calibration errors. The caption of Figure 3: lists the data and measurement error standard deviations in Y, X_1, and X_2 in training (calibration) and testing, which can be modified to mimic the effect of varying container thickness, with or without an adjustment for container thickness being different in training than in testing items.

Burr et al. [50] compare simulation-based UQ to analytical approximations of UQ for calibration data. If the operator uses some other method, such as DA, then the operator's DA measurement can be assessed using separate simulation.

Discussion

Generally in NDA applications, items emit neutrons and/or gamma-rays that provide information about the source

material, such as isotopic content. However, item properties such as density, or the distribution of radiation-absorbing isotopes, which relate to neutron and/or gamma absorption behaviour of the item, can partially obscure the relation between detected radiation and the source material; this adds a source of uncertainty to the estimated amount of SNM (Special Nuclear Material) in the item. One can express item-specific impacts on uncertainty using a model such as

$$CR/M = g(X_1, X_2, ..., X_N),$$ (6)

where CR is the item's neutron or gamma count rate, M is the item SNM mass, g is a known function, and $X_1, X_2, ..., X_N$ are N auxiliary predictor variables such as item density, source SNM heterogeneity, and container thickness, which will generally be estimated or measured with error and so are regarded as random variables. To map Eq. (4), to GUM's Eq. (2), write

$$M = CR / g(X_1, X_2, ..., X_N) = h(X_1, X_2, ..., X_M)$$ (7)

where the measured CR is now among the $M = N+1$ inputs. Note that Eq. (6) is the same as Eq. (1), but some of the X_i account for item-specific departures from reference items used for calibration.

Top-down UQ used in MBE estimates the random and short-term systematic standard deviations σ_R and σ_S, which are estimated from data sets that have items measured by each of two or more assay methods. The net random error can include variation in background that cannot be perfectly adjusted for, Poisson counting statistics effects, item-specific biases, and other random effects. In principle, each of $X_1, X_2, ..., X_N$ could be estimated for each item as part of the assay protocol. However, there would still be modelling error because the function f must be chosen or somehow inferred, possibly using purely empirical data mining applied to calibration data [51] or physics-based radiation transport codes. Typically, only some of $X_1, X_2, ..., X_N$ will be measured as part of the assay protocol. Most assay methods rely on a calibration step [52]; as mentioned, calibration is not fully addressed in the GUM [53, 47, 54, 44, 45, 46] but one GUM supplement in progress will address calibration; and, the GUM is currently being revised to include more detail on calibration [47, 53].

4. 'Bridging the gap' - Reconciliation of terminological and methodological differences

Common or mapped terminology and a common mathematical basis are prerequisites for any reconciliation process. The limited understanding between the laboratory and evaluator communities has a mathematical/logical component and also a paradigmatic component. This article aims to establish a common language between the communities by translating and mapping terms used by the two communities. 'Mapping' means here to list recurring perceived differences stemming from semantics, including a sociological component in using terminology in a particular manner within a community, ('apparent differences' – Table 1) as well as differences originating from the application of metrological/statistical concepts ('conceptual differences' – Table 2). The first can lead to 'bridging the gap' towards harmonisation of terminology and formalism between communities. The latter can lead to 'bridging the gap' towards mutual understanding, complementarity, and convergence, see sections 6, and 7, while maintaining those differences of approach that are rooted in different requirements in the problem domain. Wherever necessary, additional notes and references to publications or other sections in this article are in the third column of the Tables. 'Reconciled' in Table 1 means that there is a consensus between the laboratory and evaluator communities about the listed differences not being conceptual but rooted in terminology. Specific terms or concepts that cause recurring misunderstandings between the communities are discussed in more detail see section 4.2. Although changes as a result of the on-going revision of the GUM are not anticipated and beyond the scope of this article, the notes to some of the entries in the mapping tables are recommending a revised version of the GUM when deemed necessary [47].

1. Terminology - Apparent Differences

Terminology used by evaluators in safeguards	Terminology used by laboratories in safeguards	Notes
Observable, measurand	Measurand - quantity intended to be measured [VIM 2.3], analyte	[39]
Measured value	Quantity value representing a measurement result [VIM 2.10]	Measurement results are quantitative probabilistic statements on the measurand.
Estimate of the true value of the measurand, associated with measurement errors (random, short-term systematic, bias)	Measurement result associated with an interval of reasonable values of the measurand; best estimate of the measurand, along with an associated measurement uncertainty [VIM 2.19]	The true value of the measurand is a fixed and unknowable constant; the result of the measurement of the measurand can be quantified. The concept of true value is inseparable from the definition of the particular quantity to be measured, see section 4.2
Measurement error standard deviation	Standard measurement uncertainty [VIM 2.26]	The expression "measurement error" may be wrongly used instead of measurement error standard deviation, see section 4.2. The absolute error standard deviation is usually denoted σ.
Relative measurement error standard deviation (RSD)	Relative standard deviation (RSD)	RSD denotes a relative standard deviation (δ), i.e. the standard deviation divided by the absolute value of the mean.
Standard deviation associated with the value of a standard	Combined standard uncertainty of a reference value [VIM 5.18]	Error of a standard may be wrongly used as a synonym of uncertainty.
Total measurement error standard deviation, propagated measurement error standard deviation	Combined standard uncertainty [VIM 2.31]	
Confidence interval: a range of values that contains the true (unknown) value of a parameter, e.g. the measurand, with a given probability referred to as the confidence level; (adopting the frequentist view that in a collection of such intervals the percentage that contain the true value of the measurand should tend toward the stated confidence level as their number increases)	Coverage factor k: a multiplication factor defining the width of the coverage interval of reasonable values of the measurand. The choice of k depends on the level of confidence required for the measurement result, usually expressed as $\hat{y} \pm k u_c(\hat{y})$ k is that value satisfying the probability statement in Eq. (G.1a) in JCGM 100 and $u_c(\hat{y})$ is an estimate of the standard deviation of $(\hat{y})$	Confidence level is isomorphic but not equivalent to coverage factor: k defines an interval corresponding to a certain confidence level, see JCGM 100:2008 G 6.1 [55].
Total error standard deviation	Expanded uncertainty U	
Consistency of estimates: the difference between the values of these estimates is smaller than a multiple of the standard deviation of the difference. The level of consistency is related to the chosen multiple of the standard deviation	Metrological compatibility of measurement results: the difference between two measurement results is smaller than the expanded uncertainty of their difference [VIM 2.47], see sction 4.2	

Table 1: Mapping of 'reconciled' terminological differences

4.1 Differences in approaches

Approaches used by evalua-tors in safeguards	Approaches used by laboratories in safeguards	Difference in approach
Principal objective of uncertainty quantification: to determine the significance of observed differences between two independent determinations of a quantity or combined quantities through statistical error propagation.	Principal objective of uncertainty quantification: to completely qualify a single measurement result	It is recognized by both groups that a complete expression of a measurement result consists of a quantity value, a statement of its uncertainty and a metrological traceability statement [VIM 2.41]. However, safeguards data evaluation typically deals with algebraic combinations of measurements such as differences, not with single measurements.
Statistical measurement error model	Measurement model - rule for converting a quantity value into the corresponding value of the measurand [JCGM 104:2009, 3.10]	Measurement results are quantitative probabilistic statements on the measurand.
The preferred error model allows for long-term systematic error (=bias), short-term systematic error and random error. Measurement error model (simplified) $$Y = \mu + b + R + S$$ Y denotes the measured value μ: denotes the true (but unknowable) value b: denotes a bias (long-term systematic error) R: denotes a random error of expectation E(R)=0 and of standard deviation denoted $\sigma_{Y,R}$ S: denotes a short term systematic error of expectation E(S)=0 and of standard deviation denoted $\sigma_{Y,S}$ The total uncertainty (standard deviation) associated to Y is given by: $$\sigma_Y = \sqrt{\sigma_{Y,R}^2 + \sigma_{Y,S}^2}$$	The preferred error model allows for Type A errors (can be reduced by repetition of measurements) and for Type B errors (can be reduced by other means). In the expression of uncertainty the effects of both types of error are combined Measurement model (simplified) $$Y = \hat{y} \pm k u_c(\hat{y})$$ Y denotes the measurand $\hat{y}$ denotes the estimate of the measurand Y; Y includes a correction factor taking into account the measurement bias (the measurement model does not distinguish between bias and short-term systematic error) k: denotes a coverage factor. $u_c(\hat{y})$ denotes the estimate of the combined standard uncertainty $u_c(\hat{y})$ (Type A, Type B uncertainties propagated from all input quantities in $\hat{y}$ including the standard uncertainty u_b associated to the correction factor for bias)	The terms in the evaluator's error model are introduced to describe: S: a fluctuating error component (of random nature) often seen in the data, for example between inspections or between calibrations, which is superposed to random fluctuations between individual measurements and b: a possible long-term bias, which is not of random nature.
When combining standard deviations, random and short-term systematic components are propagated differently.	The uncertainty being related to a single value, the random and systematic components are combined in a single uncertainty estimate	SG data analysis must take the different behaviour of measurement error components through combination of measurement results into account, whereas labs deliver a single measurement result and strive to minimize any short-term systematic component.
The main method of uncertainty quantification is analysis of variance based on paired (or multiple) data from independent measurement methods	Standard methods of uncertainty quantification are repetition under controlled conditions and quality control (QC) with certified standard materials.	Labs can control the measurement conditions, can perform as many repetitions as needed and have certified reference materials available. Safeguards data evaluators on the other hand analyze operator and inspector measurements of the same items performed in conditions that they do not control.

Table 2: Mapping of conceptual differences

4.2 Discussion of recurrent terms of misunderstanding

True quantity value

The primary objective of the safeguards evaluators, approach is not to estimate the true value of a measurand but to estimate random and systematic error variances by means of ANOVA applied to operator-inspector differences as described in section 5 and in [48] or by propagation of known operator's and inspector's measurement uncertainties. The objective of a measurement following the GUM's bottom-up approach as understood by the laboratory community, (see section 3), is neither to determine a true value nor to produce separate estimates of random and systematic error from paired data, but rather to determine an interval of reasonable values of the measurand, based on the assumption that no mistakes have been made in performing the measurement [39] and that the measurement conditions have been adequately controlled. In terms of reconciliation, both approaches rely on the concept of a true value to make a measurement meaningful and characterize its performance ([56], GUM D.3.5). In that sense the GUM's measurement objective is to establish an interval of values within which the true value of the measurand (with sufficiently small intrinsic uncertainty, GUM D.3.4) is believed to lie, with a given degree of belief, based on the available information from the measurement and possibly possibly from other sources [48,57].

Metrological compatibility

Metrological compatibility of measurement results is the property of a set of measurement results for a specified measurand, for which the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than a chosen multiple of the standard measurement uncertainty of that difference [VIM 2.47,58]. Safeguards evaluation addresses a similar but different problem domain concerned with the evaluation of mass differences to determine if they are explicable by measurement uncertainties, considering detection probabilities of nuclear material diversion and the risk of false alarm. Note: the IAEA (at present) relies mostly on top-down UQ based on ANOVA independently from any uncertainties reported by the laboratories, to which the top-down uncertainty estimates can be compared to identify the existence of sources of uncertainty outside of or unaccounted for by the laboratories. On the other hand, the EC makes use of both operator's and inspector's measurement results, including the reported and validated uncertainties, to build a bottom-up uncertainty budget.

Measurment trueness

Measurement trueness is not a quantity and thus cannot be expressed numerically, but a 'trueness check' is part of a laboratory's method validation [39]. This means to compare the measured value of a measurand associated to a certified (matrix) reference material (x_m) to its certified value (x_{CRM}) and to assess their metrological compatibility in order to exclude any significant bias (Δ_m).

$$\Delta_m = |x_m - x_{CRM}| \tag{8}$$

The standard measurement uncertainty for Δ_m is given by:

$$u_\Delta = \sqrt{u_m^2 + u_{CRM}^2} \tag{9}$$

If $abs(\Delta_m) \leq 2u_\Delta$, there is no evidence that the measured and the certified value are incompatible (a hypothesis test that the bias is zero would not be rejected at the 0.05 significance level). Thus, there is no significant bias, no correction is needed and u_m is is used in subsequent data evaluation. If there is a significant difference, the laboratory preferably improves the method or, in case this is not possible must correct the measurement model for the quantified bias and propagate the uncertainty introduced by the correction.

Systematic errors and measurement bias.

A laboratory can tailor its effort depending on the available resources to do method validation, including performing measurements under repeatability and reproducibility conditions, and to check for trueness, aiming to establish reasonable combined measurement uncertainties to provide a fit-for-purpose measurement result. In safeguards verification, opportunities to perform measurements under repeatability and reproducibility conditions are severely limited by inspection schedules and practicalities - see sections 2 and 4. In addition, the safeguards authority does not determine the measurement procedures used by the operators. Therefore, one source of misunderstanding between the two communities lies in the following:

For the evaluator community, the terms 'bias' (or long-term systematic error), 'short-term systematic error', and 'random error', are integral parts of the statistical model of measurement error. When dealing with operator-inspector differences, many sources of error remain unknown. Thus one must allow for the presence of both bias and short-term systematic error. In MBE, a short-term systematic error is a random variable with expectation zero that is constant for a group of measurements (for example, a group can be a time period such as a 1-week inspection period each year) and is a component of the total error that cannot be reduced by averaging over all measurements in a group. The average (expectation) of short term systematic errors observed over a long period (a large number of shorter periods each corresponding to one systematic error observation) tends to zero.

Although the laboratory community uses a similar terminology, there are clear differences in approaches, particularly also because the GUM-based ITVs-2010 are values for

uncertainties associated with a single determination result [3]. In the measurement model as described in GUM there is no notion of 'time' or measurement group. From the laboratory viewpoint, a measurement bias is not related to a timeframe (short term – long term). In GUM a measurement bias is stated to be an estimate of a systematic measurement error; however, we anticipate that the next version of GUM will define a bias to be a true unknown quantity, not an estimate [59]. But regardless, a reference quantity value is required to quantify bias, and if the bias is significant, a correction factor with a combined standard uncertainty can be applied to the measurement to take this bias into account. Both the laboratory and MBE communities accept the possibility of performing bias adjustments; however, the laboratory preferred method if the bias is statistically significant is to return to first principles and remove or reduce the source of bias. Depending on the MBP, which is commonly one year, one can consider that the assessment by a laboratory of a bias by means of a QC chart based on a certified reference material [38] can correspond to a short term or long term systematic error according to MBE terminology. In MBE, the term bias is used to denote a long-term systematic error, a fixed effect, not modelled as a random variable, to distinguish it from the short-term systematic error, which is modelled as a random variable fluctuating with the measurement conditions. A method is unbiased if the long-term systematic error is zero. Short-term systematic and random errors always exist and propagate differently. This is a clear example where the same term 'bias' is used by the two communities but with different meanings, causing misconceptions and misunderstanding because it is related to different effects. The safeguards statistical data evaluators partition error variance into random and short-term systematic in their approach to assess whether a bias is significant. From a pure measurement point of view, a long term systematic error can only be assessed via a series of measurements during a certain time frame [25]. This approach cannot be easily implemented in MBE because metrological conditions can change across balance periods. However, laboratories can demonstrate long term measurement performance via regular participation in ILCs with independent and traceable reference values in compliance with ISO 13528:2005 [60, 61 ,62 ,63]. A recurring measurement bias of an operator or safeguards laboratory could be translated into an indication for a long-term systematic error in MBE [4, 61]. In the case of an operator measurement bias, this could also be interpreted as an indicaton of possible diversion.

5. The statistical basis of different approaches to quantification of measurement uncertainty

Reconciling GUM-based UQ and UQ via the IAEA error model empirical estimation was approached by reviewing the design basis and corresponding mathematical/ statistical formalisms of each. The full scope of this investigation can be found in [48].

Empirical approaches to UQ, such as estimation of variance components by an appropriate ANOVA, are applied in metrology to estimate a specified error variance parameter of a measurement method. When estimating variance parameters in an empirical approach, the precision conditions under which the data are collected must be clearly specified – this includes statements regarding the degree to which the sample replicates are true measurement replicates, as well as an acknowledgement of what measurement conditions may have changed when measuring the set of items (e.g. day, analyst, calibration, instrument, etc.). The statistical model and corresponding estimation approach in conjunction with the conditions under which the replicate measurements were collected imply how the resulting variance estimates are to be interpreted and used in subsequent UQ exercises.

Ideally, the term 'top-down' uncertainty should only be used in conjunction with the specific empirical approach of reproducibility studies that deal with measurement replication across many participant analytical systems, thereby covering a wide range of varying environmental conditions (this is described in ISO 21748 [64]). Therefore 'top-down' UQ involves explicit estimation of the reproducibility standard deviation as defined in ISO 5725 [65]. Estimates of variance obtained in reproducibility studies comprise a theoretically and empirically justified benchmarking of an important component of true uncertainty in a measurement method – i.e. the reproducibility standard deviation is the primary empirically derived parameter for estimating the uncertainty of a measurement method. Because of this, estimates of the reproducibility standard deviation are used to assess the correctness of an analytical laboratory's uncertainty evaluated for a measurement method.

The current 'best practices' approach for UQ of analytical methods is the GUM, JCGM:100 2008 [12]. The GUM is often referred to as a 'bottom-up' approach. The measurement method is described by a model equation where all input quantities comprising the final measurement result are stated. Each input quantity is assigned an uncertainty either through experimentation and appealing to the appropriate estimation procedure and often application of ANOVA or variance components estimation (this is referred to as Type A evaluation), or via other sources including expert knowledge, published data, reference material certificates, physics based limits, etc. (this is referred to as Type B evaluation).

The IAEA The IAEA uses many similar methods for MBE. The same statistical approaches are appealed to (most notably variance components estimation by ANOVA). The fundamental error model assumed for UQ of measurements taken for safeguards purposes includes variances

accounting for product variability, and also random and systematic error variances. The systematic error variance is historically modelled to represent the aggregate 'between inspection' shifts which can be due to many factors, including: changes in calibration, inspectors, background, and any other effects. The random error variance has been demonstrated to be the combination of pure random error (variance due to the repeatability of the measurement method) plus item specific bias because test samples are not true replicates (they are different sampled items from the facility).

Walsh et al [48] studied in detail one approach among an ensemble used by the IAEA to produce ITVs and uncertainty estimates for use in MBE (Grubbs' ANOVA applied to paired (operator, inspector) verification data obtained over multiple inspection periods) and revealed that the estimate of random error standard deviation can be almost interpreted as the inverse of method repeatability precision as defined in the international vocabulary of metrology, except for being larger by item-specific bias because test samples are not true replicates. The short-term systematic error variance estimate can be used in error propagation for MBE. MBE requires separately 4 variance components, i.e. the random and systematic error variance estimates of the operator and inspector measurement systems since (1) standard assumptions of the variance components imply that the random and systematic error variances propagate differently through an MBE calculation and (2) MBE comprises three statistical evaluations: operator's MUF, the D Statistics, and the *Inspector's estimate of Material Unaccounted For* (IMUF).

6. Complementarity of Approaches

Integrating competences across academic disciplines is a creative approach to reaching effective solutions. Going beyond disciplinarity might be remedial to problematic epistemological and political effects of excessive specialization. Therefore, disciplinary and interdisciplinary approaches should not be seen as mutually exclusive, but possibly complementary.

As discussed above, the authors are seeking reconciliation of terminological and conceptual differences in nuclear safeguards quantification of measurement uncertainties. The disciplinary UQ approaches involved are those of the measurement laboratories, metrology institutes, nuclear operators, and safeguards evaluators. These fields originally developed their approach for different specific purposes, and today the need for reconciliation is addressed.

The discussion over conceptual and terminological differences in section 4 showed that a language-mapping table could improve the understanding between communities adopting different approaches. Although the two main ways (top-down and bottom-up) of estimating uncertainty

in measurements are not contradictory, and the GUM and the standard statistical error theory are consistent in the probabilistic UQ modelling, in a number of areas they differ substantially.

A review of statistical models and computational methods [57] highlighted the value they bring to the evaluation of measurement uncertainty. A number of problems still beyond the reach of GUM such as some aspects of calibration uncertainty, multi-dimensional absorption spectra, ILCs, and attribute testing, can be addressed using top-down UQ by long-standing observation equations (measurement error models) and statistical analysis [10, 55]. The bottom-up GUM approach is therefore complemented, rather than contradicted, by top-down statistical models and associated ANOVA-based variance component estimation and by Monte Carlo methods.

Contributors[6] influential to the GUM revision, propose to regard a measurement result as a *degree of belief* probability distribution for the measurand. Descending from the pattern of dispersion of values, as well as from uncertainties estimated by expert judgement [64], the probability distribution reveals the true character of the measurement uncertainty. The distribution could then be represented, for communication purposes only, by simpler summaries, such as the mean, mode, or others, and the standard deviation.

Multidisciplinary approaches to UQ thus complement disciplinary ones, introducing elements of elicitation and prior knowledge to the distribution of measurement results. This way of thinking is formalised in the Bayesian approach to inference, and it is identified as an area of potential expansion of GUM, both for bottom-up and top-down UQ, to address the challenges that measurement science will be facing in the years to come [47, 42].

In this article, deductive and inductive logical processes are addressed respectively in the top-down and bottom-up approaches. A deductive approach to processing information focuses to the most general first, and then narrows it down to the more specific. Conversely, inductive reasoning starts with specific observations of input quantities and then broadens the concepts up to generalisations and theories. It has to be noted that statistical methods stemming from inductive reasoning are intrinsic to both, the top-down as well as the bottom-up approach [66].

In practical terms, both statistical inference and probability theory are used in the metrological approach to UQ. Nevertheless, the input quantities that form the basis of the error model will drive the determination of the overall uncertainty. To illustrate this, the uncertainties associated with UO_2 drums measurements reported in section 3 are quantified focusing on the individual input quantities X_i from Eq. 1.

---

6 International conference that celebrated the twentieth anniversary of the GUM publication.

The uncertainties associated to the Xi are then propagated bottom-up to estimate the overall uncertainty on Y. Conversely, the same example is presented in section 5 by Walsh et al [48]. concentrating on the performance of the complete method. The reproducibility standard deviation, equivalent to the overall uncertainty on Y [48], is derived in top-down fashion by estimators applied to the ANOVA on paired measurement results.

As mentioned in section 4.3, MBE requires independent estimates of random and systematic components for the measurement uncertainties affecting the material balance [48]. Both bottom-up and top-down approaches can provide these estimates, but not without specific weaknesses, pointing to the possible advantage of a hybrid approach combining the two UQ methods.

1. The bottom-up approach does not necessarily model variation in all the effects influencing the measurement result [33, 34]. For example, in the UO_2 drums example that uses the EMP, variation in drum container thickness and self-absorption due to elemental matrix and its density are not fully accounted for. As discussed in section 3, calibration items differ from measured items, because calibrations could only be performed using drums containing reference material distributed differently than in measured drums. Hence, factoring expert knowledge into Monte Carlo simulations proved to be remedial. Failing to identify significant variation of input quantities could lead to uncertainty underestimation, which in safeguards terms translates into unnecessarily high false alarm rates.

2. The top-down approach assumes that all the variances associated with the input of the mathematical model vary representatively across the reproducibility study. However, variations associated with item-specific features and/or spectrum background cannot be perfectly accounted for based on the measurement method only, and would benefit from the expert judgement of an analyst to assess their impact on the overall measurement process. Failing to ensure representative variations in the course of the 3-year exercise discussed in section 5 [48], has the potential to lead to uncertainty overestimation. Thus, uncertainties potentially tuned to conceal nuclear material diversion could be deemed acceptable by the safeguards evaluators.

Comparing the two estimates discussed in section 3 and 5 is useful at this stage to assess the completeness of the UO^2-drums measurement model. The bottom-up UQ random RSD values cited in section 3 range from less than its 0.5% target to more than 1.0%, but less than the 1.81% RSD calculated in the top-down UQ approach applied in Section 5 [48]. For practical uncertainty estimates, therefore, it is recommended to appropriately use elements of both methods, in a hybrid, interdisciplinary way of thinking.

7. Benefits of Convergence

Professionals in Safeguards work in applied science, where scientific methods are developed to solve specific problems effectively, and then operationally optimized to make them fit for efficient production (be it performing measurements, analysing samples or evaluating data). Development and rehearsal of patterns of thought and their associated notation, terminology and jargon is part of the optimization process. Other than in the realm of pure science, basic assumptions and theories underlying the practically applied methods are not continuously questioned. Professional exchanges tend to be with experts in one's own field, who "speak the same language" and follow the same thought patterns. Discussions with experts in adjacent fields tend to remain at a shallower level, because of a lack of adequate understanding of each other's problem space, preferred solutions and accepted terminology.

Nevertheless, adjacent professional groups, such as staff at an analytical laboratory and staff of a statistical data evaluation group, can successfully collaborate under the premise of mutual recognition of expertise and as long as organizational and technical interfaces (such as distribution of responsibilities for the various process steps, and data exchange formats) are well-defined and respected. Seeking deeper understanding inevitably costs additional effort and may create insecurity and friction, as longstanding practices are being examined and criticised by knowledgable outsiders.

There are, however, at least two weighty reasons for why it is worthwhile to make the effort required to understand one's professional neighbours and to make oneself understood:

The first reason is following a broad and accelerating societal trend: authority, including professional authority, no longer goes unquestioned. Institutional status, educational credentials and a slightly aloof attitude do not bestow the expert with credibility. Credibility flows from an openness to review and the willingness to explain. And the capability to review and receptiveness to learning is easiest to find in adjacent professional communities. For this reason, no group of specialized experts can nowadays afford to not reach out to their neighbours.

The second reason is the opportunity to improve one's own approaches and practices by

- accepting, seriously considering and where useful incorporating constructive criticism, of which adjacent professional groups are uniquely capable; and

- enriching one's methodological portfolio by testing and adopting methods developed by adjacent professional groups.

As an example we consider potential benefits from convergence between the bottom-up approach and the top-down approach to measurement uncertainty estimation within the framework of the evaluation of the material balance of a bulk-handling nuclear facility, for example a fuel fabrication plant or an enrichment plant. A correct and credible assessment of measurement uncertainties is critical, because the aggregated and propagated uncertainties determine the variances of the fundamental statistics MUF, D and IMUF. While the expected value of these measures is zero, their actual values for each material balance period are non-zero and the crucial question to be answered in the evaluation of the facility is, whether the deviation from zero can plausibly, i.e. with reasonable confidence, be explained by legitimate measurement errors. Should the answer be no, alternative explanations, including the possible diversion of nuclear material, would need to be considered.

Bottom-up strives to understand all sources of uncertainty from first principles, exact knowledge of measurement practices and metrological traceability, see section 4.

Top-down uses statistical analysis to estimate and allocate uncertainties from paired data analysis, see sections 4 and 5.

Bottom-up analysis usually understates the variances of the MUF, D and IMUF statistics, as only known causes of uncertainty are within scope of the analysis. Uncertainty arising from unknown causes is itself, however, far from being an unknown phenomenon; it has been termed "dark uncertainty" [33]. Such unknown causes can, for example, be sampling errors due to material heterogeneity, chemical changes to the material over time, human mistakes or uncontrollable measurement conditions. Alique et al.[67] have presented a bottom-up methodology for estimation of sigma-MUF; in his example the bottom-up sigma-MUF is a factor of 71 smaller than sigma-MUF based on the ITV 2010 [3]. In the evaluation of nuclear facilities, dark uncertainty can be a large fraction of total uncertainty, and a decision criterion built exclusively on bottom-up uncertainty will tend to result in unacceptably high false alarm rates and can be perceived as unrealistic.

Top-down analysis provides a more realistic approach as, by construction, it takes into account all sources of uncertainty. However this should not lead to accepting (purposefully or not) poor performance, which would decrease detection probability. Therefore, a top-down -approach to UQ must be combined with a comparison with ITVs and a close monitoring of trends. The causes for significant changes must be investigated by obtaining additional information about measurement conditions and procedures.

So joining forces is an attractive proposition: "Bottom-uppers" can use paired data analysis and three-lab analysis based on Grubbs [21] and subsequently improved methods to allocate uncertainty components, and to quantify

dark uncertainty not yet covered in their uncertainty budgets. "Top-downers" can feed prior knowledge on the uncertainty of certain measurement methods into their analysis and use the method of uncertainty-budgeting to identify dark uncertainty. A jointly derived decision criterion would have a good chance of striking a better balance between the twin risks of non-detection and false alarms. Subsequently, both groups can collaborate in characterizing and reducing the existing unknown sources of uncertainty and thereby increase the effectiveness of safeguards. Motivated by this perspective, the mathematical equivalence of paired data analysis as practiced by IAEA with GUM-based methods has recently been demonstrated [48].

The dialogue between metrologists, statisticians, mathematicians and laboratory professionals is in full bloom today [55], and this attempt to foster a mutual understanding between the laboratory and evaluator community in safeguards is believed to be of interest also to other measurement and evaluation disciplines [68, 69, 70]. It is even considered as a potential contribution to the ongoing process of the GUM revision.

It is the authors' wish that our article should motivate and facilitate this kind of fruitful collaboration.

8. Acknowledgements

9. References

[1] NUREG/CR-4604,Statistical Methods for Nuclear Material Management, PNL-5849, December 1988

[2] IAEA Safeguards Glossary 2001, Edition International Nuclear Verification Series No.3

[3] IAEA-STR-368 International Target Values for Measurement Uncertainties in Safeguarding Nuclear Materials, Vienna; November 2010

[4] Aregbe Y, Jakopič R, Richter S, Venchiarutti C; Conformity assessment in nuclear material and environmental sample analysis; Proceedings Symposium on International Safeguards Linking Strategy, Implementation and People; 2014; S13–01

[5] IAEA-TECDOC-261; IAEA Safeguards Technical Manual Part F: Statistical Concepts and Technique, Vol.3; Vienna; 1982

[6] Jaech J; Statistical Methods in Nuclear Material Control, TID-26298; National Technical Information Service, U. S. Department of Commerce, Springfield, Virginia, USA;1973

[7] IAEA/SG/SCT/5, Statistical Concepts and Techniques for IAEA Safeguards, Fifth Edition; IAEA, Vienna;1998

[8] Martin K; IAEA Technical Report Estimation Of Variances Of Random And Short-Term Systematic; Measurement Errors Based On Data From Two And Three Independent Measurement Methods; unpublished manuscript; 2016

[9] BIPM 1980; Report of the BIPM Working Group on the Statement of Uncertainties to the Comite International des Poids et Mesures; www.bipm.org/utils/common/pdf/WGUncertainties1980.pdf

[10] Cox M, Harris P; GUM anniversary issue; Metrologia 51; 2014; S141–S143

[11] ISO/IEC Guide 98:1993; Guide to the expression of uncertainty in measurement (GUM)

[12] ISO/IEC Guide 98-3:2008; Guide to the expression of uncertainty in measurement (GUM)

[13] ESARDA 1979; Target Values" for uncertainty components in DA methods presented to Euratom and IAEA

[14] ESARDA 1983; Target values for uncertainty components in fissile element and isotope assay – achievable uncertainties in destructive assay of nuclear material, 6th ESARDA Symposium, Venezia, Italy, May 14-18, 1984

[15] De Bièvre P, Baumann S, Gorgenyi T, Kuhn E, Deron S, Dalton J, Perrin R E, Pietri C, De Regge P; Target values for uncertainty components in fissile isotope and element assay; Journal of the Institute of Nuclear Materials Management; 15(4); 1987; p. 99-104

[16] Kuhn, E., et al.,1993 International target Values for Uncertainty Components in Fissile Isotope and Element Accountancy for the Effective Safeguarding of Nuclear Materials; IAEA STR-294; Vienna; 1994.

[17] International Target Values 2000 for Measurement Uncertainties in Safeguarding Nuclear Materials; IAEA-SM-367/5/01

[18] https://esarda.jrc.ec.europa.eu/index.php?option=com_content&view=article&id=242:report-on-the-workshop-on-reference-material-needs-and-evaluation-of-uncertainties-in-da-and-nda&catid=91&Itemid=331

[19] IAEA 2013 Meeting Minutes, IAEA International Technical Meeting on Statistical Methodologies for Safeguards; 16-18 October 2013, Vienna

[20] 36th ESARDA Annual Meeting 2014; Report on the ESARDA Joint Workshop 'Applied Metrology & Material Balance Evaluation'; Luxembourg 12 May 2014

[21] Grubbs F E; On Estimating Precision of Measuring Instruments and Product Variability, Journal of the American Statistical Association 43; 1948; 243-264

[22] ISO/IEC 17025:2017; General requirements for the competence of testing and calibration laboratories, International Organization for Standardization; Geneva

[23] ISO/IEC 17043: 2010; Conformity assessment – general requirements for proficiency testing, International Organization for Standardization, Geneva

[24] ISO Guide 34:2009; General requirements for the competence of reference material producers, International Organization for Standardization, Geneva

[25] ISO 17034:2016, General requirements for the competence of reference material producers, International Organization for Standardization, Geneva

[26] Bürger S, Essex R M, Mathew K J, Richter S, Thomas R B; Implementation of Guide to the expression of Uncertainty in Measurement (GUM) to multi-collector TIMS uranium isotope ratio metrology; 2010 International Journal of Mass Spectrometry 294 65–76

[27] Richter S, Venchiarutti C, Hennessy C, Jacobsson U, Bujak R, Truyens J, Aregbe Y; Preparation and certification of the uranium nitrate solution reference materials series IRMM-2019 to IRMM-2029 for the isotopic composition; Journal of Radioanalytical and Nuclear Chemistry 318:1359–1368; 2018

[28] ASTM C1832 – 16 2016 Standard Test Method for Determination of Uranium Isotopic Composition by the Modified Total Evaporation (MTE) Method Using a Thermal Ionization Mass Spectrometer

[29] Richter S, Alonso A, Truyens J, Kühn H, Verbruggen A, Wellum R; REIMEP-18: Interlaboratory Comparison for the Measurement of Uranium Isotopic Rations in Nitric Acid Solution; Report EUR 22529; 2006

[30] Jakopic R, Bujak R, Aregbe Y, Richter S, Buda R, Zuleger E; REIMEP-17: plutonium and uranium amount content, and isotope amount ratios in synthetic input solution; Report EUR 26667 EN; 2014

[31] Jakopic R,Bujak R, Aregbe Y, Richter S, Buda R, Zuleger E; Results of the REIMEP-17 interlaboratory comparison for the measurement of the U and Pu amount content and isotope amount ratios in the synthetic dissolved spent nuclear fuel solution; Accred Qual Assur 20; 2015; 421–429

[32] Mason P; An Evaluation of Uranium Measurement Capabilities and Comparison to State-of-the-Practice Target Values, including an Examination of Historical Performance; NBL-2011-ME Annual Report; 2011

[33] Thompson M, Ellison S L R, Dark uncertainty; Accred Qual Assur 16; 2011; 483–487

[34] Walsh S J, Venzin A, Wegrzynek D, Mansoux C; Using reproducibility to test the adequacy of GUM based uncertainty quantification; American Nuclear Society Advances in Nuclear Non-proliferation Technology & Policy Conference 25-30 September, 2016, Santa Fe NM

[35] Pendrill L R; Using measurement uncertainty in decision-making and conformity assessment; Metrologia 51; S206–S218; 2014

[36] Clark D L, Geeson D A, Hanrahan R J; Plutonium Handbook, 2nd edition, Volume 4; American Nuclear Society; 2019

[37] ASTM C1672 – 17; Standard Test Method for Determination of Uranium or Plutonium Isotopic Composition or Concentration by the Total Evaporation Method Using a Thermal Ionization Mass Spectrometer; 2017

[38] Vogl J; Characterisation of reference materials by isotope dilution mass spectrometry; Journal of Analytical Atomic Spectrometry 22; 2007; 475–492

[39] JCGM 200:2012 International Vocabulary of Metrology – Basic and General Concepts and Associated Terms: https://jcgm.bipm.org/vim/en/index.html

[40] Burr T, Sampson T, Vo D; Statistical evaluation of FRAM γ-ray isotopic analysis data Applied Radiation and Isotopes 62; 2005; 931-940

[41] Bonner E, Burr T, Guzzardo T, Krieger T, Norman C, Zhao K, Beddingfield D H, Lee T, Laughter M, Geist W; Ensuring the Effectiveness of Safeguards through Comprehensive Uncertainty Quantification, Journal of Nuclear Materials Management 44(2), 2016; 53-61

[42] Burr T, Krieger T, Norman C; Approximate Bayesian Computation Applied to Nuclear Safeguards Metrology ESARDA BULLETIN No. 57; December 2018

[43] ASTM C1514; Standard Test Method for Measurement of 235U Fraction Using the Enrichment Meter Principle; 2008

[44] Burr T, Croft S, Jarman K, Uncertainty Quantification in Application of the Enrichment Meter Principle for Nondestructive Assay of Special Nuclear Material; Journal of Sensors 15, Article ID 267462; 2015

[45] Burr T, Croft S, Krieger T, Martin K, Norman C, Walsh S, Uncertainty Quantification for Radiation Measurements: Bottom-up Error Variance Estimation using Calibration Information; Applied Radiation and Isotopes 108; 2016; 49-57

[46] Burr T, Croft S, Dale D, Favalli A, Weaver B, Williams B.; Emerging Applications of Bottom-Up Uncertainty Quantification in Nondestructive Assay; ESARDA Bulletin 53; 2015; 54-61

[47] Elster C; Bayesian uncertainty analysis compared with the application of the GUM and its supplements; Metrologia 51; 2014; S159–S166

[48] Walsh, S, Burr T, Martin K; Discussion of the IAEA Error Approach to Producing Variance Estimates 4 for Use in Material Balance Evaluation and the International Target Values, and Comparison to Metrological Definitions of Precision; Journal of Nuclear Materials Management; Volume XLV; 2017; 4-14

[49] R Core Team; R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018; Available online at https://www.R-project.org/

[50] Burr T, Croft S, Jarman K, Nicholson A, Norman C, Walsh S; Improved uncertainty quantification in nondestructive assay for nonproliferation Chemometrics and Intelligent Laboratory Systems; 2016; 159 164-173

[51] Burr T, Pickrell M, Rinard P, Wenz T; Data mining: applications to nondestructive assay data. Journal of Nuclear Materials Management 27(2); 1999; 40-47

[52] Dufour J L, Pepin N, Deyglun, Weber A-L; Optimisation and uncertainty estimation of the enrichment meter measurement technique for UF6 cylinders; ESARDA BULLETIN No. 59; December 2019

[53] Bich W; Revision of the 'Guide to the Expression of Uncertainty in Measurement'. Why and how; Metrologia 51; 2014; S155–S158

[54] Bonner E,. Burr T, Krieger T, Martin K, Norman C; Comprehensive Uncertainty Quantification in Nuclear Safeguards Science and Technology of Nuclear Installations Hindawi Science and Technology of Nuclear Installations; Volume 2017; Article ID 2679243; 2017;

[55] Possolo A; Statistical models and computation to evaluate measurement uncertainty; Metrologia 51; 2014; S228–S236

[56] Ehrlich C; Terminological aspects of the Guide to the Expression of Uncertainty in Measurement (GUM); Metrologia 51; 2014; S145–S154

[57] Hagan A O; Eliciting and using expert knowledge in metrology; Metrologia 51; 2014; S237–S244

[58] EURACHEM; Terminology in Analytical Measurement – Introduction to VIM 3, first edition 2011, https://www.eurachem.org/index.php/publications/guides/48-gdtam11

[59] Willink R; Measurement Uncertainty and Probability, Cambridge University Press Cambridge;2013

[60] https://ec.europa.eu/jrc/en/interlaboratory-comparisons

[61] Crozet M, Roudil D, Rigaux C, Bertorello C, Picart S, Maillard C; EQRAIN: uranium and plutonium interlaboratory exercises from 1997 to 2016—comparison to ITVs-2010 Journal of Radioanalytical and Nuclear Chemistry 319; 2019; 1013–1021; https://cetama.partenaires.cea.fr/

[62] https://science.osti.gov/nbl/Programs/Measurement-Evaluation

[63] ISO 13528:2005: Statistical methods for use in proficiency testing by inter-laboratory comparisons

[64] ISO 21748:2017 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation

[65] ISO 5725: Accuracy (trueness and precision) of measurement methods and results; 1994

[66] Sprenger J; Statistics between inductive logic and empirical science Journal of Applied Logic 7; 2009; 239–250

[67] Alique O, Vaccaro S, Svedkauskaite J; The use of measurement uncertainty in nuclear material accountancy and verification; ESARDA BULLETIN, No. 52; 2015; 35-40

[68] ESARDA World Café Report – Stresa 2019; Implementation of the RG2019 Actions and Roadmap: https://esarda.jrc.ec.europa.eu/index.php?option=com_content&view=article&id=392:world-cafe-report&catid=23&Itemid=101, https://esarda.jrc.ec.europa.eu/images/files/miscellaneous/ESARDA%20World%20Caf%D0%92%20report_final.pdf

[69] Lindstrom R M; 2017 Believable statements of uncertainty and believable science J Radioanal Nucl Chem 311:1019 1022

[70] Niemeyer I, Dreicer M, Stein G; Nuclear Non-proliferation and Arms Control Verification; 2020; Springer https://doi.org/10.1007/978-3-030-29537-0

Stochastic Approach to Inspection Evaluation: Methodology and Validation

Lohith Annadevula[1], S. K. Aghara[1], Kenneth Jarman[2] and Logan Joyce[1]

[1]University of Massachusetts Lowell, Lowell, MA, USA
[2]Pacific Northwest National Laboratory, Richland, WA, USA

Abstract

The paper discusses the methodology for performing International Atomic Energy Agency's (IAEA's) post-inspection analysis to assess the effectiveness of verification inspection plans using a stochastic method. Conventionally, well-established statistical distributions are employed to calculate Detection Probability (DP) which is the effectiveness metric for both planning and evaluation purposes. The detection probability here is the probability of detecting at least one "defective" item (an item from which material has been removed) from the multi-defect sample space of items. The DP, in turn depends on the probability that a defected item is randomly selected for measurement (the "selection probability") and the probability that the applied measurement identifies the defect (the "identification probability"). The stochastic method described here involves simulating the inspection process by randomly choosing a fixed number of items from a population of items and performing measurements for these samples. A detection probability value is calculated at the end of a simulation depending on the random outcome. Multiple such simulations/trials are performed on the same sample space to get multiple detection probabilities. The Final Detection Probability and its uncertainty are estimated by computing the average and standard error of all the DP values from all simulations. The stochastic model development, its verification, and benchmarking are discussed in detail.

Keywords: Stochastic Approach; Detection Probability; Selection Process

1. Introduction

The increase in computational power in modern computers and the development of pseudorandom generators have resulted in the prevalent use of state-of-the-art Monte-Carlo/Stochastic methods [1] for many applications. These methods allow us to harness the computational power to simulate real-world experiments involving probabilities and random processes. With known outcomes and outcome probability distribution function (pdf) [2], any random process can be simulated by invoking a pseudorandom generator satisfying the required pdf function. The outcomes simulated by the random generator can contain related and unrelated events to our quantity of interest. For example, in the coin-toss experiment, if the probability of getting Heads P(Heads) is our quantity of interest, then Head events are related events, and Tail events are unrelated to our quantity. Such quantities of interest that are involved in the process can be derived/computed based on the relative frequency with which the random generator simulates the quantity-related events. The true power of stochastic methods becomes apparent when dealing with complex random processes which contain multiple simple random processes embedded within the complex process. Such complex processes can be simulated by invoking multiple pseudorandom number generators, with each generator simulating one of the embedded simpler random processes. The entire complex process can be simulated by concatenating the outputs of one simple process with the input of another simple process. The real-world inspection problem is an example of such a complex random process. It has a random selection process followed by an instrumental measurement process embedded sequentially. The detection probability DP is the primary quantity of interest. In further sections, we shall describe the conventional way of deterministically evaluating DP using distributions and evaluating DP from stochastic simulations.

A probabilistic model of the IAEA's inspection problem [3] is that of random selection from a set of identical items, from some of which a proliferator has removed some amount of material. Items from which material has been removed are referred to as defects or defective items. The original set of items following proliferation, in general, contains both defects and non-defects. Depending on the proliferator's diversion strategy, multiple types of defects (each type of

defect is resulted from removing different amounts of material from the original item) can be induced in the sample space. For example, consider a sample space or a stratum containing ten items; following diversion, two of the items are transformed into Defects, and the rest remain unchanged (ND). Among the two defects, assume both are different types of Defects (D_1, D_2).

Sample space: $\{D_1, D_2, ND\} = [1, 1, 8]$; Total = 10 items.

In the remainder of this section, we illustrate a deterministic approach to calculating the probability of detecting diversion using the above example. In sections 2 & 3, we describe the stochastic approach and demonstrate its application to examples, including validation of the approach against a previously published deterministic solution.

1.1 Illustration of a Deterministic Approach

The inspection process involves randomly selecting a few items and performing measurements on the selected items using an instrument (method) from a range of choices, each with a unique measurement fidelity and uncertainty. The instrument or method's ability to detect a specific item in the sample space as a defect varies with the type of item being measured, characterized as a probability that the measurement method identifies a defective item. This probability is termed Identification probability (IP). Assume that the instrument identifies D_1 items 100% of the time as defected, D_2 items 50% of the time, and the measurement never identifies non-defects as defected, i.e., IP = 0%.

$$[IP_{D1}, IP_{D2}, IP_{ND}] = [1, 0.5, 0]$$

Analytically, the overall DP is computed by summing up individual DP components corresponding to all possible outcomes of the random selection of the set of items in the sample space. For each outcome, its DP value is given by the product of the outcome's selection probability (SP) and identification probability (IP). The conditional tree diagram in Figure 1 exemplifies the identification of all possible selection outcomes and the determination of each outcome's selection probability.

For Single Measurement Inspection, a single item is randomly sampled for measurement. The left conditional tree diagram in Figure 1 shows three possible outcomes of single measurement sampling where one of the three item types will be selected. Therefore, for a single measurement inspection, the Detection Probability DP is given by the sum of component DPs of all outcomes.

$$DP = SP_{D1}*IP_{D1} + SP_{D2}*IP_{D2} + SP_{ND}*IP_{ND} = (1/10)*1 + (1/10)*0.5 + (8/10)*0 = 0.15 = 15\%.$$
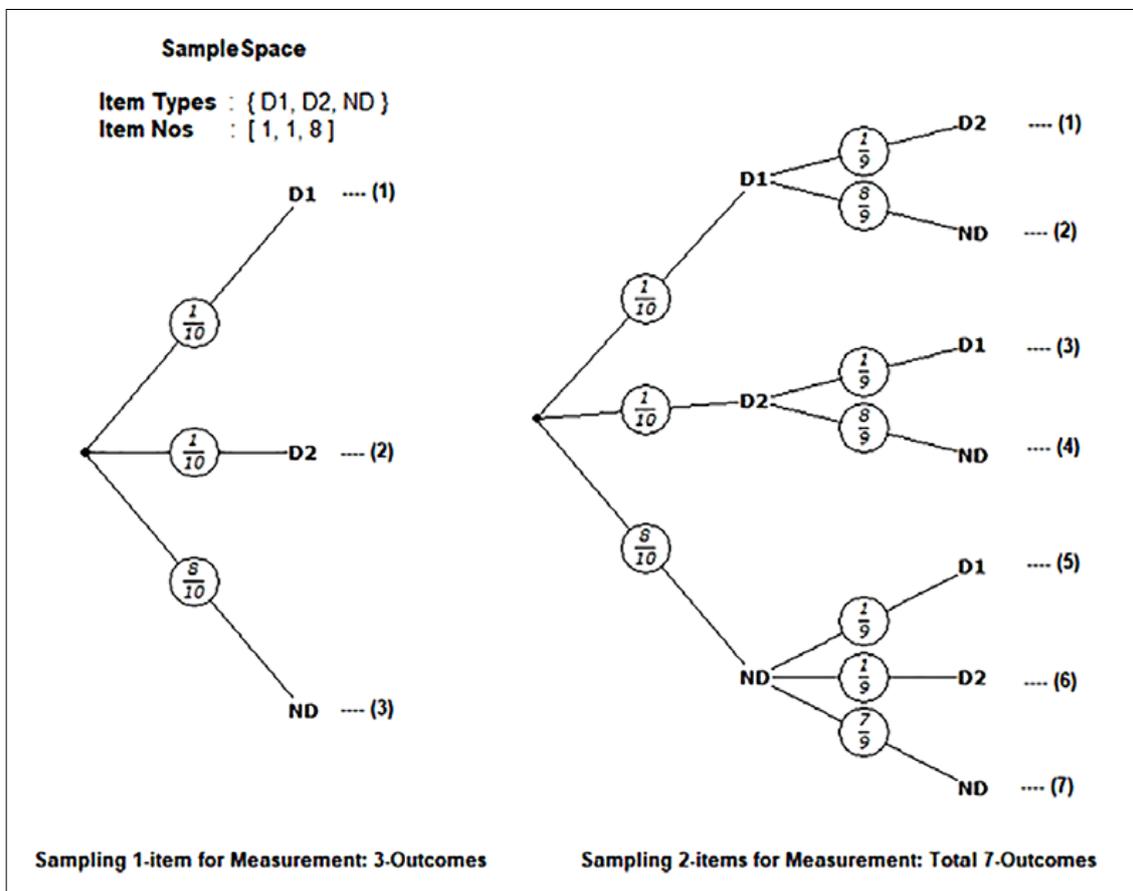
Figure 1: Conditional Tree Diagrams Depicting Various Outcomes and their Selection Probabilities.

For Double Measurement inspection, two items are randomly sampled for measurement. There are seven possible combinations described in the right conditional tree diagram in Figure 1. As the number of item types and measurements increases, the total combinations of selection outcomes required to evaluate DP quickly gets large. For a generic data set of items N and number of measurements n the multivariate hypergeometric PDF (denoted *'MVHG_PDF'* below) is used to compute selection probabilities of different combinations of outcomes as shown below. The MVHG_PDF gives the conditional probability of n draws, without replacement, from a finite population of size N that contains i types of items with I_i numbers in the population leading to the selection of x_i numbers of respective item types in the outcome upon sampling.

$$SP(x_1, x_2, x_3 \dots, x_i, I_1, I_2, I_3 \dots, I_i, N, n) = MVHG\_PDF$$

$$= \frac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}\dots\binom{I_i}{x_i}}{\binom{N}{n}}$$

$$DP(x_1, x_2, x_3 \dots) = SP(x_1, x_2, x_3, I_1, I_2, I_3 \dots, N, n) * IP$$

$$= \frac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}\dots\binom{I_i}{x_i}}{\binom{N}{n}} * \left[1 - \prod NIP_i^{x_i}\right]$$

$$TDP = \sum_{All\ combinations} DP(x_1, x_2 \dots, x_i)$$

$$= \sum_{All\ combinations} \frac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}\dots\binom{I_i}{x_i}}{\binom{N}{n}} * \left[1 - \prod NIP_i^{x_i}\right]$$

The variables and constants used in the equations above are as follows:

i = 3 (Three item types in stratum {D_1, D_2, ND})

I_1 = Total number of D1 items = 1

I_2 = Total number of D2 items = 1

I_3 = Total number of ND non-defect items = 8

N = Total number of items in stratum = $I_1 + I_2 + I_3$ = 10

n = Number of items randomly sampled from total items for inspection = 2

x_1 = Number of D_1 items in sample

x_2 = Number of D_2 items in sample

x_3 = Number of ND items in sample

NIP_i = non-identification probability of ith item type = 1 - IP_i

TDP = Total detection probability

The terms $\binom{a}{b}$ represent typical combination operation $C(a,b) = \dfrac{a!}{(a-b)!\ b!}$

Note that x_1, x_2, and x_3 values vary for different outcomes or combinations.

Calculation of Selection Probabilities:

- [D_1 D_2] and [D_2 D_1] combination: $SP = \dfrac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}}{\binom{N}{n}} = \dfrac{\binom{1}{1}\binom{1}{1}\binom{8}{0}}{\binom{10}{2}} = \dfrac{1}{45}$

- [D_1 ND] and [ND D_1] combination: $SP = \dfrac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}}{\binom{N}{n}} = \dfrac{\binom{1}{1}\binom{1}{0}\binom{8}{1}}{\binom{10}{2}} = \dfrac{8}{45}$

- [D_2 ND] and [ND D_2] combination: $SP = \dfrac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}}{\binom{N}{n}} = \dfrac{\binom{1}{0}\binom{1}{1}\binom{8}{1}}{\binom{10}{2}} = \dfrac{8}{45}$

[ND ND] combination: $SP = \dfrac{\binom{I_1}{x_1}\binom{I_2}{x_2}\binom{I_3}{x_3}}{\binom{N}{n}} = \dfrac{\binom{1}{0}\binom{1}{0}\binom{8}{2}}{\binom{10}{2}} = \dfrac{28}{45}$

The usage of the multivariate hypergeometric distribution multiple times is necessary to account for selection probabilities for various possible outcomes. With the increase in the number of measurements and item types in the sample space, the inspection outcomes increase exponentially. This exponential increase in inspection outcomes quickly limits the model's performance in terms of computational resources (CPU and Memory). The development, performance and limitations of conditional tree-based deterministic models will be discussed extensively in a forthcoming paper [9]. The illustration of DP calculations based on a deterministic approach shows how the calculation can quickly become rather complicated (in terms of identifying outcomes) even for a single stratum of material, let alone multiple strata within a facility and ultimately multiple facilities within a state. So far, the examples depicted in Figure (1) use a single instrument or measurement method, and it must be noted that the deterministic models get even more complicated in multi-instrument inspections. This is why

Combination Type	Selection Probability	Non-Detection Probability
[D_1 D_2] and [D_2 D_1]	$\dfrac{1}{45}$	$\dfrac{1}{45} * (1-1) * (1-0.5) = 0$
[D_1 ND] and [ND D_1]	$\dfrac{8}{45}$	$\dfrac{8}{45} * (1-1) * (1-0) = 0$
[D_2 ND] and [ND D_2]	$\dfrac{8}{45}$	$\dfrac{8}{45} * (1-0.5) * (1-0) = \dfrac{4}{45}$
[ND ND]	$\dfrac{28}{45}$	$\dfrac{28}{45} * (1-0) * (1-0) = \dfrac{28}{45}$
Total Non-Detection Probability		$0 + 0 + \dfrac{4}{45} + \dfrac{28}{45} = 0.711112$
Total Detection Probability		$1 - (0.71112) = 0.28889 = \mathbf{28.89\%}$

Table 1: Calculation of Total Detection Probability for Double-item Measurement Inspection

the deterministic models developed in literature are case-specific and lack universal applicability.

A stochastic approach provides an intuitive and flexible alternative. It involves simulating the inspection process, randomly selecting items from the stratum followed by instrumental measurements on each selected item, and repeating this simulation multiple times to acquire a distribution of DP values. The mean of this distribution of simulated DP values and its standard error provide estimates of the total DP and its uncertainty, respectively. The accuracy of the result increases with the increase in the number of simulated inspections. Individual inspection simulations require low computer memory requirements relative to the deterministic approach. Individual simulations are independent of each other, so the increase in computational cost is primarily in terms of CPU, which is easily manageable on a generic multi-threading and multi-core computer. We discuss the stochastic approach in detail in the next section.

2. Stochastic Approach

The stochastic approach uses a set of random simulations or trials, called an ensemble, to generate a distribution of outcomes from which the best estimate of the desired quantity is computed. The stochastic nomenclature used in this paper is summarised below:

- *Stochastic Simulation/Trial:* A single (pseudo-) random sample of a random variable or process.
- *Outcome:* A possible result of a simulation or trial.
- *Ensemble:* A set of outcomes acquired from multiple simulations or trials.
- *Ensemble Mean:* The mean of an ensemble (when outcomes are numerical values).
- *Stochastic Solution:* An estimate of the desired quantity acquired from ensemble means (this may involve multiple ensembles and is computed from the average of all ensemble means).
- *Stochastic Standard Error:* Standard error in the estimated stochastic solution computed using the ensemble means.

A simple example problem illustrates the application of the stochastic approach. Suppose we simulate an unbiased coin-toss experiment where the outcomes are Heads or Tails. To estimate the probability of getting "Heads" on a single toss, the act of tossing is simulated many times using a uniform pseudorandom number generator which yields a value of 0 or 1. In each simulation or trial, a number is sampled from the pseudorandom generator; getting 1 is equivalent to getting Heads, and 0 means Tails, respectively. Since we are looking for Head events, we assign the value 1 to the stochastic outcome when a Head turns up. If a Tail turns up, we assign 0 to the stochastic outcome. We simulate the experiment 100 times (collectively called an

ensemble) and store the outcomes as 0s and 1s. The ensemble mean or the mean of the outcomes of 100 trials will yield a value close to 0.5 with an estimate of error associated with the result. By increasing the number of simulations/trials, the ensemble mean will get closer to 0.5. Theoretically, the value will converge to 0.5 with zero error with an infinite number of trials. Thus, using the stochastic approach, the probability of the outcome "Heads" is estimated by repeatedly simulating a coin toss using a pseudorandom number generator multiple times, counting the occurrence of "Heads" and dividing by the total number of simulations (or, equivalently, averaging the numerical values assigned to Heads and Tails) represents the stochastic solution. In the following sub-section, we describe the application of the stochastic process to evaluate the effectiveness of the IAEA inspection using pseudorandom generators.

2.1 Methodology for Inspection Problem

Applying the stochastic approach to inspection involves simulation of the random selection of a specified number of items from the set of all possible items, followed by measurements on selected items. For each simulation, a DP value (the outcome) is calculated. The simulation is repeated multiple times to acquire a sufficient distribution of DP values (the ensemble). The mean of this distribution is the desired approximation to DP (the stochastic solution) for the specified inspection campaign data.

2.1.1 Selection Process

Consider the single-item and double-item inspection examples discussed in Section 1, where items are randomly selected from the set of one D1, one D2, and eight ND items. Consider $M \times N$ inspection simulations representing M independent ensembles of N trials each. For practical reasons, it is convenient to split the total number of trials into multiple ensembles. Figure 2 describes the outcomes of $M \times N$ simulations for single- and double-item inspection examples.
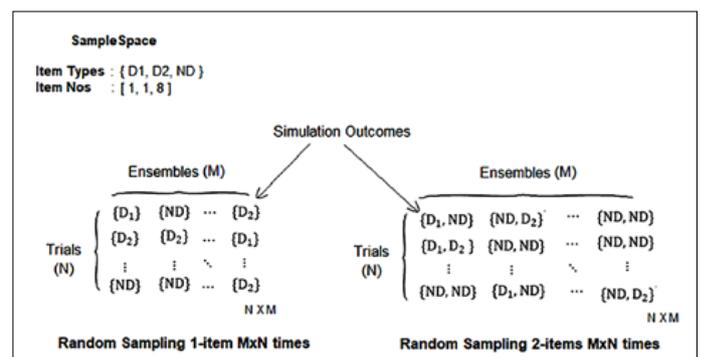
Figure 2: Random Selection Matrices Depicting Various Outcomes and their Event spaces.

2.1.2 Identification Process

The second step of the inspection process involves measuring the selected items. Measuring selected items will allow the inspector to identify defects. Using the same identification probabilities as in the deterministic treatment, [IP_{D1}, IP_{D2}, IP_{ND}] = [1, 0.5, 0] gives identification probabilities and NIP = 1-IP gives the non-identification probabilities. The identification step involves replacing the item types within the simulation matrices with their overall non-identification probabilities. The items in Figure 2 are replaced by their non-identification probabilities to get Figure 3. Then the overall outcome identification probabilities are computed in Figure 4 by multiplying item NIPs within all brackets present in Figure 3.
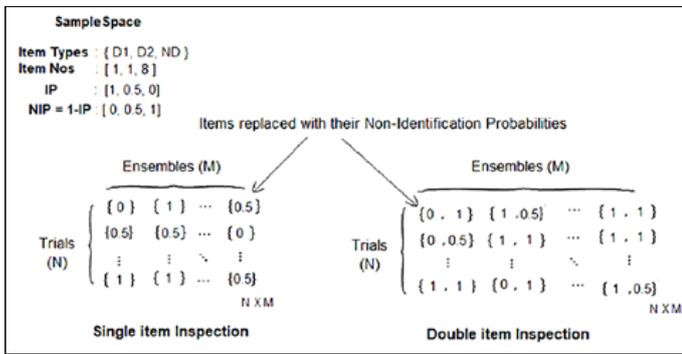
Figure 3: Non-Identification Probability Matrices
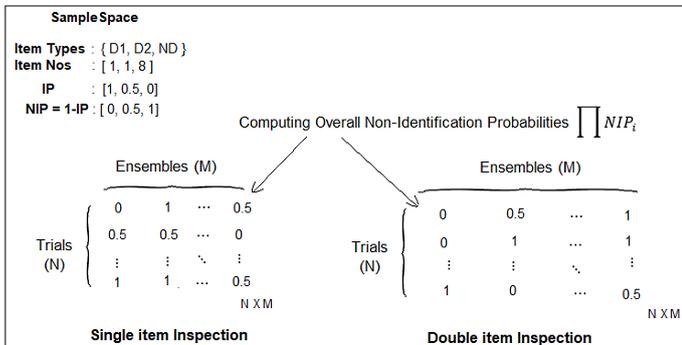
Figure 4: The Overall Outcome Non-Identification Probability Matrices

2.1.3 Computing Detection Probability DP

To compute the overall DP for a specific inspection campaign, we must first calculate the non-detection probability (NDP) corresponding to each simulated inspection. The non-identification probability (NIP) value for each outcome shown in Figure 4 is, in fact, the NDP value for the respective outcome. The detection probability is 1 - NDP, as shown in Figure 5. For each ensemble of N trials, our implementation of the stochastic approach computes an ensemble mean DP and standard error. The approach then computes the aggregate mean and standard error over the M ensembles. We shall discuss all the necessary derivation

steps to get to the final DP estimate and its standard error in detail in section 2.2.
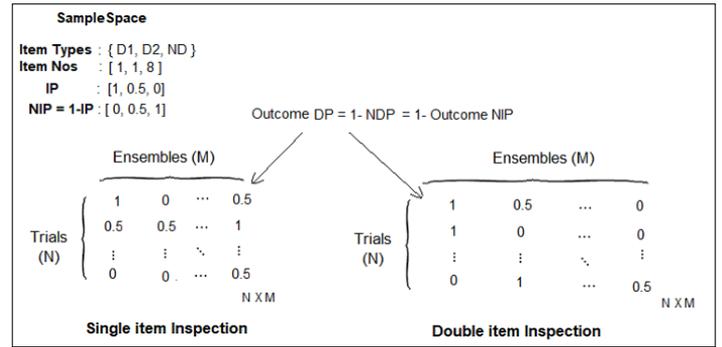
Figure 5: The Outcome Detection Probability Matrices

2.2 Stochastic Standard Error

Let X_{ij} represent DP for the ith simulation and j^{th} ensemble. The values for 'i' range from 1 to N trials, and 'j' takes values from 1 to M ensembles. Assume that all DP values are independent and identically distributed random variables with mean μ and variance σ^2. The complete set of DP values from Figure 5 can be represented in terms of the following matrix:

$$\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1M} \\ X_{21} & X_{22} & \cdots & X_{2M} \\ \vdots & \vdots & X_{ij} & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NM} \end{pmatrix}$$

2.2.1 Overall Statistics

The overall estimate of DP and uncertainty can, in principle, be calculated by the simple statistical formulas below:

$$\hat{\mu} = \bar{X}_{..} = \frac{1}{MN} \sum_{ij} X_{ij}$$

$$s^2 = \frac{1}{MN-1} \sum_{ij} (X_{ij} - \bar{X}_{..})^2 \qquad (1)$$

$$se(\hat{\mu}) = \frac{s}{\sqrt{MN}}$$

Here we make use of the typical statistical notation: $\hat{\mu}$ denotes the estimated mean, which is the average of all DP values, $\bar{X}_{..}$; s denotes sample standard deviation, and se denotes standard error in the estimated parameter. However, we first calculate these statistics for individual ensembles of N DP values and then aggregate them across ensembles to estimate a final DP value and uncertainty. The

reason for breaking the entire collection of outcomes into separate ensembles has to do with our lack of knowledge of the number of simulations/trials needed to achieve a target convergence criterion prior to simulations. It is computationally convenient to run one ensemble of N simulations/trials at a time, estimate running standard error, and decide based on the acquired error whether to run further ensembles or not. The following sub-sections consider the statistics for individual ensembles and across the ensembles of simulation averages.

2.2.2 Ensemble Statistics

Each ensemble consists of N trials that yield N DP values. For the jth ensemble, the mean, standard deviation, and standard error in the mean are as follows:

$$\hat{\mu}_j = \bar{X}_{.j} = \frac{1}{N}\sum_i X_{ij}$$

$$s_j^2 = \frac{1}{N-1}\sum_i (X_{ij} - \bar{X}_{.j})^2 \qquad (2)$$

$$se(\hat{\mu}_j) = \frac{s_j}{\sqrt{N}}$$

2.2.3 Statistics Across Ensembles

Aggregate mean, standard deviation, and standard error across the ensembles may be computed as follows:

$$\hat{\mu}_{avg} = \bar{X}_{..} = \frac{1}{M}\sum_j \bar{X}_{.j} = \hat{\mu}$$

$$s_{avg}^2 = \frac{1}{M-1}\sum_j (X_{.j} - \bar{X}_{..})^2 \qquad (3)$$

$$se(\hat{\mu}_{avg}) = \frac{s_{avg}}{\sqrt{M}}$$

2.2.4 Overall Statistics: Combining Ensemble & Across Ensemble Statistics

The breakdown of the complete set of simulations into multiple subsets of ensembles is similar to the "within-group" and "across-group" calculations used in the analysis of variance (ANOVA) [5]. We apply the same calculations used in ANOVA to compute overall statistics. First, the overall mean

is simply the average of ensemble means, as noted in Equation (3). The standard error is computed as follows [6]:

$$s^2 = \frac{1}{MN-1}\sum_{ij}(X_{ij}-\bar{X}_{..})^2 \to from\ (1)$$

$$s^2 = \frac{1}{MN-1}\sum_{ij}(X_{ij}-X_{.j}+X_{.j}-\bar{X}_{..})^2$$

$$s^2 = \frac{1}{MN-1}\sum_{ij}\left[(X_{ij}-X_{.j})^2 + 2(X_{ij}-X_{.j})(X_{.j}-\bar{X}_{..}) + (X_{.j}-\bar{X}_{..})^2\right]$$

$$s^2 = \frac{1}{MN-1}\left[\sum_{ij}(X_{ij}-X_{.j})^2 + 0 + N\sum_j (X_{.j}-\bar{X}_{..})^2\right]$$

$$s^2 = \frac{1}{MN-1}\left[\sum_j (N-1)s_j^2 + 0 + N(M-1)s_{avg}^2\right] \to from\ (2)\ \&\ (3)$$

$$s^2 = \frac{M(N-1)\overline{s_.^2} + N(M-1)s_{avg}^2}{MN-1}$$

$$s^2 = \frac{M(N-1)\overline{s_.^2}}{MN-1} + \frac{N(M-1)s_{avg}^2}{MN-1}$$

Converting Variances into Standard errors,

$$MN se(\hat{\mu})^2 = \frac{M(N-1)N\ \overline{se(\hat{\mu}_j)^2}}{MN-1} + \frac{N(M-1)M\ se(\hat{\mu}_{avg})^2}{MN-1}$$

$$se(\hat{\mu})^2 = \frac{(N-1)\ \overline{se(\hat{\mu}_j)^2}}{MN-1} + \frac{(M-1)\ se(\hat{\mu}_{avg})^2}{MN-1}$$

$$se(\hat{\mu}) = \sqrt{\frac{(N-1)\ \overline{se(\hat{\mu}_j)^2}}{MN-1} + \frac{(M-1)\ se(\hat{\mu}_{avg})^2}{MN-1}} \qquad (4)$$

$$where\ \overline{se(\hat{\mu}_j)^2} = \frac{1}{M}\sum_j se(\hat{\mu}_j)^2$$

$$\hat{\mu} = \frac{1}{M}\sum_j \bar{X}_{.j} = \frac{1}{MN}\sum_{ij} X_{ij} \qquad (5)$$

The Overall Average $\hat{\mu}$ from Equation (5) gives the best estimate of Detection Probability, with Equation (4) as the best estimate of its standard error $se(\hat{\mu})$. Based on the described stochastic approach, a python model has been developed. The model allows users to input the required

standard error in DP estimate, trials per ensemble, and case data. The model starts with a single ensemble of stochastic simulations and computes DP & standard error using equations (4) & (5). It initiates a new ensemble of stochastic simulations, recomputes running error, and repeats the process until the running standard error converges to a user-set value. The single-item & double-item inspection examples are simulated using the stochastic model with the required error set to 0.002, and the number of trials per ensemble N is set to 2000. For the single-item inspection example, the code ran 13 ensembles, and the final DP value is 0.148 with 0.002 as the standard error in the estimate. For the double-item inspection case, the code ran 21 ensembles, and the final DP value is 0.287 with 0.002 as the standard error in the estimate. By comparison, the stochastic results agree with the deterministic results in section 1.1, i.e., DP is 0.15 for the single-item inspection, and DP is 0.289 for the double-item inspection.

3. Validation of the Stochastic Approach

In the publication Krieger et al. [7], the authors investigate scenarios to develop inspection sampling plans for inventory verification of spent fuel ponds. The paper discusses probable diversion scenarios from the spent fuel storage ponds and calculates the achieved DP for the specified sampling plans. We choose this paper primarily as it defines various inspection scenarios, treats them deterministically, and computes DP, all in one place, sufficient for our benchmarking purposes. We calculate the DP for two cases mentioned in the paper [7] using our stochastic approach [4] and compare our results to the published results [7, 8].

3.1 Example: Varying Falsified Pins

In this example, the spent fuel pond contains 2500 (N) spent fuel assemblies (SFAs), with each assembly containing 96 (L) fuel pins. In terms of material, each assembly contains 2 kg or 0.25 SQ ($\bar{x}$) of Pu. A total goal amount (G) of 1 SQ or 8 Kg of Pu is chosen to be diverted by removing r_{pins} from each assembly. To acquire 1 SQ would require r_{SFA} assemblies from which r_{pins} pins are removed while the remaining N-r_{SFA} assemblies remain untouched. The r_{pins} falsified pins per assembly are varied from 1 to 96 in steps of 1. The total number of assemblies r_{SFA} required to divert 1 SQ is given by equation (6).

$$r_{pins} = [1, 2, 3, \ldots, 95, 96]$$

$$r_{SFA}(r_{pins}) = ceil\left(\frac{G * L}{\bar{x} * r_{pins}}\right) \qquad (6)$$

Out of 2500 SFAs, the inspector verifies n_1 SFAs with the ICVD, n_2 SFAs with the DCVD, and n_3 SFAs with the PGET, where per verified SFA only one measurement instrument is applied. For the given example, the values of n_1, n_2, and n_3 are taken to be 10, 65 & 25 measurements, respectively.

Each instrument's identification probability function is modeled as a step function; i.e., the identification probability is 0 or 1 when the number of pins diverted in a measured assembly is less than or greater than a certain % of total pins, respectively, as shown in Equation (7). The ICVD detects diversion only when 100% of pins are absent from the measured SFA. DCVD detects diversion when 30% of total pins are absent. PGET detects diversion when 0.38% of total pins are missing from the measured SFA. Therefore, the piece-wise function in Equation (7) gives the instrument identification probability.

$$IP(r_{pins}, L, \%Limit) = \begin{cases} 1 \ if \ \left(\frac{r_{pins}}{L} * 100\right) >= \%Limit \\ 0 \ if \ \left(\frac{r_{pins}}{L} * 100\right) < \%Limit \end{cases}$$
$$(7)$$

The following summarizes Equation (7) and the number of measurements for each instrument type:

$Using \ ICVD, \%Limit = 100; \quad IP_{ICVD} = IP(r_{pins}, L, 100); \ n_1 = 10$
$Using \ DCVD, \%Limit = 30; \quad IP_{DCVD} = IP(r_{pins}, L, 30); \ n_2 = 65$
$Using \ PGET, \%Limit = 0.38; \quad IP_{PGET} = IP(r_{pins}, L, 0.38); \ n_3 = 25$

The overall DP for this example has been discussed in Krieger et al. [7]. The piece-wise DP equation from [7] is repeated below.

$$DP(N, n_1, n_2, n_3, r_{pin}) = \begin{cases} 1 - \frac{\binom{N-r_{SFA}(r_{pins})}{n_3}}{\binom{N}{n_3}} & for \ r_{pin} < [0.3L] - 1 \\ 1 - \frac{\binom{N-r_{SFA}(r_{pins})}{n_2+n_3}}{\binom{N}{n_2+n_3}} & for \ 0.3L < r_{pin} < L - 1 \\ 1 - \frac{\binom{N-r_{SFA}(r_{pins})}{n_1+n_2+n_3}}{\binom{N}{n_1+n_2+n_3}} & for \ r_{pin} = L \end{cases}$$
$$(8)$$

We applied the stochastic approach to compute the overall detection probability and uncertainty with Equations (4) and (5). We used 2000 trials per ensemble with a target standard error set to 0.002. The code automatically generates ensembles until the running standard error is less than or equal to the set value. Convergence in standard error with number of ensembles to the set value is illustrated in Figure (6). The plot demonstrates how different falsified pin examples converge at a different rate to the set error and also shows the practicality of estimating running standard error that allowed the code to stop initiating additional ensembles when the error reaches the set value. The deterministic and stochastic results are co-plotted in Figure (7), showing the agreement between both. The same agreement is further depicted in the residual plot of Figure (8), where the difference between deterministic and stochastic DPs are computed and plotted along with set standard error limits. All the residual values plotted in Figure (8) lie within the

limits of three times set standard error, indicating the agreement of the stochastic results with that of deterministic results.

The lowest detection probability occurs at $r_{pins} = 28$. The deterministic estimate of DP is 0.1315, while the estimate using the stochastic approach is 0.1309 with 0.0019 as its standard error. The residual between deterministic and stochastic DP estimates is 0.0006, which lies within the limits of twice the stochastic standard error (± 2*SE), i.e., ± 0.0038. Therefore, the stochastic results agree with deterministic results.
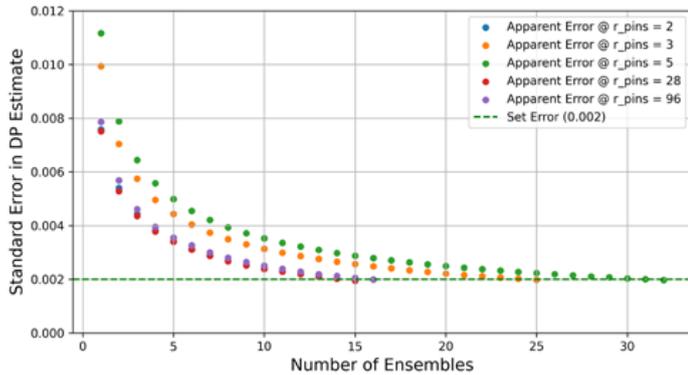
Figure 6: Convergence in standard error with number of ensembles for different r_{pin} examples.
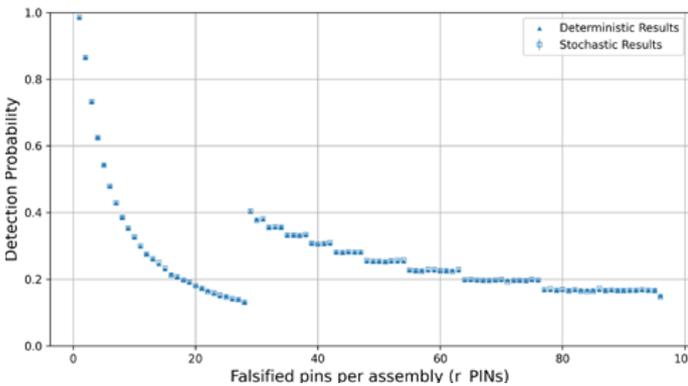
Figure 7: Co-plot of Deterministic & Stochastic Detection Probabilities for Varying r_{pin} Example.
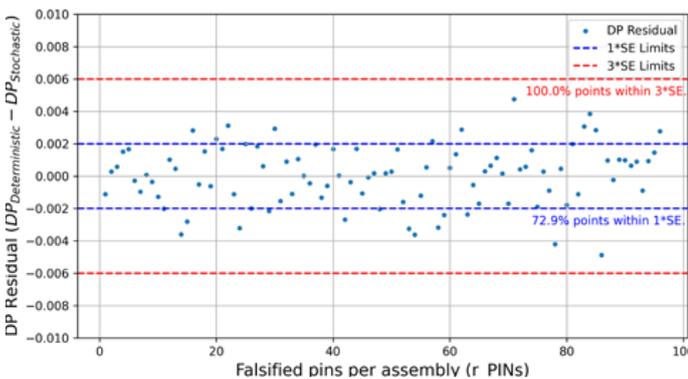
Figure 8: DP Residual plot for Varying r_{pin} example; all residuals lie within limits of three times the set error.

3.2 Example: Multi-Group Diversion

In this example, the spent fuel pond contains 2000 (N) Spent Fuel Assemblies with each assembly containing 96 (L) fuel pins, and in terms of material, each assembly contains 2 kgs or 0.25 SQ ($\bar{x}$) of Pu. A total goal amount (G) of 1 SQ or 8 Kg of Pu is chosen to be diverted by removing 4 pins from 21 SFAs and 30 pins from 10 SFAs while the remaining $N - 31$ assemblies remain untouched.

	Group1	Group2	Group3
Spent Fuel Assemblies	21	10	1969
Falsified Pins per Assembly	4	30	0
Total Material Diverted	(4*21 + 30*10 + 0*1969) *0.25/96 = 1 SQ		

Table 2: Multi-Group Diversion Example Case Information

Out of 2000 SFAs, the inspector verifies n_1 SFAs with the ICVD, n_2 SFAs with the DCVD, and n_3 SFAs with the PGET, where per verified SFA only one measurement instrument is applied. The instruments are the same as in the previous example. The only difference is the values of n_1, n_2, and n_3 are taken to be 59, 162 & 97 measurements, respectively. The identification probability functions, in this case, are as follows:

$$Using\ ICVD, \%Limit = 100;\ \ IP_{ICVD} = IP(r_{pins}, L, 100);\ n_1 = 59$$
$$Using\ DCVD, \%Limit = 30;\ \ IP_{DCVD} = IP(r_{pins}, L, 30);\ n_2 = 162$$
$$Using\ PGET, \%Limit = 0.38;\ \ IP_{PGET} = IP(r_{pins}, L, 0.38);\ n_3 = 97$$

The overall DP for this example given in [7] is 0.91

$$DP(N, n_1, n_2, n_3) = 1 - \frac{\binom{31}{0}\binom{N-31}{n_3}}{\binom{N}{n_3}} \frac{\binom{10}{0}\binom{N-n_3-10}{n_2}}{\binom{N-n_3}{n_2}} \approx 0.9133$$

(9)

The stochastic approach using the same options as the previous example produced the estimate of 0.9157 with 0.0019 as standard error, which agrees with the deterministic value of 0.9133. The residual between deterministic and stochastic DP estimates is 0.0024, which lies within limits defined by twice the stochastic standard error (± 2*SE), i.e., ± 0.0038.

4. Conclusion

The paper describes in detail the development of a stochastic approach, in section 2, to compute detection probability for inspection problems at the stratum level. In section 3, the stochastic approach is validated against two spent fuel inspection examples discussed in detail and treated deterministically in the published paper [7]. For the varying r_{pins} example, the computed DP residuals (difference between deterministic and stochastic DPs) of all the points

lie within 3*SE limits depicted in Figure (8). For the multi-group diversion example, the DP residual is 0.0024, which is within 2*SE limits. Thus, the stochastic results agree with the deterministic results. The main advantage of the stochastic model over the deterministic models is its universal applicability to any inspection scenario at the stratum level (involves multi-defect stratum and multi-instrument scenarios), and the methodology remains the same making it a versatile tool. Currently, the stochastic approach can compute DPs at the stratum level. Future work involves extending the approach to compute facility-level DP and then to the state-level DP concept.

5. Acknowledgements

6. References

[1] Kroese Dirk, Taimre Thomas and Botev Zdravko; Handbook of Monte Carlo Methods; 2011; doi: 10.1002/9781118014967.

[2] Harrison, Robert L; "Introduction to Monte Carlo Simulation." AIP conference proceedings vol. 1204 (2010): 17-21. doi:10.1063/1.3295638.

[3] International Atomic Energy Agency; IAEA Safeguards Glossary, International Nuclear Verification Series No. 3; 2003.

[4] Chris Gazze, S.K. Aghara, Ian Bleeker, Lohith Annadevula, Ahmad Nofal, Logan Joyce, James Porcello, Katherine Bachner, Jose Gomera; Stochastic Approaches for Calculating and Aggregating Detection Probabilities for Nuclear Material Diversion; 2019; BNL-211980-2019-INRE.

[5] Douglas C. Montgomery and George C. Runger; "The Analysis of Variance," in Applied Statistics and Probability for Engineers, 7th ed.: Wiley; 2018; ch. 13th, pp. 472-475.

[6] Lohith Annadevula, S. K. Aghara, Kenneth Jarman, and Logan Joyce; Statistical Analysis of Convergence and Error Propagation in Stochastic Model for Safeguards Inspection; submitted to Proc. of the INMM & ESARDA Joint Annual Meeting; 2021.

[7] Thomas Krieger, Katharina Aymanns, Arnold Rezniczek and Irmgard Niemeyer; Optimal Sampling Plans for Inventory Verification of Spent Fuel Ponds: Article 01; ESARDA Vol 58; June 2019; ISSN:1977-5296.

[8] Logan Joyce, Lohith Annadevula, S. K. Aghara, Thomas Krieger, and Kenneth Jarman; Stochastic Model Simulation for Evaluation of Spent Fuel Pond Inventory Verification Sampling Plans; submitted to Proc. of the INMM & ESARDA Joint Annual Meeting; 2021.

[9] Lohith Annadevula, S. K. Aghara; Universal deterministic modeling to compute stratum-level detection probability based on conditional tree diagram; submitted concurrently to Proc. of the INMM 63rd Annual Meeting; 2022 (forthcoming).

Effort Bounded Inspections

Rudolf Avenhaus[1], Morton J. Canty[2] and Thomas Krieger[3]

[1]University of the German Federal Armed Forces Munich, Germany
[2]Heinsberger Str. 18, D-52428 Jülich, Germany
[3]Corresponding author. Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany, E-Mail: t.krieger@fz-juelich.de

Abstract:

Given an Inspectorate with the task of verifying the adherence of an Operator of a group of facilities to an agreement on permitted activities within those facilities, how large should the inspection effort be and how should it be distributed among the facilities? A game-theoretical approach is described which addresses these important questions, generalizing and extending the applicability of earlier inspection models, which either treated inspection effort as extrinsic, or which imposed special assumptions. A solution of the inspection game, i.e., a Nash equilibrium, is presented in quite general terms, and two applications are presented.

Keywords: game theory, inspection games, resource optimization

1. Problem formulation

The problem of distributing inspection effort across different locations or facilities has been the subject of various analyses in the past. In cases where inspections serve the purpose of deterring an organization or State from violation of an agreement or treaty, game theoretical models involving inspection resource distribution in space and over time have been applied. These models attempt to formulate inspection goals in terms of some objective function such as detection probability, expected time to detection of illegal behaviour, or deterrence.

The allocation of some – continuously divisible – inspection effort was explicitly the subject of analyses by [1], [2] and [3]. The latter work was the stimulus for the present contribution. While Deutsch et al. imposed very specific assumptions, it will be demonstrated that their approach, related to earlier work in [4], can be applied to much more general situations.

Problems of distributing inspection effort across different locations or facilities have also been discussed for some time in the context of applying nuclear material safeguards under the *State Level Approach* by the International Atomic Energy Agency (IAEA) in Vienna. In partial fulfilment of the Non-Proliferation Treaty (NPT) the IAEA verifies the peaceful use of nuclear material in the Treaty's member States [5]. Within this rather general context, the following important questions must be answered: How much inspection effort shall the IAEA allot to a given State? How should that effort be distributed over the individual nuclear facilities within the State? Since major studies along this direction have not been forthcoming from other fields of application, nuclear safeguards in particular and arms control in general have stimulated original work which has become to be known as the field of inspection games; see, e.g., [6] and [7].

In section 2 a general inspection model is developed, that is, a set of assumptions which permits the analysis of the inspection problem in quantitative terms. In section 3 a Nash equilibrium of the resulting non-cooperative two-person game is presented. Since the game and its solution are expressed in rather general terms, two applications are given in sections 4 and 5. The concluding section 6 remarks on further applications and future extensions of the results are discussed.

2. The Model

We consider K facilities operated by an organization or State which are subject, under a verification agreement, to control by an Inspectorate or Inspector. An illegal activity, in the sense of the agreement, is assumed to take place *in at most one facility,* since this way the number of staff members involved in the illegal activity is kept small.

If the illegal activity is performed in facility i, $i = 1, ..., K$, then it will be detected with some probability p_i. Note that the p_i, $i = 1, ..., K$, do not necessarily sum to unity. For example, if an a-priori required detection probability of $p_i = 1/2$ is assumed in all facilities, then for $K \geq 3$ the sum of the p_i's exceeds 1.

At IAEA and other inspection authorities the inspection measures z_i spent at the i-th facility, $i = 1, ..., K$, depend on the a-priori required detection probabilities p_i for any of the facilities, i.e., $z_i(p_i)$. For example, $p_i = 0.9$ for 'high' and $p_i = 0.2$ for 'low' probability level; see [5]. Since however, values of this kind can hardly be justified by formal means, we consider in this paper the reverse case: At the i-th facility, the inspection measure z_i is taken which then results in a detection probability p_i for that facility, i.e. $p_i(z_i)$, under the condition that the illegal activity takes place in facility i. Note that the $p_i(z_i)$ are conditional probabilities which do not need to add up to one when summing with respect to the conditional event (see the example above). The probabilities $p_i(z_i)$ are related to each other only via the effort boundary condition; see (1) below. Also note that according to $(z_1, ..., z_K)$, in any of the facilities with $z_i > 0$ an inspection is performed by applying the inspection measures $z_i, i = 1, ..., K$. Assumptions on $p_i(z_i)$ are given in (7).

Also assume that the unit inspection measure in the i-th facility requires the inspection effort $w_i (> 0)$, $i = 1, ..., K$ and that the total available inspection effort is fixed,

$$\sum_{i=1}^{K} w_i z_i = w .$$

Therefore, the Inspector's strategy set is

$$Z := \left\{ \boldsymbol{z} = (z_1, ..., z_K): 0 \leq z_i, i = 1, ..., K, \sum_{i=1}^{K} w_i z_i = w \right\}. \quad (1)$$

Let q_i, $i = 1, ..., K$, be the probability that the illegal activity takes place in the i-th facility and let q_0 be the probability for legal behaviour. These probabilities sum to unity, because as stated above, the Operator will act illegally in at most one facility. Therefore, the Operator's strategy set is

$$Q := \left\{ \boldsymbol{q} = (q_0, q_1, ..., q_K): 0 \leq q_i \leq 1, \right. \\ \left. i = 0, ..., K, q_0 + \sum_{i=1}^{K} q_i = 1 \right\}. \quad (2)$$

In a non-cooperative two-person game formulation of this inspection problem, the payoffs to the Inspector (player 1) and to the Operator (player 2) are given by

$(-a_i, -b_i)$ for detected illegal activity of the Operator in facility i

$(-c_i, d_i)$ for undetected illegal activity of the Operator in facility i ,   (3)

$(0,0)$ for legal behavior of the Operator in all facilities

where we have for all $i = 1, ..., K$

$$0 < a_i < c_i \quad \text{and} \quad 0 < Min(b_i, d_i) .$$

Note that $a_i > 0$ since the highest priority of the Inspector is to deter the Operator from illegal behaviour.

By (3), the expected payoff to both players, conditional on the facility i, $i = 1, ..., K$, at which the illegal activity is performed, is, for all $i = 1, ..., K$, given by

$$-a_i p_i(z_i) - c_i(1 - p_i(z_i)) \quad \text{and} \\ -b_i p_i(z_i) + d_i(1 - p_i(z_i)). \quad (4)$$

Define for all $i = 1, ..., K$

$$A_i := b_i + d_i \quad \text{and} \quad B_i := c_i - a_i . \quad (5)$$

Because the Operator behaves illegally in at most one facility (see above), and because the probability of behaving illegally in facility i is q_i, the (unconditional) expected payoffs to both players are, using (4) and (5), for all $\boldsymbol{z} \in Z$ and for all $\boldsymbol{q} \in Q$ given by

$$In(\boldsymbol{z}, \boldsymbol{q}) := \sum_{i=1}^{K} (-c_i + B_i p_i(z_i)) q_i \quad \text{and} \\ Op(\boldsymbol{z}, \boldsymbol{q}) := \sum_{i=1}^{K} (d_i - A_i p_i(z_i)) q_i . \quad (6)$$

By (1), (2) and (6), a non-cooperative two-person game (Z, Q, In, Op) is defined.

The functional dependence $p_i(z_i)$ is assumed to be strictly monotonically increasing and strictly concave, i.e., for $z_i \geq 0$ and for all $i = 1, ..., K$,

$$0 \leq p_i(z_i) \leq 1, \quad p_i(0) = 0, \quad \text{and} \\ \frac{dp_i(z_i)}{dz_i} > 0 \qquad \frac{d^2 p_i(z_i)}{dz_i^2} < 0 . \quad (7)$$

Justification of (7): One can assume reasonably that the higher the inspection measures z_i in facility i, the higher the conditional detection probability $p_i(z_i)$. Thus, $p_i^{-1}(z_i)$ must be monotone increasing. The strictly monotone behaviour of $p_i(z_i)$ assures the existence of its inverse $p_i^{-1}(z_i)$, i.e., $p_i\left(p_i^{-1}(z_i)\right) = p_i^{-1}\left(p_i(z_i)\right) = z_i$ for all $z_i \geq 0$ and all $i = 1, \dots, K$. The strict concavity of $p_i(z_i)$ is needed to assure a global maximum of the Inspector´s expected payoff; see the proof of the Theorem.

In chapter 6 of [4] an inspection model is considered which is, with respect to modelling, very different to the inspection model described in this paper, but its game theoretical solution, i.e., the Nash equilibrium, is a special case of the game theoretical solution of the inspection game presented in this paper. Therefore, for reasons of comparisons, we present the inspection game of chapter 6 of [4] in some detail.

The Inspector chooses the facility in which the inspection is performed with probability $z_i, i = 1, \dots, K$, and only one facility is inspected. Therefore, the Inspector's strategy set in [4] is given by

$$\tilde{Z} := \left\{ \boldsymbol{z} = (z_1, \dots, z_K) : 0 \leq z_i, i = 1, \dots, K, \sum_{i=1}^{K} z_i = 1 \right\}.$$

If we put $w = w_i = 1$ for all $i = 1, \dots, K$, then we have, using (1), $\tilde{Z} = Z$ and the Inspector's strategy sets coincide. To illustrate the difference in the meaning of z_i in both inspection models, consider $K = 3$ facilities with $w = w_1 = w_2 = w_3 = 1$ and assume that w and w_i are measured in hours. If the Inspector plays $\boldsymbol{z} = (z_1, z_2, z_3) = (1/3, 1/3, 1/3)$, then he performs in the inspection model described in this paper an inspection in all three facilities each one lasting 20 minutes. In the inspection model in [4], however, only one of the three facilities is inspected, and each one is selected with probability $1/3$.

In [4], the Operator chooses the facility in which the illegal activity will take place with probability $q_i, i = 1, \dots, K$, whereby only illegal behaviour is considered. Thus, the Operator's strategy set in [4] coincides with the Operator's strategy set (2) if we assume $q_0 = 0$ in (2).

Regarding the payoffs to both players, in [4] it is assumed that if the inspection is performed in the same facility in which the illegal activity takes place, then detection happens with detection probability $1 - \beta_i, i = 1, \dots, K$. In [4] it is shown that the (unconditional) expected payoffs to both players are then given by

$$\widetilde{In}(\boldsymbol{z}, \boldsymbol{q}) = \sum_{i=1}^{K} (-c_i + B_i(1 - \beta_i)z_i)q_i \quad \text{and}$$

$$\widetilde{Op}(\boldsymbol{z}, \boldsymbol{q}) = \sum_{i=1}^{K} (d_i - A_i(1 - \beta_i)z_i)q_i \,, \tag{8}$$

where A_i and B_i are given by (5). Comparing the payoffs (6) and (8) we see that if we put $p_i(z_i) = (1 - \beta_i)z_i$ for all $i = 1, \dots, K$, then the payoffs in (6) simplify to the payoffs in (8). Therefore, the inspection model in this paper is a far more general inspection model than that in chapter 6 of [4], and in case of $p_i(z_i) = (1 - \beta_i)z_i$ and $w = w_i = 1$ for all $i = 1, \dots, K$, both inspection games formally coincide. However, as mentioned above, both inspection models describe very different inspection problems, and it is surprising that they lead to the same forms for the expected payoffs, which then result in corresponding Nash equilibria.

Note that in [2] and [4] the special case $p_i(z_i) = z_i$, $i = 1, \dots, K$, is also considered in a $(K + 1)$-person game with K independently acting Operators, each responsible for one facility only. We will come back to this model in section 6.

3. Nash Equilibria

In this section we solve the non-cooperative two-person game (Z, Q, In, Op) by determining the so-called Nash-equilibrium; see [8]. A Nash equilibrium is a pair of strategies with the property that unilateral deviation of one player from its equilibrium strategy does not improve the deviator's payoff. Formally, the pair of strategies $(\boldsymbol{z}^*, \boldsymbol{q}^*)$ with $\boldsymbol{z}^* \in Z$ and $\boldsymbol{q}^* \in Q$ constitutes a Nash equilibrium of the game (Z, Q, In, Op) if and only if the Nash the equilibrium conditions

$$
\begin{aligned}
In^* &:= In(\boldsymbol{z}^*, \boldsymbol{q}^*) \geq In(\boldsymbol{z}, \boldsymbol{q}^*) && \text{for all} \quad \boldsymbol{z} \in Z \\
Op^* &:= Op(\boldsymbol{z}^*, \boldsymbol{q}^*) \geq Op(\boldsymbol{z}^*, \boldsymbol{q}) && \text{for all} \quad \boldsymbol{q} \in Q
\end{aligned} \tag{9}
$$

are fulfilled. Because of (7), the existence of a Nash equilibrium for the game (Z, Q, In, Op) can be assured using the Theorem by Nikoida-Isoda; see [9].

The Nash equilibrium of the game (Z, Q, In, Op) is given in

Theorem. Given the non-cooperative inspection game (Z, Q, In, Op) and assume that (7) is fulfilled. Without loss of generality assume

$$d_1 > d_2 > \cdots > d_K > d_{K+1} =: 0 \,. \tag{10}$$

Let $1 \leq k \leq K + 1$ be chosen so that

$$\sum_{i=1}^{k-1} w_i p_i^{-1}\left(\frac{d_i - d_k}{A_i}\right) < w \leq \sum_{i=1}^{k} w_i p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right) \tag{11}$$

where $p_i^{-1}()$ is the inverse function of $p_i()$ for

$i = 1, \dots, K$, and the first inequality is to be ignored for $k = 1$ and the second for $k = K + 1$.

(i) For $1 \leq k \leq K$ equilibrium strategies for the two players are

$$
z_i^* = \begin{cases} p_i^{-1}\left(\dfrac{d_i - Op^*}{A_i}\right) & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases}, \tag{12}
$$

and

$$
q_i^* = \begin{cases} 0 & for \quad i = 0 \\[2mm] \dfrac{\dfrac{w_i}{B_i \dfrac{dp_i(z_i)}{dz_i}\Big|_{z_i = z_i^*}}}{\sum_{j=1}^{k} \dfrac{w_j}{B_j \dfrac{dp_j(z_j)}{dz_j}\Big|_{z_j = z_j^*}}} & for \quad i = 1, \dots, k \\[4mm] 0 & for \quad i = k+1, \dots, K \end{cases} \tag{13}
$$

The equilibrium payoff Op^* to the Operator is given implicitly by

$$
w = \sum_{i=1}^{k} w_i p_i^{-1}\left(\frac{d_i - Op^*}{A_i}\right), \tag{14}
$$

and it satisfies the condition

$$
d_{k+1} \leq Op^* \leq d_k. \tag{15}
$$

The equilibrium payoff In^* to the Inspector is given by

$$
In^* = \sum_{i=1}^{k} \left(-c_i + B_i \frac{d_i - Op^*}{A_i}\right) q_i^*, \tag{16}
$$

where q_i^* and Op^* are given by (13) and (14).

(ii) For $k = K + 1$, i.e., with (10) and (11) for

$$
\sum_{i=1}^{K} w_i p_i^{-1}\left(\frac{d_i}{A_i}\right) < w, \tag{17}
$$

the set of equilibrium strategies of the Inspector is, for all $i = 1, \dots, K$, given by

$$
z_i^* \geq p_i^{-1}\left(\frac{d_i}{A_i}\right) \quad with \quad w = \sum_{i=1}^{K} w_i z_i^*. \tag{18}
$$

The equilibrium strategy of the Operator is $q_0^* = 1$ and $q_i^* = 0$ for all $i = 1, \dots, K$, i.e., legal behaviour of the Operator. The payoffs to both players are zero.

Proof. 1) We show that the inequalities in (11) cover the whole parameter space. The proof goes along the same lines as in the proof of Theorem 6.2 in [4]: We show that, for given values of d_i/A_i, $i = 2, \dots, k$, the inequalities in (11)

cover all values of d_1/A_1. For $2 \leq k \leq K$ and with (7), both inequalities in (11) are equivalent to

$$
\frac{d_{k+1}}{A_1} + p_1\left(\frac{w}{w_1} - \sum_{i=2}^{k} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right)\right)
$$

$$
\leq \frac{d_1}{A_1}
$$

$$
< \frac{d_k}{A_1} + p_1\left(\frac{w}{w_1} - \sum_{i=2}^{k-1} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_k}{A_i}\right)\right)
$$

Thus, we have to show that

$$
\frac{d_{k+1}}{A_1} + p_1\left(\frac{w}{w_1} - \sum_{i=2}^{\kappa} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right)\right)
$$

$$
< \frac{d_k}{A_1} + p_1\left(\frac{w}{w_1} - \sum_{i=2}^{k-1} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_k}{A_i}\right)\right).
$$

This is equivalent to

$$
p_1\left(\frac{w}{w_1} - \sum_{i=2}^{k} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right)\right)
$$

$$
- p_1\left(\frac{w}{w_1} - \sum_{i=2}^{k-1} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_k}{A_i}\right)\right) \tag{19}
$$

$$
< \frac{d_k - d_{k+1}}{A_1}.
$$

We show that the left-hand side of (19) is less than zero and hence, as the right-hand side is by (10) larger than zero, that the inequality holds. By (10) we have $d_k - d_{k+1} > 0$ as well as $d_i - d_{k+1} > d_i - d_k$ and thus, the monotonicity of $p_i^{-1}()$ and $w_i > 0$, $i = 1, \dots, K$, implies

$$
\sum_{i=2}^{k} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right) > \sum_{i=2}^{k-1} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_{k+1}}{A_i}\right)
$$

$$
> \sum_{i=2}^{k-1} \frac{w_i}{w_1} p_i^{-1}\left(\frac{d_i - d_k}{A_i}\right),
$$

which implies that the left-hand side of (19) is less than zero. For $k = 1$ the second inequality in (11) is equivalent to

$$
\frac{d_1}{A_1} \geq \frac{d_2}{A_1} + p_1\left(\frac{w}{w_1}\right),
$$

and for $k = K + 1$, the first inequality in (11) is, because of $k = K + 1$, equivalent to

$$\frac{d_1}{A_1} < p_1 \left(< \frac{w}{w_1} - \sum_{i=2}^{K} \frac{w_i}{w_1} p_i^{-1} \left(\frac{d_i}{A_i} \right) \right),$$

which completes the first part of the proof.

2) We show that (15) holds for $1 \le k \le K$. Assume $Op^* > d_k$. With (14) this implies

$$w = \sum_{i=1}^{k} w_i p_i^{-1} \left(\frac{d_i - Op^*}{A_i} \right) < \sum_{i=1}^{k} w_i p_i^{-1} \left(\frac{d_i - d_k}{A_i} \right)$$

$$= \sum_{i=1}^{k-1} w_i p_i^{-1} \left(\frac{d_i - d_k}{A_i} \right),$$

which is a contradiction to the first inequality in (11). Assume $Op^* < d_{k+1}$. Then (14) implies

$$w = \sum_{i=1}^{k} w_i p_i^{-1} \left(\frac{d_i - Op^*}{A_i} \right) > \sum_{i=1}^{k} w_i p_i^{-1} \left(\frac{d_i - d_{k+1}}{A_i} \right),$$

which is a contradiction to the second inequality in (11).

3) Because the functions $p_i(z_i)$ are monotone increasing and $p_i(0) = 0$ (12) is equivalent to

$$p_i(z_i^*) = \begin{cases} \dfrac{d_i - Op^*}{A_i} & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases} . \qquad (20)$$

Note that $p_i(z_i^*) \ge 0$ is equivalent to $d_i \ge Op^*$ for all $i = 1, \dots, k$ which holds because of (10) and (15), and $p_i(z_i^*) \le 1$ is equivalent to $Op^* \ge d_i - A_i$ which holds because of $Op^* > 0 \ge d_i - A_i = -b_i$ for all $i = 1, \dots, k$. Thus, we have, with (2), (6), (10) and (20),

$$Op(z^*, q) = \sum_{i=1}^{K} (d_i - A_i p_i(z_i^*)) q_i$$

$$= Op^* \sum_{i=1}^{k} q_i + \sum_{i=k+1}^{K} d_i q_i$$

$$\le Op^* \sum_{i=1}^{k} q_i + \sum_{i=k+1}^{K} Op^* q_i$$

$$\le Op^* \sum_{i=1}^{K} q_i \le Op^*$$

for all $q \in Q$, i.e., the second inequality in (9) is satisfied.

Before we derive the Inspector's equilibrium strategy, we note that for the Operator's equilibrium strategy we must have $q_i^* = 0$ for all $i = k+1, \dots, K$: Suppose $q_j^* > 0$ for a

facility $j, j = k+1, \dots, K$, then the Inspector would have to allocate some of the inspection effort in facility j. However, because of $z_i^* = 0$ for all $i = k+1, \dots, K$, he does not allocate inspection effort in facility j, we must have $q_i^* = 0$ for all $i = k+1, \dots, K$.

Using the first inequality in (9), we determine q^* such that $In(z, q^*)$ is maximized with respect to z, and we apply the Lagrange formalism. Using the Lagrange function $L(z, q)$ given by

$$L(z, q) = In(z, q) + \lambda \left(w - \sum_{j=1}^{k} w_j z_j \right), \qquad (21)$$

we determine q^* such that for all $i = 1, \dots, k$

$$\left. \frac{\partial}{\partial z_i} L(z, q^*) \right|_{z=z^*} = 0 ,$$

where the Lagrange parameter λ is determined with the help of the normalization of the q^*. Using (6) and (21), the condition

$$\frac{\partial}{\partial z_i} L(z, q^*) = B_i \frac{dp_i(z_i)}{dz_i} q_i^* - \lambda w_i = 0$$

implies for all $i = 1, \dots, k$

$$q_i^* = \lambda w_i / \left(B_i \frac{dp_i(z_i)}{dz_i} \right),$$

and leads, by using the normalization in (2), to (13). The condition $0 \le q_i^* \le 1$ for all $i = 0, 1, \dots, K$ is obvious, and we even have $0 < q_i^*$ for all $i = 1, \dots, k$. The Hessian $HL(z, q)$ of the Lagrange function $L(z, q)$ is a diagonal matrix:

$$HL(z, q) = \begin{pmatrix} B_1 \frac{d^2 p_1(z_1)}{dz_1^2} q_1 & 0 & \cdots & 0 \\ 0 & B_2 \frac{d^2 p_2(z_2)}{dz_2^2} q_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k \frac{d^2 p_k(z_k)}{dz_k^2} q_k \end{pmatrix} .$$

Therefore, the eigenvalues of the Hessian $HL(z, q)$ are the diagonal elements. Because $B_i > 0$ and $p_i(z_i)$ are all assumed to be strictly concave functions (see (7)), all eigenvalues of the Hessian $HL(z, q^*)$ are smaller than zero, and thus, the Hessian is negative definite. Therefore, $L(z, q^*)$ is a concave function and Theorem B in [10] implies that z^* is even (because of the concavity of $In(z, q^*)$) a global maximum of $In(z, q^*)$ with respect to z.

4) For $k = K + 1$ we see immediately that (18) and $q_0^* = 1$ satisfy the equilibrium conditions (9). This completes the proof.

Three comments on the Theorem. First, as mentioned at the end of section 2, the inspection model in chapter 6 of [4] can be seen as a special case of the inspection model considered in this paper: Put $p_i(z_i) = (1 - \beta_i)z_i$ and $w = w_i = 1$ for all $i = 1, \dots, K$. Because we have here $d^2 p_i(z_i)/dz_i^2 = 0$ for all $i = 1, \dots, K$, the last condition in (7) is not fulfilled, and thus, the Theorem cannot be applied. Surprisingly, however, the Theorem also covers this case. This can be seen by comparing the Nash equilibrium obtained in case (i) of the Theorem with the Nash equilibrium presented in Theorem 6.2 in [4]. They coincide. Note that because in [4] only illegal behaviour of the Operator is considered, i.e., $q_0 = 0$, case(ii) of the Theorem is not part of Theorem 6.2 in [4].

Second, there are some general features of this solution which are typical for inspection games of this kind, for example, the fact that the equilibrium strategies depend only on the system parameters of the adversaries, or the so-called *cone of deterrence* (18); see [11]. Since, however, it is difficult to discuss more properties of the solution in general terms, we look at an inspection regime based on attribute sampling procedures in section 4 and analyse a time constrained inspection model in section 5.

Third, the Theorem presents a Nash equilibrium but does not address the issue whether there are further Nash equilibria. Indeed, the uniqueness of Nash equilibria in this inspection game is an open question. In section 4 we make a short comment on the uniqueness in case of $K = 2$ facilities and that $p_i(z_i), i = 1,2,$ depends linearly on z_i.

4. First Application: Attribute Sampling

Consider the problem of safeguarding nuclear material in connection with the NPT, in which the role of the Inspector, or player 1, in the model of this paper is played by the IAEA in Vienna: There are K storage facilities for spent nuclear fuel elements in a State (or community of States such as the European Union), operated by an Operator, or player 2. The i-th storage facility contains N_i fuel elements, $i = 1, \dots, K$, the inspection of one of which requires the effort w_i. Thus, if n_i items in the i-th facility are verified, the total inspection effort is

$$w = \sum_{i=1}^{K} w_i n_i. \tag{22}$$

For $w_i = w_1, i = 2, \dots, K$, (22) means that the total number of elements to be verified is fixed. Of course, the number $n_i, i = 1, \dots, K$, of verified items in facility i is a nonnegative integer by their very nature. To be able to apply our Theorem, we have to consider n_i as continuous variable. In the applications we have in mind, the n_i may go into the hundreds, therefore, we assume that in these cases the error is small, if nonnatural n_i^* are rounded to natural ones such

that the boundary condition is maintained. In the second application (see section 5) there does not exist such a problem.

Furthermore, we assume that in the sense of the NPT it is necessary to detect the diversion of at least one fuel element in one of the facilities. Let us first consider the case that, in order to acquire a so-called *significant quantity* of nuclear material (see [5]), just one fuel element needs to be diverted. The diversion strategy involves replacing the removed fuel element by a dummy.

If the diversion takes place in facility i, then the conditional detection probability (see section 2) in case of drawing without replacement is for all $i = 1, \dots, K$, given by

$$p_i = 1 - \frac{\binom{N_i - 1}{n_i}}{\binom{N_i}{n_i}} = \frac{n_i}{N_i}. \tag{23}$$

Because we have all $i = 1, \dots, K$

$$p_i(0) = 0, \quad \frac{dp_i(n_i)}{dn_i} = \frac{1}{N_i} > 0 \quad \text{and} \quad \frac{d^2 p_i(n_i)}{dn_i^2} = 0,$$

(7) is fulfilled except the last condition. It can be shown, however, that also in this specific case the results of the Theorem are valid; see the first comment after the Theorem. Because (23) is equivalent to $n_i = N_i p_i$ for $i = 1, \dots, K$, we get in case of $1 \leq k \leq K$ by (12) through (16) the equilibrium strategy of the Inspector

$$n_i^* = \begin{cases} N_i \dfrac{d_i - Op^*}{A_i} & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases}, \tag{24}$$

where k according to (11) is given by

$$\sum_{i=1}^{k-1} w_i N_i \frac{d_i - d_k}{A_i} < w \leq \sum_{i=1}^{k} w_i N_i \frac{d_i - d_{k+1}}{A_i},$$

and where with (14) the equilibrium payoff Op^* to the Operator is given by

$$w = \sum_{i=1}^{k} w_i N_i \frac{d_i - Op^*}{A_i}.$$

The equilibrium strategy of the Operator is

$$q_i^* = \begin{cases} 0 & for \quad i = 0 \\ \dfrac{w_i N_i}{B_i} \dfrac{1}{\sum_{j=1}^{k} \dfrac{w_j N_j}{B_j}} & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases}, \tag{25}$$

and the equilibrium payoff In^* to the Inspector is

$$In^* = \sum_{i=1}^{k} \left(-c_i + B_i \frac{d_i - Op^*}{A_i} \right) q_i^*, \qquad (26)$$

where q_i^* is given by (25). For $k = K + 1$ and with (17) the condition for legal behaviour of the Operator is

$$\sum_{i=1}^{K} w_i N_i \frac{d_i}{A_i} < w, \qquad (27)$$

and with (18) the set of equilibrium strategies of the Inspector is for all $i = 1, \dots, K$, given by

$$n_i^* \geq N_i \frac{d_i}{A_i} \quad \text{and} \quad w = \sum_{i=1}^{K} w_i n_i^*. \qquad (28)$$

At this point we make a remark on the uniqueness of the Nash equilibria given by the Theorem. For $K = 2$ facilities it can be shown that (24) through (28) represent the only Nash equilibrium of the game. Also, it can be shown that in case of legal behaviour of the Operator, i.e., (27) holds, (24) is not an equilibrium strategy of the Inspector for $k = 1$ whereas this is so for $k = 2$.

Now we assume that, in order to acquire a significant quantity of nuclear material, it is necessary to divert not one but two fuel elements, again by replacing them by dummies. If the diversion takes place in facility i, then the conditional detection probability (see section 2) is, for all $i = 1, \dots, K$, given by

$$1 - \frac{\binom{N_i - 2}{n_i}}{\binom{N_i}{n_i}} = 1 - \left(1 - \frac{n_i}{N_i} \right) \left(1 - \frac{n_i}{N_i - 1} \right),$$

or, for our purposes in case of $N_i \gg 1$,

$$p_i = 1 - \left(1 - \frac{n_i}{N_i} \right)^2. \qquad (29)$$

(7) is fulfilled, because

$$p_i(0) = 0, \quad \frac{dp_i(n_i)}{dn_i} = \frac{2}{N_i} \left(1 - \frac{n_i}{N_i} \right) > 0 \quad \text{and}$$

$$\frac{d^2 p_i(n_i)}{dn_i^2} = -\frac{2}{N_i^2} < 0.$$

From (29) we get $n_i = N_i \left(1 - \sqrt{1 - p_i} \right)$ for all $i = 1, \dots, K$, and thus, for $1 \leq k \leq K$ we get from (12) through (16) the equilibrium strategy of the Inspector

$$n_i^* = \begin{cases} N_i \left(1 - \sqrt{1 - \dfrac{d_i - Op^*}{A_i}} \right) & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases} \qquad (30)$$

where k according to (11) is given by

$$\sum_{i=1}^{k-1} w_i N_i \left(1 - \sqrt{1 - \frac{d_i - d_k}{A_i}} \right)$$

$$< w$$

$$\leq \sum_{i=1}^{k} w_i N_i \left(1 - \sqrt{1 - \frac{d_i - d_{k+1}}{A_i}} \right),$$

and where with (14) the equilibrium payoff Op^* to the Operator is given by

$$w = \sum_{i=1}^{k} w_i N_i \left(1 - \sqrt{1 - \frac{d_i - Op^*}{A_i}} \right).$$

The equilibrium strategies of the Operator are

$$q_i^* = \begin{cases} 0 & for \quad i = 0 \\ \dfrac{w_i N_i / (B_i (N_i - n_i^*))}{\sum_{j=1}^{k} \dfrac{w_j N_j}{B_j (N_j - n_j^*)}} & for \quad i = 1, \dots, k \\ 0 & for \quad i = k+1, \dots, K \end{cases} \qquad (31)$$

and the equilibrium payoff In^* to the Inspector is

$$In^* = \sum_{i=1}^{k} \left(-c_i + B_i \left(1 - \left(1 - \frac{n_i^*}{N_i} \right)^2 \right) \right) q_i^*,$$

where n_i^* and q_i^* are given by (30) and (31). For $k = K + 1$ and with (17) the condition for legal behaviour is

$$\sum_{i=1}^{K} w_i N_i \left(1 - \sqrt{1 - \frac{d_i}{A_i}} \right) < w, \qquad (32)$$

and the set of equilibrium strategies of the Inspector is, for all $i = 1, \dots, K$, given by

$$n_i^* \geq N_i \left(1 - \sqrt{1 - \frac{d_i}{A_i}} \right) \quad \text{and} \quad w = \sum_{i=1}^{K} w_i n_i^*. \qquad (33)$$

Let us compare conditions (27) and (32) for legal behaviour: Because of

$$\frac{d_i}{A_i} > 1 - \sqrt{1 - \frac{d_i}{A_i}}, i = 1, \dots, k,$$

the inspection effort w according to (27) has to be larger than that according to (32) which is reasonable: In the former case the number of manipulated fuel element is smaller than in the latter, therefore it is more difficult to detect them.

5. Second Application: Time Constrained Inspections

Let us consider next the problem of drug control at a large international seaport. Assume that K ships of varying sizes have arrived from South American ports of origin and are being unloaded. The port Customs Authority has in total T man-hours at its disposal for inspection of the cargoes for concealed drugs. We might model the detection probability for the i-th ship, $i = 1, \ldots, K$, as a function of allotted control time as

$$p_i(t_i) = 1 - \exp(-t_i/\lambda_i) \text{ for } t_i \geq 0 \tag{34}$$

with parameters $\lambda_i > 0$. The expected detection time λ_i will increase with the size of the ship (or of its cargo). We have for all $i = 1, \ldots, K$

$$0 \leq p_i(t_i) \leq 1, \quad p_i(0) = 0, \quad \frac{dp_i(t_i)}{dt_i} = \frac{1}{\lambda_i} \exp(-t_i/\lambda_i)$$

$$\frac{d^2 p_i(t_i)}{dz_i^2} = -\frac{1}{\lambda_i^2} \exp(-t_i/\lambda_i) < 0 \, ,$$

so that (7) is fulfilled. According to our assumptions, we have

$$\sum_{i=1}^{K} t_i = T \, .$$

We assume additionally that, if drugs are actually being smuggled, it is under the control of a single organization, in the following called Smuggler. In section 6 we will sketch the case that there are several independent Smugglers.

From the Theorem we get with $w_i = w_1$ for $i = 2, \ldots, K$ and $w/w_1 = T$ the following equilibrium strategies and payoffs. For $1 \leq k \leq K$, the equilibrium strategy of the Customs Authority is by (12)

$$t_i^* = \begin{cases} \lambda_i \, \mathrm{Ln}\left(\dfrac{b_i + d_i}{b_i + Op^*}\right) & for \quad i = 1, \ldots, k \\ 0 & for \quad i = k+1, \ldots, K \end{cases} \, ,$$

where according to (11) k is given by

$$\sum_{i=1}^{k} Ln\left(1 - \frac{d_i - d_{k+1}}{A_i}\right) \leq -T < \sum_{i=1}^{k-1} Ln\left(1 - \frac{d_i - d_k}{A_i}\right) ,$$

and where the equilibrium payoff Op^* to the Smuggler is given by

$$T = Ln\left(\prod_{i=1}^{k} \left(\frac{b_i + d_i}{b_i + Op^*}\right)^{\lambda_i}\right) .$$

Furthermore, the equilibrium strategy of the Smuggler is by (13)

$$q_i^* = \begin{cases} 0 & for \quad i = 0 \\ \dfrac{\dfrac{\lambda_i}{B_i} \dfrac{b_i + d_i}{b_i + Op^*}}{\sum_{j=1}^{k} \dfrac{\lambda_j}{B_j} \dfrac{b_j + d_j}{b_j + Op^*}} & for \quad i = 1, \ldots, k \\ 0 & for \quad i = k+1, \ldots, K \end{cases} ,$$

and the equilibrium payoff In^* to the Customs Authority is given by (16). For $k = K + 1$, and with (17), the condition for legal behaviour is

$$\sum_{i=1}^{K} \lambda_i Ln\left(1 + \frac{d_i}{b_i}\right) < T \, , \tag{35}$$

and the set of equilibrium strategies of the Customs Authority is, for all $i = 1, \ldots, K$, given by

$$t_i^* \geq \lambda_i Ln\left(1 + \frac{d_i}{b_i}\right) \quad with \quad T = \sum_{i=1}^{K} t_i^* \, . \tag{36}$$

Here, the lower limit for t_i^* is proportional to the expected detection time and increases monotonically with the ratio d_i/b_i, i.e., with the ratio of the smuggler's incentive to its punishment in the event of detection.

6. Summary and outlook

As already mentioned at the end of section 2, in [2] and in [4] the special case $p_i(z_i) = z_i$ for all $i = 1, \ldots, K$, is also considered in a $(K + 1)$-person game with K independently acting Operators, each responsible for one facility only.

Without going into all details of sections 2 and 3, this inspection problem can be analysed in the same way as before under the same assumptions for the Inspector: Whereas the Inspector's strategy set and its (unconditional) expected payoff are again given by (1) and the first equation of (6), the strategy set of the i-th Operator, $i = 1, \ldots, K$, and the corresponding (unconditional) expected payoff are given by

$$Q_i = \{q_i : 0 \leq q_i \leq 1\} \quad and \quad Op_i(z, q) := (d_i - A_i p_i) q_i \, .$$

We will not formulate the solution of this $(K + 1)$-person game as a Theorem. We just report that – not surprisingly – the condition for legal behaviour of all K Operators is again given by (17), respectively (27), (32) and (35), and the cone of deterrence, i.e., the set of strategies of the Inspector in case of legal behaviour of all Operators, again by (18), respectively (28), (33) and (36).

We think that these results, together with those presented in the Theorem and of earlier work, e.g., in [11], describe a universally valid structure of the problem of deterring persons, organizations or even States from illegal behaviour by appropriate inspections.

As a future activity we plan to contact an oversea port authority to discuss according to which criteria its inspection

resources are distributed and whether the use of the detection probability (34) is appropriate in the time constrained inspection model of section 5.

7.   Disclaimer

Neither the authors of this paper, nor the organization or industrial company they are affiliated with, nor the Federal Government of Germany assume any liability whatsoever for any use of this paper or parts of it. Furthermore, the content of this paper does not reflect any policy of the Federal Government of Germany.

8.   References

[1]   Avenhaus, R., Kilgour, D.M.; Efficient distributions of arms-control inspection effort; Naval Research Logistics; vol. 57; 2004; pp. 1-17.

[2]   Deutsch, Y., Golany, B., Rothblum, U.; Determining all Nash Equilibria in a (bi-linear) Inspection Game; European Journal of Operational Research; vol. 215; 2011; pp. 422-430.

[3]   Deutsch, Y., Golany, B., Rothblum, U.; Inspection games with local and global allocation bounds; Naval Research Logistics; vol. 60; 2013; pp. 125-140..

[4]   Avenhaus, R., Canty, M.J.; Compliance Quantified - An Introduction to Data Verification; Cambridge; UK: Cambridge University Press; 1996.

[5]   IAEA; IAEA Safeguards Glossary, 2001 Edition; (IAEA International Nuclear Verification Series No. 3); Vienna; IAEA; 2002.

[6]   Avenhaus, R., von Stengel, B., Zamir, S.; Inspection games; in J. Aumann and S. Hart (Eds.): Handbook of Game Theory with Economic Applications; Elsevier North Holland; 2002; pp. 1947-1987.

[7]   Avenhaus, R., Canty, M.J.; Inspection Games; in Meyers, Robert A. (Ed.): Encyclopedia of Complexity and Systems Science; New York, NY; Springer; 2009; pp. 4855-4868.

[8]   Nash, J.F.; Non-Cooperative Games; Annals of Mathematics; vol. 54; no. 2; 1951; pp. 286-295.

[9]   Nikaido, H., Isoda, K.; Note on noncooperative convex games; Pacific Journal of Mathematics; vol. 5; 1955; pp. 807-815.

Examining Autonomous Inspection of Geologic Repositories

Jacob Benz[1], Camille Palmer[2], HaliAnne McGee-Hilbert[1], Chris (Yu Hsuan) Lee[2], Ravi Kumar[2], Geoffrey Hollinger[2], Nicole Benker[1]

[1]Pacific Northwest National Laboratory, Richland, WA 99352
[2]Oregon State University, Corvallis, OR 97333
E-mail: Jacob.Benz@pnnl.gov; Camille.Palmer@oregonstate.edu; HaliAnne.McGee-Hilbert@pnnl.gov;
LeeYuh@oregonstate.edu; Geoff.Hollinger@oregonstate.edu;

Abstract

Geological repositories for nuclear waste, including spent nuclear fuel, present a significant challenge for traditional International Atomic Energy Agency (IAEA) safeguards tools due to their inaccessibility and demanding operational conditions. The IAEA has been working closely with Member State organizations currently involved in repository construction and planning including Euratom, the Finnish and Swedish regulatory authorities, and relevant facility operators. A verification challenge for the IAEA is to verify that no nuclear material is diverted from a repository environment. The challenge is also not static as activities must encompass verification of the design prior to and during the construction/operation phase, and post backfill. Throughout these various phases, it is imperative that the IAEA maintains a continuity of knowledge (CoK) of all material, including information on material inventory and flow. Monitoring via autonomous robotic vehicles may augment current and envisioned IAEA safeguards approaches for geologic repositories. Implementing automated solutions for underground nuclear repository inspection may be a new venture for the IAEA but carries the potential to greatly enhance the efficacy and efficiency of inspections and allow inspectors' time and expertise to be directed where needed most. This paper highlights these challenges and outlines how they might be addressed by using remote or autonomous vehicles. Specifically, it discusses the current state of the art in robotic autonomy for known or partially known environment mapping and patrolling, as well as shared autonomy, where humans collaborate with closed loop autonomation to complete tasks. The feasibility of using rovers for these verification tasks is explored, along with the challenges associated with system implementation. Hardware and software suggestions are provided based on the adoption of similar technologies in other comparable areas and ability to close technical gaps. Lastly, human-robotic interactions are considered based on the challenges of the environment of the repository and effective deployment and continued operation of the robot system.

Keywords: Geologic Repository, Autonomous Monitoring.

1. Introduction

This paper explores possible robotic technologies to aid and augment IAEA inspection of deep geological repositories. The identified technologies can help maintain continuity of knowledge (CoK) of spent fuel stored in buried canisters during the repository's operation phase. The functional lifecycle of a repository will transition through several phases: pre-operation, operation phase, and post-operation. The pre-operation phase involves geological assessment of a spent fuel repository site. The operation phase, the focus of this paper, is the most complex and involves the construction, processing, emplacement, and backfill of canisters containing spent fuel. The post-operation phase addresses closure of the facility and the long-term monitoring and maintenance. The KBS-3 repository is currently at various stages of development and use in Sweden and Finland [1] and its design, based on the KBS-3 method developed by SKB, is further along than any other repository. It will serve as the model repository template for this paper, although the technological evaluation and applicability of the approach are not limited to this particular repository design.

Given that the design and construction of future deep geological repositories will occur over many decades and in a multitude of countries and geological regions, the potential for variability of repository designs is quite likely. These variations coupled with limitations for routine inspector-access and the complications posed by continual State activities occurring within and requiring access to the repository during the operation phase create significant challenges to traditional safeguards solutions for maintaining CoK over spent fuel. During the operation phase, the repository consists of a surface facility (above-ground) and a sub-surface facility (below-ground). The surface facility can serve a variety of functions by simply acting as an entrance to the sub-surface facility, housing an encapsulation plant, and offering temporary storage before emplacement. Due to the accessibility that IAEA inspectors now have to the above-ground facilities, this paper will not focus on technologies to help augment or automate above-ground inspection processes; rather, it will focus on maintaining CoK over the spent fuel in the sub-surface portion of the repository. However, solutions highlighted in this paper may apply to

above-ground challenges such as temporary storage of encapsulated spent fuel.

This paper is the result of a research collaboration between Pacific Northwest National Laboratory and Oregon State University, supported by the U.S. National Nuclear Security Administration's Office of International Nuclear Safeguards. The research explores two aspects of autonomous monitoring critical to identify a feasible deep geologic repository monitoring solution: 1) the level of autonomy of the robotic inspection vehicle; 2) potential technologies which have been, or can be, demonstrated as usable in the challenging repository environment. The following sections provide a high-level introduction to the KBS-3 design, discuss differing degrees of autonomy and potential monitoring technologies, and conclude with example robotic inspection systems that could be evaluated for future use in safeguards applications.

2.   The KBS-3 System

A typical KBS-3 repository system in operation phase consists of underground openings, nuclear waste canisters, buffers, backfill, and engineered barriers, as seen in Figure 1. According to the KBS-3 production report, construction of additional drifts in the underground sections can all

occur concurrently during the operation phase of the nuclear waste facility [1]. A brief summary of the construction process is provided in section 2.1 to provide context for the environments in which the robot systems would operate and to highlight the applicability of the suggested monitoring automation technology; more detailed information is contained in KBS-3 reports [2, 3].

2.1   Repository Construction

The model KBS-3 repository is located deep within bedrock and accessible only through access tunnels and shafts which lead to disposition tunnels, each housing multiple holes or cavities along its length. The spent fuel is encapsulated in copper canisters, described in more detail in Section 2.2, and emplaced in the cavities. Bentonite clay is packed around the canisters and the tunnels are backfilled and plugged to stabilize the sealing. This also prevents access to the canisters. Once the tunnels and shafts are filled, the repository will be closed. The IAEA safety standard SSR-5 gives the safety requirements, the passive safety must be demonstrated but no monitoring shall be left beyond the plug. might be referred here

While the repository is in the operational phase, it is assumed that there will be many tunnels in different stages; construction will continue for portions of the repository
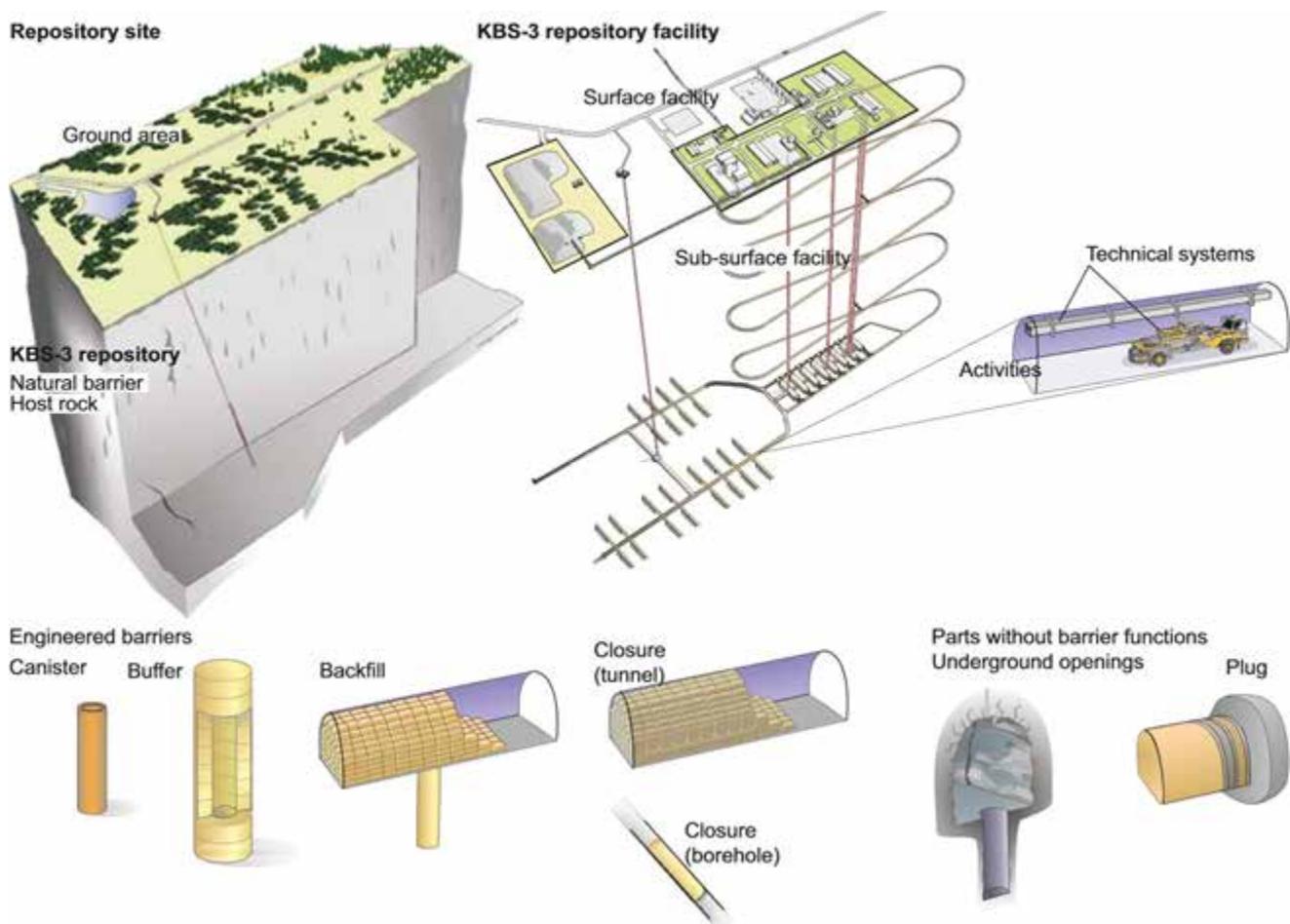
Figure 1: CKBS-3 repository layout [1]

while emplacement and backfill occur in others. As a result, the underground facility will be constantly evolving and it will be important to monitor changes in order to detect whether any undeclared activities of concern have occurred, such as the removal of backfill material or excavation in unauthorized locations, as these activities could indicate attempts to divert previously emplaced material. The emplacement process itself is also valuable to monitor to provide independent assurance that a canister was emplaced in the declared location.

2.2 Spent Fuel Canister

The KBS-3 methodology uses a spent fuel canister design that involves a corrosion-resistant copper shell that encapsulates the spent nuclear fuel assemblies placed in steel-iron internal. The canister is designed to withstand corrosion and mechanical loads anticipated from the surrounding bedrock. Although the canister is designed to accommodate multiple fuel assemblies, safety considerations related to long-term storage of the waste canister inside the repository impose limits on the maximum decay power and radioactivity at the canister's surface – subsequently limiting the number of fuel assemblies allowed. The burn-up and age of the spent fuel assembly inform the calculation of the radioactivity and decay heat of the fuel assembly which governs the number of fuel assemblies that can be encapsulated within a single canister. The KBS-3 canister has specific guidelines for the acceptable decay power and radioactivity at the canister surface, such as:

- Maximum permissible decay power: The KBS-3 safety guidelines explicitly state that the total decay power in each canister should not exceed 1,700W [1]. This limitation should maintain the temperature in the buffer less than 100°C. Temperatures exceeding this value may have adverse impacts on the properties of the engineered barriers and the surrounding rock.

- Radiation dose rate: Similarly, the radiation dose rate at the surface of the canister must be less than 1Gy/h as high radiation levels may lead to the formation of nitric acid and other corrosive species at the canister surface [4].

The canister is deemed ready to be emplaced in the repository once these safety conditions are met.

3. Levels of Autonomy

Autonomous robotics is currently a research topic which is drawing a lot of attention. The field of study delves into various aspects of autonomy such as autonomous mining, unknown environment navigation, task allocation and scheduling, robotic localization, path planning, and multi-robot coordination. A robotic rover can augment in-person/onsite IAEA inspections of the geological waste repositories by providing inspectors with an extension of sensing capabilities to areas that might otherwise be challenging to

access on a regular basis. Some key aspects of rover design and deployment that should be explored are the automation capabilities that allow it to maneuver around the repository environment.

Three levels of automation are discussed in this paper, which could be used for inspection implementation: full autonomy, shared autonomy, and manual teleoperation. The advantages and disadvantages of each, as well as the considerations for state-of-the-art automation implementation, are discussed and described within this section. The feasibility of each approach for IAEA inspection tasks will depend on external factors such as inspection timelines, repository environment, budget, and technological availability and future research is encouraged to fully evaluate their applicability to safeguards objectives. Beyond the scope of this paper, but of relevance to a viable solution, is the inclusion of Euratom where many repositories will be located. Here, implementation of solutions should be considered, as all safeguards equipment must be authorized for common use across Euratom authority.

3.1 Full Autonomy

Fully autonomous operations encapsulate the common perception of robotics—where a robot can self-maneuver and accomplish tasks with minimal to no human intervention. In practice, this requires a large amount of research and development to implement a truly hands-off system. A fully autonomous system can allow an IAEA inspector to deploy several robot systems at once to complete the inspection without overloading the attention capabilities of the inspector. In the case of remote inspection implementation, multiple systems can be deployed at once with the operator simply monitoring the status of each robotic system. This can provide improvements in efficiency over current inspection techniques, especially in those locations where access poses great risk to human inspectors.

This paper discusses potential situations where fully autonomous capabilities may be generally applied to IAEA inspection methods for underground waste repositories, but further detailed research specific to each repository is recommended. Exploration of unknown environments requires algorithmic solutions that process sensor information (e.g., LiDAR scans, camera images) and return a direction of desired travel. Even with the declared ground-truth information of repository design layouts, construction deviations and potentially undocumented drifts may be present which require impromptu assessment and decision-making approaches.

A popular method, frontier-based exploration algorithms, has proven successful and has been commonly applied in the robotics community [5, 6, 7, 8]. Another example solution would be to apply task allocation and scheduling to organize sub-objectives alongside the main objective of nuclear inspection. The deployed system would be capable of

navigation through the environment and may encounter a variety of sub-tasks such as plug inspection scheduling or deposition hole verification [9]. Desired mission outcomes would be programmed beforehand by the operators and algorithms would have to balance operational parameters (e.g., remaining mission duration time and power supply), with inspection goals (i.e., confirming integrity of drift closure versus canister ID and integrity verification in deposition hole) in order to determine which sub-tasks to perform.

The final example considers the case where multiple robot rovers may be deployed simultaneously into the underground repository to maximize the area covered in a single excursion. These robots would require scheduling and coordination between units to optimize efficiency in addressing competing inspection goals [10, 11]. Additionally, this kind of multi-robot coordination would need to consider scenarios in which communication disruptions could occur between one or more units, requiring contingencies to allow each unit to adapt to such situations and make plan adjustments in a decentralized manner [12, 13].

3.2 Shared Autonomy

Shared autonomous operation can incorporate desirable elements from fully autonomous or manual approaches in a modular fashion, without the level of painstaking development required for full automation, and with less demand on operators than required by manual teleoperation. This approach allows situational adaptability, at the cost of certain capabilities. Furthermore, a shared autonomous solution can be implemented as a developmental midpoint between a manual and fully autonomous approach, with incremental features developed, tested, and implemented at separate times.

In addition to research topics for fully autonomous robotics, shared autonomy research topics include fields like human-robotic collaboration and hybrid control schemes. In the case of nuclear repository inspection, an example of human-robot collaboration could be graphical user interface (GUI) control designs to help maximize the productivity of the IAEA inspector without overloading them with inspection results and data. Additionally, verification technologies return information in different forms, such as radiation measurement spectra or images indicating the quality of deposition tunnel backfill. The inspectors' data needs can be studied, and the display options modified for desired traits, output, and controls, directing the inspectors' attention where it will be most effective [14, 15].

An example of a shared autonomous application can involve navigational waypoints with task commands issued to a unit by the inspector. The unit may then execute the commands in an autonomous fashion, or revert to manual teleoperation for more complex, sensitive, or difficult tasks. Additionally, autonomous sub-routines may be installed for

the robot to take independent action if communication with the operator is lost. These routines could be as simple as performing recovery behaviors like backtracking to the last known position within communication range or reinitializing communications channels, or extended behaviors to enable the inspection of areas of interest that fall outside of communication range, such as the exploration of undeclared tunnels or sections. In both instances, the successful execution of the sub-routines will return the robot to a location within communication range, allowing the operator to resume manual teleoperation.

Understanding and developing the human-robotic collaboration will necessitate the creation and use of control systems to allow inspectors to interface with a unit during planning activities, navigation, movement, and task execution. A control system must allow for smooth shifting between levels of autonomous function and operator control; for example, when an inspector observes an area of interest, they must be able to designate tasks for the robot to execute which allow for closer inspection. This alternation between autonomous self-guidance and manual control can require a flexible control scheme which is capable of many modes of operation and a variety of input types [16].

3.3 Manual Teleoperation

Lastly, manual robotic teleoperation has been demonstrated to be fully viable in a variety of high consequence applications [17]. In this operation mode, the inspector has full and direct control of the robot's motions and planning capabilities. An example of manual teleoperation would involve an inspector at the base station sending motion commands to a unit based upon feedback obtained through the unit's onboard sensors. This base station can be either on-site at the geologic repository or remote from another location, provided constant communication is maintained.

The teleoperation option is the easiest to implement based on current developments in technology, however, the efficiency of this option is much lower than that of other options. The performance of the system is highly dependent upon communication quality between the inspector and the rover. It also requires substantially more inspector training and full attention of the inspector at all times during operation. Furthermore, this limits an inspector's ability to deploy multiple units at once. Practical implementation of this technology may require a permanent communication network to be deployed along the main and deposition tunnels within the underground facilities, which would need to be expanded as construction progresses. Another option for implementation could utilize a temporary communication network, consisting of retrievable communication nodes placed by the unit as it travels along the inspection path [18, 19].

4. Autonomous Technologies for Verification of Geologic Repositories

This section introduces sensor and robotic technologies which could be employed by IAEA inspectors to maintain CoK of spent fuel in geological repositories. Robotics have been utilized in environments like underground repositories for goals such as inspection, search and rescue, and exploration [16]. There is active development, including government-funded efforts, in progressing the capabilities of autonomous robotics [20]. The authors recognize that there is ongoing research being performed in this area and related to some of the technologies listed below, by IAEA member countries. For a few key examples, please see [70 – 76].

4.1 Robot Sensing Technologies

The verification technologies recommended for implementation on autonomous robots require general considerations of portability, power consumption, current state-of-the-art, and the ability to augment IAEA inspection. The verification technologies are targeted towards implicit and indirect inspection techniques to maintain CoK since most of the sub-surface portion of the operation phase does not involve direct access or visual line-of-sight to the spent fuel canisters.

4.1.1 LiDAR Mapping

Large-scale mapping of complex environments has been effectively accomplished by long range light detection and ranging (LiDAR) sensors and point cloud methods for a variety of uses. LiDAR has been applied in underground mine environments like those proposed for deep geological repositories [21], and the Joint Research Centre recently employed the use of backpack mounted LiDAR for IAEA usage in nuclear facility design information verifications (DIV) [22].

The integration of these use cases with robotic rovers can provide a means to address IAEA DIV safeguards criteria

for geological repositories during construction and operation stages. The LIDAR technology can be mounted on a mobile robot and combined with odometry information to create detailed digital maps which can be compared to reference facility designs during each routine inspection.

LiDAR hardware is commercially available and widely supported. This sensing technology benefits from a large market and wide variety of applications. There are multiple hardware manufacturers (e.g., Velodyne, Waymo, Sick) who also provide commercial off-the-shelf software to fuse the data together from different viewpoints. An example of a commercial Velodyne LiDAR Puck is shown in Figure 2. For robotic integration, there may be more specific work necessary to integrate the LiDAR information with odometry information. Simultaneous Localization and Mapping (SLAM) algorithms are commonly used in robotic applications to combine the data streams to provide real time maps and floor plans [24]. An example of SLAM is shown in Figure 3.
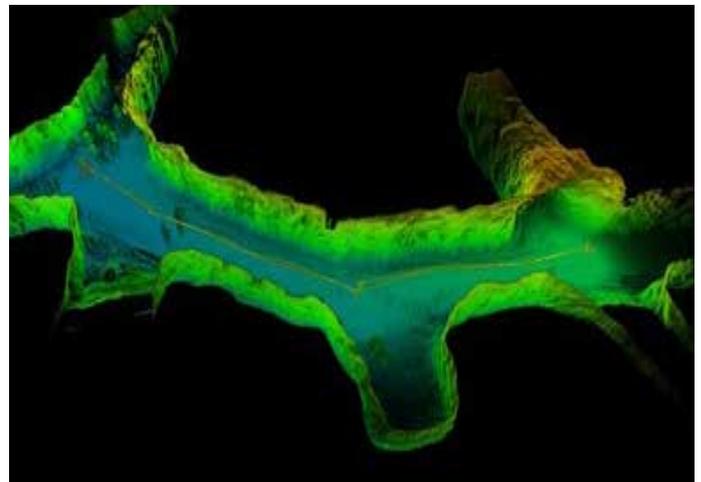
Figure 3: Map of underground mine created using LiDAR and a robot [25]

4.1.2 Optical Imaging

Alongside long-range perception capabilities like LiDAR, short range feature identifications can also support inspection routines through change detection capabilities. Optical camera sensors can be an option for short range analysis of defects or disturbances to different features of the geological repositories. These sensors have a shorter sensing range compared to the LiDAR scanning systems but have increased resolution and optical distinguishing capabilities (i.e., red, green, blue (RGB) color models). A commercial example of a RGB-D (color + depth) camera is shown in Figure 4. As a potential capability, such a camera can be used to visually inspect the surface of a bentonite plug at the end of a deposition tunnel and compare it with past image data to ensure no deviations or disturbances, in both its placement and appearance, are present. Feature-based identification methods are also available and image recognition of defects and surface cavities can be conducted during inspections.

Figure 2: Velodyne LiDAR Puck [23]

Figure 4: Intel Realsense D435i RGBD Camera [26]

Like LiDAR, imaging technology is readily available and is supported in a variety of industries. A few examples of camera sensor manufacturers are Intel, Sick, and Keyence. Much of the novel implementation for this application would be in software development. While image displaying software can be configured off-the-shelf with little effort, machine vision and recognition software will likely be required for more intensive inspection tasks, such as plug inspection and backfill surface disturbance validation, and may require more customization. Algorithms like Scale-Invariant Feature Transform (SIFT) have been used readily in applications which require image feature comparisons [27, 28]. In a KBS-3 plug imaging example, these algorithms can be used during each inspection to compare against the images taken at the time of plug installation. Research will need to be conducted to establish disturbance and deviation thresholds of image matching scores for IAEA safeguards baselines.

4.1.3 RFI On Metal Tagging

Traditionally, radio-frequency identification (RFID) tags are used to uniquely identify objects. Tags are attached to an object and later identified with the help of an electronic reader or interrogator. RFID systems have an integrated circuit and a transponder that communicates to an RFID interrogator through radio waves. However, traditional RFID tags may not be applicable in a subsurface repository since canisters like the KBS-3 are made of copper, and the depth of the buffer layer may prove to be a barrier for the communication between the transponder and interrogator. This challenge is due to the higher frequency of traditional RFID tags and readers which are severely attenuated in these types of environments. A potential solution is a magnetic alternative to RFID, RFID-on-metal, which uses a packet-based wireless technology. For example, the IEEE1902.1, or RuBee, is an RFID-on-metal tagging protocol that is accurate even when attached to a metal surface [29]. The RuBee tag's wireless signal has been shown to travel through solid materials [30] and even works when embedded inside a steel plate, as shown in Figure 5.

The integration of an RFID reader suitable for the detection of an RFID tag with a robot system can be accomplished [31] and could be used to verify the continued presence or movement of material within a repository, particularly before any back-fill activity is initiated. As the RuBee's

communication protocol is based on magnetic field waves [32], the low frequency and long wavelength characteristics might be suitable for transmission through both the bentonite clay and copper canister in which spent nuclear fuel is encapsulated.

RuBee comes with a long battery life of more than 15-25 years [32], which supports monitoring during the operational phase of KBS-3 better than normal RFID tags [33]. Additionally, since RuBee communicates at a low frequency of 131KHz and has low power characteristics, it is an attractive technology for operation in harsh environments on or near steel. As shown in Figure 6, RuBee tag has even been demonstrated to be able to communicate through stainless-steel, something that many other RFID solutions cannot do [32]. If RuBee is left in place for ultimate disposal, it would be important to evaluate the long-term risk to canisters in proximity to RuBee, or other types of RFID.

4.1.4 Ground Penetrating Radar Systems

Ground-penetrating radar (GPR) is a non-destructive geophysical technique to investigate the underground surface. This method can provide a high resolution 3-D subsurface
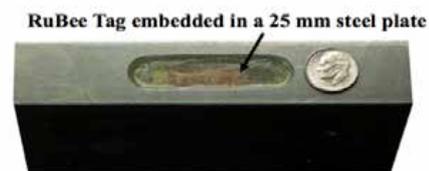
Figure 5: Embedded RuBee Tag inside steel plate
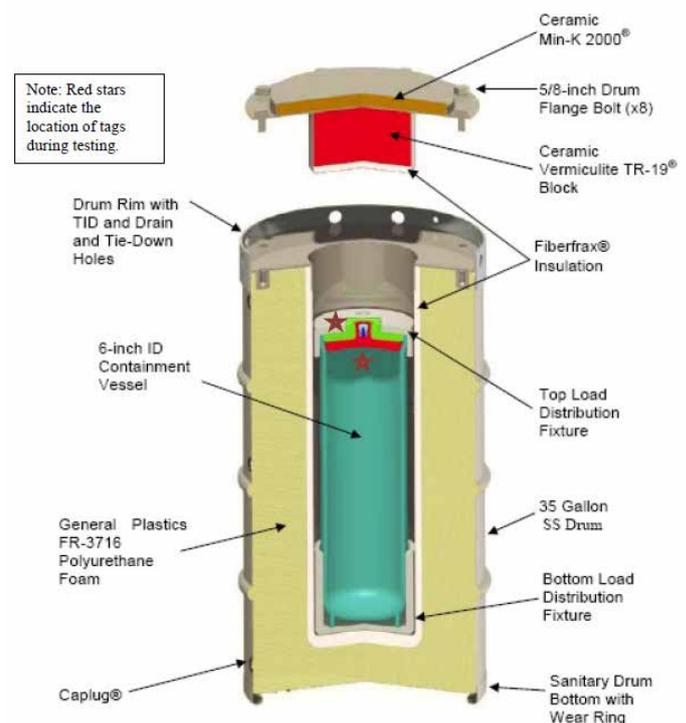
Figure 6: RuBee Tag inside US-DOE certified Type-B Model 9977 drum

image using radar pulses in the microwave band of the radio spectrum. For the creation of a subsurface image, a radar pulse is transmitted through the surface material, and the strength and time of reflected signals are recorded (Figure 7). Reflections are produced based on the electrical conduction properties and dielectric permittivity of the material from which they occur. Metals act as a complete reflector and thus do not allow any amount of signal to pass through. The frequency of the radar pulse could be optimized to the reflective characteristic of the buffer material. If an area is to be scanned, then a series of pulses will be sent throughout the surface [34].

This subsurface imaging capability of ground-penetrating radars is actively utilized to detect metallic landmines buried in the ground [35]. For the KBS-3 repository design, the spent fuel canister is to be emplaced inside the bentonite buffer. Therefore, GPR technology could be used to verify the continued presence of the canister emplaced under the buffer surface due to the electrical conduction properties and dielectric permittivity difference between bentonite clay and copper canister. Challenges exist which would need to be understood prior to use, including the structure and formation of the repository strata to allow for appropriate analysis of GPR results.
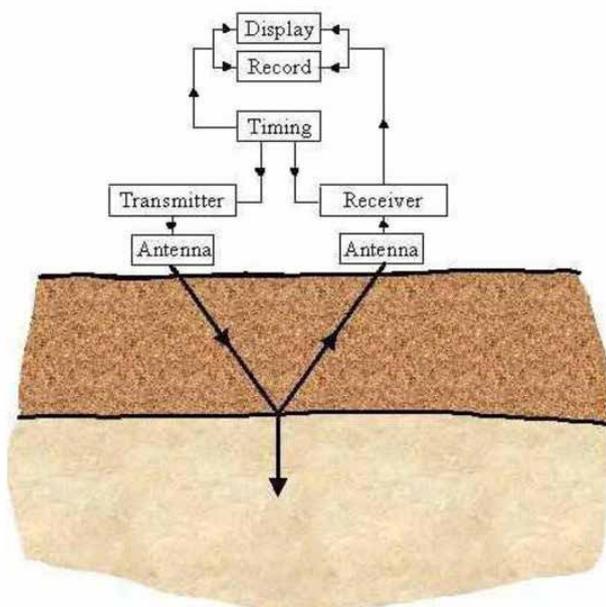
Figure 7: Working principle of GPR [36]

GPR systems are available commercially from various manufacturers such as Geophysical Survey Systems Inc. (GSSI) and GeoSearches Inc., among others. The relative simplicity of the GPR system allows for easy installation on a robot rover system [37, 38]. For instance, CRUX-GPR was developed by the NASA JPL for the "Construction and resource utilization explorer" (CRUX) project. This CRUX-GPR was also mounted under the chassis on the K10 "Black" robot system for underground mapping, as illustrated above in

Figure 8: K10 "Black" CRUX GPR [33]

Figure 8 [37]. To maneuver a rover mounted with GPR inside a deposition tunnel, the system will need a control unit to operate subordinate components and signal processing, a display unit to generate a cross-sectional profile for the scanned area, and a continuous power supply [39].

GPR technology could be an attractive option for inspection activities at an underground repository since it provides a continuous real-time cross-sectional profile without drilling or digging beyond the surface. Also, it can detect unspecified underground voids and trenches in the repository which, apart from surveillance activities, would be useful for exploring the buffer material or host geology beyond the surface. It is operable across a considerable range of frequencies (1MHz-5GHz) depending on the competing requirements of resolution and depth of penetration, allowing for optimization of use case. Potential challenges associated with operating GPR in an underground repository may include power system requirements and maneuverability. Ultimately, the method still needs validation under realistic conditions before GPR can be realized for autonomous monitoring purposes.

4.1.5 Measurement and Detection of Radioactivity

It is expected that IAEA inspectors will perform a series of non-destructive analysis (NDA) measurements on the spent fuel at the above-ground facility. If desired, after a canister is emplaced, NDA measures could also be used to maintain CoK of material in the canister. Passive NDA technologies for gamma-ray and neutron measurement are widely available from many commercial vendors and can be successfully integrated into robotic systems. For example, a radiation detecting robot system mounted with a Geiger counter, camera, LCD screen, and Xbee modems controlled by an Arduino microcontroller has been used to travel through and characterize highly radioactive areas [40].

It is unlikely that a high-resolution detector such as high purity germanium (HPGe) would be viable for use on an autonomous unit. More realistic solutions could be gamma detection techniques for gross counting using a Geiger counter or spectroscopy using a low-resolution sodium iodide (NaI) detector or a medium-resolution cadmium zinc telluride (CZT) detector. Alternatively, small neutron detectors (e.g., He-3, BF3, or Li-6 and ZnS(Ag)) could be deployed. One benefit of small neutron detectors over gamma systems is that neutron detection may help differentiate special nuclear material from other benign or nuisance sources and would not rely on spectroscopy.

Significant questions remain about the feasibility of using passive NDA techniques. Although it is the most direct method to establish the integrity of nuclear material inside a waste canister, one disadvantage is that the post emplacement radiation signature is likely to be reduced to very low levels due to the sheer volume of surrounding material. For example, if a nuclear material that emits 662 keV gamma rays is surrounded by 70 cm of buffer material (polyvinyl polymer-coated bentonite clay with a density of 2.8 g/cc) then the gamma-ray signal will be attenuated by at least a factor of 108 [41]. Computational models based on modern radiation transport codes such as MCNP [42] can be used to assess the impact of the attenuation of the encapsulation on the radiation signature of spent fuel and confirm the viability of this approach for maintaining CoK.

4.1.6 Temperature Profile Measurement

A thermographic camera can detect infrared radiation wavelengths as long as 14,000 nm [43] which are emitted by an object. An example thermal image is shown in Figure 9. Like a characteristic radiation signature, the decay power (heat) of a fuel assembly depends on its burn-up, age, and mass of radioactive material, which can be estimated using SCALE/ORIGEN or other similar software packages. With knowledge of the decay heat source term of the emplaced material, the temperature in the buffer region of the repository can be estimated using modeling and simulation software such as ANSYS [44]. Inspectors can combine computer modelling and thermal detection measurements with qualitative material accountancy techniques during inspections on emplaced material. In this type of measurement, the expected temperature at the buffer's surface (based on computer modelling) would be compared with the actual measured temperature to determine whether a heat source is present. Although this measurement cannot confirm the presence of radioactive material, the approach would serve as a consistency check and could add a degree of confidence in the CoK for the emplaced spent fuel canister, in place of or in complement to using NDA techniques if so desired. In both cases, measurements are only available for the time window before the emplacement tunnel is filled in.
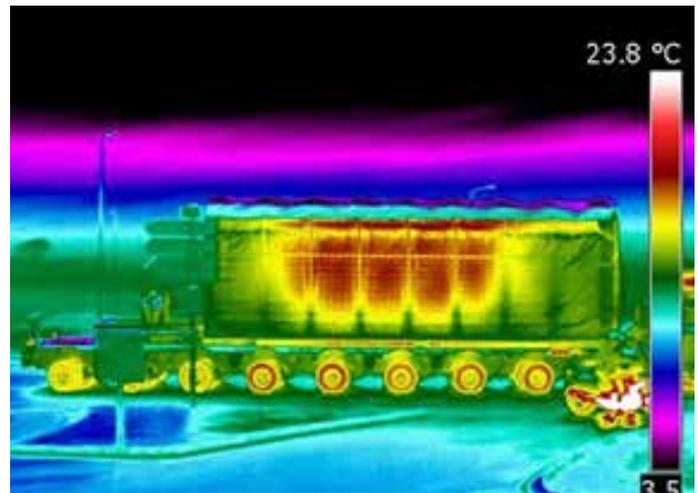
Figure 9: Thermal image of railcar CASTOR system [45]

5. Technology Implementation Considerations

This section addresses considerations in combining available sensor technologies with operations aided by automation within the levels outlined in Section 3 to fulfill safeguards requirements.

5.1 Robot-Assisted Inspection System Design Considerations

An inspection system using any level of remote inspection capability is assumed to be designed and tailored to the IAEA inspection process. Environmental variations between repositories will require flexibility in dealing with facility-specific challenges and an effective inspection system must be facility-agnostic. Solutions specifically tailored to inspection (rather than the needs of another industry) would better allow for system expansion, including accommodation of future sensor designs and modifications to inspection tasks. Current commercial-off-the-shelf underground and mining inspection equipment may provide a framework for inspection solutions but would need to be adapted to suit mission needs. Units must be resistant to environmental elements like water and dust, be capable of traversing rocky terrain with obstacles, and potentially communicate in a bandwidth-restricting environment. Notably, any custom modification, adaptation, or integration of equipment developed for other environments would require operational tests to ascertain equipment suitability and reliability for inspection requirements, safety for operation in inspection environments, system power performance, and operational longevity. Additionally, data security, and hardware and software integrity must be provided for in any solution. The former could be achieved through data authentication, while the latter through tamper indicating enclosures and/or conduits.

The following discussion of automated inspection systems for safeguards use in geologic repositories is derived from currently available robots which operate in similar

environments [46, 47, 48], thereby supporting the selected specifications. Focus is placed on hardware considerations, as opposed to software, because much existing hardware is currently available which could be adapted for use. However, from the software perspective, it is unlikely that an entire software stack would need to be written, as many packaged algorithms available could be adapted and integrated to serve as the basis for this system, though customized software development for this novel robotics application will be required. The creation of full or shared autonomy software is still part of on-going research and development and extensive testing would be needed once the initial software development is completed.

5.1.1 Baseline Hardware

Baseline hardware can be divided into three major subsystems: motion hardware, electronic and computing devices, and power supplies.

Motion hardware includes a unit's chassis, motor controllers, actuators, and locomotion hardware (e.g., wheels, treads, etc.) it uses to move through a space and manipulate objects. Motion hardware must contain protections from the physical environment, and supply effective locomotion, whether a unit uses wheels, continuous tracks, treads, or Mecanum wheels (which allow omni-directional movement). Ultimate selection depends on facility and environmental requirements. Higher complexity motion hardware may make maintenance tasks more difficult or more frequent, and may result in difficulties with uneven terrain, resulting in operational delays or failures. Motion hardware will also likely constitute much of the unit's weight, having a significant impact on power requirements.

Electronic and computing hardware includes central processing units (CPU), graphics processing units (GPU), communications equipment, navigational sensors, and on-board measurement devices such as radiation monitors. A unit's power consumption increases with the number of active on-board electronic devices, necessitating a balance between inspection capability and operational power requirements.

Units will require an on-board power supply to operate both its motion hardware and electronic and computing devices. Power requirements depend on the unit's weight, electronic/computing power consumption, and operational parameters (e.g., operational time and travel distance between recharges). Batteries are preferred to combustion engines in the underground operational environment as batteries do not emit fumes or particulate exhaust. For a point of reference, deep-cycle batteries have been used on autonomous robots in the DARPA Subterranean Challenge and provided a few hours of operational time [46] with sensor payloads of around 50kg [47].

5.1.2 Communication Challenges

For remote robot operation, the IAEA inspector will be situated at a base station, which could be positioned near the repository (e.g., in the above-ground portion of the facility) or in a remote location (e.g., in Vienna). A deep geologic repository is an environment which presents challenges to communication between the base station and the robot system. These potential challenges are a result of a combination of the distance between the unit and the base station, and the amount of geological and man-made material interposed between the underground repository and above ground facilities. It is assumed that existing communication infrastructure will be limited and there will be areas without readily available communication systems. In these cases, the autonomous inspection system can create a self-deployed communication network as it traverses the repository [49, 50]. There have been practical applications of robot systems dropping nodes along the mission route to continually establish communication with the base station while exploring unknown environments [46]. In underground repositories, these nodes would need to be dropped in locations which do not interfere with normal repository operations and allow retrieval by the unit at the end of the inspection mission.

5.1.3 Size and Footprint Considerations

The size and footprint of autonomous units must be considered when discussing the feasibility of robotic augmentation and automation of IAEA inspection processes. Autonomous system design must consider the trade-offs between the size and weight of a unit and its inspection capabilities. A larger system can accommodate more inspection hardware, however transportation between and within facilities becomes more challenging as the size and weight of a unit increase. A robot the size of an automobile would likely not have the fine maneuverability to be effective in the constrained environment of underground repositories. A comprehensive evaluation is recommended optimize the number and type of inspection sensors while meeting inspection capability requirements.

Optical sensors like LiDAR, RGBD, and thermal cameras are relatively small and lightweight, which would allow multiple sensors to be coupled together in a smaller inspection unit. The robots competing in the DARPA Subterranean Competition [46, 47] are examples of units that successfully incorporated multiple devices into rovers with minimal footprints. The width of the emplacement tunnels is reported to be around 4.2m with a similar dimension for height [51], which provides a first-order constraint for design of the robot system dimensions. Deploying small, agile inspection units will allow for high maneuverability through the inspection environment while being inconspicuous to repository workers and preventing disruption to facility operations.

5.2 Inspection Implementation Considerations

According to the IAEA, the primary safeguards objective for geologic repositories is the detection of diversion. DIV will be used to confirm repository design, and to detect potential undeclared activities [52, 53]. This section discusses how a robotic system capable of performing surveillance activities while traversing the repository may support IAEA objectives and augment the DIV process, serving as the primary surveillance mechanism for DIV during an IAEA inspection.

At the highest level, any selected robotic system must be able to meet primary IAEA safeguards measures for a geologic repository. This applies to any level of autonomy desired in the final system. Considering additional challenges, one important consideration for a robotic system is the status of currently available technologies applicable to the task or application. Other important considerations are the time and cost spent in development and testing, which are driven by the expected complexity and novelty of the required algorithms. For example, for software solutions to support fully autonomous DIV mapping, algorithms like SLAM are well-studied, numerous libraries/packages are readily available, and many applications are being researched and developed. However, defect detection for plug surfaces is a novel application and a likely approach would involve deep learning tools like convolutional neural networks (CNNs), which have provided the best results in object detection and feature recognition. Since no direct applications of deep learning currently involve defect recognition for repository tunnel plugs, the networks need to be trained before inspectors can have practical confidence in a network's capability. Finally, nuclear-specific applications involving sensor fusion of different sensing abilities (e.g., temperature, radiation, etc.) may require more research, training, and software development before implementation in autonomous applications. As this would require a cross-disciplinary effort involving nuclear domain knowledge, for now, it is likely that these methods are best suited for shared autonomy mode of operation by trained IAEA inspectors.

5.2.1 Technology Implementation – DIV

DIV is a process that verifies design information provided by the state. In the context of repositories, DIV may also support the detection of undeclared activities. According to Fritzell [52], the DIV for a geologic repository must provide the assurance that:

1. Design information of repository with access routes and other features is verified.

2. Backfill of emplacement tunnels is completed as declared with no voids or other means (e.g., softer fill material).

3. Sealing of back-filled areas are completed as declared.

4. Integrity of repository sealed areas has been maintained through construction phase.

5. Access routes to back-filled areas are filled during the final stages of operational life of the repository.

6. Decommissioning is completed as declared with removal of all surface equipment and facilities.

7. There are no undeclared excavations or boreholes around the repository within a given distance and that none are active during operation or after sealing

There are no undeclared excavations or boreholes around the repository within a given distance and that none are active during operation or after sealingThere are no undeclared excavations or boreholes around the repository within a given distance and that none are active during operation or after sealingThe novel, restrictive, and expansive environment of the underground repositories may present significant challenges to traditional DIV approaches. For instance, the IAEA currently uses 3-Dimensionsal Laser Rangefinders (3DLR) to conduct DIV in surface facilities because of the technology's high resolution and comparison capability [54]. Inspectors rely on 3DLRs on to identify anomalies and potential areas of interest; however, as inspector-carried equipment, 3DLRs are heavy, unwieldy, and would be cumbersome for inspector-use in assessing large geological repositories. Robotic units equipped with instruments like the 3DLR could be utilized to minimize inspector burden during initial and routine DIV inspections. Due to the required training and operational difficulty of 3DLR, the technology is potentially better suited to initial inspections during the pre-operation phase of a repository to establish a baseline and highlight areas of interest for future inspection activities, rather than during the operational phase. However, as 3D range finding technologies evolve and advance, this could change and could find broader application and use.

Robotic inspection systems have already been considered for DIV processes [54] and could serve as the primary means for safeguards inspections in areas that are hazardous or inaccessible. Regular DIV inspections during the operational phase can utilize combinations of technology, a few case examples of which are listed below, to provide assurances of repository design and function. These technologies can address the DIV objectives outlined above and ensure that design deviations and modifications are verified through comparison of observations to baseline information.

LiDAR + Full Autonomy: Addressing DIV requirements (1, 3, 4, 5): Large scale mapping is critical to satisfying DIV requirements. Time-series maps can be generated during each autonomous inspection and compared to historical data. LiDAR technology can be combined with odometry to create high fidelity maps suitable for DIV, on the order of +/- 2cm within the actual position [54]. LiDAR has been used

previously in applications and environments similar to underground nuclear repository inspections [46, 47, 48, 54, 55] and has precedence and synergies within fully autonomous robotic applications [56, 57].

Optical Imaging + Shared Autonomy: Addressing DIV requirements (2, 3, 4, 5, 6): Optical sensors (e.g., cameras) can provide real-time feedback to operators and allow for visual detection of anomalies and aberrations that require additional scrutiny. Object recognition and flaw detection are topics of machine vision research which receive attention from academia and industry alike. Automatic identification of defects (e.g., voids, cracks) has been used in autonomous applications in other domains [58, 59, 60, 61], however the development of customized solutions specific to IAEA inspections would be required to make this a shared autonomous routine. For example, in addressing DIV assurance requirement 2, it might be challenging to identify differences in strength or composition of materials and validate proper material characteristics from optical recognition or imaging alone. Additional capabilities using tactile sensors, or hyperspectral imaging may be needed. There are also opportunities for development where inspection areas are larger than standard fields of view (FOV), requiring imaging capabilities that allow for multiple viewing angles or minute positioning of the robotic inspection system.

In initial implementation, control of the unit could transition to the inspector to facilitate dexterous robot movement during specific inspection activities. Later implementations can transition to a fully automated routine using active perception [62, 63] based on information maximization.

LiDAR + Optical Imaging + Full Autonomy/Shared Autonomy: Addressing DIV requirements (1, 6): Accounting for declared equipment throughout the facility [52, 54] is another important aspect of DIV. Robotic inspection systems can enhance this process by visually identifying equipment through optical imaging-enabled object detection, determining object dimensions, and tagging object location data on facility maps. It is possible for undeclared equipment to be used for diversionary activities [52], and automating the identification and accounting of equipment could be used to help inspectors verify that that the allotted occupied area falls within guidelines. An example of object detection using the You Only Look Once (YOLO) algorithm [66] is shown in Figure 10.

These activities can be performed with full autonomy, where the unit identifies, categorizes, and flags suspect equipment, or with shared autonomy where control reverts to inspectors if items or areas of interest are discovered [64, 65]. The software would require customized development to detect, identify, and track accountable equipment.
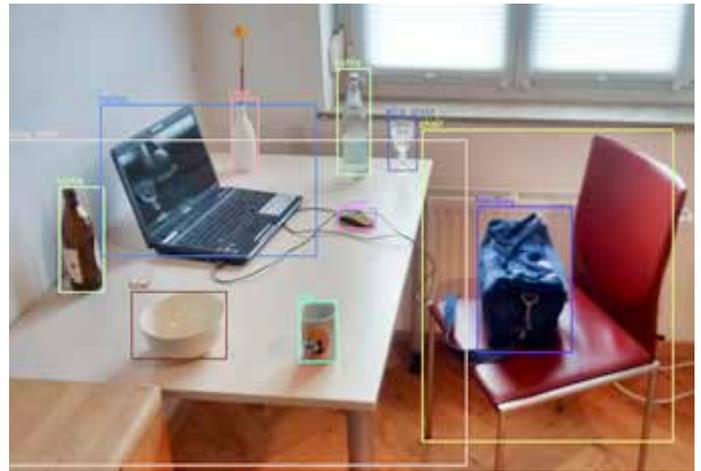
Figure 10: Object detection using the YOLO algorithm [66]

5.2.2 Technology Implementation – C/S of Canisters

This section explores other sensor implementations which are directly applicable to C/S methods. The following list of C/S system requirements are derived from Fritzell [52] and Mongiello et al. [67]:

1. Ensure CoK from above ground activities.
2. Systems are designed for independent operation and remote monitoring.
3. Redundancy within the C/S system is employed.
4. Coverage of all credible diversion paths.
5. Ability to report health status to safeguards authorities.
6. Devices should be tamper-resistant and be capable of indicating tampering has occurred.
7. Devices should have low "false alarm" frequencies.

The focus of the research presented in this paper is on below ground CoK options, therefore consideration of C/S system requirement 1 above is out of scope. CoK of above ground activities can and should utilize existing IAEA C/S measures. Regarding C/S system requirement 6, it is likely that aspects of existing IAEA tamper indicating devices, enclosures, and data security methods can be employed and would need to be implemented for any technical solution described below. A few example cases of C/S technologies in autonomous settings are discussed below.

- **Ground Penetrating Radar + Shared Autonomy:** Addressing C/S requirements (2, 3, 5): Autonomous inspection with ground penetrating radar (GPR) has been accomplished by NASA using their portable CRUX GPR technology [33]. The technology was capable of scanning depths up to 5m and resolutions within 15cm [67] in lunar environments. NASA has also developed a GPR for use with the Martian environment which was capable of depths up to 50m. These portable technologies can be used by the inspection robot system in conjunction with

thermal sensors to verify location and presences of emplaced canisters for a dual C/S system.

- **Thermal Imaging + Full Autonomy Addressing C/S requirements (2, 3, 4, 5):** A thermal camera mounted on a mobile unit can capture thermal information as the unit traverses a facility and is able to map out areas inaccessible by the static C/S surveillance approaches [68]. Heat maps of the facility can be matched against historical data with thresholds that are pre-determined to calculate potential diversionary activities. This passive sensing technique can be fully autonomous and set to either collect continuous thermal data during the inspection operation or perform readings at a determined frequency to reduce power draw.

- **Thermal Imaging + Shared Autonomy: Addressing C/S requirements (2, 3, 5):** In addition to fully autonomous thermal sensing, the thermal camera can also be used at the inspector's discretion to examine areas of interest. Specifically, it could potentially be used to match thermal profiles of emplaced canisters underneath the Bentonite clay cap at the end of the emplacement tunnel. During a shared autonomy inspection, the operator could take control of the unit to inspect these emplaced canisters prior to backfill to provide additional verification that the canister is present, thereby maintaining CoK. This verification process would require extensive modelling and testing to be approved as a sound inspection method. In the future, fully autonomous inspection routines can be conducted in a similar vein to the second suggestion in the DIV implementation scenarios, optical imaging using shared autonomy.

- **Radiation Monitoring + Full Autonomy: Addressing C/S requirements (2, 3, 4, 5):** Post emplacement, small low-power gamma and neutron counters can be used to detect off-normal repository conditions, given that the background levels of radiation will be very low due to the presence of the buffer material. Comparing the expected signal with the measured signal could then be used for real-time safeguards verification.

6. Conclusion

Implementing automated solutions for underground nuclear repository inspection will be a new venture for the IAEA but carries the potential to greatly enhance the efficacy and efficiency of inspections and allow inspectors' time and expertise to be directed where needed most. Many different factors must be considered before integration of autonomous solutions into inspection processes. This paper presents a variety of ways in which autonomous units can augment IAEA inspection of underground geologic repositories during the pre-operation and operation phases under varying levels of autonomy and inspector interaction. Sensor technologies for autonomous monitoring are described and hardware systems outlined which can maintain CoK of nuclear waste, adhering to the IAEA safeguards. The

permanence of these nuclear repositories allows time for the development of technologies and approaches for augmented inspections. Technology applications that are not yet fully developed can mature by the time these geological repositories are fully functional.

The technology recommendations and implementation scenarios of robotic inspection of underground nuclear repositories in Section 5.2 of this paper provide multiple avenues for follow-on work. There are still many questions to be answered and testing to be conducted. For example, the DIV map constructed via LiDAR and full robot autonomy using SLAM can suffer from error propagation due largely to odometry deviations caused by encoder slippage and other environmental variables. The extent of the acceptable deviations must be tested to gain full confidence as a DIV safeguards approach.

Equipment and data integrity are key requirements for IAEA equipment. Integrity of the data collected by the robotic system and the communication between the rover and the base station must be validated and warrants further examination. The physical unit requires tamper resistance and must be capable of tamper indication. Data streams and samples gathered during inspection must be authenticable. Encryption protocols for data and communication are not extensively considered in this paper and should be explored. Furthermore, research into spoofing of various features that are utilized as inspection criteria and thresholds is necessary. These concerns should be explored early in development to ensure adherence to system requirements and to prevent late-stage changes.

Lastly, additional domain knowledge can be leveraged to maximize the potential of on-board sensors. The efficacy of C/S surveillance via thermal profiles of the canisters under the bentonite cap can be explored using modelling software. Radiological signatures in the environment can be catalogued and tested to identify capable environment sampling sensors which can be mounted on mobile units. This paper can serve as a starting point to the exploration of robotics for underground nuclear repository inspection.

7. References

[1] S. K. AB, "Design and production of the KBS-3 repository," in SKB Technical Report TR-10-12. S v e n s k K¨arnbr¨anslehantering AB Stockholm, Sweden, 2010.

[2] A. Hedin and J. Andersson, "SKB's safety case for a final repository license application," Tech. Rep., 2014.

[3] L. Boerjesson, D. Gunnarsson, L.-E. Johannesson, and E. Jonsson, "Design, production and initial state of the backfill and plug in deposition tunnels," Swedish Nuclear Fuel and Waste Management Co., Tech. Rep., 2010.

[4] P. Grahn, L. Moren, and M. Wiborgh, "Spent nuclear fuel for disposal in the KBS-3 repository," Swedish Nuclear Fuel and Waste Management Co., Tech. Rep., 2010.

[5] D. Holz, N. Basilico, F. Amigoni, S. Behnke et al., "A comparative evaluation of exploration strategies and heuristics to improve them." in ECMR, 2011, pp. 25–30.

[6] B. Yamauchi, "A frontier-based approach for autonomous exploration," pp. 146–151, 2002.

[7] C. Dornhege, A. Kleiner, and A. Kolling, "Coverage search in 3D," in Proc. IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2013, pp. 1–8.

[8] W. Tabib and N. Michael, "Simultaneous localization and mapping of subterranean voids with gaussian mixture models," in Conf. on Field and Service Robot., Tokyo, Japan, 2019.

[9] G. A. Korsah, A. Stentz, and M. B. Dias, "A comprehensive taxonomy for multi-robot task allocation," Int.J. Robotics Research, vol. 32, no. 12, pp. 1495–1512, 2013.

[10] L. Liu and D. A. Shell, "Multi-level partitioning and distribution of the assignment problem for large-scale multi-robot task allocation," Robotics: Science and Systems VII; MIT Press: Cambridge, MA, USA, pp. 26–33, 2011.

[11] B. P. Gerkey, "On multi-robot task allocation," Ph.D. dissertation, University of Southern California, 2003.

[12] G. Best, O. M. Cliff, T. Patten, R. R. Mettu, and R. Fitch, "Dec-MCTS: Decentralized planning for multi-robot active perception," Int. J. of Robotics Research, vol. 38, no. 2-3, pp. 316–337, 2019.

[13] G. A. Hollinger, S. Yerramalli, S. Singh, U. Mitra, and G. S. Sukhatme, "Distributed data fusion for multirobot search," IEEE Trans. on Robotics, vol. 31, no. 1, pp. 55–66, 2014.

[14] A. Bauer, D. Wollherr, and M. Buss, "Human–robot collaboration: a survey," International Journal of Humanoid Robotics, vol. 5, no. 01, pp. 47–66, 2008.

[15] L. Milliken and G. A. Hollinger, "Modeling user expertise for choosing levels of shared autonomy," in Proc. IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 2285–2291.

[16] Y. Li, M. Li, H. Zhu, E. Hu, C. Tang, P. Li, and S. You, "Development and applications of rescue robots for explosion accidents in coal mines," J. of Field Robotics, vol. 37, no. 3, pp. 466–489, 2020.

[17] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," Automatica, vol. 42, no. 12, pp. 2035–2057, 2006.

[18] K. Derr and M. Manic, "Wireless sensor network configuration—part II: adaptive coverage for decentralized algorithms," IEEE Trans. on Industrial Informatics, vol. 9, no. 3, pp. 1728–1738, 2013.

[19] D. B. Licea, M. Bonilla, M. Ghogho, S. Lasaulce, and V. S. Varma, "Communication-aware energy efficient trajectory planning with limited channel knowledge," IEEE Trans. on Robotics, 2019.

[20] B. Allen and T. Chung, "Unearthing the subterranean environment," 2020. [Online]. Available: https://www.subtchallenge.com/

[21] R. Zlot and M. Bosse, "Efficient large-scale 3D mobile mapping and surface reconstruction of an underground mine," in Field and service robotics. Springer, 2014, pp. 479–49--R.

[22] W. Juchmann, "Nuclear inspections made portable—with LiDAR," LiDAR Magazine, vol. 5, no. 7, pp. 25–27, 2015.

[23] V. Puck, "Velodyne puck," Available athttps://velodynelidar.com/products/puck/(2020/08/12).

[24] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," in Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), 2016, pp. 1271–1278.

[25] DJI, "Lidar dji," Available at https://content.dji.com/drones-for-good-exploring-mines-with-autonomous-drones/(2020/08/12).

[26] Intel, "Intel realsense d435i," Available athttps://www.intel.com/content/www/us/en/architecture-and- technology/realsense-overview.html?wapkw=realsense(2020/08/12).

[27] E. Karami, S. Prasad, and M. Shehata, "Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images," arXiv preprint arXiv:1710.02726, 2017.

[28] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in Proc. Int. Conf. on Pattern Recognition (ICPR). IEEE, 2016, pp. 378–383.

[29] J. K. Stevens, "An Introduction to RuBee (ieee p1902. 1) and its use in real-time visibility networks," IEEE RFID, http://caneus. org/workshopphp/index temp. php, 2007.

[30] S. Dantas, A. N. Barreto, L. Aguayo, A. J. Braga, L. S. Silva, and L. G. Uzeda Garcia, "Simulation of IEEE 1902.1 (RuBee) protocol for communication with buried assets," in IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017, pp. 1–6.

[31] M. Soleimanifar, D. Beard, P. Sissons, M. Lu, and M. Carnduff, "The autonomous real-time system for ubiquitous construction resource tracking," in Proc. Int. Symposium on Automation and Robotics in Construction, vol. 30, 2013, p. 1.

[32] D. Shull, "9977 type b packaging internal data collection feasibility testing-magnetic field communications," Savannah River Site (United States). Funding organisation: US Department of Energy, Tech. Rep., 2012.

[33] C. Clementi, L. Capineri, and F. Littmann, "Innovative method to authenticate copper canisters used for spent nuclear fuel based on the ultrasonic investigation of the friction stir weld," IEEE Access, vol. 5, pp. 7843–7852, 2017.

[34] H. M. Jol, Ground penetrating radar theory and applications. elsevier, 2008.

[35] Z. Zyada, T. Matsuno, Y. Hasegawa, S. Sato, and T. Fukuda, "Advances in gpr-based landmine automatic detection," Journal of the Franklin Institute, vol. 348, no. 1, pp. 66–78, 2011.

[36] J. L. Davis and A. P. ANNAN, "Ground-penetrating radar for high-resolution mapping of soil and rock stratig- raphy 1," Geophysical prospecting, vol. 37, no. 5, pp. 531–551, 1989.

[37] T. Fong, M. Allan, X. Bouyssounouse, M. G. Bualat, M. Deans, L. Edwards, L. Flu¨ckiger, L. Keely, S. Y. Lee,D. Lees et al., "Robotic site survey at haughton crater," in Proceedings of the 9th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS), Los Angeles, CA, 2008.

[38] T. D. Barfoot, G. M. D'Eleuterio, and A. P. Annan, "Subsurface surveying by a rover equipped with ground- penetrating radar," in Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), vol. 3, 2003, pp. 2541–2546.

[39] J. Ralston and C. Hargrave, "Forward looking radar for early sensing of ground conditions," 2006.

[40] J. Jiang, D. Seniff, D. Callahan, L. Tan, and Y. Lee, "Development of a radiation detecting rover system," in 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON). IEEE, 2017, pp. 106–108.

[41] R. Y. O. H. e. a. Hager, I.Z., "Nano-structured natural bentonite clay coated by polyvinyl alcohol polymer for gamma rays attenuation." J Theor Appl Phys, vol. 13, pp. 141–153, 2019. [Online]. Available: https://doi.org/10.1007/s40094-019-0332-5

[42] MCNP, "Monte carlo team, MCNP–a general purpose monte carlo n-particle transport code, version 5," LA-UR-03 1987, Los Alamos National Laboratory, April 2003, The MCNP5 code, Tech. Rep., 5.

[43] T. Omar and M. L. Nehdi, "Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography," Automation in Construction, vol. 83, pp. 360–371, 2017.

[44] S. Moaveni, Finite element analysis theory and application with ANSYS, 3/e. Pearson Education India, 2011.

[45] N. Geographic, "Thermal Image of CASTOR Railcar," Available at https://www.nationalgeographic.com/news/2011/1/110119-nuclear-waste-train-castor-anti-nuclear-protest-germany-power-energy-pictures/ (2020/08/12).

[46] CMU/OSU, "SubT Team Explorer Website CMU/OSU," Available at https://www.subt- explorer.com/ (2020/08/12).

[47] T. Rou˘ček, M. Pecka, P. C˘ı˘zek, T. Petˇrı˘cek, J. Bayer, V. S˘alansky`, D. Heˇrt, M. Petrlı́k, T. B´a˘ca, V. Spurny` et al., "Darpa subterranean challenge: Multi-robotic exploration of underground environments," in International Conference on Modelling and Simulation for Autonomous Systesm. Springer, 2019, pp. 274–290.

[48] S. Thrun, S. Thayer, W. Whittaker, C. Baker, W. Burgard, D. Ferguson, D. Hahnel, D. Montemerlo, A. Morris, Z. Omohundro et al., "Autonomous exploration and mapping of abandoned mines," IEEE Robotics & Automation Magazine, vol. 11, no. 4, pp. 79–91, 2004.

[49] R. Elhabyan, W. Shi, and M. St-Hilaire, "Coverage protocols for wireless sensor networks: Review and future directions," Journal of Communications and Networks, vol. 21, no. 1, pp. 45–60, 2019.

[50] Y. Wang, S. Wu, Z. Chen, X. Gao, and G. Chen, "Coverage problem with uncertain properties in wireless sensor networks: A survey," Computer Networks, vol. 123, pp. 200–232, 2017.

[51] S. K. AB, "Design, production and initial state of the backfill and plug in deposition tunnels," in SKB Technical Report TR-10-16. Svensk K¨arnbr¨anslehantering AB Stockholm, Sweden, 2010.

[52] A. Fritzell, "Concerns when designing a safeguards approach for the back-end of the swedish nuclear fuel cycle," Swedish Nuclear Power Inspectorate, Tech. Rep., 2006.

[53] IAEA, "Technological Implications of International Safeguards for Geological Disposal of Spent Fuel and Ra- dioactive Waste," International Atomic Energy Agency, Tech. Rep., 02 2010.

[54] P. C. Durst, R. Metcalf, R. Bean, T. Bjornard, B. Boyer, M. Burks, A. Dougan, M. Ehinger, D. Few, D. Hanks et al., "Report of the workshop on nuclear facility design information examination and verification for safeguards," 2009.

[55] R. R. Murphy, J. Kravitz, S. L. Stover, and R. Shoureshi, "Mobile robots in mine rescue and recovery," IEEE Robotics & Automation Magazine, vol. 16, no. 2, pp. 91–103, 2009.

[56] J. Aulinas, Y. R. Petillot, J. Salvi, and X. Llad'o, "The slam problem: a survey." CCIA, vol. 184, no. 1, pp. 363–371, 2008.

[57] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," IPSJ Transactions on Computer Vision and Applications, vol. 9, no. 1, p. 16, 2017.

[58] Z. Liu, Y. Cao, Y. Wang, and W. Wang, "Computer vision-based concrete crack detection using u-net fully convolutional networks," Automation in Construction, vol. 104, pp. 129–139, 2019.

[59] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure," Advanced Engineering Informatics, vol. 29, no. 2, pp. 196–210, 2015.

[60] A. Mohan and S. Poobal, "Crack detection using image processing: A critical review and analysis," Alexandria Engineering Journal, vol. 57, no. 2, pp. 787–798, 2018.

[61] T. Dawood, Z. Zhu, and T. Zayed, "Machine vision-based model for spalling detection and quantification in subway networks," Automation in Construction, vol. 81, pp. 149–160, 2017.

[62] G. A. Hollinger, B. Englot, F. Hover, U. Mitra, and G. S. Sukhatme, "Uncertainty-driven view planning for underwater inspection," in International Conference on Robotics and Automation. IEEE, 2012, pp. 4884–4891.

[63] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," The International Journal of Robotics Research, vol. 30, no. 11, pp. 1343–1377, 2011.

[64] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," Acm computing surveys (CSUR), vol. 38, no. 4, pp. 13–es, 2006.

[65] G. P. Zhang, "Neural networks for classification: a survey," Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 4, pp. 451–462, 2000.

[66] W. Commons, "Detected-with-yolo–schreibtisch-mit-objekten," 2019. [Online]. Available: https://commons. wikimedia.org/wiki/File:Detected-with-YOLO–Schreibtisch-mit-Objekten.jpg

[67] R. Mongiello, R. J. Finch, and G. Baldwin, "Safeguards approaches for geological repositories: Status and gap analysis," Sandia National Laboratories Technical Report SAND2013-5185P, 2013.

[68] S. S. Kim, S. R. Carnes, A. F. Haldemann, E. H. W. Ng, C. Ulmer, and S. A. Arcone, "Miniature ground penetrating radar, crux gpr," in Aerospace Conference. IEEE, 2006, pp. 7–pp.

[69] S. Lagü̈ela, L. D́iaz, D. Roca, H. Lorenzo et al., "Aerial thermography from low-cost uav for the generation of thermographic digital terrain models," Opto-Electronics Review, vol. 23, no. 1, pp. 78–84, 2015.

[70] W. Jansens et.al. Advanced safeguards measurements monitoring and modelling laboratory (AS3ML), IAEA Symposium on International Safeguards, October 2014

[71] Wolfart E., Sanchez-Belenguer C., Sequeira V. (2021). Deep Learning for Nuclear Safeguards. Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting August 23-26 & August 30-September 1, 2021

[72] Wolfart E, Sequeira V, Murtezi M, Zein A, Turzak P, Rocchi S, Enkhjin L, Ceriani S, Sanchez Belenguer C, Taddei P. Mobile 3D Laser Scanning for Nuclear Safeguards. ESARDA 37th Annual Meeting Proceedings; Luxembourg (Luxembourg): Publications Office of the European Union; 2015. p. 686-698. JRC94409

[73] Sanchez Belenguer, C., Wolfart, E. and Sequeira, V., RISE: A Novel Indoor Visual Place Recogniser, In: IEEE International Conference on Robotics and Automation (ICRA), 31 May - 31 August 2020, Paris, France, PROCEEDINGS - IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2020, ISSN 1050-4729, p. 265-271, JRC120295.

[74] Wolfart E; Ceriani S; Puig Alcoriza D; Sanchez Belen-guer C; Taddei P; Sequeira V; Murtezi M; Turzak P; Zein A; Enkhjin L; Ingegneni M; Rocchi S; Yudin Y. Mobile 3D Laser Scanning for Nuclear Safeguards. Luxembourg (Luxembourg): Publications Office of the European Union; 2015. p. 71-/ 81. JRC104573

[75] Sanchez Belenguer, C., Taddei, P., Wolfart, E. and Se-queira, V., Sensor Tracking and Mapping (STeAM), EUR 29628 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-98982-7, doi:10.2760/762493, JRC114245.

[76] Surmann, Hartmut & Nuchter, Andreas & Hertzberg, Joachim. (2003). An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitaliza-tion of indoor environments. Robotics and Autono-mous Systems. 45. 181-198. 10.1016/j.robot.2003.09.004.

A Novel Approach for Detection of Illicit Nuclear Activities Using Optically Stimulated Dosimetry

Egemen M. Aras and Robert B. Hayes

North Carolina State University, Burlington Laboratories,
2500 Stinson Dr, Raleigh, NC 27607, United States
E-mail: emaras@ncsu.edu, rbhayes@ncsu.edu

Abstract

The enemy always looks to defeat the detection systems. To handle this tackle, defense in depth in detection is a must. In this work, we propose a complementary detection mechanism for illicit nuclear activities in nuclear facilities in addition to current detection techniques. If one of the detection systems cannot detect illicit nuclear activity, at least one more system is supposed to catch the enemy's action. This action can be either an internal or external thread.

Optically stimulated luminescence dosimetry is used in personal, environmental, retrospective, space, neutron, and medical areas. This system can be a complementary measurement system with the advantages of using without electricity and having those in any nuclear facility. A model is defined with the function of system parameters and the background dose to use OSLDs for the proposed purpose. The model enables us to evaluate the background dose, the initial dose, and the bleaching constant for the reader, including the uncertainty. A case study is worked to prove the model.

According to the model and the case study, we can flag the illicit nuclear activity in the proposed nuclear facility by using optically stimulated luminescence dosimeters (OSLDs).

Keywords: radiation detection; nuclear safeguards; nuclear non-proliferation; optically stimulated luminescence dosimetry

1. Introduction

This work aims to bring a novel approach to detecting illicit nuclear activities using commercially available optically OS-LDs produced by LANDAUER[®] [1], leading to well-known and well-established techniques to determine the dose. OSLDs are used in many fields of radiation dosimetry, including personal, environmental, retrospective, space, neutron, and medical dosimetry [2].

In addition to using a commercial OSLD for retrospective dosimetry, lots of options are available like fired building materials, cementitious building materials, chalk-based plaster, calcium silicate bricks, portable and personal objects, including certain types of telephone cards that contain micro-electronic chips, dental ceramics in the forms of crowns [3]. Moreover, every facility that handles radioactive sources or radiation sources like x-ray above the exempt limit must purchase personal dosimetry services elsewhere. The outcome is that plenty of luminescence materials can be found anywhere; specifically, OSLDs can be found in any facility with a radiation source, including nuclear power plants (NPPs).

Although the ultimate goal of using nuclear technology is to benefit from it by operating for peaceful purposes, someone may turn the nuclear materials into a nuclear weapon. The source of the threat can either be inside or external. This work will focus on trying to address an internal thread in NPPs.

To detect any thread or illicit activity, we need an instrument. This instrument can be used for detection, verification, localization, and identification of the nuclear source and can be pocket-type, hand-held, or fixed installed [4]. Radiation pagers, radiation portal monitors, radioactive isotope identification devices, and radiographic imaging systems are the current detection technologies to detect illicit nuclear activities [5].

An enemy could use various means to defeat detection systems, for example, shielding. The gamma rays from weapons-grade Plutonium are sufficiently energetic and plentiful. They are difficult to shield; however, a layer of lead would shield gamma rays from highly enriched uranium [5]. Background radiation from naturally occurring material, cosmic rays, and even some commercial goods containing
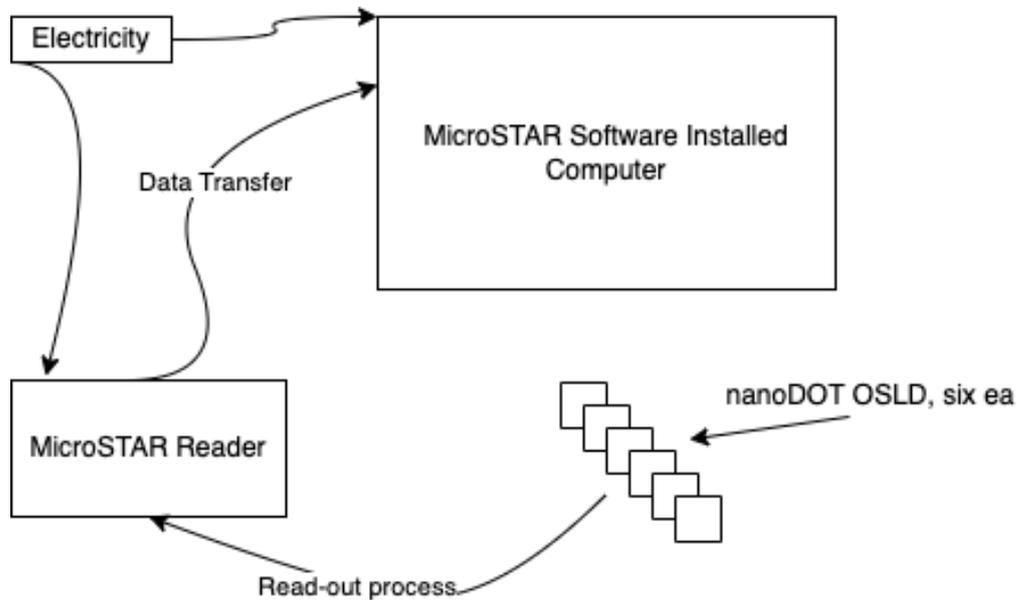
Figure 1: Experimental setup

radioactive material make detection complicated. Multiple detection systems should be used to ensure that the instrumentation system can detect the enemy's activities. For example, if an enemy shields a bomb with a lead, that will create a large, opaque image on a radiograph. Putting a multi-system available can be called defense in depth in detection. In this sense, we proposed a passive measurement method that accompanies the already existing instruments. Our approach will help detect the radiation dose greater than the background and flag the activity as illicit. The approach does not aim to take the place of the current ones; instead, a supportive method that does not require any power during the measurement. Moreover, almost every nuclear facility uses OSLDs to track personal doses.

2.   Materials & Methods

The proposed method to flag illicit nuclear activity consists of two steps. The first step is the experimental part, which includes obtaining the background dose, the bleaching constant, and the initial dose to be used in the second part of the study, which is the case study and validation of the proposed method.

The background dose is the amount deposited from the background radiation within the determined time. The bleaching constant is the parameter for the readers that represents how much dose equivalent light is removed with each read-out step. The final result from the first part of the work is the initial dose, the average dose representing the value before the read-out starts.

2.1   Experimental Design and Analysis

The experimental part of the study includes six nanoDOT™ OSLDs produced by LANDAUER® exposed only to

background radiation and MicroSTAR® Medical Dosimetry System [1] installed in a regular laptop, as shown in Figure 1.

The consecutive measurements in this system can be done in two different ways, as summarized in Table 1. In addition to getting the necessary values mentioned above, we intend to investigate whether there is a statistically significant difference in different measurement methods or not. A well-known and well-established technique called analysis of variance (ANOVA) is used to compare the results. Details for the ANOVA can be reached elsewhere; however, we follow Chapter 10 of Devore [6].

ANOVA has three primary assumptions, 1) the responses for each factor level have a normal population distribution, 2) these distributions have the same variance, and 3) the data are independent. We test the first condition by looking at the residuals and QQ-Plot of the data, and we know each measurement is independent, which is the third assumption of ANOVA. To check the variance, we have a method called Bartlett's test [7] is used to check whether the measurements have equal variances or not. When the variances are not equal, Welch's ANOVA [8] should be used instead of ANOVA.

The analysis part covers evaluating the dose from the measurement and propagation of uncertainty.

The system does four LED exposure measurements of the dosimeter crystal when conducting a measurement for a nanoDot™ OSLD. It takes the mathematical average to evaluate the expected dose in the unit of mGy. Equation (1) below shows how to assess the average dose. It depends on average raw counts (C), corrected background counts (B), calibration factor (CF) in units counts/dose (mGy), sensitivity (S) as a fractional value, and sensitivity adjustment

Measurement Method	NOREP	REP
Steps for the Measurement	1. Open the system	1. Open the system
	2. Do the calibration	2. Do the calibration
	3. Take one nanoDOT™ OSLD and put it in the reader	3. Take one nanoDOT™ OSLD and put it in the reader
	4. Do measurement	4. Do measurement
	5. Save the result	5. Save the result
	6. Do the same steps (4 and 5) 30 times for every nanoDOT™ OSLD	6. Take the nanoDOT™ OSLD from the reader and
	7. Then take the nanoDOT™ OSLD from the reader	7. Put a new one and do the measurement, save the result and get it back
	8. Put a new nanoDOT™ OSLD to the reader and repeat the steps from 4 to 6	8. Do the same steps (6 and 7) for all six nanoDOT™ OSLDs
	9. Complete measurement	9. When the first measurements are done, do the same procedure 30 times
	10. Get all results from the laptop to analyze	10. Complete measurements
		11. Get all results from the laptop to analyze

Table 1: Road map for two different measurement methods

factor (SAF) is a unitless parameter. The average raw clount is the mathematical average of each dosimeter's four simultaneous measurements. Corrected background counts are obtained as part of the initial instrument start-up process, and the sensitivity adjustment factor is a parameter that the user can modify when needed.

The calibration, which can be done by reading dosimeters with known radiation dose levels and characterizing the relationship between measured raw counts and exposed dose levels, is crucial for any measurements, including the average raw count. So the conversion factor from average raw counts to average dose is called the calibration factor. The sensitivity is the manufacturer's parameter for each dosimeter that refers to the relative light count per dose to the reference nanoDot™.

$$D = \frac{(C) - (B)}{(CF)(S)(SAF)} \tag{1}$$

Each term in Equation (1) contributes to the average dose's uncertainty. First-order uncertainty propagation to Equation (1) gives Equation (2) below.

$$\sigma_D^2 = D^2 \left[\frac{(\sigma_C^2) + (\sigma_B^2)}{(C - B)^2} + \frac{(\sigma_{CF}^2)}{(CF)^2} + \frac{(\sigma_S^2)}{(S)^2} + \frac{(\sigma_{SAF}^2)}{(SAF)^2} \right] \tag{2}$$

The average net count uncertainty $\sigma_N^2 = \sigma_C^2 + \sigma_B^2$ can be used in Equation (2), calculated as the standard deviation of four follow-up measurements for each dosimeter per measurement step. Although no method is applied to correct either background counts or SAF in this study, it is assumed that B's uncertainty was 1. The uncertainty on SAF is 0.01 since the software does all calculations by setting B as 0 and SAF as 1.00, respectively. The chosen uncertainty is based on significant digits given by the software. The parameter related to manufacturing, S, is specific for each dosimeter though similar to all others. However, S's uncertainty is assumed to be the same for each dosimeter, that is, 0.01 as a fractional value. The sensitivity takes a value of either 0.92 or 0.96, which justifies our assumption is valid for the uncertainty on S. The conversion factor from counts to dose in the units of counts/dose (mGy), named calibration factor, has two values: low dose measurement and high dose measurement. The uncertainty is 0.001 for low dose measurement, which is our concern in this work. The calibration factor for low dose measurement is 408.081 for our case given by the software.

2.2 Case Study and Validation

A scenario is developed to verify our proposed method. According to the scenario, a single used ^{235}U fuel assembly in a corridor with the assumed positions and geometry as shown in Figure 2. Angle 1 is θ, and Angle 2 is φ in our calculations. The nanoDOT™ OSLD is located in the wall of the corridor. The detail for the scenario is that it is used for 30,000 hours, has a thermal power of 20 MW, then removed from the service and placed in storage for one year.

The goal is to evaluate the dose rate at a 5 meters distance for unshielded and shielded with concrete circumstances. Although the numbers are subject to change from facility to facility, the calculation method is supposed to be identical.
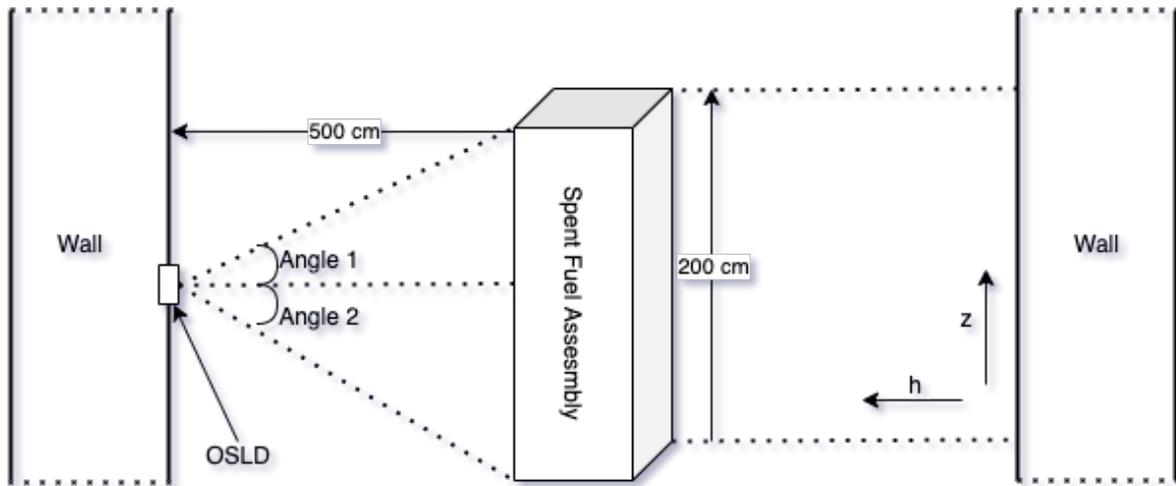
Figure 2: An example case for applying nanodot OSLD as a detector for flagging illicit nuclear activity

Evaluating dose rate with the given conditions [9] will follow the evaluation steps listed below:

1. Source strength (S)
2. Unshielded - uncollided gamma-dose rate (D0)
3. Shielded gamma-dose rate (D)

The calculation starts with Equation (3) by the evaluating $\Gamma_j(t_0, t_s)$ [MeV/fission] is the rate of energy release in group j at time t_s (sec) following an operation at a constant fission rate (sec^{-1}) for time t_0(sec). i is the number of parameters listed for a_{ij} (MeV/sec) and λ_{ij} (s^{-1}). The average energy in each group is assumed as the arithmetic average of the range provided in the same table. The data used for calculation is available in Appendix G.1 of Shultis & Faw [10]. P_0 (fission/sec) is evaluated assuming that 200 MeV energy is released per fission and N_j is the group number.

$$\Gamma_j(t_0, t_s) = P_0 \sum_{i=1}^{N_j} \frac{a_{ij}}{\lambda_{ij}} e^{-\lambda_{ij}t_0}[1 - t_s] \quad (3)$$

We ignore the attenuation and build-up in the air while evaluating the dose rate as given in Equation (4). The response function can be assessed as seen in Equation (5) by applying Appendix C.7 of Shultis & Faw [10] for the $\left(\frac{\mu_{en}}{\rho}\right)_{tissue}$. Note that a point source is moved between two positions along a line at a constant rate that would be equivalent to a linear source of the same total activity distributed over the travel distance.

$$D^0(P) = \frac{1}{4\pi h} \sum_{i=1}^{6} S_{l,i} R_i \left[\tan^{-1}\left(\frac{200 - z}{h}\right) + \tan^{-1}\left(\frac{z}{h}\right) \right]$$

$$(4)$$

$$with \begin{cases} \theta = \tan^{-1}\left(\frac{200 - z}{h}\right) \\ \varphi = \tan^{-1}\left(\frac{z}{h}\right) \end{cases}$$

$$R = 1.602x10^{-10}E \left(\frac{\mu_{en}}{\rho}\right)_{tissue} \quad (5)$$

When considering shielding material between the source and the detector, the data for the parameters *A1, A2, α1,* and *α2* is taken from Table E.5 of Shultis & Faw [10] for Equation (6).

$$D(t) = \frac{1}{2\pi h} \sum_{j=1}^{6} S_{l,i} \bar{R}_J \sum_{i=1}^{2} A_{i,j} F\left(\theta, [1 + \alpha_{i,j}]\mu_j t\right) \quad (6)$$

3. Results & Discussion

This section has two main pillars; in the first pillar, we present experimental results with the statistical analysis, and in the second pillar, we present the case study results.

3.1 Experimental Results with Statistical Analysis

We have thirty measurements for each method mentioned in Table 1. In addition to comparing these two measurements, we also use a complete set of sixty measurements as a one-batch measurement dupped as ALL, and we compare NOREP, REP, and ALL measurements.

The average trap population's expected behavior in terms of light output per read-out is given in Equation (7) and depends on the number of read-outs n. Here, m1 is the total noise, and m2 is the bleaching constant.

$$\frac{dD(n)}{dn} = m_1 - m_2 D(n) \tag{7}$$

The solution of this Equation, given in Equation (8), provides a fitting function for the experimental data set. The new parameter m3 is the initial dose in Equation (8), representing the dose before any read-out. Moreover, m1/m2 is the background dose, one of the main outputs of this experiment.

$$D(n) = \frac{m_1}{m_2} - \left(m_3 - \frac{m_1}{m_2}\right)e^{-m_2 n} \tag{8}$$
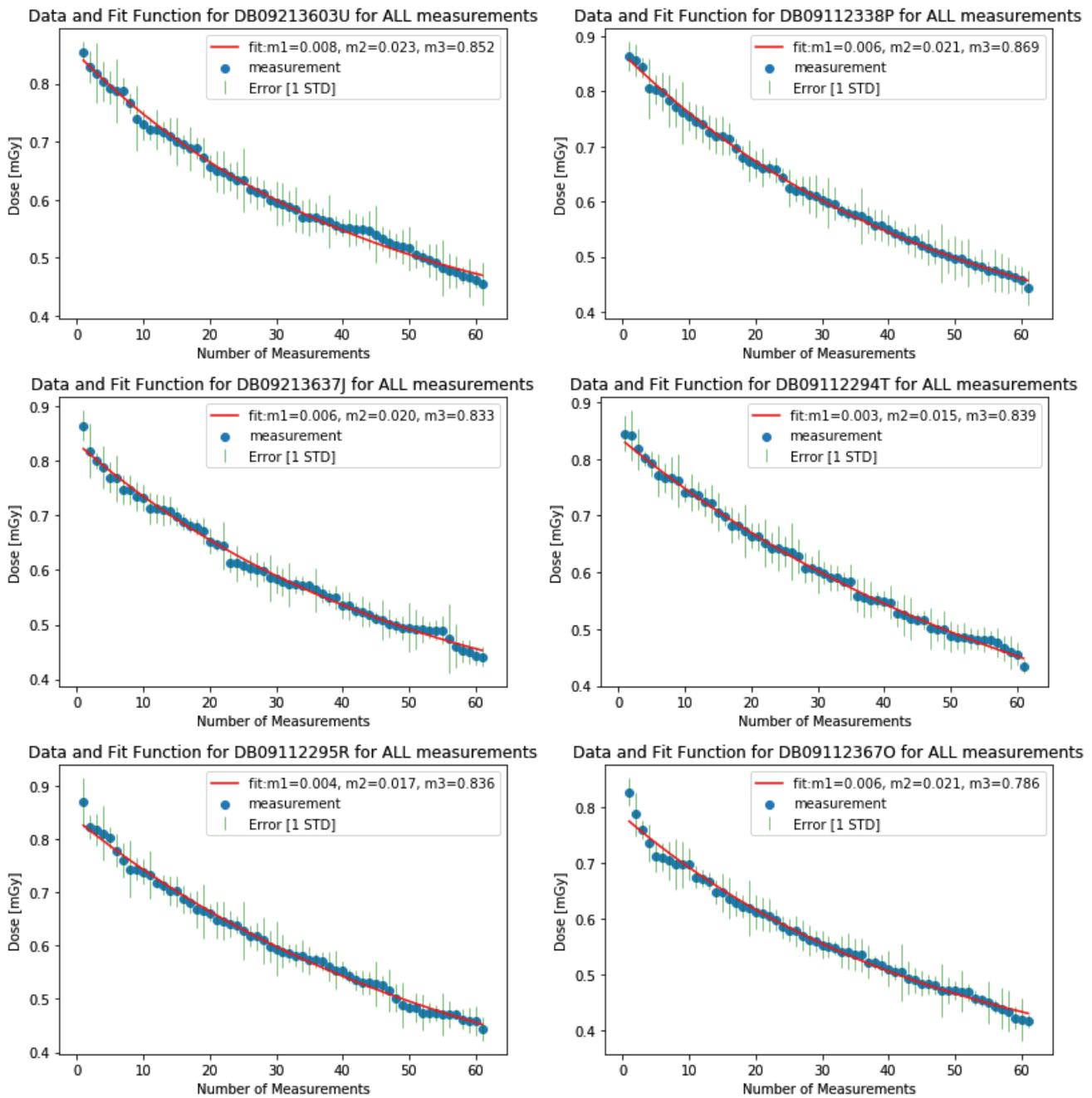
Figure 3: All measurements for six nanoDOT™ OSLD, including uncertainty and fitting model

Figure 3, Figure 4, and Figure 5 show the measurements and fit function for the measurement set, including uncertainty, for six nanoDot™ OSLD dosimeters. These include the whole set, the without removal set, and the removed set, respectively. Combining experimental results with Equation (8) gives background dose, bleaching constant, and initial dose with the uncertainty for six nanoDOT™ OSLD for two measurement methods and all measurements.

The background dose rate is assumed constant during the measurement. The obtained result for the background dose is shown in Table 2.

#	Bkg dose-All [mGy]	Bkg dose-NoRep [mGy]	Bkg dose-Rep [mGy]
1	0.345±0.034	0.424±0.077	0.750±0.114
2	0.298±0.022	0.264±0.139	0.282±0.066
3	0.293±0.045	0.082±0.289	0.151±0.294
4	0.189±0.039	0.279±0.123	0.267±0.188
5	0.242±0.046	0.338±0.164	0.359±0.101
6	0.294±0.042	0.440±0.122	0.035±0.250

Table 2: Evaluated background dose [m1/m2] for each dosimeter and each experiment part
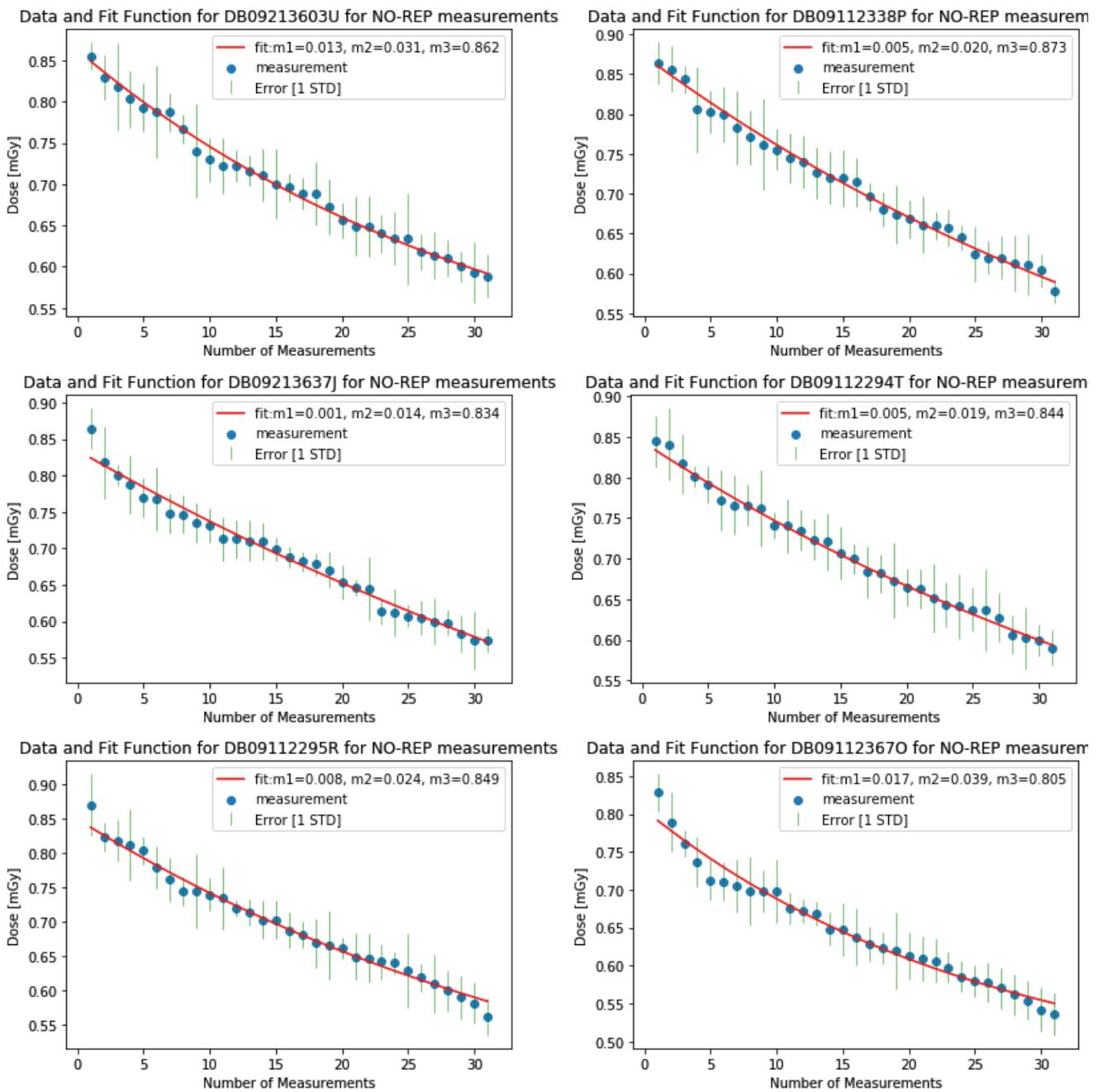
Figure 4: Measurements for six nanoDOT™ OSLD, NOREP method, including uncertainty and fitting model
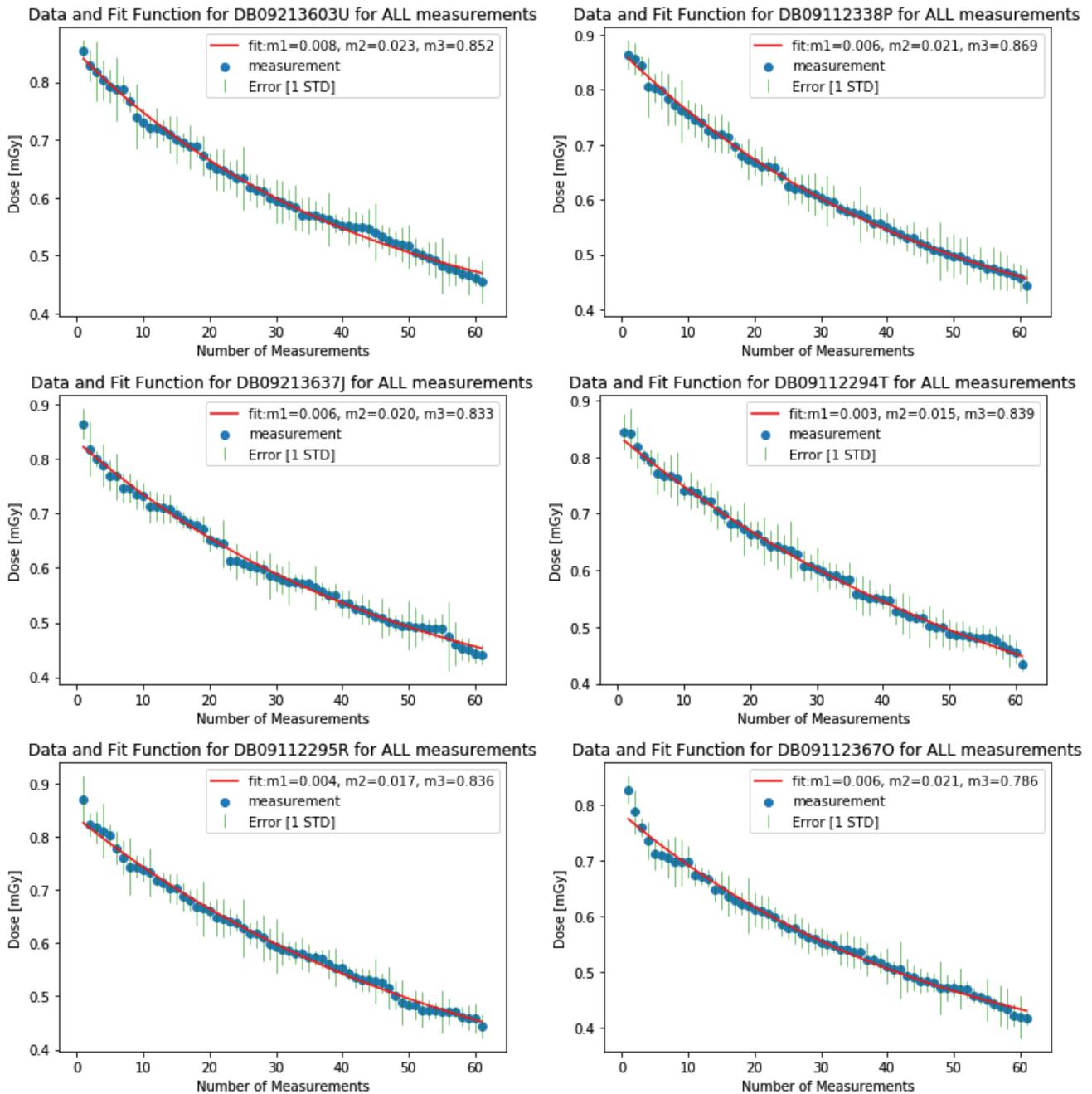
Figure 5: Measurements for six nanoDOT™ OSLD, NOREP method, including uncertainty and fitting model
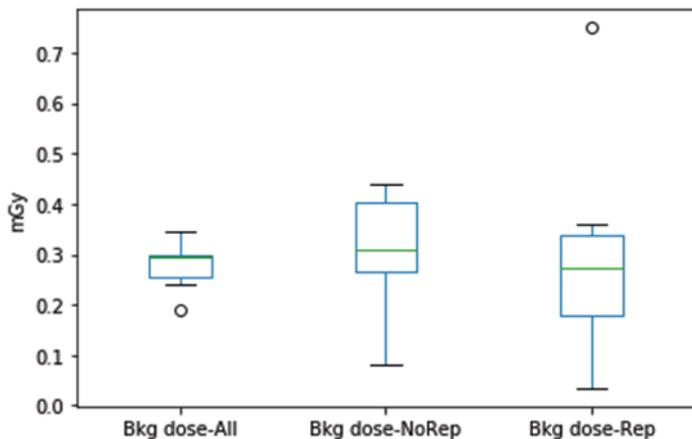
Figure 6: Comparison of background dose values for each case.

The first goal is to compare the REP, NOREP, and ALL results by applying ANOVA. Table 2, Table 3, and Table 4 are the results for the background dose, the bleaching constant, and the initial dose, respectively. According to Figure 7, Figure 9, and Figure 11, the residuals are normally distributed with a mean zero that obeys the first assumption of the ANOVA. Moreover, all the measurements are independent, meaning that the second assumption for ANOVA is also satisfied. On the other hand, the Bartlett test [7] for each parameter suggests that the variances are not homogeneous. Welch's ANOVA applies to the data set instead of ANOVA. As a result of Welch's ANOVA, we fail to reject the null hypothesis that the background dose, the bleaching constant, and the initial dose are equal. These leads using
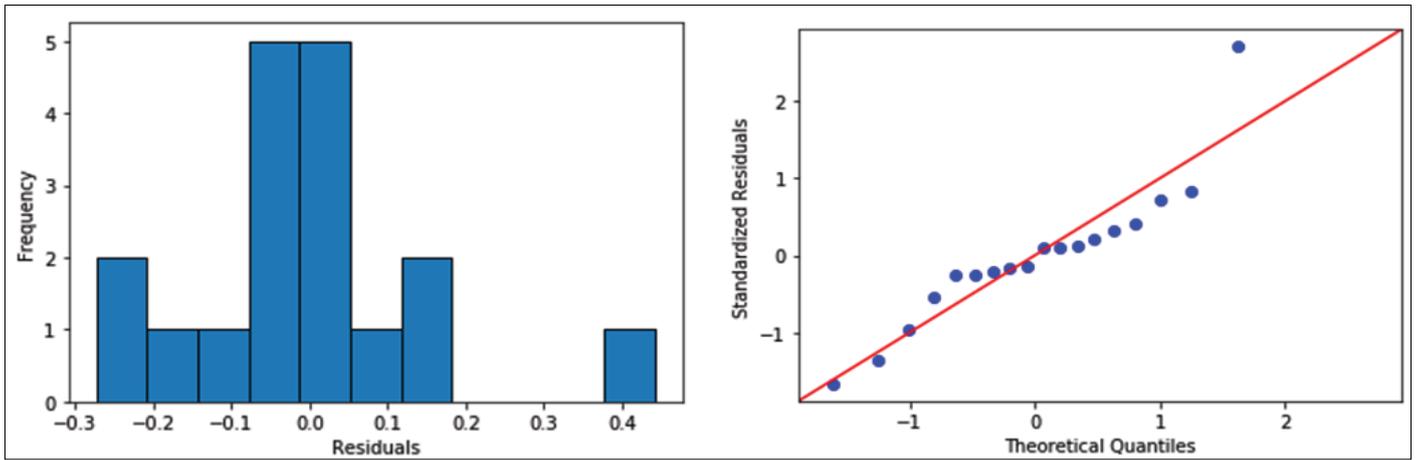
Figure 7: Histogram of residuals and QQ plot for background dose ANOVA analysis

#	Bleaching constant -All [1/read-out]	Bleaching constant -No Rep [1/read-out]	Bleaching constant -Rep [1/read-out]
1	0.023±0.001	0.031±0.003	-0.020±0.004
2	0.021±0.001	0.020±0.004	0.021±0.003
3	0.020±0.002	0.014±0.006	0.013±0.008
4	0.015±0.001	0.019±0.003	0.021±0.008
5	0.017±0.001	0.024±0.006	0.035±0.007
6	0.021±0.002	0.039±0.007	0.010±0.005

Table 3: Evaluated bleaching constant [m2] for each dosimeter and each experiment part

Figure 8: Comparison of bleaching constant values for each case
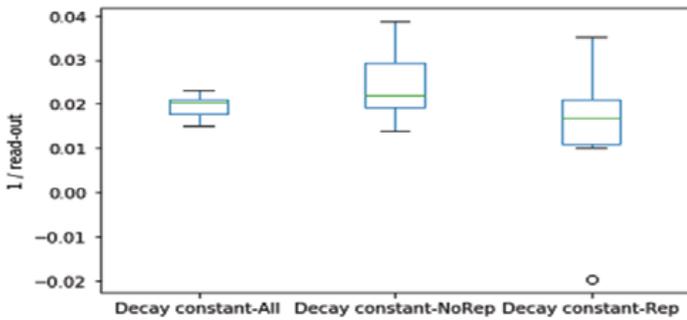
either REP or NOREP methods for measurement do not have a statistically significant difference.

Figure 6, Figure 8, and Figure 10 are for the comparison of the evaluated parameters. In every box plot, we have an outlier for the REP method. This region was the second suite of measurements so it would have been the most bleached. If it were sufficiently flat, an exponentially decaying exponential with sufficient noise can have the best fit give it a negative slope, just as would occur with a linear fit.

Figure 9: Histogram of residuals and QQ plot for bleaching constant ANOVA analysis

#	Initial dose -All [mGy]	Initial dose - NoRep [mGy]	Initial dose - Rep [mGy]
1	0.852±0.004	0.862±0.004	0.660±0.002
2	0.869±0.003	0.873±0.005	0.892±0.001
3	0.833±0.005	0.834±0.007	0.803±0.004
4	0.839±0.003	0.844±0.003	0.893±0.005
5	0.836±0.004	0.849±0.006	1.111±0.005
6	0.786±0.005	0.805±0.008	0.749±0.003

Table 4: Initial dose values for each case
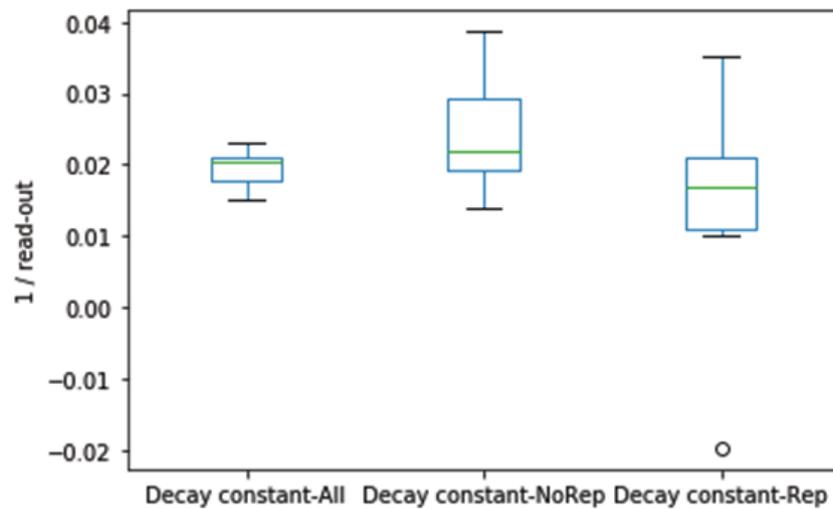
Figure 10: Comparison of bleaching constant values for each case
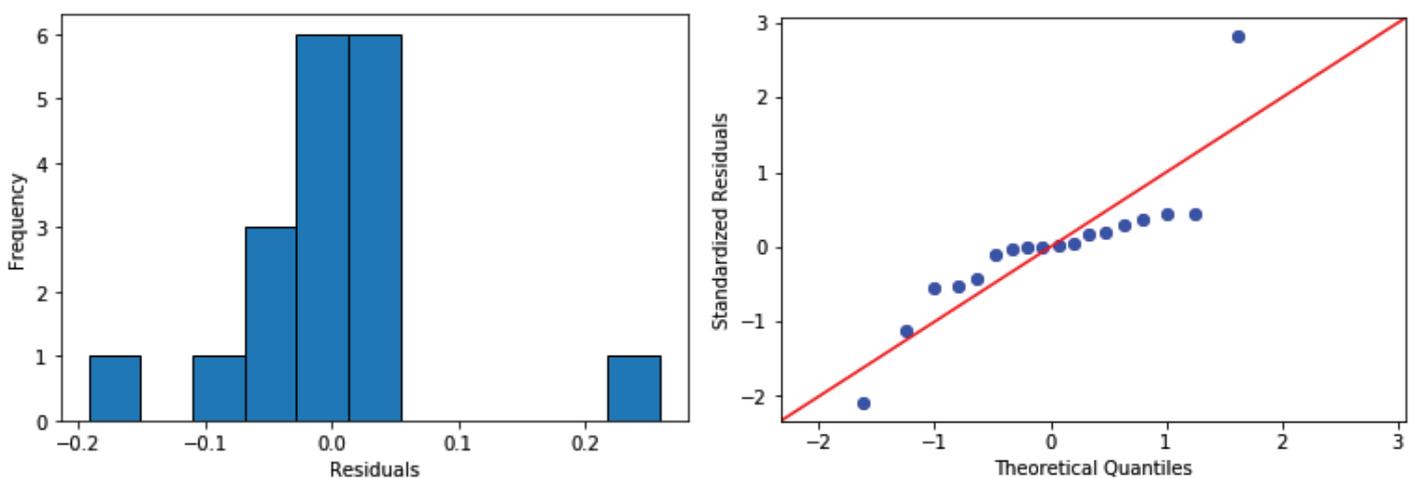
Figure 11: Histogram of residuals and QQ plot for initial dose ANOVA analysis

Energy [Mev], [1]	Kappa [Mev/sec] [2]	Average Energy [MeV] [3]	Kappa [photons/ sec] [4]=[2]/[3]	Source Strength [photons/ cm-sec] =[4]/[Length]
Group 1 [5-7.5]	5.1E+02	6.25	8.2E+01	4.1E-01
Group 2 [4-5]	3.93E+06	4.5	8.7E+05	4.4E+03
Group 3 [3-4]	1.98E+08	3.5	5.7E+07	2.8E+05
Group 4 [2-3]	2.35E+14	2.5	9.4E+13	4.7E+11
Group 5 [1-2]	9.99E+13	1.5	6.7E+13	3.3E+11
Group 6 [0-1]	4.61E+15	0.5	9.2E+15	4.6E+13

Table 5: Evaluated source strength for each group

3.2 Results of Case Study

The source strength for each group is evaluated as an initial step in Table 5.

Using the source strength, we obtain the uncollided - unshielded dose per length of the spent fuel assembly given in Table 6. Group 6 is responsible for more than 90% of the total dose rate. We ignore other groups' contributions to the dose rate for the remaining calculations.

Figure 12 shows the uncollided – unshielded dose rate for Group 6 only with the function of the spent fuel assembly's height. As expected, the maximum dose rate at the center is about 8.2 mGy/sec, much more than the background dose of about 0.3 mGy.

Figure 13 is the final result that we intend to obtain, the total dose rate for the concrete shielded spent fuel. As shielding thickness increases, the dose rate decreases as a function of the thickness of the shield. The total gamma-dose rate is about one mGy per second beyond the 22 cm concrete shield, which is even distinguishable from the approximate background dose, about 0.3 mGy.

Energy [Mev]	Average Energy [MeV]	Source Strength [photons/cm-sec]	$(\mu en/\rho)$tissue [cm²/g]	Response Function [Gy-cm²]	SI*R [Gy-cm]
Group 1 [57.5]	6.25	4.1E-01	1.77E-02	1.77E-11	7.2E-12
Group 2 [4-5]	4.5	4.4E+03	1.97E-02	1.42E-11	6.2E-08
Group 3 [3-4]	3.5	2.8E+05	2.15E-02	1.21E-11	3.4E-06
Group 4 [2-3]	2.5	4.7E+11	2.42E-02	9.69E-12	4.6E+00
Group 5 [1-2]	1.5	3.3E+11	2.81E-02	6.74E-12	2.2E+00
Group 6 [0-1]	0.5	4.6E+13	3.27E-02	2.62E-12	1.2E+02
TOTAL					1.3E+02

Table 6: Calculated uncollided dose rate as a function of the length of the spent fuel assembly
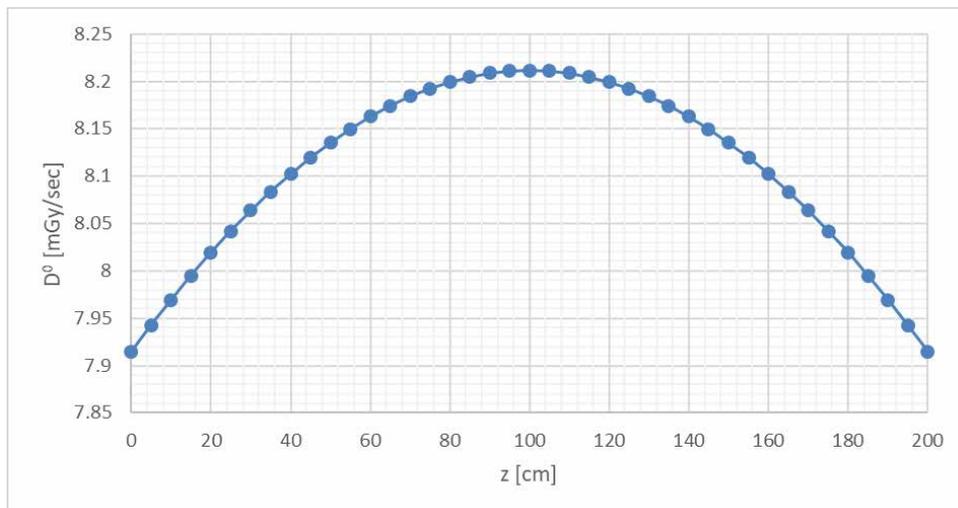
Figure 12: Dose rate distribution of spent fuel assembly for uncollided and unshielded case
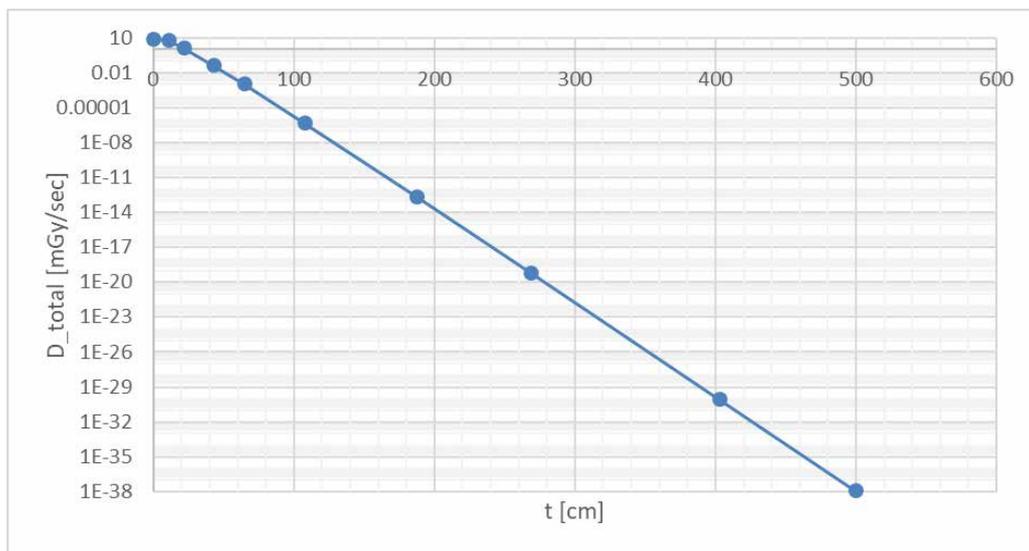
Figure 13: Total gamma dose rate for different shielding thickness

4. Conclusion

We propose a model to flag the illicit nuclear activities in nuclear facilities by using OSLDs as a complementary system to current detection systems.

We first set an experiment to obtain model variables to make the model work. Then we demonstrate a case study on a used nuclear fuel assembly in unshielded and concrete shielded cases to prove our model.

The result of this work allows us to get the background dose, the bleaching constant for the reader, and the initial dose with their uncertainties as a first step. The first step enables us to demonstrate our scenario. In the end, we flag the illicit nuclear activity in the proposed facility.

As follow-up work, we plan to redo the experimental part by using OSLDs from the same producer exposed to a high radiation level of more than 500 mGy. This high dose level will prove the model more realistically. The follow-up work will allow us to generalize our model for nuclear facilities. Some specific cases are intended to be picked as a scene rather than traditional light water reactors' spent fuel as additional works.

5. Acknowledgements

6. References

[1] Landauer, "MicroSTAR User Manual." Landauer, 2019. [Online]. Available: https://www.landauer.com/sites/default/files/2020-01/MICROSTAR%20ii%20USER%20MANUAL.pdf

[2] S. W. S. McKeever, *"Optically stimulated luminescence: A brief overview,"* Radiation Measurements, vol. 46, no. 12, pp. 1336–1341, Dec. 2011, doi: 10.1016/j.radmeas.2011.02.016.

[3] H. Y. Göksu and I. K. Bailiff, *"Luminescence dosimetry using building materials and personal objects,"* Radiation Protection Dosimetry, vol. 119, no. 1–4, pp. 413–420, Sep. 2006, doi: 10.1093/rpd/nci699.

[4] International Atomic Energy Agency, Ed., *Detection of Radioactive Materials at Borders.* Vienna: IAEA, 2003.

[5] J. Medalia, *"Detection of Nuclear Weapons and Materials: Science, Technologies, Observations,"* Congressional Research Service, R40154, Jun. 2010.

[6] J. L. Devore, *Probability and statistics for engineering and the sciences.* Boston, MA: Boston, MA : Cengage Learning, [2016], 2016.

[7] G. W. Snedecor and W. G. Cochran, *Statistical Methods.* Ames: Ames : Iowa State University Press, [1967], 1967.

[8] H. Liu, "Comparing Welch's ANOVA, a Kruskal-Wallis test and traditional ANOVA in case of Heterogeneity of Variance," Master of Science Thesis, Virginia Commonwealth University, Richmond, Virginia, 2015.

[9] R. J. LaBauve, T. R. England, D. C. George, and C. W. Maynard, "Fission Product Analytic Impulse Source Functions," *Nuclear Technology,* vol. 56, no. 2, pp. 322–339, Feb. 1982, doi: 10.13182/NT82-A32861.

[10] J. K. Shultis and R. E. Faw, *Radiation Shielding*. La Grange Park, IL: La Grange Park, IL : American Nuclear Society, [2000], 1996.

Developing a Big Data Framework for Processing Sentinel-2 Data in the Context of Nuclear Safeguards
Evaluation of Apache Airflow, Rasdaman and Google Earth Engine

Lisa Beumer and Irmgard Niemeyer

Forschungszentrum Jülich GmbH, Germany
E-mail: l.beumer@fz-juelich.de, i.niemeyer@fz-juelich.de

Abstract:

In the last years, Earth observation (EO) satellites have generated big amounts of geospatial data. Many providers offer their satellite data at low cost or even for free. For example, initiatives such as the Copernicus program, the European Union's Earth observation program, have revolutionized the market. The growing archives of satellite imagery open up a wide range of satellite EO applications, also in the field of nuclear verification where satellite imagery represents a key source of information for the implementation and verification of nuclear non-proliferation treaties [1]. The data collected, processed, analyzed, and managed for monitoring purposes is not only increasing in volume, but also becoming more and more heterogeneous, unstructured, and complex. However, Big Data is also accompanied with several issues related to capturing the data, sharing, transferring, updating, processing, and analyzing. To meet these demands, novel technologies have been developed. Apache Airflow for example has become a popular tool for defining, scheduling, visualizing, and monitoring Big Data related workflows [2]. For storing and accessing multidimensional raster data, such as satellite imagery, an array database management system, called Rasdaman, has become well established [3]. To analyze these large amounts of data effectively and efficiently, Google has developed a free-to-use cloud computing platform, known as Google Earth Engine (GEE) [4]. In this research an automated procedure for collecting, storing, processing, and analyzing satellite images based on the tools mentioned above was developed. Hereby, the strengths of Airflow in terms of the creation of dynamic workflows with high granularity and the log entries of execution became evident. Furthermore, Rasdaman provides indispensable advantages such as the open standards-based data-cube analytics possibilities. The usability and benefits of GEE with respect to big EO data management and analysis were evaluated through an analysis of two different machine learning algorithms, namely Random Forest (RF) and Classification and Regression Trees (CART). Regarding the target land over classes, the classification results of manual generation were compared with two by GEE provided land cover maps from the years 2017 and 2019. The overall accuracy of the RF and CART classifiers for the Sentinel-2 images was in the range of 87% to 98%, and 68% to 83%, respectively.

Keywords: Satellite Imagery; Big Data; Data Science; Airflow; Copernicus Hub; Rasdaman; Google Earth Engine

1.　Introduction

Geospatial data, satellite imagery in particular, represents a key source of information for the implementation and verification of nuclear non-proliferation. In 1998, the IAEA started to investigate the potential use of commercial satellite imagery to support the safeguards implementation and nowadays it "[…] has become a very important source of information [...] especially with respect to sites to which the IAEA does not have access." [1]. Many applications of satellite imagery in the field of nuclear verification have been identified over time. With commercial satellite imagery available to the public, new opportunities are emerging to monitor nuclear activities at both known and undeclared nuclear facilities in a more proactive manner to verify compliance with non-proliferation agreements. As numerous studies have shown, satellite imagery provides analysts with clear insights into nuclear facilities and nuclear activities worldwide, for example, to confirm the status of an inoperable facility or declared production without having to visit the sites in person [1,5,6,7]. Moreover, use cases such as the recognition and monitoring of small-scale features for instance the construction of buildings, plant expansions or the preparation of underground facilities are also considered. The amount of available and heterogeneous satellite EO data is steadily increasing , as no longer only a few operators and government sources offer the data as primary source, but private companies are also investing in EO satellites, driven by technological advances that allow for higher resolution sensors and a higher return rate capacity. Many provider offer their satellite data at low cost or even for free. For example, initiatives such as the Copernicus program, the European Union's Earth observation program, have revolutionized the market. The demand for their immense amount of data is huge, as the Copernicus Sentinel Data Access Report of 2020 shows [8]. In 2020, a total data volume of 7.65 PiB was published, which is significantly more compared to the European Space Agency's (ESA's) entire collection of EO data from the pre-Copernicus era which amounts to 5.6 PB [8]. The average daily download volume of the Sentinel Data Access System was

405 TiB, resulting in a total of 82.8 PiB of products downloaded just in 2020 [8].

Due to the sheer volume and the velocity at which the amount of data is increasing, remote sensing data is referred to as Big Data. However, aspects such as diversity, complexity and trustworthiness also make this type of data Big Data. But what exactly does Big Data mean and to what extent does it apply to satellite imagery? The term Big Data refers to large data sets, whether structured, unstructured, or complex, that are difficult or even impossible to store, process and analyze using conventional methods. To define the term more precisely, several multi-V models were introduced in the last years starting with the 3-V model. In this paper, the 5-V model, characterized by the following propertied, is taken into consideration: Volume, velocity, variety, veracity and value. The term volume simply refers to the quantity of an existing and fast-moving amount of data. In this context, there is no upper boundary at which data is considered to be "big". The speed at which the data accumulates is summarized under the second V, namely velocity. Depending on the data source, a different data type is present, for example, data can be of an unstructured, semi-structured or structured nature. This characteristic is represented by the term variety. Veracity describes the data quality and its accuracy. Only data with known origin and quality, such as correctness and completeness, are generally considered to be reliable and can be trusted. In addition, data analysis can only provide a meaningful result if high quality data is available as input data. The last characteristic, referred to as value, refers to the usefulness of data. This raises the question of the benefit of high quality data if there is no use case in terms of a concrete example. So, one has to weigh whether to store all data or only useful data, the so-called smart data.

The data quantity and quality keeps moving forward with the aim of offering high and medium spatial resolution images on a daily basis. However, Big data is also accompanied with several issues related to capturing the data, storing, processing and analyzing it. In turn, this will create new challenges for the analyst to use the datasets appropriately and in a timely manner. No longer can visual interpretations of single satellite image scenes be expected to address the analysis requirements for such large repositories of satellite imagery datasets. To meet these demands, novel technologies have been developed. Apache Airflow for example enables the optimization of data processing and workflow management processes [2]. For storing and accessing multidimensional raster data, such as satellite imagery, an array database management system, called Rasdaman, has become well-established [3]. To analyze large amounts of data effectively and efficiently, Google has developed a free-to-use cloud computing platform called Google Earth Engine (GEE). The platform provides access to publicly available remote sensing imagery and machine learning algorithms [4]. In this research, these tools have been utilized to develop a semi-automated procedure for collecting, storing, processing and analyzing satellite images. The project plan is shown in Figure 1.

Within the scope of this work, Sentinel-2 data is obtained from the Copernicus program. Due to the diversity of possible data, a comprehensive preparation of the data in process usable formats is necessary to be able to use appropriate analysis algorithms. The data source is integrated into the Apache Airflow workflow management system capable of downloading, validating, preprocessing and storing the data into a Rasdaman database. Finally, the efficiency of the Google Earth Engine to effectively execute Big Data workflows using Google's provided machine learning techniques is explored. The potential of the developed framework is tested using case studies concerning nuclear fuel cycle related sites. Hereby the objective is to classify land cover use, as these features provide essential information for recognizing and monitoring for example changes of the operational status, constructions of new buildings and roads, plant expansions, etc.

2. Tool Fundamentals

2.1 Airflow

Apache Airflow is an open-source workflow management platform written in Python that enables the creation and management of data pipelines, as well as their automatic
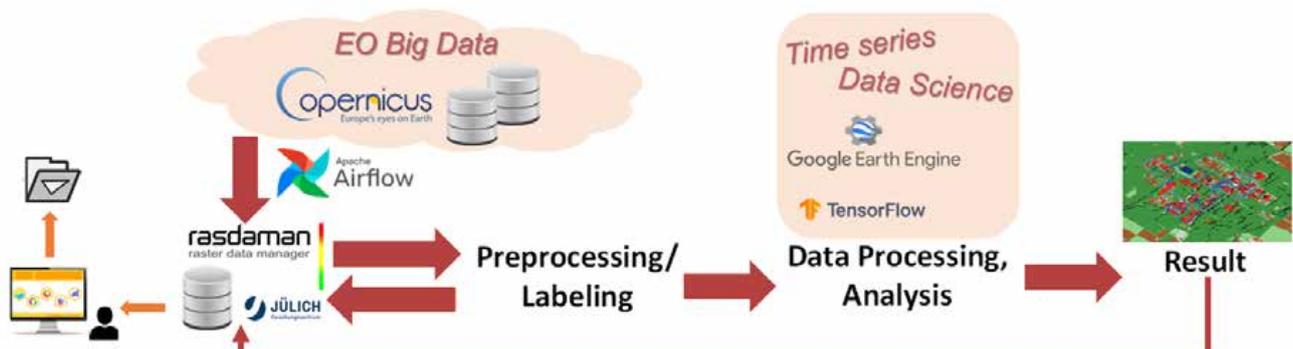
Figure 1: Project overview.
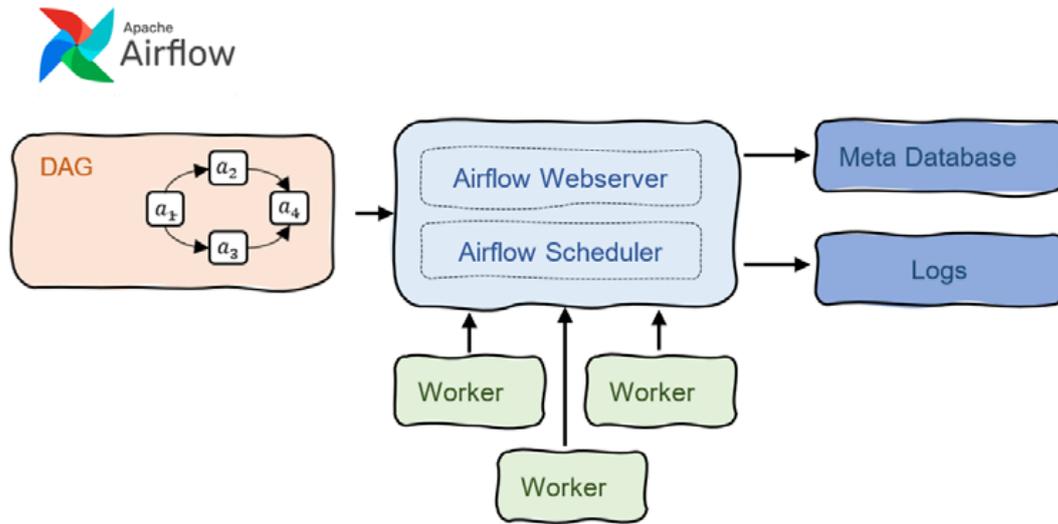
Figure 2: Terminology used in Airflow. The key concept are direct acyclic graphs (DAGs) managed by a webserver and a scheduler.

execution [1]. The key terminology used in Airflow is shown in Figure 2.

The core concept of Airflow is represented by Directed Acyclic Graphs (DAGs), which collect tasks together. A DAG forms an abstract structure consisting of nodes and edges. The nodes represent the individual work tasks and the edges the connections between them, having a direction. According to [9] a direct graph is defined as $G = \{V, E\}$, where $V = \{a_1, a_2, ..., a_N\}$ is a finite set of nodes and E a finite set of directed edges. It holds that $E = \{a_j \rightarrow a_{j'} | a_j, a_{j'} \in V, a_j \neq a_{j'}$ This graph is called acyclic if there does not exist

$$(j, j^{(1)}, j^{(2)}, ..., j^{(n)}) \; s.t. \{a_j \rightarrow a_{j(1)}, a_{j(1)} \rightarrow a_{j(2)}, ..., a_{j(n-1)} \rightarrow a_{j(n)}, a_{j(n)} \rightarrow a_j\} \subset E$$

An example of a DAG is given in Figure 3.
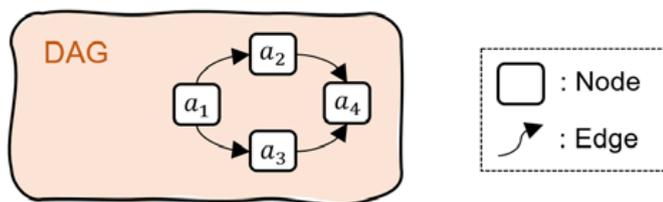
Figure 3: Simple example of a DAG $G = \{V, E\}$ with the set of events $V = \{a_1, a_2, a_3, a_4\}$ and
$E = \{a_1 \rightarrow a_2, a_1 \rightarrow a_3, a_2 \rightarrow a_4, a_3 \rightarrow a_4\}$ representing the set of directed edges.

The management of the whole system is performed by a graphical user interface (webserver) and a scheduler. The webserver enables the creation of workflows and their management. The detailed status of each workflow can be displayed, thus making a live monitoring possible. Apache

Airflow manages and controls workflows via schedulers, whereas both sequential and parallel schedulers are supported. Workflows are executed according to a predefined schedule or trigger events. Once the schedule is created, according to which the tasks of the defined rules can be processed, the scheduler assigns them to the workers, which are responsible for the actual processing of the individual tasks according to their respective Python description. For documentation purposes, all task information is stored in the meta database. Log files can be used for debugging, error analysis or documentation.

2.2 Copernicus Data Hub

Copernicus, formerly known as GMES (Global Monitoring for Environment and Security), is a European Union program aimed at establishing a European capacity for global environment and security monitoring [10]. The Program is funded, coordinated, and managed by the European Commission in cooperation with partners such as ESA (European Space Agency) and EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites). The program provides data from its own fleet of satellites, called Sentinel, in-situ data and data from national and commercial satellites [11]. The Sentinel satellites consist of six missions: Sentinel -1 (High Resolution Radar), -2 (Optical for Vegetation), -3 (Optical/Thermal for Oceans), -5P (An Eye for Air) and -6 (Sea Level Elevation) are stand-alone satellites, while Sentinel-4 and -5 are dedicated measurement instruments installed on EUMETSAT. There are currently eight Sentinels in space, namely Sentinel-1A & -1B (2014, 2016), Sentinel-2A & -2B (2015,2017), Sentinel-3A & -B (2016, 2018), Sentinel-5P (2017), Sentinel-6 (2021). The Sentinel data and Copernicus services are free of charge and are provided through ESA's Copernicus Open Access Hub, previously known as Sentinels Scientific Data Hub.

One has the possibility to download the data through a graphical interface or via two different application programming interfaces (APIs), OData and Open Search (Solr). In this work, OData, a data access protocol built on the HyperText Transfer Protocol (HTTP) and the Representational State Transfer (REST), was used because it can be easily integrated in Python using Client for URLs (cURL) or Wget. The data resources to be queried are uniquely identifiable via so-called Uniform Resource Identifiers (URIs) and can be requested via HTTP messages. As shown in Figure 4, a URI is composed of up to three components: (1) Service Root URI, (2) Resource path and (3) Query options that control the amount and order of the data.

Figure 4: Example of an OData URI. Source [12].

Since the queries used in this work are too complex and thus vulnerable to cURL and Wget errors, a generic bash script was implemented. This is based on the script dhusget.sh provided by Copernicus, which is a simple demo script illustrating how to use OData and OpenSearch APIs to query and download products from any Data Hub Service [13].

2.3 Rasdaman

Originally, databases were developed to store, manage, and query alphanumeric data efficiently. However, data storage requirements have changed over the years. For example, when looking at satellite imagery, it is necessary to be able to store multi-dimensional data. Array database management systems (array DBMSs) provide database services specifically for raster data and aim to provide a flexible and scalable management of this kind of data. In the context of this work, the array DBMS Rasdaman is reviewed, which aimed to form a comprehensive DBMS

support for raster data of arbitrary size and dimension over arbitrary base types, so-called multidimensional discrete data (MDD) [3]. Before the system architecture is described, the logical data model used by Rasdaman is explained whereby the declaration of [14] is used.

2.3.1 Logical data model

Multidimensional array data, also known as multidimensional discrete data (MDD) is located in a discrete space $\mathbb{z}^d$. Figure 5 illustrates a three-dimensional MDD in a discrete space $\mathbb{z}^3$.

A multidimensional object a is the mapping of a value of the base type to each vector of its domain, i.e., the multidimensional interval it takes:

$\alpha := \{(x, v(x)) | v(x) \in T, x \in D\}$, where x describes the cell and $v(x)$ the corresponding cell value. The domain is spanned by an interval D of dimensionality d, where each dimension i has a lower bound l_i and an upper bound j_i:

$$D := \underset{i=1}{\overset{d}{X}} \{x | 1_i \le x \le h_i, x \in \mathbb{z}\} = [1_l : h_l] \times \dots \times [1_d : h_d],$$

$$l, h \in \mathbb{z}^d \text{ and } l_i \le h_i \forall i \in \{1, \dots d\}.$$

A single cell value $v(x)$ can be assigned a base type T, which may be of atomic or composite data types

$$T := \underset{i=1}{\overset{n}{X}} \{t_i | t_i \in \zeta \cup T\}; n \in \mathbb{N}; \zeta \in \mathbb{N}_0 \cup \mathbb{Z} \cup \mathbb{R} \cup \mathbb{B} \dots.$$

A cell can represent a single value such as a gray value or a composite value, for example the red, green, and blue components of a color image. In addition to the basic data type T, an MDD has a data type M, which is described by D and T, $M=<D,T>$. A set of MDD of the same type M are called collection, defined by $C \subset \{\alpha | type(\alpha) = <D, T>$. If operations are now applied, a distinction is made between those on MDD and those on collections. Geometric
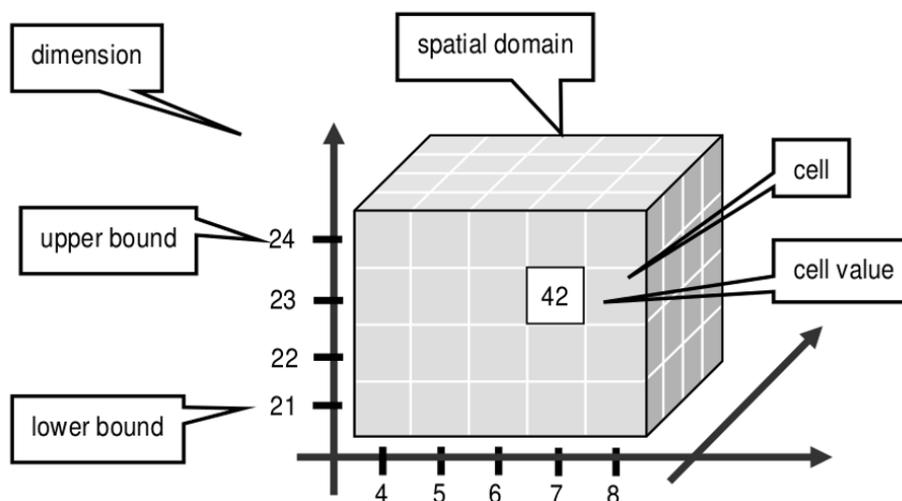
Figure 5: Constituents of a three-dimensional MDD. Source [15].

operations, induced operations, aggregate operations, and cell operations can be applied to multidimensional objects. In case of collections, relational operations such as application, selection, cross product, among others, can be applied.

Big Data cubes can be created in Rasdaman via an OGC (Open Geospatial Consortium) Web Coverage Service - Transaction Extension (WCS-T) standard interface that allows users and machines to insert, update, and delete data via simple web requests. A Python tool wcst_import is provided for this purpose. This tool is based on two concepts: (1) Recipes and (2) Ingredients. The recipe defines how data files are combined into a coverage. All information needed to create a data cube are specified in an ingredient file. This is a JavaScript Object Notation (JSON) file based on a recipe which translates the files and information specified in the ingredient file into the data cubes. If wcst_import is run again with a different set of files to be imported, the data cube will be updated at the correct positions.

2.3.2 Rasdaman architecture

The Rasdaman storage concept relies on a separate data storage. The raster data is stored in the file system and the metadata in a separate database. The client-server system can be summarized as a four-layer architecture as shown in Figure 6.
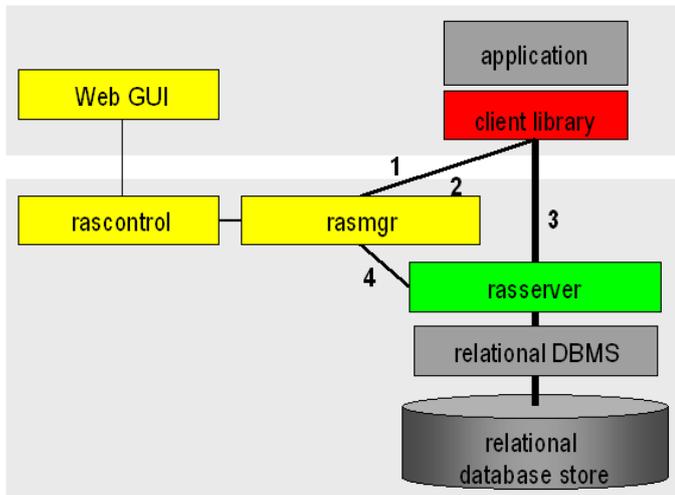
Figure 6: Rasdaman architecture. Source [16].

The foundation is formed by a conventional relational DBMS, which allows efficient storage of large volumes of data. The second layer is the Rasdaman server (rasserver), which provides various functions such as an interface to the relational database, metadata management and query processing. Furthermore, the server interacts with a Rasdaman manager (rasmgr). This manager handles tasks such as security functions, authentication, or multi-user operation by allocating requests to different Rasdaman servers. The fourth layer is the client. An application initially

builds a connection to the Rasdaman manager, which then establishes communication between a server and the application.

2.4 Google Earth Engine

In 2010, Google Inc. launched the development of a platform called Google Earth Engine (GEE), offering cloud computations for EO products, initially focusing on forest monitoring using satellite imagery, but later expanding to a variety of applications related to Earth Observation [4]. These include case studies such as Map of Life, Global Surface Water, or Collect Earth [17]. There exist a series of applications in the field of Earth surface analysis [18] but very few suitable for our purposes [19]. The Google Earth Engine is free for research, education, and nonprofit use. Since the platform is a browser-based IDE (Integrated Development Environment), no separate software needs to be downloaded and maintained. To use GEE, a JavaScript-based code editor is provided. Furthermore, data can be requested and analyzed using the Earth Engine (EE) Python API. To process the data Google infrastructure is provided consisting of a large pool of servers, co-located with the data that allows for fast data processing. In addition to the cloud computational capabilities, GEE offers an exhaustive catalog of remote sensing datasets including multispectral, radar, aerial, climate, land cover and vector data including data from satellite missions such as Landsat, Sentinel, MODIS as well as high-resolution imagery data sets [20]. The data is updated and expanded daily. When working with data from the GEE data catalog, three dataset types have to be distinguished: features, images and collections. A feature *(ee.Feature)* is a geometric object containing a list of properties. Images *(ee.Image)* are like features but may include several bands. A combination of features or images is called collection *(ee.ImageCollection)*. Machine Learning is supported via EE API methods and export and import functions for TensorFlowRecord files. The EE API provides methods such as *ee.Classifier, ee.Clusterer or ee.Reducer*. After performing the analysis, it is then possible to export a resulting *ee.Image* as a Geo-TIFF to Google Drive or the local machine.

2.4.1 Google Colaboratory

Google Colaboratory is a cloud based hosted Jupyter Notebook service developed by Google specifically for machine learning applications. It allows users to develop, execute and share python code within Google Drive. It provides limited and up to a certain point free access to central processing units (CPU), graphical processing units (GPU), and tensor processing units (TPU). The EE Python API can be easily deployed in a Google Colaboratory notebook.

3. Case Study

3.1 Data Acquisition using Apache Airflow

The starting point of the developed framework is the workflow management implementation using Apache Airflow. To ensure flexibility and scalability, it has been implemented dynamically. Dynamic DAGs are usually better for dynamically loading configuration options or changing operator options. In this case, the DAG is built dynamically based on two configuration files. The first one contains an overview of areas of interest (AOIs), defined by a name, the corresponding polygon, and the satellite from which the data should be requested. The corresponding satellite configuration is stored in another file, where its name, provider, and product type define each satellite. The Open Access Hub offers data starting from Level-1C. As a product type, Level-2A was chosen for this project because only atmospheric corrected and orthorectified data are only available for Level-2A (see Fig. 7).

(a) Level-1C

(b) Level-2A

Figure 7: Two different processing stages of Sentinel-2 data. Source [21].

In addition, the maximum cloud cover percentage that the image may have can be specified as well as it's geometric resolution. This satellite provides a set of 13 spectral bands spanning from the visible (4 bands) and near infrared (6 bands) to the shortwave infrared (3 bands) with a resolution of 10m, 20m, 60m depending on the wavelength. As test examples, we defined six different areas from which we only request Sentinel-2 bands having a geometric resolution of 10m or 20m and a cloud cover percentage less than ten percent. For each satellite provider, one DAG is created. The resulting structure can be seen in Figure 8.

The DAG consist of as many tasks as specified AOIs. Since Airflow allows the execution of parallel tasks, the graph adopts a tree structure. However, since Copernicus Hub can only handle two server requests at a time, the node CopernicusHub_start has only two child nodes. In the user interface (UI), the blue nodes represent task groups, which are a UI grouping concept and useful for creating repeating patterns. In each task group, the same sequence of tasks is executed accordingly to the different areas. The tasks are shown in Figure 9.

At the beginning, it is checked whether the corresponding data entry already exists in the database. If so, the execution of the task group is terminated. Otherwise, it is verified whether the data is available on the platform. For this purpose, the modified dhusget.sh script is executed. If no data is available, no further tasks of the group will be executed. If the data is available, the task *checkDownload* passes the output of the executed script to the next task by using a cross-communications message (xcom). The script response may look like *id('2b17b57d-fff4-4645-b539-91f305c27c69')* which represents an individual entity given by the UUID (Universally Unique Identifier) *'2b17b57d-fff4-4645-b539-91f305c27c69'*. The next step is to determine whether the data is available for download or not. The availability of online products on the Data Hub can be identified by means of an OData query. If the data is online, it is downloaded directly. Otherwise, the download request automatically triggers the request for restoring the data from

Figure 8: Dynamically created DAG based on the two configuration files.

Figure 9: The actual tasks to be performed per area organized into Task Groups.

the archive. Restored data is then kept online for at least three days. In case of a download, the file is then checked for completeness. For this, the MD5 (Message-Digest Algorithm 5) checksum provided by the Data Hub is compared with the MD5 value of the download. Sentinel-2 data is provided in the form of data packages (tiles) with a size of 100x100 km2. Therefore, it may happen that several zip files are downloaded for one query which consequently have to be merged. The result is one image per geometric resolution cropped to the AOI. The last step is the insertion of the images into the Rasdaman database using the tool *wcst_import.* The corresponding recipe is dynamically filled with all file related information and is shown in Listing 1.

This recipe contains all the necessary information, such as the size of the image data, the associated coverage ID and the resolution. The input section contains information about the source files to be considered. The structural information such as the data cube type, the timestamp and metadata are part of the recipe section.

3.2 Analysis using Google Earth Engine

To analyze this data, artificial intelligence (AI), in particular machine learning (ML) is applied, aiming at building and improving a generalizing system based on relevant data that automatically identifies patterns of data not previously introduced. There exist different types of machine learning algorithms such as supervised, unsupervised, semi-supervised, and reinforcement learning. A number of studies have already been conducted on the application of supervised and unsupervised methods with respect to our use case [22]. In this paper, we provide an overview of two machine learning algorithms, namely Random Forest (RF) and Classification and Regression Trees (CART) provided by GEE. The decision tree algorithm CART developed by [23] provides decision trees for classification, as well as for regression. The key to this algorithm is to find an optimal binary separation. For this purpose, a univariate binary decision tree is built by the algorithm. The Gini index is used as an impurity measure and minimal cost-complexity pruning is

```
"input": {
  "coverage_id": "CopernicusHub_Sentinel_2",
"paths": [
  "masked_202109*.tif"
  ]
},
"recipe": {
  "name": "general_coverage",
  "options": {
    "coverage": {
    "crs": "OGC/0/AnsiDate@EPSG/0/4326",
    "metadata": {
     "type": "xml",
     "global": {
      "area_or_point": "Area",
      "resolution": "10m",
      "provider": "Copernicus Open Access Hub",
      "product_type": "S2MSI2A",
      "level_description": "The SENTINEL-2
      Level-2A processing provides Level-2
      products (BOA reflectance)
      from Level-1C products (TOA reflectance)."
    },
},

"slicer": {
    "type": "gdal",
    "bands": [Band 1, Band 2, Band 3, Band 4,
             Band 5, Band 6],
    "axes": {
    "ansi": {
    "min": "datetime(regex_extract('${file:name}',
    '(.*)_(.*)\\.(.*)', 2), 'YYYYMMDD')",
    "type": "ansidate",
    "irregular": true,
    },
    "Lat": {
    "min": "${gdal:minY}",
    "max": "${gdal:maxY}",
    "resolution": "${gdal:resolutionY}",
    "gridOrder": 2,
    "crsOrder": 1
    },
    "Lon": {
    "min": "${gdal:minX}",
    "max": "${gdal:maxX}",
    "gridOrder": 1,
    "crsOrder": 2,
    "resolution": "${gdal:resolutionX}"
```

Listing 1: Sentinel-2 L2A recipe.

Figure 10: Methodology for classification on the GEE platform using the machine learning algorithms RF and CART.

used after the tree is built. The GEE library provides the technique classifier.smileCart. Random Forest [24] is a well-known supervised machine learning method, which is based on decision trees and is used for classification and regression tasks. In Random Forests, many decision trees are created randomly based on so-called bootstrap datasets. Each tree makes individual decisions on its own. Classification is done by repeatedly applying a learning procedure to bootstrap samples of the training data and then aggregating the individual results. Since the individual decision trees can be built and trained quickly and in parallel, the overall algorithm also trains fast. In this study GEE's technique classifier.smileRandomForest was used. Their performance is compared using accuracy assessment. The methodology used for training the classifier is shown in Fig. 10.

As previously mentioned, Sentinel-2 imagery with a cloud cover of less than ten percent is collected for six different AOIs. The median was used to compose the Sentinel-2 images for the entire years of 2017, 2019 and 2021. The bands B2-B8A as well as the normalized difference water index (NDWI), the normalized difference vegetation index (NDVI) and the bare soil index (BSI), calculated as follows

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)}, NDWI = \frac{(Green-NIR)}{(Green+NIR)}, BSI = \frac{((Red+SWIR)-(NIR+Blue))}{((Red+SWIR)+(NIR+Blue))}$$

with near infrared (NIR), and short-wave infrared (SWIR), were used as input features.

Since the selected machine learning methods are supervised algorithms and the Sentinel-2 data do not contain labels, training data must be collected. Two different methods are compared for this purpose. On the one hand, *FeatureCollections* for two AOIs are created manually using the GEE drawing tool and on the other target labels extracted from two different tagged land cover datasets [25, 26] provided by GEE were used for all six AOIs. An overview of available labeled land cover datasets can be found in [27].

For the manually extracted features, 72 feature polygons were selected distributed throughout the first study area and 89 for the second, covering six different classes, seven respectively. The features and the corresponding study areas are shown in Figure 11.

(a) FeatureCollection AOI1

(b) FeatureCollection AOI2
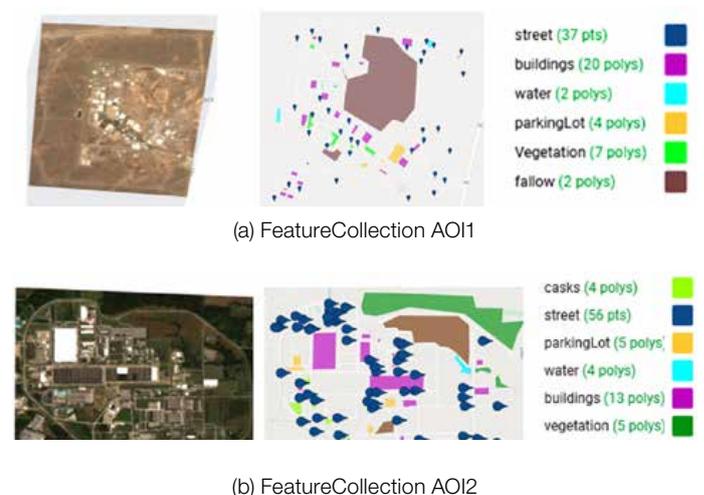
Figure 11: Manually created FeatureCollections consisting of 72 and 89 features
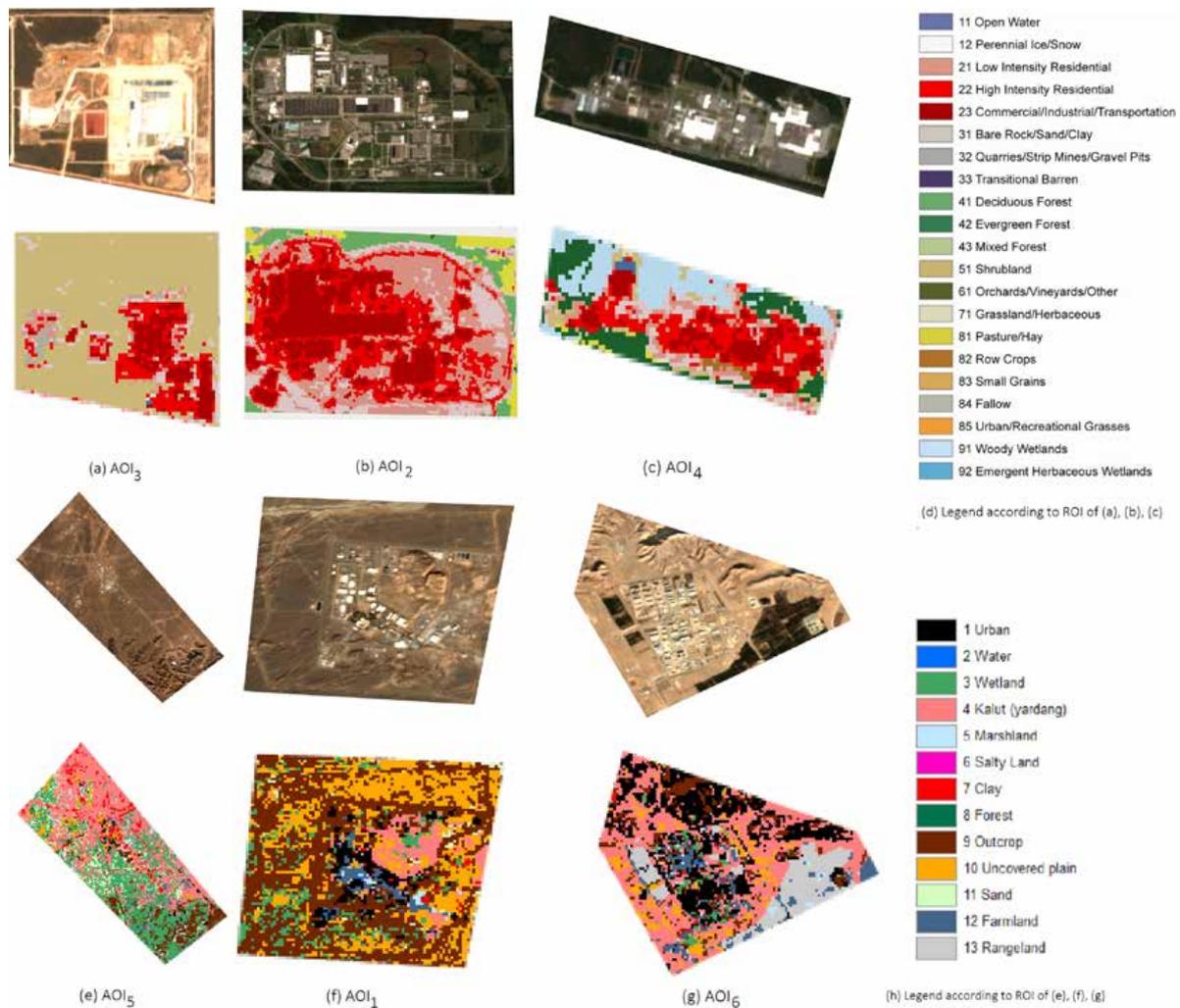
Figure 12: Class labels based on two different land cover datasets provided by GEE.

For the first area, a land cover of GEE was used, which spans eight different epochs and contains 20 different land cover classes. The 2019 release was used in this study. A single mosaic dataset from 2017 containing 13 classes was used as the second land cover. The land cover maps according to their AOI are shown in Figure 12.

Training and validation datasets were generated using a stratified random sampling approach which were then used to build, train, and classify a RF and a CART classifier. In this study, a total of 170 trees combined with a minimum leaf population of 3 and a fraction of input to bag per tree of 0.9 yielded good results for RF. In terms of the CART algorithm, the best cross validation factor was determined to be 5.

3.3 Results

Figure 13 shows the classification results obtained by both classifiers based on the manually created *FeatureCollections.*

Figure 13 (b) shows that for the year 2017 the classification by CART resulted in a misclassification of fallow land to asphalt roads and parking lots in a less extend. In addition, much of the fallow land was misclassified as vegetation.

Water was incorrectly declared as vegetation. Furthermore, road sections were classified as buildings. Also in the 2019 case, there were more misclassifications of CART compared to RF. Again, vegetation was misclassified as parking lot. Here, the wasteland has been identified as roads and parking lots. In both cases, Random Forest performed well compared to the second algorithm. Almost all buildings and vegetation are correctly classified. But also, here some road sections were misclassified. Figure 14 shows the classification maps of all six AOIs where the target classes were assigned to the predefined land cover datasets.

Since pasture/hay areas had very few pixels and thus insufficient for efficient training, this class was misclassified as a developed class. Furthermore, several issues were encountered with the classification of woody wetlands and shrub/shrub classes, which were classified as forest in the first case and developed area in the second case.

Comparing the classification results with the underlying land cover map, which is shown in Figure 12, the RF algorithm again provides significantly better results than CART. The effectiveness of the different classifiers was evaluated based on accuracy. The most used metrics for evaluating
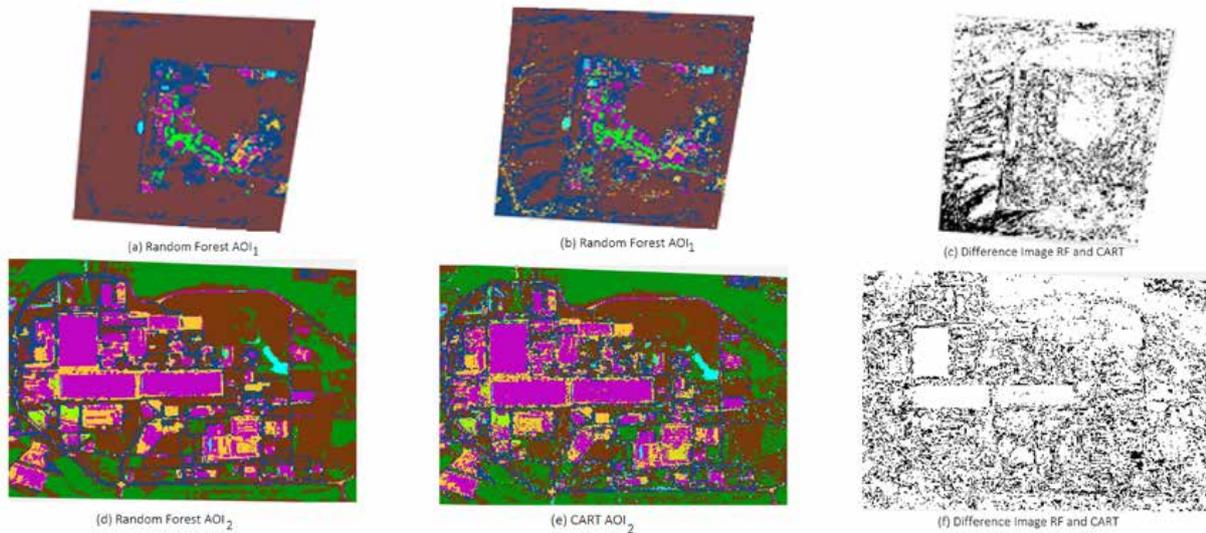
Figure 13: Classification maps using RF and CART classifiers for the years 2017 of AOI$_1$ and 2019 of AOI$_2$.

the accuracy and effectiveness of each classifier is the overall accuracy (OA) representing the percentage of correctly classified instances out of all instances and the Kappa coefficient used to test reliability.

The RF classifier outperformed the CART classifier with an average overall accuracy of 91.40 % in contrast to 74.57 %. The average kappa coefficients for RF, and CART classifiers were 0.85 and 0.64, respectively.

In addition, there are a few issues that should be investigated if the proposed method is applied not only for verification purposes. First, the spatial resolution of the Sentinel-2 satellite is limited to 10m, which results in mixed pixels containing different surface classes. This has an impact on the FeatureCollection creation and the classification. If these images had been used for feature selection, the low resolution would have caused problems regarding its concrete location. Therefore, a basemap of high-resolution reference imagery available directly within GEE was used. The downside is that this is a mosaic of images, with no information available on the date of acquisition. On the other hand, the classification performance is not accurate enough for these pixels. So, existing land cover maps were used for feature collection. However, since these maps capture only the land cover situation for one specific year the application for verification purposes is limited.

4. Conclusion

Today satellite imagery is an integral part of the IAEA's monitoring and verification efforts. The images can be used in a variety of ways to confirm that a country's nuclear facilities are in compliance with the specifications in internationally signed treaties and declarations made by the member states. Remote sensing data is for example well suited for planning on-site inspections and recognizing as well as monitoring features of interest within nuclear facilities in order to detect suspicious activities at an early stage. Thanks to the ongoing development of new satellite systems and the amount of data which will further increase in the coming years, even more applications are conceivable. Many providers offer their satellite data at low cost or even for free. For example, initiatives such as the Copernicus program, the European Union's Earth observation program, have revolutionized the market. Remote sensing has thus arrived in the Big Data era posing challenges regarding data management, processing, and analysis. The ever growing quantity of data and its properties require the further automation of processing and the development of quantitative techniques that have the potential to improve upon traditional techniques in terms of computational cost, reliability and objectivity. Several novel technologies have been developed to meet these challenges . In this research, three tools namely Apache Airflow, Rasdaman and google Earth Engine have been utilized to develop a

	AOI1		AOI2		AOI3		AOI4		AOI5		AOI6	
	CART	RF	CART	RF	CART	RF	CART	RF	CART	RF	CART	RF
Overall Accuracy (%)	83.07	91.02	76.63	98.34	68.95	86.65	72.05	89.78	73.67	92.61	73.05	89.98
Kappa Coefficient	0.65	0.76	0.70	0.92	0.63	0.81	0.59	0.85	0.60	0.89	0.64	0.86
Correct Area (%)	78.35	84.76	58.12	81.39	61.86	73.95	53.45	64.77	55.13	72.24	57.16	75.12

Table 1: Overall accuracy and Kappa statistic of CART and RF classifiers based on land cover tagged maps.
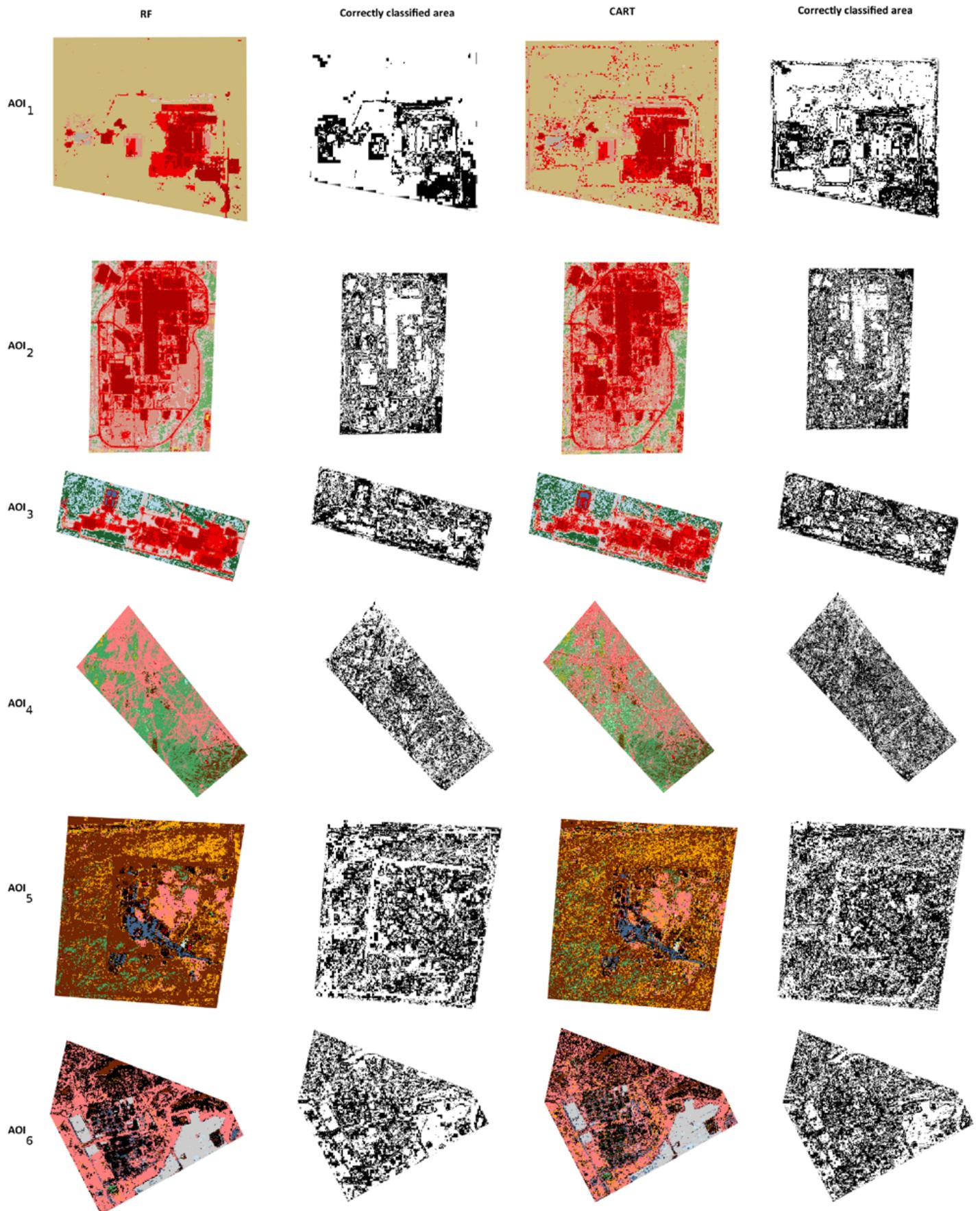
Figure 14: Classification maps using RF and CART classifiers based on two different predefined land cover classes provided by GEE. The corresponding land cover map legend is shown in Figure 13.

semi-automated procedure for collecting, storing, processing and analyzing satellite images. The potential of the developed framework is tested using case studies concerning nuclear fuel cycle related sites. Hereby the objective is to classify land cover use, as these features provide essential information for recognizing and monitoring for example changes of the operational status, constructions of new buildings and roads, plant expansions, etc.

It has been shown that Apache Airflow enables the optimization of data processing and workflow management processes and can be easily adapted in Python. The well-established array database management system Rasdaman offers the benefits of using OGC standards and storing as well as functionality scalability, among others. Finally, the potential of GEE in terms of data analysis capabilities was evaluated. The platform offers powerful capabilities in handling large volumes of remote sensing imagery and provides several machine learning algorithms. Here, the use of Google Earth Engine was demonstrated using two supervised learning algorithms, namely Random Forest and Classification and Regression Trees, provided along with GEE to classify Sentinel-2 images. To evaluate the classifiers performance, accuracy assessment was carried out. It has been shown that RF is the most suitable classifier for any given scenario reaching an overall accuracy of 91.40%. In a next step, we will evaluate the application of unsupervised learning methods that do not require training data with respect to the classification of Sentinel-2 images.

5. Acknowledgements

6. References

[1] Quevenco R; *Completing the picture: using satellite imagery to enhance IAEA safeguards capabilities;* IAEA Bulletin 2016 57-2; 2016; pp. 24-25–405.

[2] Apache Airflow Documentation – Version 2.2.4; URL: https://airflow.apache.org/docs/apache-airflow/stable/index.html.

[3] Baumann P, Dehmel A, Furtado P, Ritsch R, Widmann N; *The Multidimensional Database System RasDaMan*; Sigmod Record 27; 1998; pp. 575-577; DOI: 10.1145/276304.276386.

[4] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R; *Google Earth Engine: Planetary-scale geospatial analysis for everyone;* Remote Sensing of Environment; 2017.

[5] Glaser A, Niemeyer I; *Nuclear Monitoring and Verification Without Onsite Access.* In: Toward Nuclear Disarmament: Building up Transparency and Verification; Ed. by M. Göttsche and A. Glaser; German Federal Foreign Office, 2021, pp. 86–115; DOI: 10.18154/RWTH- 2021- 04949; URL: https://publications.rwth-aachen.de/record/819269.

[6] Jo S.Y, Shin J.S; *Status and Application of Spatial Information Technology as a Nuclear Verification Tool;* Transactions of the Korean Nuclear Society Spring Meeting Pyeongchang, Korea; 2010.

[7] Zhang H; *Strengthening IAEA safeguards using high-resolution commercial imagery*; Symposium on International Safeguards: Verification and Nuclear Material Security, Vienna, Austria; 2001.

[8] Castriotta A.G, Volpi F; *Copernicus Sentinel Data Access Annual Report 2020*; URL: https://sentinels.copernicus.eu/web/sentinel/-/copernicus-sentinel-data-access-annual-report-2020/1.0?redirect=%2Fweb%2Fsentinel%2Fsentinel-data-access.

[9] Dehmer M, Emmert-Streib F, Chen Z, Li X, Shi Y; *Mathematical Foundations and Applications of Graph Theory*; 2017; ISBN: 9783527339099.

[10] *Copernicus*; URL: https://www.copernicus.eu/en.

[11] *Copernicus in detail*; URL: https://www.copernicus.eu/en/about-copernicus/copernicus-detail.

[12] *Copernicus Open Access Hub - OData API; URL: https://scihub.copernicus.eu/userguide/ODataAPI

[13] *Copernicus Open Access Hub – Batch Scripting;* URL: https://scihub.copernicus.eu/userguide/BatchScripting

[14] Baumann P; *A Database Array Algebra for Spatio-Temporal Data and Beyond*; Proceedings of the 4th International Workshop on Next Generation Information Technologies and Systems (NGITS), pp. 76 – 93; 1999.

[15] *Rasdaman Documentation* – Query Language Guide; URL: https://doc.rasdaman.org/04_ql-guide.html

[16] *Rasdaman Documentation* – Server Architecture; URL: https://doc.rasdaman.org/02_inst-guide.html#server-architecture.

[17] *Google Earth Engine – Case Studies*; URL: https://earthengine.google.com/case_studies/.

[18] Canty M.J, Nielsen A.A. Conradsen K, Skriver H; *Statistical Analysis of Changes in Sentinel-1 Time Series on the Google Earth Engine*; Remote Sens. 2020; pp. 12, 46; DOI: 10.3390/rs12010046.

[19] Rutkowski J, Canty M.J, Nielsen A; *Site Monitoring with Sentinel-1 Dual Polarization SAR Imagery Using Google Earth Engine*; Journal of Nuclear Materials Management; 2018; pp. 48-59.

[20] *Google Earth Engine – Data Catalogue*; URL: https://developers.google.com/earth-engine/datasets/catalog

[21] Ehlert I, Schweitzer C; *Copernicus für das Umweltmonitoring – Eine Einführung*; p. 62; 2018; URL: https://www.bsh.de/DE/PUBLIKATIONEN/_Anlagen/Downloads/Meer_und_Umwelt/Weitere_Publikationen/Copernicus-fuer-das-Umweltmonitoring.html.

[22] Rutkowski J, Niemeyer I; *Remote Sensing Data Processing and Analysis Techniques for Nuclear Non-proliferation*; In: Niemeyer I, Dreicer M, Stein G (eds); Nuclear Non-proliferation and Arms Control Verification, Springer; Mar. 2020; pp. 339–350; ISBN: 978-3-030-29536-3; DOI: 10.1007/978-3-030-29537-0_23.

[23] Breiman L, Friedman J.H, Olshen R.A, Stone, *C.J; Classification And Regression Trees;* Routledge; 1984; DOI: 10.1201/9781315139470.

[24] Breiman L; *Random Forests; Machine Learning 45*; 2001; pp. 5–32; DOI: 10.1023/A:1010933404324.

[25] Ghorbanian A, Kakooei M, Amani M, Mahdavi S, Mohammadzadeh A, Hasanlou M; *Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples. ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 276-288; 2020; DOI: 10.1016/j.isprsjprs.2020.07.013.

[26] Dewitz J and U.S. Geological Survey; *National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021): U.S. Geological Survey data release; 2021*; DOI: 10.5066/P9KZCM54.

[27] *Google Earth Engine – Datasets tagged landcover in Earth Engine*; URL: https://developers.google.com/earth-engine/datasets/tags/landcover.