Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

MUON SCATTERING TOMOGRAPHY FOR INVENTORY VERIFICATION IN ARMS CONTROL

Robert J Hughes **Chris Thompson** **Mat Budsworth**

AWE, Aldermaston, Reading, Berkshire, RG7 4PR

ABSTRACT

The verification of declarations made about an inventory of fissile material stocks will form a key aspect of any future nuclear weapon treaty. The ability to confirm, in a treaty accountable item, the presence or absence of some identifying characteristic forms the basis of a testable declaration. To be able to do so passively, and without revealing detailed information about the contents, presents an attractive verification method. We present the initial results of a viability study into Muon Scattering Tomography (MST) for use in such a manner. Employing a templating approach, our study uses MST to identify changes in geometries inside sealed containers, whilst protecting sensitive data. The study is based on *Geant4* simulations and provides some promising results.

INTRODUCTION

Any future nuclear weapon related arms control agreement will likely require an agreed set of verification measures and provisions in order to demonstrate treaty compliance. Developing an effective verification system requires a holistic approach, considering all aspects of a nuclear enterprise and utilising the most appropriate technological solutions to achieve the verification objectives. It is, therefore, imperative that a suite of verification technologies and strategies that verify different aspects of a declaration are available to those tasked with designing the verification system.

Muon Scattering Tomography (MST) is a technique that uses cosmic ray muons to passively generate volumetric reconstructions of objects inside containers. This paper assesses the feasibility, and applicability, of MST as a unique tool for inventory verification. Specifically, this report investigates the ability of MST to be used in verifying that the geometry of an item inside a container has not changed, without revealing any information about the geometry itself which might be sensitive or proliferative. The work is based upon MST simulations using the *Geant4* code, and employs a statistical test to quantitatively evaluate whether two geometries are a “match” to each other.

The paper begins by discussing cosmic ray muons and the tomographic process used in this work, followed by an outline of the “template matching” approach to inventory verification within nuclear weapon arms control. The test concept, statistical test and scenarios are then outlined before the data are presented. The paper finishes with a discussion on possible future work.

COSMIC RAY MUONS

Muons are negatively charged, fundamental particles; similar to electrons but around 200 times more massive. They are produced from interactions of cosmic rays within the upper atmosphere and reach the surface of the Earth at a rate of approximately 1 per square centimetre, per minute (Clarkson, et al., 2015). Muons are able to penetrate deeply into materials and can provide a viable method for probing the internal contents of items, even when shielded, without the need for an active source, as is required for x-ray imaging, for example. The use of active sources is generally undesirable in nuclear weapons verification due to the stringent safety criteria that apply when nuclear material or explosives are present.

Muon interactions

When traversing a material, muons undergo multiple small-angle scatters dominated by Coulomb scattering from the nuclei in the target material. The net scatter angle produced by many small-angle scatters can be represented by a Gaussian distribution, with non-Gaussian tails appearing at higher angles due to less frequent large-angle scatter events (Bichsel, Groom, & Klein, 2018). A simplified Gaussian approximation of the scattered angles can be calculated with a width dictated by the "radiation length", X_0, as shown in Eq. 1 (Lynch & Dahl, 1991).

$$\sigma = 13.6\frac{\sqrt{\frac{X}{X_0}}}{p\beta c}\left[1 + 0.088\log_{10}\left(\frac{X}{X_0}\right)\right] \qquad \textit{Eq. 1}$$

In Eq. 1, X is the path length (in g/cm^2, or cm, as long as X_0 is in the same units) of the muon through the material, X_0 is the radiation length, p is the momentum (in MeV/c), and βc is the particle velocity. The radiation length is a convenient unit of thickness that represents the mean thickness required for a high-energy charged particle to lose all but 1/e of its energy (Bichsel, Groom, & Klein, 2018). The radiation length for muons through a material can be approximated by Eq. 2 (Frazão, Velthuis, Thomay, & Steer, 2016).

$$X_0 \sim \frac{716.4A}{\rho Z(Z+1)\ln\left(\frac{287}{\sqrt{Z}}\right)} \qquad \textit{Eq. 2}$$

In Eq. 2, A is the mass number, Z is the atomic number, and ρ is the density. From these two equations, it can be seen that the distribution of scattering angles will depend upon the path length each muon travels through the material (X), as well as the density (ρ), atomic number (Z), and atomic mass (A) of the material. The dependence upon the path length through the material also implies some sensitivity to the shape of the material relative to the incident muon direction. One important implication of the Gaussian approximation to the multiple Coulomb scattering is that the width of the distribution is inversely proportional to the momentum of the incident muon. One method for estimating the momentum is based on the amount of scattering within the detector layers, with a 50% precision in the momentum possible from a basic muon detector setup (Borozdin, et al., 2002).

Tomography

In MST an object is typically placed in a detector volume with planar detectors above and below the detector volume. The trajectories of incident and outgoing muons are calculated from interactions with the detectors above and below the detector volume respectively. These tracks are then used to identify the "point of closest approach" within the detection volume where a scattering event occurred, i.e. where the incident and outgoing tracks "meet". The detection volume is discretised into voxels, which receive a score when a scattering event is deemed to have occurred within their volume. These voxel scores are then used to reconstruct a 3D image of the object in the detection volume.

Clearly, this process is not physically accurate as the muon is much more likely to undergo multiple small-angle scattering events in the material, rather than one single scattering event. An alternative method for tomographic reconstruction that has been seen to produce better tomographic reconstructions than a simple point of closest approach method is the "Angle Statistics Reconstruction" (ASR) algorithm (Stapleton, Burns, Quillin, & Steer, 2014). This method is akin to back-projection in traditional tomography, with a score assigned to each voxel within a certain distance of a muon's path. The score assigned to each voxel is related to the two-dimensional projection angles of the difference between the incoming and outgoing muon tracks. The muon momentum is also included in the ASR score. The ASR process results in a distribution of scores for every pixel. The final reconstruction value of each voxel is then based upon a chosen quantile from the distribution of the accumulated scores at each voxel. For instance, a choice of the 50% quantile would yield the median score for every pixel. The ASR algorithm is used in this work.

ARMS CONTROL VERIFICATION

The ability to verify an inventory of accountable items at given locations, and their absence from other locations, is a cornerstone of many arms control agreements. Inventory verification measurements should be capable of accurately diagnosing specific characteristics to ascertain whether a tested item is consistent, or not, with declared characteristics of accountable items. By using multiple verification measurements that test for presence of different declared characteristics of an accountable item it is possible to minimise the risk of treaty noncompliance by limiting the parameter space in which a credible spoof item could exist.

This work investigates the ability of MST to test whether the geometry of a tested item is consistent with declarations about accountable items. In particular, the method adopts a "template matching" approach where a unique signature for an accountable item is generated and compared to the signatures of subsequent items to determine whether they have the same geometry (to within a predefined statistical tolerance). The template, against which subsequent items are tested, needs to be generated from a trusted, genuine accountable item. This is often referred to as the "golden warhead" problem – how can one be sure that the template is recorded from a genuine accountable item? There are several methods that have been proposed for this, including random selection from declared deployed nuclear warheads on missiles (Yan & Glaser, 2015). The comparison of subsequent items to the original Template Item could then be used for two potential scenarios: it could check whether a single item remains the same after a given time period, or whether a different test item belongs to the same "class" as the Template Item.

Template matching for radiation signatures

Traditionally, the signature proposed for use in template matching in nuclear weapon arms control verification is the gamma spectrum emitted by the special nuclear material. The gamma spectrum may be considered sensitive and thus would not be disclosable to an Inspector undertaking the verification measurement. Rather than analysing the gamma spectrum from the tested object, the Inspector instead relies upon a comparison of the spectrum recorded from the test item and the template previously recorded from an authentic, accountable item. The only information released to the Inspector is an indication of whether the two spectra are a match (to within a pre-agreed statistical uncertainty). Previous examples of template matching methods developed for use in nuclear weapon verification include the CIVET (Controlled Intrusiveness Verification Technology) (Sastre, 1988) and the TRIS (Trusted Radiation Identification System) (Merkle, et al., 2010).

However, these methods only confirm (or otherwise) that the gamma radiation signatures are the same. They do not confirm anything about the geometry of the item. Thus, one could potentially add, subtract, reconfigure or replace material within the box of the accountable item so long as the gamma radiation signature is not modified beyond a pre-agreed tolerance. By adding a test for consistency of geometry, the tested item must match the radiation signature *and* the geometry, thus further constraining the space of potential spoof items that could incorrectly pass the verification measurement.

Template matching for geometric signatures

The most obvious way to compare geometries of items in boxes is to take radiographs of the boxes and compare the images directly. However, there are challenges with this process:

- Active radiation sources with sufficient endpoint energies required for radiography may be difficult to certify for use in proximity to nuclear weapons, components, or materials.
- It would require the template to be a radiograph image of a sensitive object, thus requiring severe security measures in handling and storage.

The first challenge does not apply to MST as it is an entirely passive measurement. The second issue can be addressed by using a comparison method that compares geometric configurations without storing any positional information, thus reducing the sensitivity of the process from a security perspective.

THE HISTOGRAM METHOD

Concept

A template generated directly from the MST 3D item reconstruction would clearly suffer from similar security issues as a radiograph image template. An alternative to comparing the full reconstructions would be to jettison the positional information (akin to "anonymisation" of the data) and compare histograms of the ASR scores attributed to the voxels. The template would then be a histogram of scores, which should be significantly less sensitive than a radiograph image or reconstruction. It should also be largely agnostic to translational (although not rotational) item movements within the detection volume. Since the ASR scores are dependent (to some extent) on the material density, atomic number and atomic mass, as well as the predominantly vertical path length through a material, it should be extremely difficult to accurately reverse-engineer the geometry from a histogram of scores.

The Histogram Method lends itself to a template-matching process, as a signature similar to a radiation spectrum is generated which can be saved and compared to histograms from later objects. An item is then said to "match" the template if the histograms are the same to within some agreed tolerance using a specified statistical test. In existing template-matching methods, such as the TRIS, an object's gamma signature is compared to the template with a goodness of fit judged by a Chi-square Test. A similar method is applied in this work, and is referred to throughout the remainder of this paper as the "Histogram Method".

Statistical test

The "Chi-square Test", which measures the spread of observations compared to the expected spread, has been chosen for comparing the histograms. For each bin in the histogram, if the observed value (i.e. from the tested item) is within the expected spread of the template value, then each bin would provide a contribution to the chi-square statistic of approximately one. The chi-square statistic would then be $\chi^2 \approx n$, where n is the number of histogram bins. In the general case, when comparing observed values to modelled values, the expectation value for chi-square is actually the number of degrees of freedom $v=n-n_c$ where n_c is the number of parameters used in describing the model. In this work, the "observed data" from the test item is being compared to previously observed data (the template), and thus no parameters are used. The expectation value for chi-square in this work is therefore just the number of bins, n.

In this work, the Chi-square Test is being used to decide whether two sets of data (the histograms) are drawn from the same parent population (i.e. the two histograms are essentially equivalent to within the expected variance). To achieve this, it is possible to test the data sets directly, independent of the parent population, using the following equation:

$$\chi^2 = \sum_{i=1}^{n} \frac{[g(x_i) - h(x_i)]^2}{\sigma^2(g) + \sigma^2(h)} \qquad \text{Eq. 3}$$

The two measured distributions in *n* bins are $g(x_i)$ and $h(x_i)$, whilst the denominator term is the variance of the difference $g(x_i)$-$h(x_i)$. The value of interest is actually the Reduced Chi-square, χ_v^2, where *v* is the number of degrees of freedom, equivalent to *n* in this case.

$$\chi_v^2 = \frac{\chi^2}{n} \qquad \text{Eq. 4}$$

If the value of Chi-square is much larger than one, there is strong evidence against the two data sets being drawn from the same population (i.e. the two histograms represent significantly different objects). If the value of Chi-square is roughly one, it cannot be claimed that the two data sets are drawn from different parent distributions (i.e. there is no evidence to suggest these two distributions are from significantly different objects). It is important to note, however, that a Reduced Chi-square value close to unity does not guarantee that the two data sets are from the same parent distribution. There always exists the possibility of two different, but closely similar, parent distributions that the two data sets are not sensitive enough to distinguish (Bevington & Robinson, 2003).

SCENARIO

As an initial proof-of-concept, six scenarios have been compared against a single Template Item using the Histogram Method. The geometries are simple, but the materials are representative of potential use case scenarios.

Item	Material	Shape	Density (g/cm^3)	Mass (kg)
Matching Item	10% ^{238}U, 90% ^{235}U	Sphere, r = 6.838cm	18.67	25
Empty Box				0
Geometry Change: Larger sphere	Lead	Sphere, r = 8.071cm	11.35	25
Geometry Change: Pucks	10% ^{238}U, 90% ^{235}U	5 piles of 5 close-packed 1kg cylinders	18.67	25
Material Change: Tungsten	Tungsten	Sphere, r = 6.762cm	19.3	25
Material Change: Natural Uranium	99.7% ^{238}U, 0.3% ^{235}U	Sphere, r = 6.810cm	18.9	25

Table 1: The six items tested against the template

The template comes from a box containing a sphere of uranium (enriched to 90% ^{235}U) weighing 25kg. The six items are given in Table 1 and are intended to broadly test scenarios where a monitored State claims that a box contains the matching item when, in fact, it contains a geometrically different item or an item of a different material in a similar geometry. The items are constrained to fit in the same box and, with the exception of the Empty Box scenario, the items are constrained to match the same mass (25kg). The scenarios can be thought of in three distinct categories:

- Presence Test
    - Matching Item – This is the exact same item, in the same box
    - Empty Box – This is the same box, but filled only with air
- Geometry Change (mass is the only common constraint)
    - Larger Sphere – Constructed from Lead, sphere volume 60% larger than Template
    - Pucks – Five towers of five pucks, same material and mass as Template
- Material Change (mass and shape are constrained, volume within 4% of the template)
    - Tungsten – Constructed from Tungsten, ~4% different to Template volume
    - Natural Uranium – Unenriched material, volume within ~1% of Template

MODELLING

The model used to simulate the generation of muon tracks was written in *Geant4*, based upon the detector system at AWE which comprises an upper and lower detector plane, each made of six alternating layers of drift tubes arranged in a grid pattern. These drift tubes detect a muon hit, and then provide either the *x* or *y* position, depending on the orientation of the tube. For the modelling, the detector resolution was set at 0.5mm. When these coordinates are combined over the layers, an estimate of the muon path can be recreated. A table surface sits between the detector planes, upon which the observed object is placed, and this is included in the model. Each of the objects described in Table 1 were simulated as being inside a cubic aluminium box (density 2.7g/cm^3) of external length of 300mm and with a wall thickness of 25mm. The surrounding air in the detection volume was simulated as dry air. The reconstructed volume was a 60cm sided cube, split into 1cm cubic voxels, with the bottom of the aluminium box positioned a few centimetres above the bottom of the reconstruction.

The *Geant4* model recreates the drift tubes and stores the position of any muon "hits" in the tubes as the model output. Upon completion of a simulation, a post-processing script is used to convert the muon hit data into track data. These track data are then used as the input for the ASR reconstruction algorithm. The modelling captured two exposure lengths for each item: 4 hours and 18 hours.

DATA

Each item from Table 1 has been compared to the Template Item using the Histogram Method. The histogram "signatures" of the items from simulations of 18-hour exposures are plotted in Figure 1, Figure 2, and Figure 3. There is a clear and obvious difference between the Empty Box and the Template Item in Figure 1, and the Matching Item is a qualitatively close match to the Template Item. This is an important, although seemingly trivial, result: the Histogram Method can determine a match.

The "Geometry Change" scenarios, shown in Figure 2, are qualitatively obviously different. It is perhaps not surprising that the Larger Sphere, composed of less-dense lead, produces a significantly different histogram to that of the Golden Template, as all the variables that affect muon scattering are different: atomic number, atomic mass, density, and volume (i.e. path length through the material). However, the visible difference between the Pucks and the Template is perhaps more intriguing. These items have the same atomic number, atomic mass, and density, with only the geometry being different and yet produce qualitatively significantly different histograms. Both of these items would pass a mass test, and the Pucks might even be able to pass a gamma radiation signature test, but the Histogram Method appears to clearly observe a difference.

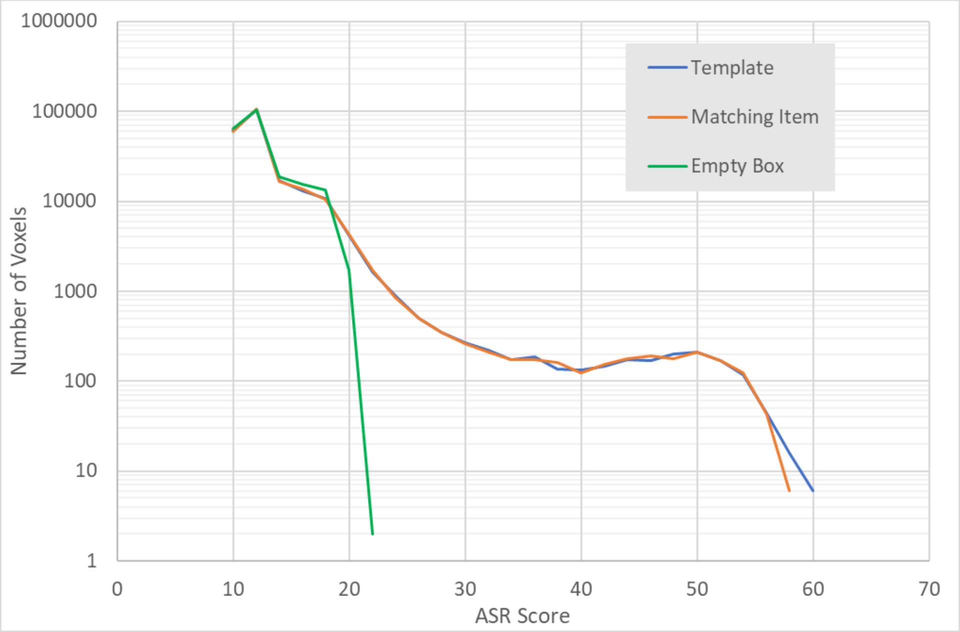

Figure 1: 18-hour exposures of the "Presence Test" scenarios that identify whether a Matching Item can be recognised

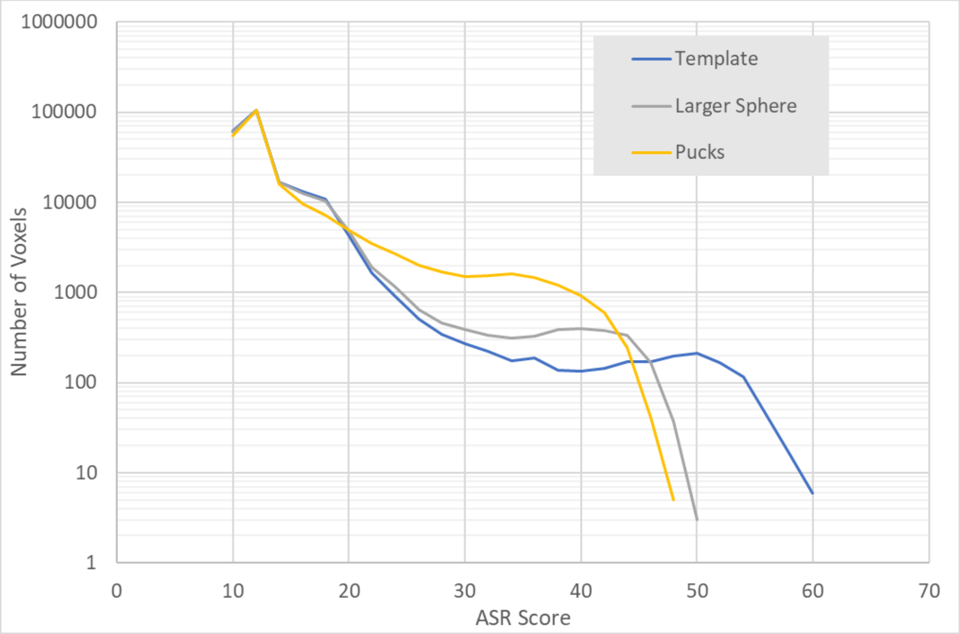

Figure 2: 18-hour exposures of the "Geometry Change" scenarios

The "Material Change" scenarios, however, do not show such clear differences to the Template Item in their histograms. This should not be a surprise, as the differences in the variables that affect muon scattering are very small. In both cases, the volume is less than 4% different to the Template Item and therefore the path length through the material will be very similar. In the case of the Tungsten item, the atomic number and atomic mass are different, but with the Natural Uranium item even the atomic number is identical. From Figure 3 it appears unlikely that the Histogram Method will be able to reliably differentiate between the Natural Uranium and the Template Item, but there appears to be a qualitative small difference to the Tungsten item.

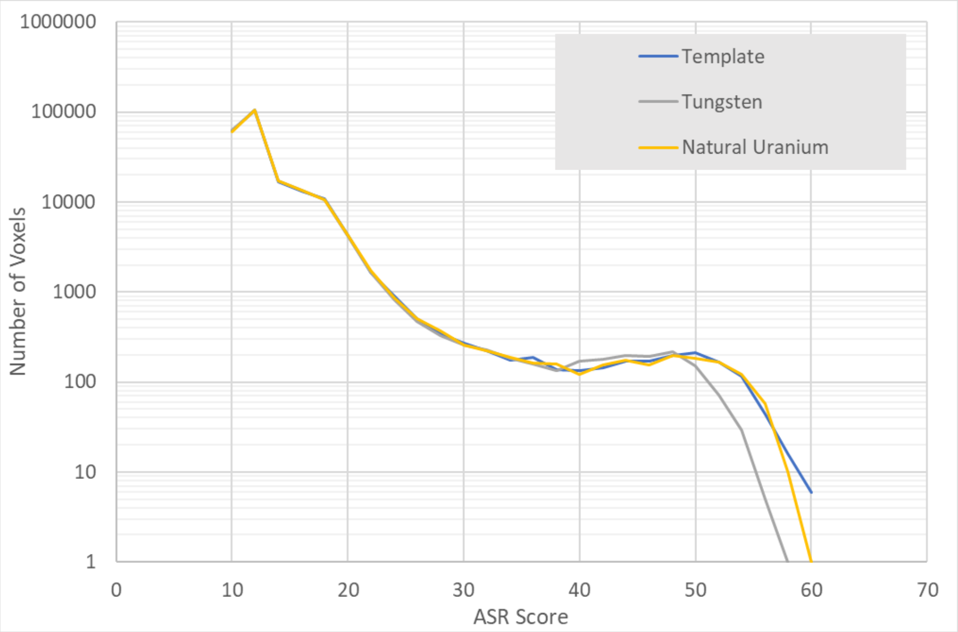

Figure 3: 18-hour exposures of the "Material Change" scenarios

Discriminator Value

Whilst the histograms can be seen to be qualitatively different in some cases, it is important to use a quantitative method of discriminating between the items that are a match for the Template Item and those which are not. As mentioned previously, the Reduced Chi-square statistic is employed in this work to determine quantitative matches. The resulting Reduced Chi-square statistics for a comparison of each item against the Template Item are shown in Table 2, along with the equivalent p-value (the probability that two matching items would produce a comparison at least as extreme as that observed). If the null hypothesis of the Reduced Chi-square test is that the items are a match, and the significance level is set to α=0.01, then a p-value less than 0.01 would indicate a less than 1 in 100 chance that the observed comparison is from items that are a match and the null hypothesis could be rejected. Given a significance level of α=0.01, the Matching Item and the Natural Uranium are unable to reject the null hypothesis and so are considered to be "matches" to the Template Item. All of the other tested items in an 18-hour exposure produce significantly lower p-values and so are deemed to be "different" to the Template Item.

Item	**Reduced Chi-square statistic**	**p-value**	**Decision**
Matching Item	1.499	0.049	Match
Empty Box	284.33	0	Different
Geometry Change: Larger Sphere	47.91	2.4E-246	Different
Geometry Change: Pucks	452.02	0	Different
Material Change: Tungsten	7.40	2.5E-27	Different
Material Change: Natural Uranium	1.737	0.011	Match

Table 2: Quantitative comparisons of the histograms for the 18-hour exposures

If the exposure time is reduced to 4-hours, the Reduced Chi-square statistic values for the comparisons are lower, as might be expected given that there will be more noise in the histograms. To investigate this effect, the 4-hour exposures were compared 80 times against the Template Item (the Matching Item was only compared 35 times), and the subsequent Reduced Chi-square statistic recorded. The results are shown as histograms of the Reduced Chi-square score for the Geometry Change scenarios in Figure 4. It is clearly possible to pick a discriminator value that correctly passes the Matching Items and fails all of the other "Geometry Change" items (for instance, a value of $\chi_v^2=5$). Interestingly, the Pucks perform even worse than a completely empty box, producing the worst mean Reduced Chi-square value.

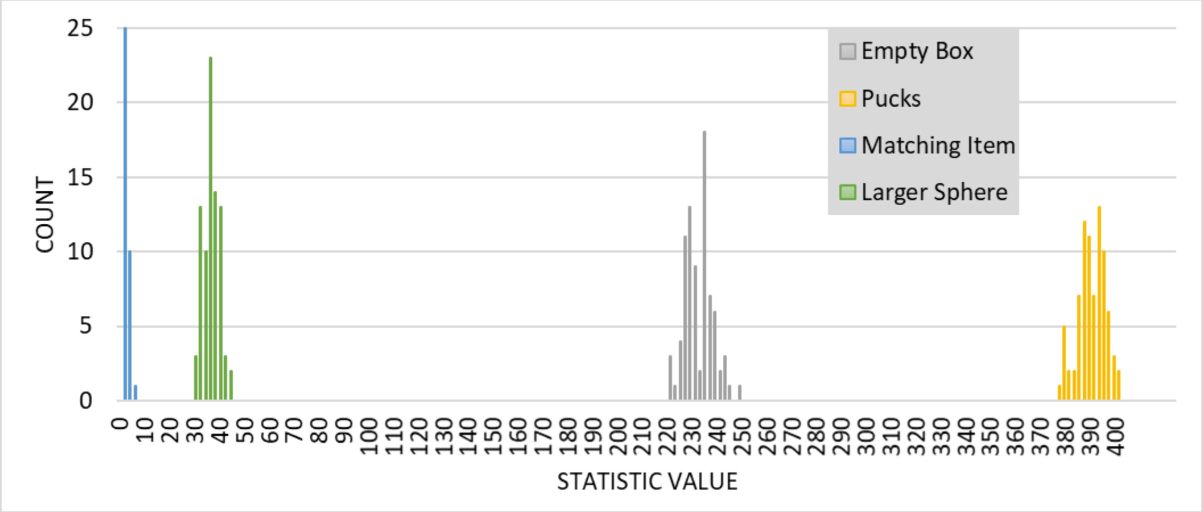

Figure 4: Distribution of Reduced Chi-square statistics for 4-hour exposures of the "Geometry Change" scenarios, plus the Empty Box

DISCUSSION

The Histogram Method has been demonstrated, from modelling, as being able to effectively discriminate between a Template Item and an item with a significant change to the geometry (the "Geometry Change" scenarios). This has been demonstrated for the same mass of a different material with a ~60% larger volume, and for the same mass of the same material in a different geometric configuration. The Histogram Method has been shown to work robustly against the "Geometry Change" scenarios even at only a 4-hour exposure, with the ability to define a discriminator value for the Reduced Chi-square statistic that would pass all of the Matching Items and reject all of the other "Geometry Change" scenarios. This could provide a potential method of differentiating between different classes or types of accountable item, particularly when the protection of sensitive information is required.

For the "Material Change" scenarios, when the geometry remains very similar (in these cases, the spherical volume changes by less than 4%) but the material is changed, the Histogram Method's ability to discriminate appears less capable. For the Tungsten item, where there is a slightly larger difference in atomic number, atomic mass, and density, there appears to be greater capability to discriminate it from the Template Item, especially for the longer, 18-hour, exposures. It seems unlikely however, that the method could ever robustly discriminate against the Natural Uranium item. Nonetheless, the templating approach does impose strict geometric constraints, making it potentially more difficult to masquerade an item as that of another type.

Use in Arms Control

The work shown in this paper indicates that an exposure of four hours would be sufficient to discriminate against significantly different geometries with essentially zero false-positives (items passed as “matching” when they are not) and to pass all the Matching Items correctly. Arms Control verification is one of the few contexts in which an exposure time of several hours is not prohibitively long, unlike cargo screening for example.

A key benefit of the Histogram Method is that the template itself retains none of the geometric information used in the tomographic reconstruction. Whilst it may be possible to make some broad inferences regarding the presence of highly-scattering material from the histogram, the comparison against the Pucks shows that even for the same volume of the same material, a different geometric configuration produces a significantly different histogram. Since there is essentially no geometric information in the template it is unlikely to be considered highly sensitive or proliferative information and thus could be handled and stored with far fewer restrictions than the potentially more sensitive radiation-based templates or radiographs.

FUTURE CONSIDERATIONS

The work presented in this paper demonstrates the potential viability of muon scattering tomography, and particularly the Histogram Method, for use in verifying geometries in arms control monitoring scenarios. However, several areas require further investigation. The choice of bin-width for the histogram of ASR scores may have an effect upon the comparison, which may well be geometry dependent. For instance, a larger bin-width may produce better comparisons for simple objects but may miss geometric features for more complex items. The method also needs to be tried against more complex, realistic items, and eventually be applied to actual physical measurements, rather than relying purely on modelling.

REFERENCES

Bevington, P., & Robinson, D. (2003). *Data Reduction and Error Analysis for the Physical Sciences, 3rd ed.* New York: McGraw-Hill.

Bichsel, H., Groom, D., & Klein, S. (2018, August). Passage of Particles Through Matter, in "Review of Particle Physics" by M Tanabashi et al. (Particle Data Group). *Physical Review D, 98*(3), 13.

Borozdin, K., Hogan, G., Morris, C., Priedhorsky, W., Saunders, A., Schultz, L., & Teasdale, M. (2002). *Radiographic Imaging with Cosmic Ray Muons.* Los Alamos National Laboratory, United States Department of Energy. New Mexico: Los Alamos National Laboratory.

Clarkson, A., Hamilton, D., Hoek, M., Ireland, D., Johnstone, J., Kaiser, R., . . . Zimmerman, C. (2015, March). Characterising Encapsulated Nuclear Waste Using Cosmic-Ray Muon Tomography. *Journal of Instrumentation, 10*(3), 1.

Frazão, L., Velthuis, J., Thomay, C., & Steer, C. (2016, July). Discrimination of High-Z Materials in Concrete-Filled Containers using Muon Scattering Tomography. *Journal of Instrumentation, 11*(7), 2.

Lynch, G., & Dahl, O. (1991). Approximations to Multiple Coulomb Scattering. *Nuclear Instruments and Methods in Physics Research, B58*, 10.

Merkle, P., Weber, T., Strother, J., Etzkin, J., Flynn, A., Bartberger, J., . . . Anderson, L. (2010). Next Generation Trusted Radiation Identification System. *Institute of Nuclear Materials Management Meeting July 2010.* INMM. Retrieved from https://media.nti.org/pdfs/SNL-1_FINAL_INMM_2010_Next_Generation_Trusted_Radiation_System.pdf

Sastre, C. (1988). *CIVET: A Controlled Intrusiveness Verification Technology.* Brookhaven National Laboratory, United States Department of Energy. Upton, New York: Brookhaven National Laboratory.

Stapleton, M., Burns, J., Quillin, S., & Steer, C. (2014, November). Angle Statistics Reconstruction: A Robust Reconstruction Algorithm for Muon Scattering Tomography. *Journal of Instrumentation, 9*(11), 1-15.

Yan, J., & Glaser, A. (2015). Nuclear Warhead Verification: A Review of Attribute and Template Systems. *Science & Global Security, 23*(3), 159.