Cosmic-ray muons for imaging: Monte Carlo simulation studies into assaying nuclear waste packages, deep geological repositories, and carbon reservoirs

Michael J. Weekes

A thesis submitted for the degree of Doctor of Philosophy

March 2023

This place is not a place of honour... no highly esteemed deed is commemorated here...
nothing valued is here. The danger is still present, in your time, as it was in ours.

-Proposed nuclear waste repository warning for future generations

Abstract

Cosmic-ray muons are a powerful, highly versatile and increasingly widely used tool for imaging and assay. Three studies into the use of these muons for applications in the energy sector are here presented. Each study has been conducted with Monte Carlo computational simulations and explores many aspects of the problem. The first is assay of stored nuclear waste; a method based on machine learning to identify materials inside waste drums is developed and applied. The second study focuses on verifying the design information of long-term nuclear waste repositories. Tests are made into the detectability of unknown voids inside a waste repository using muon radiography, and an imaging algorithm is implemented. Finally, the problem of continuous monitoring of an undersea storage site for carbon dioxide is considered, making use of data from a real, active storage site to test muon radiography's efficacy to detect and monitor movement of the stored carbon dioxide.

Contents

Chapter 1

Introduction

Cosmic rays from extreme astrophysical events collide with the Earth constantly. Their interactions with the atmosphere produce muons, which reach the surface with a flux of approximately 1 cm^{-2}min^{-1}; the flux is strongest towards the zenith and reduces towards the horizon. These naturally occurring muons are of great value for science. They are highly penetrating, non-hazardous, easily detectable particles with well-understood interactions in matter. Taking measurements of their attenuation when passing through some structure gives information about the structure's density, much like an X-ray photograph. Alternatively, measurements of muon scatterings through some sealed container, for example, gives information about the materials contained, without having to open or damage it.

The field of applications of cosmic-ray muons is very diverse. Some of the first studies exploiting the muon flux for imaging were in archaeology, searching ancient pyramids for hidden voids using spark chambers. Over the intervening decades, particle detector technology and the power of computational modelling have advanced hugely, making muon studies of a vast range of targets and scenarios possible. Today, muon science is also applied to volcanology, civil engineering, mining, geology, nuclear waste assay, biology and the energy sector, among others. The potential of the technique remains enormous. The main limitation is the relatively low magnitude of the muon flux, which corresponds to long exposure times often being required; the effect is compounded when taking measurements underground. However, almost any imaging or assay challenge for which long exposure times are acceptable will have some amenability to muon radiography or tomography.

In this work, three novel applications of the cosmic ray muon flux are explored via computational simulation studies. Firstly, chapter 2 will describe the origin of the cosmic ray muon flux, the interactions of the muons through matter, and describe some of the detector technologies and computational techniques commonly used in the field. Chapter 3 then elucidates the two key relevant techniques: muon radiography and muon scattering tomography, and explores key examples of their uses.

Chapter 4 is focused on a study into the use of muon scattering tomography for non-destructive assay of nuclear waste drums. Specifically, a method is developed that incorporates machine learning techniques acting on detector hit data in order to first identify and isolate

objects stored inside the drum, then determine their most likely material composition. The chapter describes previous work utilising machine learning for nuclear waste assay with muons, then explains the simulation software used and the algorithms involved in the method, shows key results and conclusions, and discusses possible future directions for this work.

Chapter 5 changes focus to a much larger volume of interest: a nuclear waste storage repository, located hundreds of metres underground. The appropriate technique for this scenario is muon radiography. The goal of the study was to assess muon radiography's potential as a technique for verifying the known design of such a repository, focused specifically on the detection of potentially hazardous voids in the surrounding rock. The chapter describes the development of an appropriate simulation framework and its application to void detection in a repository model, then extends this work to test a full 3D imaging algorithm using muon data to directly image features in the repository.

Finally, chapter 6 departs from nuclear waste applications to focus on another aspect of the energy sector: carbon capture and storage. The study described in this chapter makes use of a simulated model obtained from seismic data of a real, functioning storage site in the North Sea. The model is sampled at several different timesteps, each corresponding to a different volume of stored carbon dioxide in the site. The study then tests an application of muon radiography in which the site is continuously monitored over many years, determining the statistical significance of the detection with different detector configurations, and predicting the necessary exposure times to monitor the movement of the CO_2.

A list of publications related to the work presented in this thesis is shown in Appendix A.

Chapter 2

Cosmic-ray muons

2.1 Cosmic rays

Cosmic rays are highly energetic particles that originate primarily from extreme astrophysical events such as supernovae and Active Galactic Nuclei (AGN) (there is also a solar component of lower-energy cosmic rays). They were discovered almost concurrently in 1911-12 by Victor Hess [1], who demonstrated that the rays originated from the atmosphere and not the surface through electroscope measurements in a balloon, and Domenico Pacini [2], who performed similar measurements at shallow depths underwater. Cosmic rays are composed mostly of protons and helium nuclei (combined $\sim 99\%$) with the remaining fraction being electrons and heavier nuclei. Upon encountering the Earth, they interact with atoms in the upper atmosphere at an altitude of around 15 km; the interactions result in a shower of various particles collectively known as 'secondary' cosmic rays (see figure 2.1). The decays of charged pions and kaons within the showers result in the generation of muons ($\mu^{\pm}$), via the decays:

$$\pi^- \to \mu^- + \nu_\mu \text{ (branching ratio 99.99\%)} \tag{2.1}$$

$$\mathrm{K}^- \to \mu^- + \nu_\mu \text{ (branching ratio 63.50\%)} \tag{2.2}$$

and their antiparticle equivalents.

2.2 The muon flux

Muons are elementary leptonic fermions, very similar to electrons but with a larger mass ($m_\mu = 105.66$ MeV $\sim 207\ m_e$). They are unstable, decaying with a relatively long half-life of 1.56 μs into an electron or positron via

$$\mu^- \to e^- + \nu_\mu + \bar{\nu}_e \tag{2.3}$$

and

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu \tag{2.4}$$
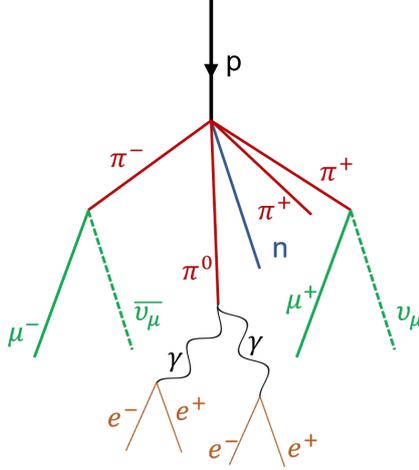
Figure 2.1: Schematic of a shower originating from a cosmic ray proton. The decay produces a neutron and several pions. The decay of the charged pions into muons and muon neutrinos is one of the primary sources of the muon flux.

where $\nu_\mu/\bar{\nu}_\mu$ are the muon's associated neutrino and antineutrino. The relatively long muon lifetime is a consequence of them not interacting via the strong force.

Muons were originally discovered in 1936 by Neddermeyer and Anderson [3]. Working at Caltech with a cloud chamber equipped with a strong magnetic field, they observed singly-charged particles traversing from above to below that had masses in between those of the electron and proton. These were in fact the cosmic ray-produced muons; initially known as the 'mesotron' or 'μ-meson', they were at first grouped with the hadronic mesons before their leptonic character was deduced.

The muon flux is the most significant remaining product of the initial cosmic rays at sea level. This is due to the exceptional penetrating power of muons; their high masses and relatively long lifetime means they pass easily through the atmosphere and a portion of the flux is detectable even under kilometres of rock. For many applications (e.g. measurements underground), the lower-energy component of the muon spectrum can be disregarded as these muons would be strongly attenuated by the overburden. In this scenario, the muon flux can be approximated using the Gaisser formula [4]:

$$\frac{dI_\mu}{dE} \approx \frac{0.14 E_\mu^{-2.7}}{\text{cm}^2 \text{ s sr GeV}} \left[\frac{1}{1 + \frac{1.11 E_\mu \cos\theta}{115 \ GeV}} + \frac{0.054}{1 + \frac{1.11 E_\mu \cos\theta}{850 \ GeV}} \right] \tag{2.5}$$

Where I_μ is the muon differential flux, E is the muon energy. The factors of 115 GeV and 850 GeV parameterize the contributions respectively of pions and kaons. Generally, this formulation is appropriate when considering muons with (zenith-weighted) energy above $100/\cos\theta$ GeV and zenith angles $\lesssim 70\,\text{deg}$, due to the assumptions of, respectively, negligible muon decay and a flat Earth. An alternative model of the sea-level flux is given by [5], which makes use of a parameterisation of $\cos\theta$ from [6] that accounts for the curvature of the Earth (see figure 2.2):

$$\cos\theta^* = \sqrt{\frac{\cos^2(\theta) + P_1^2 + P_2(\cos\theta)^{P_3} + P_4(\cos\theta)^{P_5}}{1 + P_1^2 + P_2 + P_4}} \qquad (2.6)$$

where the factors P_{1-5} are $P_1 = 0.102573$, $P_2 = -0.068287$, $P_3 = 0.958633$, $P_4 = 0.0407253$, $P_5 = 0.817285$.
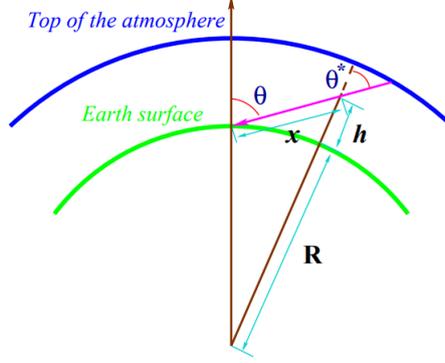
Figure 2.2: From [6], schematic demonstrating the angle θ^*, which is a modified form of the zenith angle θ that accounts for the Earth's curvature. Here, R is the Earth's radius, h is the height of the muon generation, and x is the distance between the muon generation and the detector.

The final flux formula, which also accounts for lower energy muons through an additional term, is then given by

$$\frac{dI_\mu}{dE} \approx 0.14 \left[\frac{E_\mu}{\text{GeV}} \left(1 + \frac{3.64 \text{ GeV}}{E_\mu(\cos\theta^*)^{1.29}}\right) \right]^{-2.7} \left[\frac{1}{1 + \frac{1.11 E_\mu \cos\theta^*}{115 \text{ } GeV}} + \frac{0.054}{1 + \frac{1.11 E_\mu \cos\theta^*}{850 \text{ } GeV}} \right] \qquad (2.7)$$

A comparison of these two muon flux models (evaluated at $\theta = 0$) against experimental data is shown in figure 2.3.

2.3 Muon interactions in matter

Muons passing through matter are subject to several energy loss processes. They can be grouped into two categories: electronic losses, comprising energy lost to the ionisation and excitation of electrons; and radiative losses, including bremsstrahlung, pair production, and photonuclear interactions. Each of these four key processes is explored in more detail below. The overall mean energy loss can be described by the expression

$$\left\langle -\frac{dE}{dx} \right\rangle = a(E) + b(E)E \qquad (2.8)$$

where the functions a and b correspond to the electronic and radiative energy losses respectively [8]. These functions are heavily dependent on the material in question, with its atomic number Z and atomic mass A (or their abundance-weighted averages for non-elemental materials) being of particular importance. The muon critical energy $E_{\mu c}$ is then defined as the energy at

9
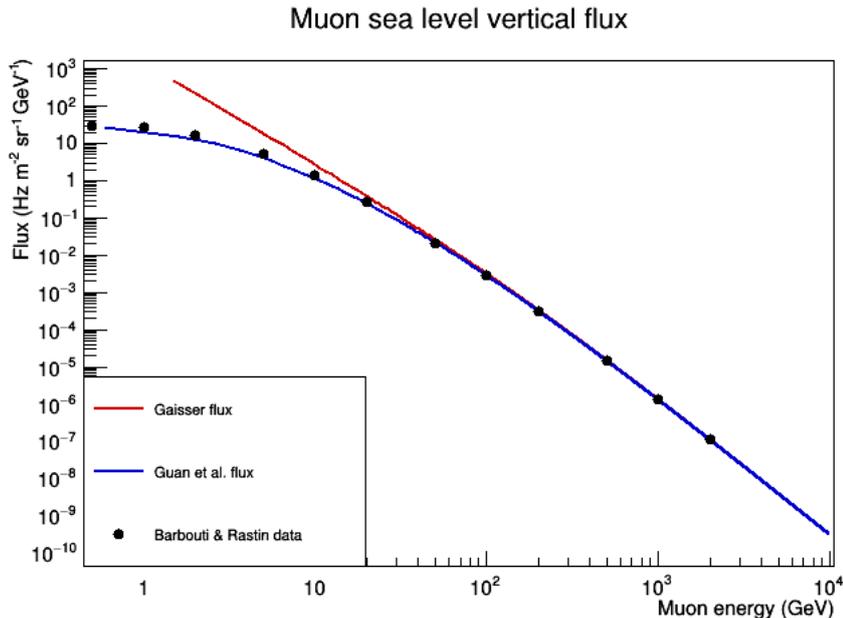
Figure 2.3: Comparison of the [4] (red) and [5] (blue) sea level vertical muon flux models compared to experimental data from [7] (black).

which losses from electronic and radiative losses are equal, i.e. $E_{\mu c} = a/b$.

The E proportionality of the radiative losses $b(E)E$ corresponds to these processes being a much smaller component of the energy loss than the electronic component $a(E)$ for non-high-energy muons. For muon energies up to ~ 100 GeV, $b(E)E < 0.01a(E)$. This is illustrated clearly in figure 2.4, which shows the mean stopping power of each of the four processes as a function of the muon energy. The three radiative process stopping powers each increase approximately linearly with E; this is due to the relevant cross sections σ approximately scaling with the energy transferred, η, via $\eta\sigma(E,\eta) = \nu\sigma(E,\nu) \approx f(\nu)$ [9]. Here ν is the fraction of the muon energy transferred.

The mean energy loss $\left\langle \frac{dE}{dx} \right\rangle$ cannot be used to precisely predict the range of any particular muon of known energy as the energy loss processes are stochastic. Hence the range of a beam of monochromatic muons will vary, an effect known as 'range straggling'. These fluctuations are particularly significant for the radiative bremsstrahlung and photonuclear energy loss processes [10] and hence are of greater importance for higher energy muons. For a deep measurement at ~ 3 km, range straggling leads to a factor of ~ 2 increase in the muon flux than would be predicted from simply inverting equation 2.8.

2.3.1 Electronic losses

A muon moving through matter will lose energy to the material's electrons, exciting them from lower to higher states. If the transfer is sufficiently energetic, the electron's binding energy is overcome and an ionisation occurs; if the resulting free electron has sufficient energy to ionise further atoms it is labelled a δ-electron. For muon energies below $\lesssim 100$ GeV, the

10
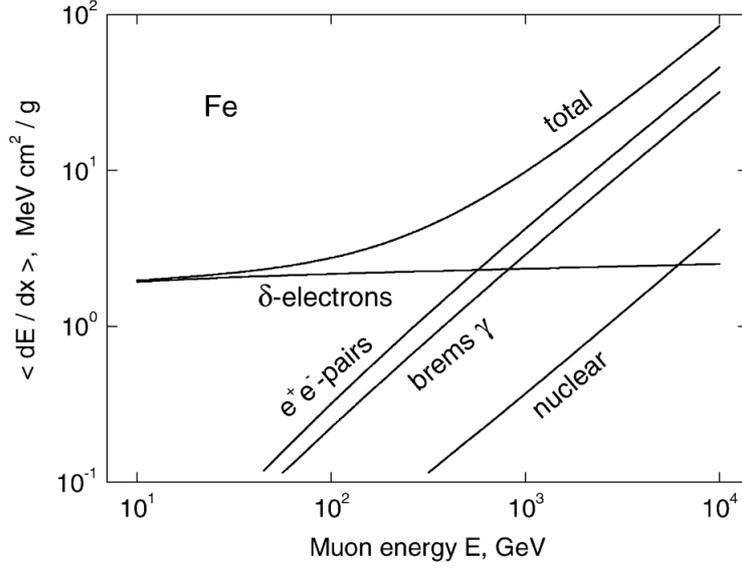
Figure 2.4: Mean stopping power of the four key muon energy loss processes in matter vs muon energy E. The ionisation component (labelled δ-electrons) dominates the three radiative processes (pair production, bremsstrahlung, and photonuclear) up to $E \sim 100$ GeV. From [9], © 2006 IEEE.

electronic energy loss is the dominant process.

The electronic component a of the energy loss is given by the Bethe-Bloch formula, first calculated in [11] by considering the contributions of all possible inelastic scatterings. In the form used in [8], it is

$$a(E) = \left\langle -\frac{dE}{dx} \right\rangle_{\text{elec}} = K \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \log \left(\frac{2m_e c^2 \beta^2 \gamma^2 Q_{\max}}{I^2} \right) - \beta^2 - \frac{\delta}{2} + \frac{Q_{\max}^2}{8(\gamma M_\mu c^2)^2} \right] + \Delta \left| \frac{dE}{dx} \right|. \tag{2.9}$$

Here, c is the speed of light in a vacuum, $\beta = v/c$ where v is the muon speed, and γ is the Lorentz factor $= 1/\sqrt{1 - \beta^2}$. Additionally, the prefactor K$=4\pi N_A r_e^2 m_e c^2$, Z and A are atomic number and mass respectively, m_e and r_e are the electron rest mass and classical radius respectively, M_μ is the muon mass and $Q_{\max}$ is the maximum possible kinetic energy of the excited electrons, given by

$$Q_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2} \tag{2.10}$$

δ is the density-effect correction [12] and I is the mean excitation energy of the electrons. Finally, the factor $\Delta \left| \frac{dE}{dx} \right|$ accounts for an additional electronic energy loss factor: bremsstrahlung interactions with electrons. Included for completeness, it is given by

$$\Delta \left| \frac{dE}{dx} \right| = \frac{K}{4\pi} \frac{Z}{A} \alpha \left[\log \frac{2E}{M_\mu c^2} - \frac{1}{3} \log \frac{2Q_{\max}}{m_e c^2} \right] \log^2 \frac{2Q_{\max}}{m_e c^2}. \tag{2.11}$$

11

2.3.2 Radiative losses

The radiative component of muon energy loss in matter, b, can be split into a sum of the contributions from the three physical processes:

$$b(E) = b_{\text{bremsstrahlung}} + b_{\text{pair production}} + b_{\text{photonuclear}} \tag{2.12}$$

For different muon energies, the b functions vary in their contributions to the total energy loss, as shown in figure 2.4. This is due to the energy dependence of the underlying interaction cross sections. Figure 2.5 shows the interaction cross sections for the four muon energy loss processes, as a function of ν, for muons of 100 GeV (left) and 10000 GeV in iron. Of the radiative processes the dominant contribution is from electron-positron pair production up to $\nu \sim 10^{-1}$.
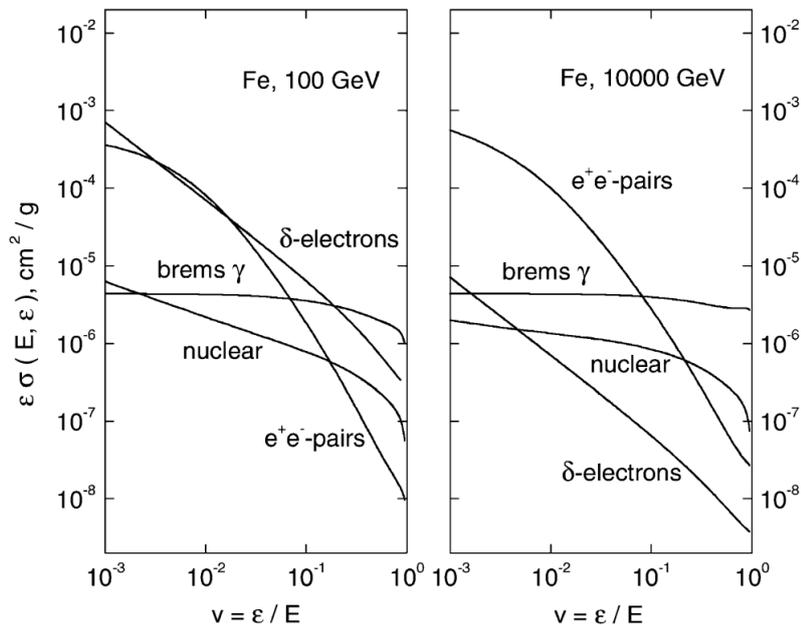
Figure 2.5: Energy-weighted differential cross sections σ for the muon energy loss processes in iron, for muon energy $E = 100$ GeV (left) and $E = 10000$ GeV (right). Here, η is the amount of energy transferred. These cross sections are those used in the GEANT4 particle simulation software [13]. From [9], © 2006 IEEE.

Details of the three radiative processes are described below. The full cross sections for the interactions can be found in [8].

Bremsstrahlung

The bremsstrahlung interaction involves the deflection of a muon by another charged particle, causing it to decelerate and emit a photon to account for the lost kinetic energy. In matter, the deflection can be caused by either a nucleus or atomic electrons, and both possibilities must be accounted for when calculating the full cross section. The former cross section was first calculated in [14], and the latter in [15]. Figure 2.6 shows the nuclear component of the

differential cross section as a function of ν, comparing several different muon energies, for muons in copper. The higher energy muons have a larger cross section for the bremsstrahlung interaction and hence are more likely to experience energy loss via this route.
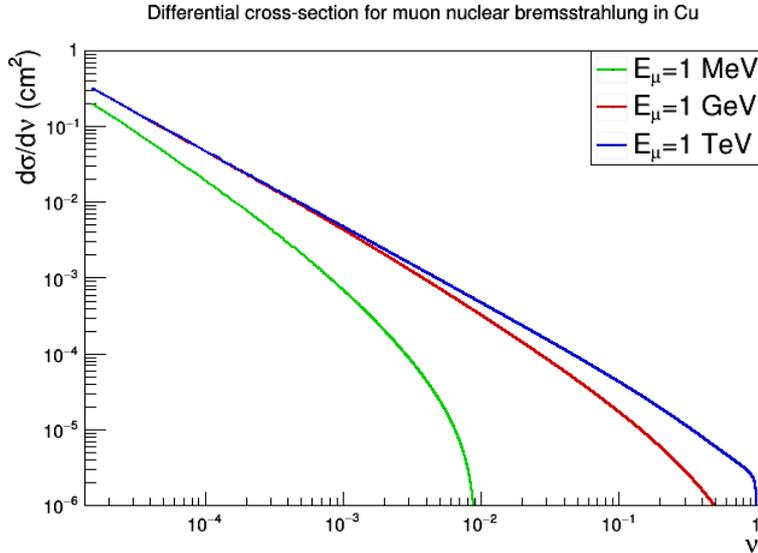
Figure 2.6: The differential cross section for muon energy loss via nuclear bremsstrahlung, for muons of different initial energies in copper.

Pair production

The largest radiative contribution to muon energy loss is electron-positron pair production [16], in which a virtual photon is emitted that decays into an electron and a positron with opposite momenta. Although tau-antitau and muon-antimuon pair productions are also possible, they are suppressed due to the larger mass of these leptons.

Photonuclear contribution

The smallest radiative contribution to muon energy loss is loss via the photonuclear interaction $b_{\mathrm{photonuclear}}$; it becomes significant only at muon energies $\gtrsim 1$ TeV. A muon interacts inelastically with an atomic nucleus, producing a photon. The cross section for this interaction was calculated in [17].

2.4   Detector technology

As charged, high-mass particles with relatively long lifetimes, muons are readily detectable by particle detectors. A wide variety of detector technologies have been used over the past seven decades of cosmic-ray muon imaging studies. Some of the earliest significant results in the field (see section 3.2.2) were obtained using spark chambers: simple gas (usually neon)-filled chambers in which a high voltage is induced across the gas when triggered by the passage

of a charged particle, which then becomes visible as the energised gas de-ionises. Some of the detector types in use currently include scintillating materials, nuclear emulsion films, and gaseous drift chambers; as muon studies can vary hugely in the environmental conditions and available exposure times, the choice of detector technology for a particular study can be tuned to the problem.

2.4.1 Scintillator

Scintillating materials have the property that they emit electromagnetic radiation, usually visible or near-visible light, after encountering a charged particle or high-energy photon. The particle deposits a portion of its energy into the material's electrons, which subsequently re-emit the energy as photons, either instantly (in which case the process is termed fluorescence) or after a delay of μs-hours (phosphorescence). When coupled to a Photo-Multiplier Tube (PMT), a device that converts small quantities of detected light into measurable electronic signals, they make ideal particle detectors. Scintillator-type detectors widely used for muon studies in particular due to their versatility and relatively low cost.

There are two main categories of scintillating material: inorganic and organic [18]. The former are often solid crystals of an alkali halide, the most common being sodium iodide, combined with a deliberately added impurity called an 'activator', such as tellurium. The scintillation mechanism in inorganic crystals arises from the electronic band structure of the crystal. The passage of an energetic particle excites electrons from the valence to the conduction band, creating a free electron and a corresponding free hole. These can move freely through the material. The presence of the activator material atoms creates localised additional energy levels in the forbidden gap; the free electrons can de-excite by dropping to these levels and emitting a photon. Although inorganic crystals can be shaped with some freedom, they can be brittle and often must be sealed in airtight containers to prevent water damage. They are therefore not ideal for muon applications which often require detectors that are highly durable due to harsh environmental conditions.

Organic scintillators are quite distinct in their physical mechanism of scintillation. They consist of organic molecules containing benzene-ring structures (e.g. anthracene, naphthalene) and hence free electrons; the scintillation occurs at the molecular level and so is preserved for different states of the material, unlike inorganic crystals for which a solid crystal structure is required. Therefore organic scintillators can be solid or liquid, and can also be dissolved in a plastic solvent that is then shaped and solidified, creating plastic scintillator. In all cases, the underlying mechanism is that a transiting particle transfers some of its energy to an electron, exciting it from a ground state S_0 to an excited state S^*. The electron then decays to an intermediate fine-structure level, created by the vibrational modes of the molecule as a whole (see figure 2.7). The emitted photon therefore has less energy than the gap between the ground and excited states and so is not re-absorbed by the material but passes through.

Plastic scintillators in particular are highly suitable for muon detection and are widely used. Commonly used plastic solvents include polyvinyltoluene and polystyrene, with the
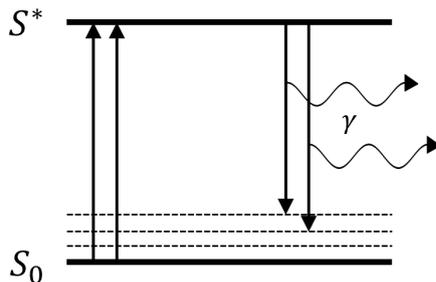
Figure 2.7: Schematic of fluorescence in an organic scintillator material. A charged particle or photon excites electrons from the ground state (S_0) to an excited state (S^*). The electronic fine structure contains additional modes close to S_0 that the excited electron can decay to, releasing a photon to which the material is transparent.

phenyl-bearing organic solute often being PPO (2,5-diphenyloxazole) or POPOP (1,4-bis(5-phenyloxazol-2-yl) benzene). For example, the CRIPT muon detection system described in [19] used plastic scintillator made of polystyrene solvent and a combination of PPO and POPOP solutes. Plastic scintillator can be easily shaped to match the available detector site, for example into square tiles that are then combined into large planes, or narrow cylinders that can be placed into a excavated borehole. They are also relatively rugged with higher tolerances of heat and mechanical stress than other materials, inorganic crystals in particular.

Figure 2.8 shows a plastic scintillator-based muon detection system used by Geoptic [20] for infrastructure studies. Each case contains two long cuboid paddles coupled to PMTs; the cross-detector allows tracks to be constructed from the detector hits. Another commonly used arrangement is to use large-area alternating x and y tracking planes. For example, [21] makes use of two 12.25 m^2 square planes formed of many narrow strips of polystyrene scintillator, separated by a 1 m air gap. The gap between the planes decreases the uncertainty on the muon trajectory measurement.

2.4.2 Nuclear emulsion films

Nuclear emulsion films are an alternate class of particle detector with applications in muon detection. They consist of silver bromide crystals, less than 1μ in size, held in some kind of support such as a frame of gelatin [22]. The passage of charged particles creates 'latent images': a concentration gradient of metallic silver atoms, reduced from the Ag^{2+} cations. When developed, more silver cations are reduced at a rate following the concentration gradient, leading to the formation of many orders of magnitude more metallic silver atoms and hence visible tracks (see figure 2.9). The process is very similar to the formation of images on photographic film. The resulting tracks can then be found via automatic optical microscopy and used to calculate the muon flux distribution.

One drawback of nuclear emulsion-based detectors is that they provide no real-time flux data. Whereas a detector based on scintillator, for example, can be directly monitored for its detected flux, a nuclear emulsion film can only be read after the survey is complete and the

Figure 2.8: A scintillator-based muon radiography detector system. Each case contains two plastic scintillator paddles coupled to PMTs. The system is set to trigger when signals are recorded in the middle plane and the base detector within a narrow time window, with the top detector used to provide additional tracking. The 10 detectors allows for 28 total angular bins, with 12 formed from 'double' hits (in the middle and base planes) and 16 from 'triple' hits.

film developed. Also, the automatic optical microscopy equipment necessary to read emulsion film data is quite specialised and costly. They do however possess excellent spatial resolution, comparable to the crystal size; the detectors developed in [22] had crystal size 0.2 μm and spatial resolution 0.5 μm. Nuclear emulsion films were also successfully used for assaying the Great Pyramid of Giza [24], see section 3.2.2.

2.4.3 Gaseous detectors

Ionisation-based gaseous detectors are often used for muon detection, and can vary greatly in design and mode of operation. The basic design is to create a strong electric field through a volume gas with a pair of electrodes [18]. Charged particles passing through the gas will then create electron-cation pairs; if the electric field is strong enough, these will separate and move towards the electrodes creating a detectable current. For example, one common chamber construction uses a cylindrical cathode surrounding a thin anode wire along its axis, to which a positive voltage is applied (see figure 2.10). The cylinder is filled with gas. The resulting electric field in the gas, as a function of the distance from the wire r, is then given by

$$E(r) = \frac{V}{r \log(b/a)} \tag{2.13}$$

16
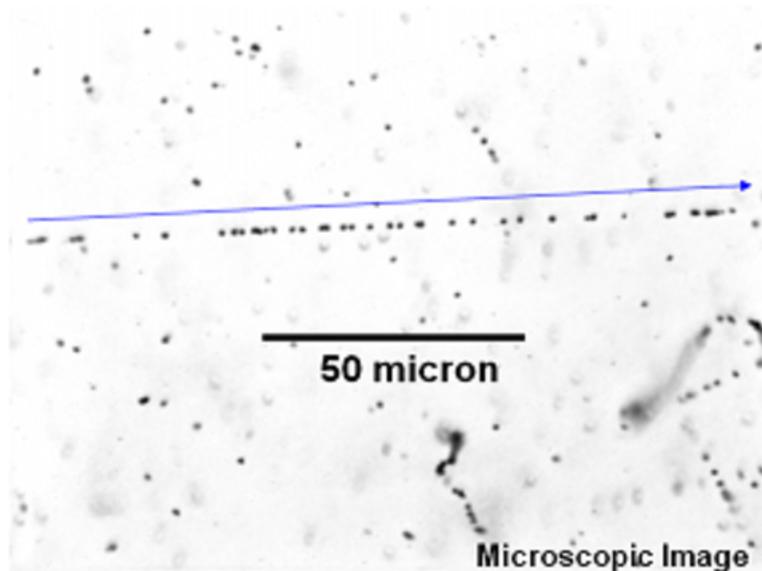
Figure 2.9: From [23], example of a charged particle track in a nuclear emulsion film. The tracks can only be found with an optical microscope, necessitating specialised equipment to interpret the results.

where V is the voltage applied to the wire, a is the radius of the anode wire, and b is the internal radius of the cylinder.

Multiwire Proportional Chamber

A particularly useful aspect of ionisation detectors is that the operational mode can be changed by adjusting the provided voltage V. If the voltage is below a threshold, the electron-ion pairs created by the passage of a charged particle will recombine, as the electric field is not strong enough to overcome their mutual attraction. If V is increased, some of the electrons start to move towards the anode wire, creating a detectable signal; eventually a plateau is reached when none of the created pairs are recombining. Detectors designed to operate in this range are called 'ionisation chambers'. A further increase in V results in a new effect: the electrons are now sufficiently energetic that they can ionise additional atoms in the gas, which can then themselves ionise more atoms, resulting in a huge increase in electron production called an 'avalanche'. The size of the avalanche is directly proportional to the number of 'primary' electrons originating from the charged particle. Detectors of this type are thus called 'proportional chambers'.

The Multiwire Proportional Chamber (MWPC) [25] is a powerful detection system based on the cylindrical proportional chamber. Rather than a single axial wire, a MWPC consists of a series of evenly spaced narrow anode wires, between two parallel flat plates that comprise the cathodes. The electric field inside the chamber then has two distinct regions: away from the wires, the field lines are parallel to each other and the field is almost constant (see figure 2.11. Near any of the wires the field has a $1/r$ dependence as for the cylindrical chamber described above. When a charged particle enters the chamber and encounters ionises atoms in the gas, the electron will move along the parallel field lines towards the closest anode wire.

17
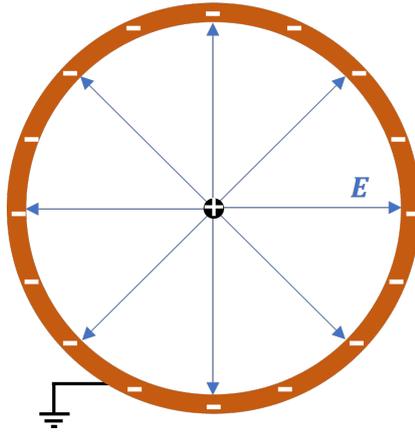
Figure 2.10: Schematic of the basic design of a cylindrical ionisation-based detector. A solid cylinder is filled with gas (e.g. argon). A positive voltage is applied to a thin anode wire along the cylinder's axis, creating a radial electric field **E**. When a charged particle enters the gas, it ionises atoms; the resulting electrons move towards the anode wire creating a detectable signal.

When close to the wire, the increasing field strength causes the electron to accelerate, once again producing an avalanche of further ionisations and a detectable signal on that wire.

A single MWPC will only give spatial information along the axis parallel to the anode wires. For track reconstruction, the usual method is to use pairs of MWPCs orientated perpendicular to each other, hence providing x and y coordinates of the particle position. Combining several such layers then allows the tracks of muons to be fitted to the hits from each detector layer. For example, [27] describes the use of a muon telescope based on layers of MWPCs for a volcano imaging study.

Drift chamber

Finally, a particularly potent form of proportional chamber is the drift chamber, which is a further evolution from the MWPC. A drift chamber provides additional spatial information about the charged particle by considering the speed at which the free electrons created by its path move towards the anode wires, the 'drift speed' u. As u is approximately constant in the region away from the wire, the distance d between the charged particle's track and the wire can be calculated if u, t_0 (the time at which the particle entered the chamber) and t_1 (the time at which the signal is detected) can be calculated as $d \approx u(t_1 - t_0)$. u can be calculated from the properties of the chamber gas, whilst t_1 is supplied by the signal itself. t_0 is determined by using a triggering system, such as a scintillator plane above the chamber. Unlike a MWPC, for which the spatial resolution is limited by the separation of the anode wires, a drift chamber can achieve much finer resolution, down to < 100 μm [18].

18
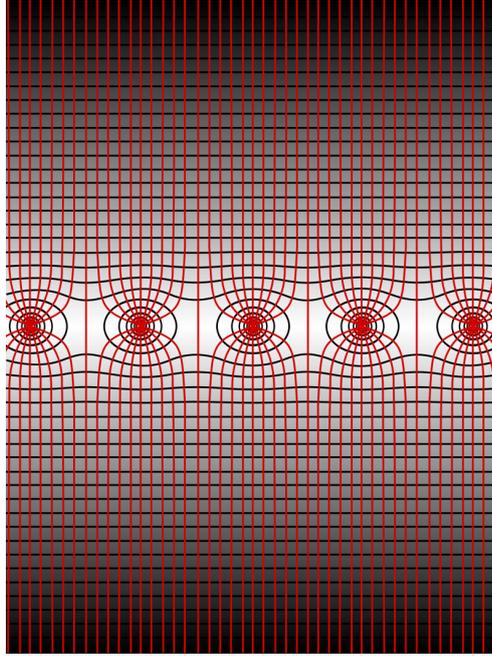
Figure 2.11: From [26]. Schematic of the electric field lines (red) and equipotentials (black) inside a MWPC. The field is almost constant in most of the region, but takes on a $1/r$ strength when close to any of the wires. An electron in the field will drift towards the closest wire then accelerate, causing an avalanche.

2.5 Monte Carlo muon simulations

Computational simulations are an extremely useful tool in muon science. For many applications of muon imaging, the sites involved are difficult or hazardous to access, such as the inside of a pyramid or at a remote volcano. Safely testing a proposed experimental programme via an accurate simulation, taking into account the distributions of cosmic-ray muons and modelling their passage through matter, is essential to ensure that the project is feasible. For example, preparation for a imaging study of a volcano would involve choosing the number, type and location of the detection systems; a simulation programme utilising an accurate geological model of the volcano could be then used to determine the minimum number and size of detectors needed to achieve acceptable results without the expense and difficulty of experimental testing. Other muon applications, such as nuclear waste assay, involve hazardous materials. Simulation studies can then be used to test algorithms and detector systems before they are applied practically.

Additionally, for muon radiography in particular (see section 3.2), simulations are often a required step in the experimental study itself. A digital model of a structure is created under the assumption of some null hypothesis (e.g. a pyramid containing no unknown chambers) and a simulated muon radiography is performed, matching the experimental configuration as closely as possible. The experimentally detected muon flux distributions can be compared to the results from the simulation. If there is a region in which the experimental results record a muon excess relative to the simulation, it suggests the presence of an unknown void. For example, the radiography of the Great Pyramid performed in [24] made use of a simulation of

19

the Pyramid (see figure 2.12) with only known structures (see section 3.2.2).

Figure 2.12: Detailed view of a simulated model of the interior of the Great Pyramid, used to produce a 'null hypothesis' simulated muography to which the experimental data could be compared, to identify differences indicating unknown features. From [24], reproduced with permission from Springer Nature.

The relevant category of simulation for muons is a Monte Carlo simulation. In general, these are used for problems involving random variables, and operate by sampling from probability distributions to determine the outcomes of stochastic events. This is the opposite of a deterministic simulation. For example, the emission of gamma rays from a piece of radioactive material could be modelled with a Monte Carlo simulation: the individual decays and the interactions of the rays with matter are stochastic processes. This is opposed to a deterministic scenario, such as the trajectory of a spacecraft in the Solar System, for which a Monte Carlo simulation would not be appropriate.

A Monte Carlo simulation of cosmic-ray muons begins with some model of the muon states at a particular altitude. From this, distributions of the muon energies and trajectories are determined; each particle to be transported then has an initial state drawn randomly from these distributions. Commonly used models are the CRY library [28], which provides muon angular and energy distributions at sea level or other altitudes as calculated by a Monte Carlo transport of primary protons from the top of the atmosphere; and the 'Shukla' flux, derived from the empirical flux model of [29].

The next stage of the simulation - transportation of the initialised muons - is handled by dedicated transport software. The GEANT4 [13] software is an extremely powerful and flexible toolkit, widely used in particle physics, that can accurately transport particles of all species through complex and intricate geometries. GEANT4 can be used for simulating muon imaging applications, but for cases where secondary particles are not relevant (e.g. deep underground), codes dedicated specifically to muon transport may be more appropriate. Examples of these include MUSIC [30] and PUMAS [31]. Whichever software is used, some kind of model describing the material distribution must be created and supplied. GEANT4, for example, uses a set of geometric primitives that can be combined and manipulated into very complex structures. The material at each point in space is defined, with the key parameters

being its density and the atomic number, mass and number density of its constituent elements.

From the material information, the appropriate differential cross-sections for the four key interactions (ionisation, pair production, bremsstrahlung and photonuclear) for muons of various energies in matter can be calculated. By integrating the cross-sections, tables of the mean energy loss $\langle dE/dx \rangle$ can be produced [9]; the energy loss during transport can then be sampled from these distributions. The usual procedure is to model each muon's energy loss as a combination of continuous and discrete processes, with the former for low-energy interactions and vice-versa. A 'detector' can be defined, shaped and positioned to match an experimental system, and set up to record particles that encounter it. The angular and spatial resolution can be implemented by applying a smearing factor to the Monte Carlo truth, and the detector's efficiency can be implemented by disregarding some fraction of the encountered particles.

Figure 2.13 shows an example GEANT4 Monte Carlo simulation of muons traversing a layer of rock, in an interactive mode that displays the modelled geometry and particle tracks visually. The path of each muon is shown in yellow. The muons with sufficient energy to penetrate deep into the rock are not greatly scattered, and so most tracks appear as approximately straight lines. The deposited energy from the muon interactions with the atoms in the rock result in the creation of secondary particles, shown in cyan. These particles, for example electrons and positrons created via pair-production, cannot penetrate the rock to the same degree as muons and are quickly attenuated.
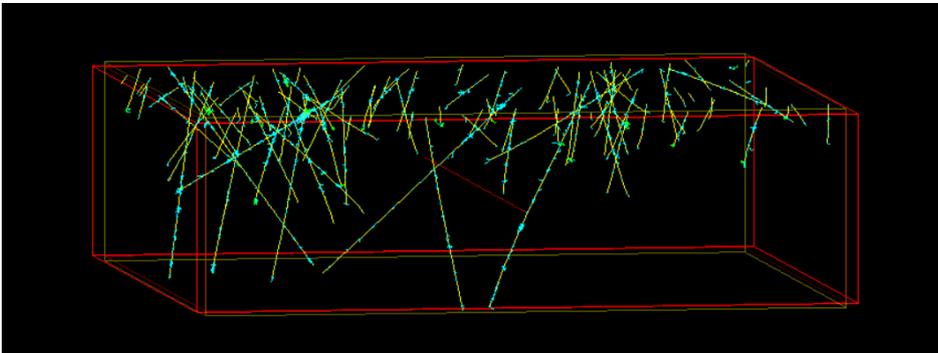
Figure 2.13: Interactive view of a GEANT4 [13] simulation of muons in rock. The muons (yellow) are initialised in a thin plane on the top of the world volume (red) with their states drawn from the CRY library [28]. They are then transported forward; generated secondaries (cyan) are quickly attenuated by the dense material.

Chapter 3

Muon radiography and muon scattering tomography

3.1 Overview

The cosmic-ray muon flux is a natural, well-characterised source of easily detectable, highly penetrating particles. It is therefore usable for a huge range of imaging and assay applications. There are two main families of techniques making use of cosmic-ray muons: muon radiography and Muon Scattering Tomography (MST). The former relies on the attenuation of muons as they pass through some volume of interest; it is most commonly used for imaging large $(10 - 1000$ m) scale objects. Fields of application for muon radiography include mining [32], civil engineering [33][34], nuclear safeguarding (see chapter 5), volcanology [35][36][37][38], biology [39] and archaeology [24][40]. Imaging can be performed using a single detector, in which a 2D projection of the density can be obtained, or a network of multiple detectors in which case a 3D density map can be calculated by combining the projections. The required exposure time can vary hugely based on the application, from short surveys of only minutes up to over a year for particularly deep locations.

MST by contrast is more suitable for smaller length scales $(1 - 10$ m) and, in general, is compatible with shorter exposure times (minutes-hours). In an MST experiment, at least two detector modules are used, in order to determine the trajectories of muons before and after encountering some volume of interest in between the detectors. Information about the contents of the volume is then obtained from the distribution of the muon scattering angles, as calculated from the two trajectories. Applications of MST include nuclear waste assay, cargo scanning, and imaging of hazardous environments - most famously, the Fukushima nuclear reactor interior [41].

3.2 Muon radiography

3.2.1 Technique

A muon radiography experiment involves deploying one or more particle detectors with angular detecting capabilities underneath or alongside some volume of interest. This allows the muon flux as a function of direction to be measured. By comparing the measured flux to some other flux profile e.g. from a Monte Carlo simulation (see section 2.5), any excesses or deficits of muons along particular lines of sight suggests a respectively a lower or higher material density along that line than the null hypothesis would suggest. This is because muons are attenuated to a greater degree in denser material. Figure 3.1 shows a schematic of a muon radiography taken inside a pyramid to look for a hidden chamber. In this case, a larger muon flux would be recorded from the directions intersecting the chamber than in the opposite direction.
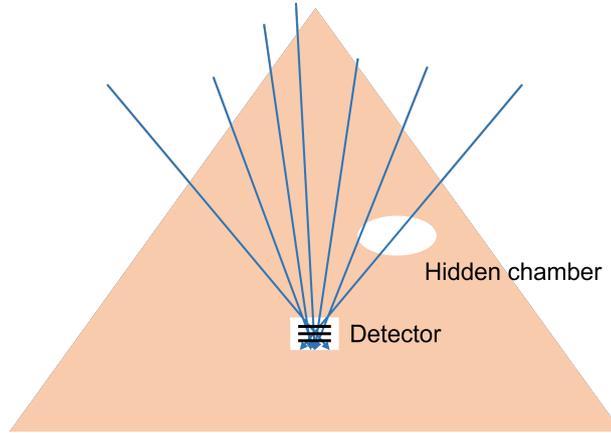
Figure 3.1: Schematic demonstrating the principle of muon radiography, in the context of searching for hidden voids in pyramids. The muon flux is less attenuated in the directions corresponding to the air-filled chamber than would be expected in the null hypothesis of the volume containing solid rock.

More specifically, the quantity accessible in a muon radiography measurement is the *opacity* ϱ, defined as the line of sight integral of the material density:

$$\varrho = \int_{\text{LOS}} \rho(x) \, \mathrm{d}x \tag{3.1}$$

ϱ is often expressed in terms of 'metres water equivalent' (m.w.e.): the thickness of water that would result in the same opacity (1 m.w.e. $\equiv$ 100 g cm^{-2} $\equiv$ 1 hg cm^{-2}). For example, the opacity of 100 cm of homogeneous concrete with density 2.3 g cm^{-3} is $\varrho = 230$ g cm^{-2} $\equiv$ 2.3 m.w.e.. The total opacity of the atmosphere between the mean height of muon production and sea level is around 10 m.w.e. [7].

An opacity ϱ can be deduced from a measured muon flux $I(\theta, \phi)$ in a number of ways. One possibility is to make use of an empirical relation that relates the two quantities, derived from many experimental measurements taken at different depths. A commonly used such

relation was derived in [7], which relates I to an opacity ϱ in m.w.e. via

$$I(\varrho) \left[\text{cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}\right] = \frac{Ke^{-\eta\varrho}}{(\varrho^{\alpha} + a)(\varrho + H)} \tag{3.2}$$

where $K = 270.7$ hg cm^{-2}, $\eta = 5.5 \times 10^{-4}$, $\alpha = 1.68$, $a = 75$ and $H = 200$ hg cm^{-2}. The form of this curve, with comparison to some of the experimental data used in its derivation, is shown in figure 3.2.

Muon vertical intensity vs depth
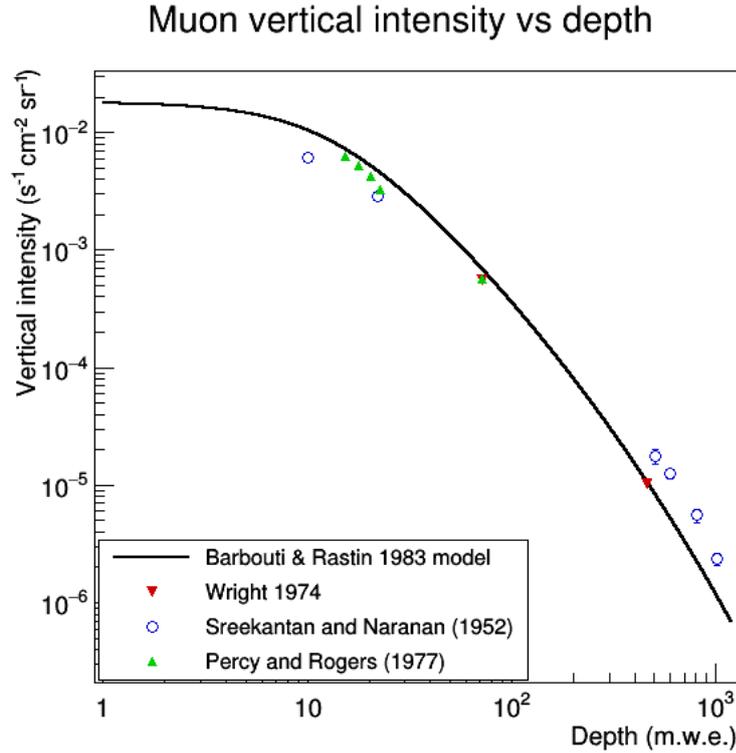
Figure 3.2: Empirical vertical muon intensity vs depth model from [7], compared to selected experimental data [42][43][44].

The experimental data used to derive this curve has been normalised to a depth of 'standard rock': a fictitious material with useful properties. Standard rock has, by definition, $\langle Z \rangle = 11$ and $\langle A \rangle = 22$. The assumption made by using this relation is that the only relevant factor affecting the muon flux is the density of the overburden materials, with its chemical nature not taken into account. However, it has been shown [45] that this is not the case and that the measured muon flux under a particular opacity of one mineral will be different than would be expected under an equivalent opacity of standard rock, with the relative difference increasing with depth. For shallow depths the effect is small. At 1 km depth under, for example, limestone, the measured flux would be only $\sim 85\%$ of that calculated from the relation in equation 3.2. The specific quantity affecting the flux is the material's $\langle Z^2/A \rangle$. This topic is explored in more detail in chapter 6.

An alternative method to calculate the opacity is to consider the muon energy loss. As

described in section 2.3, the mean muon energy loss in matter is given by $\langle -dE/dx \rangle = a(E) + b(E)E$. If this equation is inverted an integrated up to an energy E, a range $R(E)$ is obtained; this range has dimensions of opacity and is called the Continuous Slowing Down Approximation (CSDA) range [8]. It is an approximation to the true opacity that muons of energy E can pass through before being stopped. As an example, figure 3.3 shows the CSDA range for muons in granite.
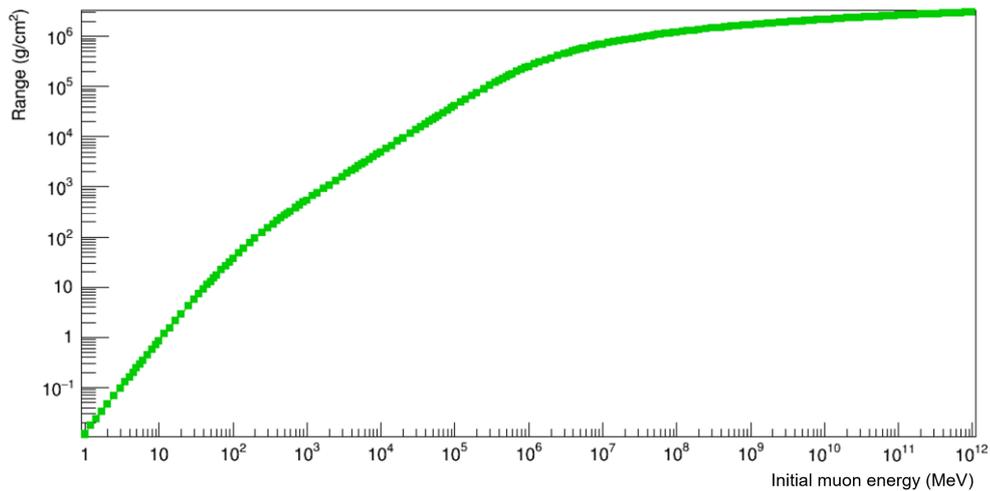
Figure 3.3: Muon range in granite as a function of initial muon energy, as calculated using the PUMAS library [31].

3.2.2 Applications

Search for hidden chambers in the Second Pyramid of Giza

The first known use of muon radiography for density measurements was in 1955, for measuring the thickness of the overburden above a tunnel in an Australian power station [33]. 15 years later, a famous and highly influential muon radiography experiment was performed in Egypt: an attempt to determine the presence of any hidden chambers inside the Second Pyramid of Giza (also called the Pyramid of Chephren) [40]. The Second Pyramid was not known to contain a King's Chamber or Grand Gallery as the larger Great Pyramid does. If such a chamber existed, it is plausible that it could have remained undetected and undisturbed during the almost 4600 years since the Pyramid's construction; the contents would be archaeologically priceless. Muon radiography provided a non-invasive and highly sensitive method to search for the density deviation of a hidden chamber.

The researchers made use of spark chambers, an early form of particle detector with which visible tracks are obtained by applying a high voltage across a volume of gas as a charged particle passes through and ionises the gas (using scintillator trigger panels to identify events and initiate the high voltage). Placed in the Pyramid's 'Belzoni' chamber, their detection system (consisting of 4 stacked spark chambers coupled to 14 scintillator counters in between

the layers) had good angular resolution and they were thus able to record muon counts as a function of angle.

Their results were compared to a simulated model of the expected muon counts, then the differences calculated and expressed in terms of the closest integer number of standard deviations of separation. The results (see figure 3.4) indicated no regions of significant muon excess beyond what would be expected from fluctuations. Although no hidden chamber was
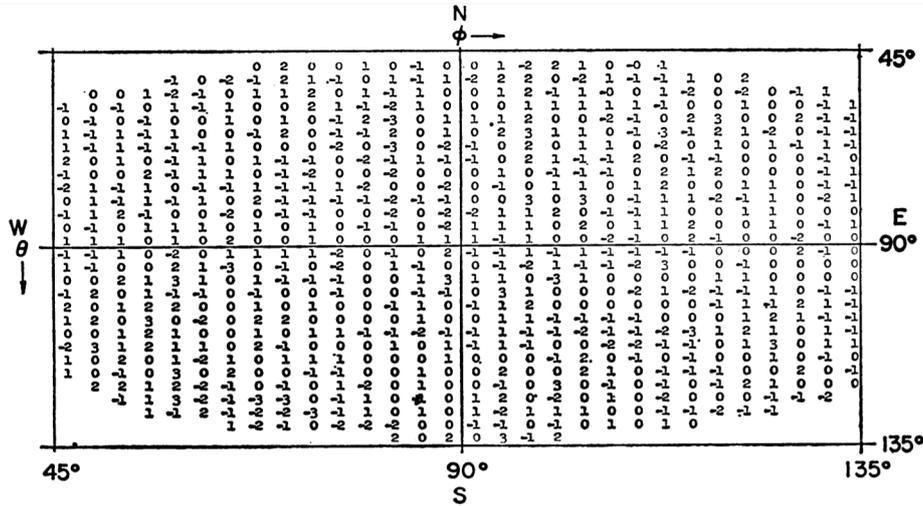
Figure 3.4: The closest integer number of standard deviations between experimental muon counts under the Pyramid and the counts expected from simulation. From [40], reproduced with permission from AAAS.

found, this study was hugely significant, both for finding to a high level of confidence that there are no large hidden chambers inside the Second Pyramid, and for demonstrating the power of muon radiography for assaying the density distribution of large structures. Later radiography studies (see below) would return to the Pyramids of Giza and resume the search for hidden chambers, with more success.

Volcano imaging

Some of the most iconic and stunning results in the field are from studies applying muon radiography to the imaging of volcanoes, in order to monitor their internal structures and watch for signs of impending eruptions. Most volcano imaging studies make use of near-horizontal muons; the intensity of these muons is much lower but the mean energy is higher (see figure 3.5) than lower-zenith-angle muons, and so they are able to penetrate even through a km-scale object. It is possible to obtain sufficiently accurate density maps on timescales less than those relevant for eruption predictions.

The first study into the feasibility of volcano imaging with muons was in 1994 [46]. At the non-volcanic Mount Tsukuba in Ibaraki, Japan, they demonstrated using a three-layer scintillator detector that the flux of near-horizontal muons is sufficient for radiographic measurements on timescales of a few weeks; an example of their results is shown in figure 3.6. They went on to calculate that an an enlarged version of their hodoscope, with a sensitive area
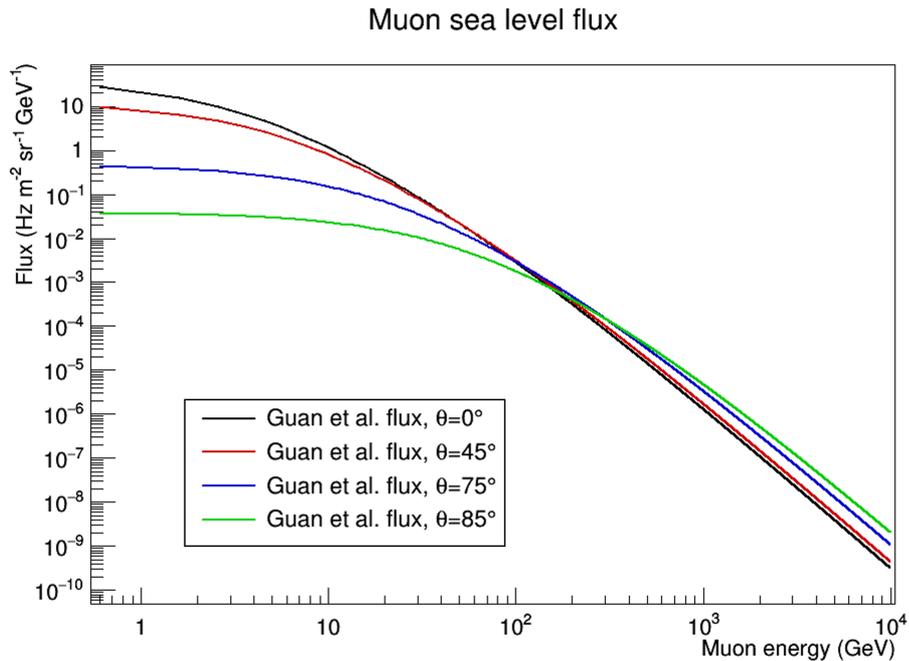
Figure 3.5: Muon sea level differential flux (equation 2.7) for different zenith angles θ. The mean muon energy is higher for close to horizontal muons.

of 20 m², could detect the formation of a 20 m-scale cavity inside a similarly sized volcano within a few days.

Great Pyramid chamber detection

One of the most prominent muon radiography results of recent years has been the detection of a hidden chamber in the Great Pyramid of Giza [24]. Radiographic measurements were taken using several different detector technologies, and all confirmed the presence of a large unknown chamber inside the Pyramid. The initial results were collected using a nuclear emulsion film-based detector (see section 2.4), stored in the Queen's chamber inside the pyramid; when compared to radiographs generated via a detailed Monte Carlo simulation (see figure 3.7) an excess corresponding to a 30 m long void was observed to $> 10\sigma$ significance.

3.3 Muon scattering tomography

3.3.1 Technique

Muon Scattering Tomography (MST) is an alternative imaging technique that exploits instead the elastic Coulomb scatterings of muons in matter. The distribution of the scattering angles is highly sensitive to the atomic number Z of the material, and so calculating the scattering angles of muons passing through some volume of interest will yield information about the volume's material composition (see figure 3.8).
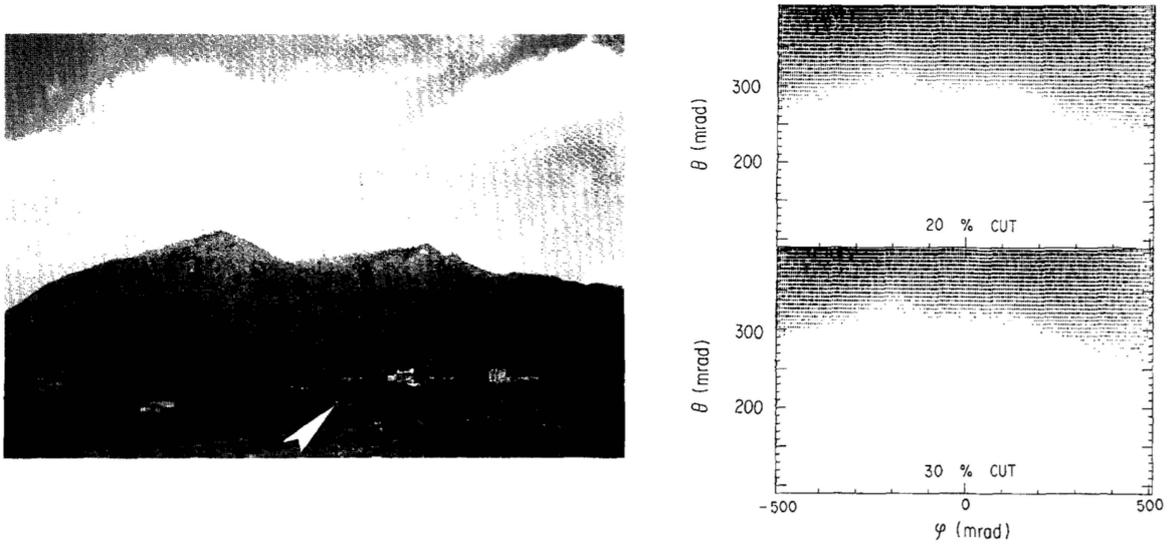
Figure 3.6: A photograph of the profile of Mt. Tsukuba (left) compared with muon radiographs of the mountain (right), with different background rejection cuts chosen. The profile of the mountain is clearly visible. From [46], reproduced with permission from Elsevier.

The scattering angle distribution is approximately Gaussian for small angle scatterings. Its width, obtained via a fit [47] to the Molière theory, is approximately given by

$$\sigma = \frac{13.6 \text{ MeV}}{\beta c p} \sqrt{\frac{x}{X_0}} [1 + 0.038 \log(x/X_0)] \tag{3.3}$$

where p is the muon momentum, βc is the muon speed, x is the thickness of the material and X_0 is the radiation length, which is dependent on the elemental composition of the material. This quantity is the length-scale of electron energy losses via bremsstrahlung in matter, and is defined as the distance over which a high-energy electron will lose $1/e$ of its energy for that material. For an elementary material, the radiation length (see figure 3.9) is given by [48]

$$X_0 = \frac{716.4 \text{ g cm}^{-2} \, A}{Z(Z+1) \log(287/\sqrt{Z})} \tag{3.4}$$

where Z and A are respectively the atomic number and mass of the material. For a compound material containing n elements, the radiation length can be calculated via

$$\frac{1}{X_0} = \sum_i^n \frac{w_i}{X_i} \tag{3.5}$$

where the X_i are the radiation lengths of the elements and the w_i are their material fractions by weight. For example, the radiation lengths of elemental iron ($Z = 26$, $A = 55.8$) and oxygen ($Z = 8$, $A = 16.0$) are 13.8 g cm^{-2} and 34.2 g cm^{-2} respectively, hence the radiation length of the compound ferrous oxide (FeO) is 15.9 g cm^{-2}.

28
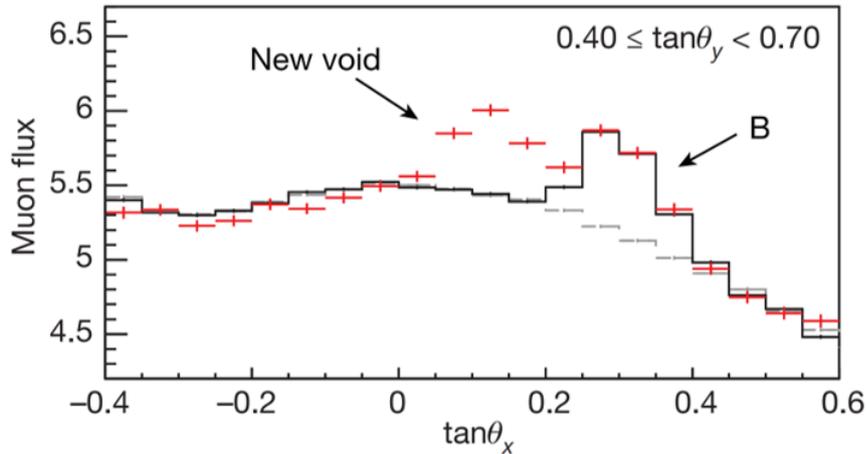
Figure 3.7: Muon flux (in $cm^{-2}sr^{-1}day^{-1}$) as a function of zenith angle, with a clear excess corresponding to the previously unknown void. Here, black and grey are results from Monte Carlo simulations of the pyramid with and without the known cavities (e.g. the King's Chamber) and red is the experimental data. From [24], reproduced with permission from Springer Nature.

3.3.2 MST algorithms

PoCA algorithm

In practice, one of several MST algorithms is used to create an image from muon scattering angle data. The simplest and most widely used is the Point of Closest Approach (PoCA) algorithm [49], which assumes that each muon scatters at only a single point, called a 'scattering vertex':

1. The volume of interest is divided into cubic voxels to store the 3D image values.

2. For each muon, the detected incoming and outgoing muon tracks are calculated through linear interpolation of the detector hits.

3. The calculated tracks are extrapolated into the volume of interest.

4. The point at which the two tracks come closest together is calculated and the weighting for that voxel is updated based on the scattering angle of the muon.

PoCA is effective for distinguishing regions of high contrast in Z, such as a large high-Z object inside a waste drum (see figure 3.10, left). However, the assumption that each muon scatters only at a single point limits the algorithm's effectiveness.

Binned clustering algorithm

To partially mitigate the weaknesses of PoCA, some algorithms build on PoCA by incorporating additional information from the muon scattering angles. The binned clustering algorithm [50] uses the PoCA but also exploits the spatial density of scattering vertices within voxels; a higher density of vertices in a region indicates more high-angle scatterings and hence indicates
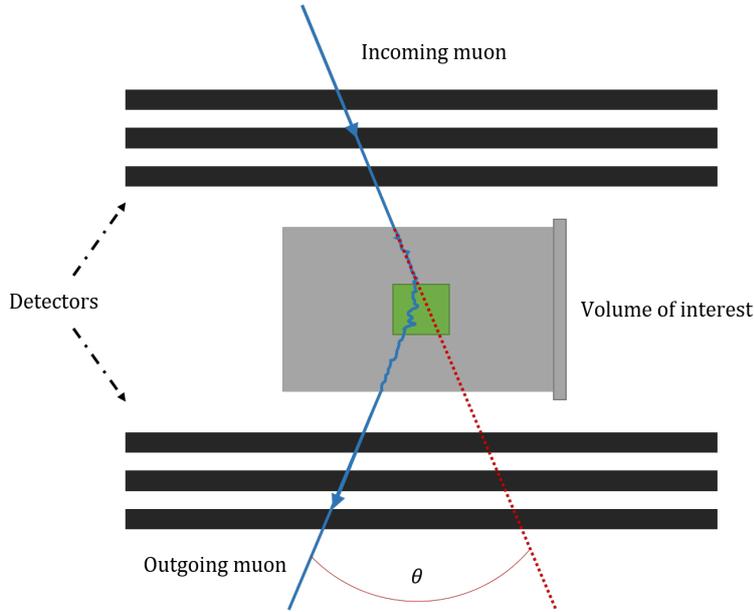
Figure 3.8: Schematic demonstrating the principle of MST: a muon passes through a detector module, scatters in the waste drum (with larger scatterings in the block of high-Z material), and is then detected by the second module. The scattering angle θ is greatly exaggerated here.

the presence of high-Z material. This algorithm is used extensively in chapter 4. A full description of the binned clustering algorithm can be found in section 4.2.

ASR algorithm

An alternative method is to not make use of the point of closest approach at all, but instead to assign weights to voxels based on their proximity to the incoming and outgoing muon tracks. One algorithm that uses this method is the Angle Statistics Reconstruction (ASR) algorithm [51]:

1. The volume of interest is divided into cubic voxels. Each voxel j is associated with an initially empty set, S_j.

2. Each muon's detected incoming and outgoing tracks are extrapolated into the volume of interest.

3. Calculate the x- and y-projected scattering angles, θ_x and θ_y.

4. Identify every voxel j that has its centre within some threshold distance (e.g. the voxel size) of the incoming or outgoing tracks.

5. Append scores $p|\theta_x|$ and $p|\theta_y|$ (if the muon momentum p is known) to S_j.

6. Determine the final image values as some quantile q of each voxel's list S_j.

An example ASR MST image of a waste drum is shown in figure 3.10 (right). The improvement in the visibility of the stored object as compared to the equivalent PoCA image (left) is clear.
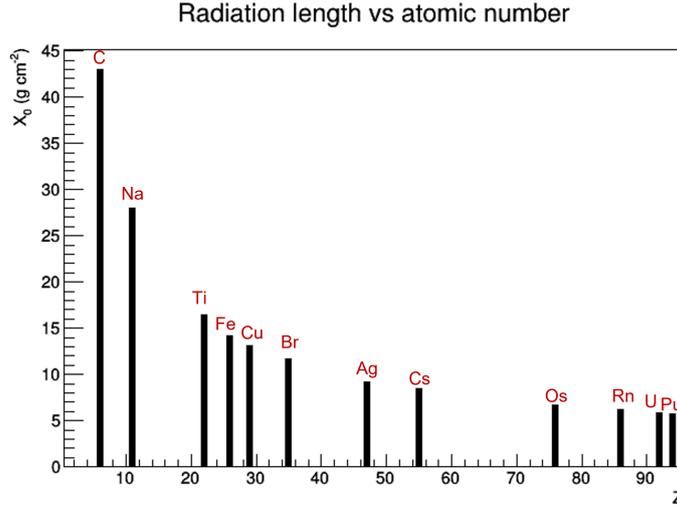
Figure 3.9: Radiation lengths of some example elements, as calculated using equation 3.4.

3.3.3 Applications

Cargo scanning for illicit nuclear material

The sensitivity of MST to high-Z materials in particular makes the technique highly suitable for applications in the nuclear industry, in which assaying the presence and quantity of the elements uranium and plutonium in various scenarios is vital. One such context is the scanning of cargo containers at border crossings or ports. Any smuggled Special Nuclear Material (SNM, a collective term for plutonium and uranium enriched in U^{233} or U^{235}[52]) is hugely dangerous; MST has the potential to identify such materials without the hazards and expense of either manual container searches or other scanning methods, such as X-ray imaging, that require introducing particles artificially.

The application introduces unique challenges: a practical system must be able to make a 'decision' (i.e. the container passes or is held for further inspection) very quickly ($\lesssim 1$ minute) to ensure that the facility can operate efficiently. Indeed the goal for cargo scanning is a rapid decision using a criterion with the maximum (minimum) possible true (false) positive rate: a full 3D image of the volume of interest is unlikely to be possible within the time constraints. Additionally, the detectors used must also be larger than the containers being investigated (or else physically moved around the detector, if this can be done within the time limits), and have sufficient spatial resolution to reconstruct muon tracks (and hence scattering angles) with enough accuracy to identify the targets.

Therefore there have been several studies, both simulation-based and practical demonstrations, into cargo scanning with MST. One of the earliest studies [50] investigated the feasibility of an RPC (see section 2.4)-based system for scanning shipping containers. They developed an algorithm based on the commonly-used PoCA algorithm, but incorporating additional information from the spatial density of the clustering vertices (this algorithm has also been applied for nuclear waste assay, see Chapter 4). Using Monte Carlo simulations of a
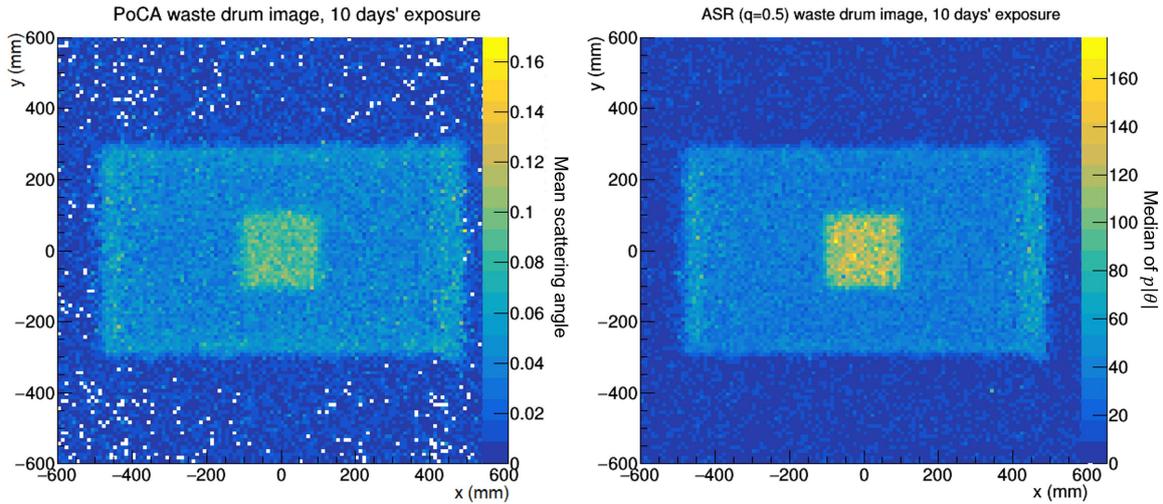
Figure 3.10: Comparison of two MST imaging algorithms: PoCA (left) and ASR (right). Each plot is 2D slice of a simulated MST image taken of a concrete-filled steel waste drum containing a 20 cm side length cube of uranium. The exposure time in both cases was 10 days.

shipping container containing a high-Z target block with various other materials present, it was demonstrated that a $10 \times 10 \times 10$ cm^3 block of tungsten or uranium could be detected with very high efficiency after $\approx$1 minute of muon exposure. A longer time of $\gtrsim$3 minutes was required when scrap iron shielding, modelled as a 1/10th density steel filling the container, was included. Figure 3.11 shows the Receiver Operating Characteristic (ROC) curves for this scenario. A ROC curve is a quantitative measure of a diagnostic test's performance. It is a plot of the true positive rate against the false positive rate for different choices of the diagnostic criteria; a cut that is too high will accept all true positives but an unacceptable rate of false positives, whereas a too-low cut will fail to detect a proportion of true positives. The optimum choice for the cut is that which maximises the true positive rate and minimises the false positive rate, which on the curve is the point of maximum curvature.

Fukushima reactor imaging

Muon imaging - both radiographic and MST - has also been applied to a unique problem: imaging the internal structure of a stricken nuclear reactor. In March 2011, the 9.1 magnitude Tōhoku earthquake and an accompanying tsunami caused devastation across large areas of eastern Japan. The disaster claimed tens of thousands of lives, and is considered the most costly natural disaster in history [53]. This is partially due to the effects of the disaster on the Fukushima Daiichi nuclear power station. The tsunami was higher than the plant's protective sea wall and so parts of the facility were flooded, disabling the emergency diesel generators required to keep the reactor core cooling systems operational. Three of the plant's six reactors consequently melted down, and huge quantities of radioactive material were released requiring the evacuation of over 150,000 people.

The reactor site now requires decontamination and decommission. However, the high
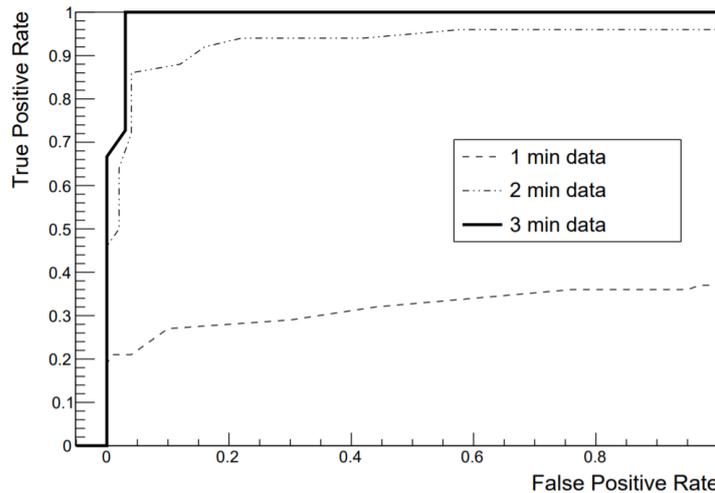
Figure 3.11: The ROC curve for the detection of a $10 \times 10 \times 10$ cm^3 uranium block in a shipping container when shielded by scrap iron, for different exposure times. From [50], © IOP Publishing, reproduced with permission.

radiation levels still present make these processes extremely hazardous; it is vital that the current state of the reactors (in particular the location of the melted fuel) are established before work can begin safely. Muon imaging techniques, with their sensitivity to material distributions and/or high-Z materials, and ability to operate passively for long periods without human intervention, are ideal for the task. Special care needs to be taken to account for the additional background on particle detectors due to the high radiation environment. Around the site, the dose is around 1 mSv hr^{-1} and primarily consists of γ rays from the isotopes ^{137}Cs and ^{134}Cs [54].

Some initial studies were based on a combination of simulation studies and practical demonstrations on passive models. [54] compared applying muon radiography and MST to the problem via GEANT4 simulation studies. Using three versions of a model of Fukushima reactor unit 1 - with an intact core, with no core, and with a damaged core - it was observed that MST was able to clearly image the core condition with only a few days' exposure, using the PoCA algorithm. Radiography however did not perform well, and was not able to image the core even after 6 weeks' exposure; after the same time MST could identify a core with only 1% of material missing.

This study was extended to additionally incorporate an experimental test [55], and additional simulation work based on Fukushima unit 2. To account for the ambient radiation in the environment the detector was modelled as being encased in a 50 cm thick concrete layer, sufficient to reduce the γ flux by a factor of 50. Practical tests were performed using a structure of concrete and lead blocks, with a central conical void as an imaging target. The detectors were assemblies of drift tubes with a detection area of 1.2 $\times$ 1.2 m^2; the void was successfully imaged after 3 weeks' exposure time.

Despite these promising results for MST, when muon imaging was applied to the reactor itself, radiography was the imaging mode used [41]; this is possibly due to the technique being

very well-established in Japan from previous volcano imaging studies. The detection system consisted of layers of scintillator bars with 10 cm-thick iron shielding; three layers of X-Y planes were used to reduce the probability of accidental coincidences from the background γ rays. The system was first tested by imaging the functional Japan Atomic Power Company (JAPC) reactor [56]. Subsequently, muon radiography was applied to all three of the damaged Fukushima reactors. At Unit-1 [57], the reactor was successfully imaged and it was determined that the melted fuel assemblies had likely dropped below the pressure vessel, but could not be directly observed. Similar conclusions were drawn for the assemblies of Unit-2 [58], for which an excess of mass consistent with that of the fuel assemblies was observed at the bottom of the pressure vessel; and at Unit-3 [59], for which a mass anomaly of around 60 tons was observed at the bottom of the pressure vessel. Figure 3.12 shows an example radiography image of Unit-1 after 90 days' exposure.
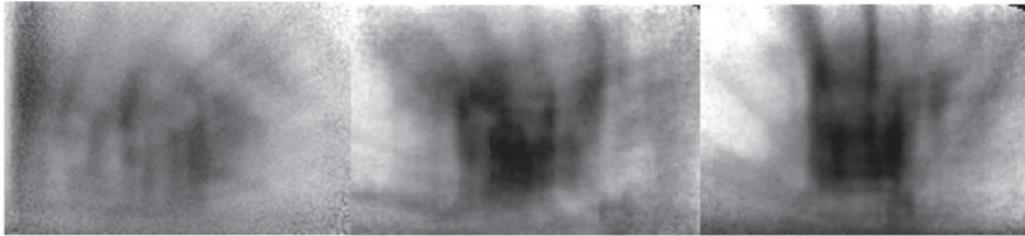
Figure 3.12: From [57], radiographic images of the Fukushima Unit-1 reactor obtained from three different detector locations, after 90 days' exposure.

Chapter 4

Material identification of objects in heterogeneous waste drums

4.1 Introduction

This chapter describes a Muon Scattering Tomography (MST) simulation study in which machine learning is applied to MST to improve its ability to determine the materials of stored objects inside concrete-filled nuclear waste drums. The technique is based on exploitation of a powerful MST imaging algorithm. Starting with a discussion of nuclear waste and available techniques for non-destructive assay of sealed waste containers, the development of the method from previous studies using machine learning for MST to the algorithm used and its adaptation for a classification problem is elucidated. Results of applying the technique to a variety of waste drum scenarios are then presented. Finally the limitations of the method and possible avenues for further study are discussed.

4.1.1 Nuclear waste storage

Nuclear energy is a vital component of the energy sources of many countries. The sector made up 30.5% of the energy generated in the European Union in 2022 [60], with most of the remainder consisting of fossil fuels (25.5%) and renewable energy sources (40.8%). The window to avert catastrophic climate change in the near future is rapidly closing, requiring a rapid transition away from fossil fuel energy production. Ultimately, renewable energy sources represent the best possible medium-term solution for the world's energy needs. However, it is difficult to see how a transition to an entirely renewables-based energy market can occur without the use of nuclear energy, making it a necessary evil in the short term at least.

The nuclear energy industry has two main characteristic problems: the waste generated from power plants (inevitable, and long-term), and the risk of nuclear catastrophe. The latter risk is illustrated vividly by the Chernobyl (1986) and Fukushima (2011) disasters; in each case unforeseen circumstances (compromises in reactor design [61] and loss of power following a natural disaster [41], respectively) led to crises that threatened millions of lives and

required enormous resources and ongoing efforts to control and stabilise. Although modern reactor designs are far safer, the risk of a disaster threatening many lives cannot be completely eliminated.

Waste material from nuclear power plants is also extremely dangerous if not properly handled and stored. For the most dangerous nuclear waste, most countries intend to build huge storage facilities deep underground ('geological repositories') [62], in which the waste is stored indefinitely, with the repository eventually back-filled with absorbant clay and abandoned. Less hazardous waste can be stored at shallower depths or on the surface; storage plans can differ significantly by country, as can waste classification schemes.

In the broadest terms, waste can be classified as High Level Waste (HLW) or Low Level Waste (LLW). The former category includes hazardous isotopes extracted during reprocessing of spent fuel, as well as the spent fuel rods themselves [63]. Initially, these are metallic tubes containing pellets of uranium oxide fuel; before use in a reactor they are only mildly radioactive. During reactor operation a controlled fission chain reaction occurs in the fissile ^{235}U isotopes, producing the unstable fission products ^{90}Sr and ^{137}Cs. These isotopes have half-lives of around 30 years and so are extremely hazardous. After removal from a reactor, fuel rods must be stored underwater for several years, for both cooling and radiation shielding.

LLW, however, is harder to define; one definition from U.S. Public Law is is "*radioactive waste not classified as high-level radioactive waste, transuranic waste, spent nuclear fuel, or byproduct material*" [64]. Examples include contaminated construction materials (e.g. metal piping), biological waste from medical applications, and miscellaneous general waste. It does not require water storage or deep geological disposal. It is often stored in steel, concrete-filled waste drums and disposed of at surface level or shallow underground sites (see figure 4.1).

Figure 4.1: Nuclear waste drums inside a concrete vault at the ANDRA CSA waste disposal facility in Aube, France. These drums contain short-lived low- and intermediate-level waste. The vault is back-filled with concrete as new layers of drums are added; when full it is sealed for a minimum of 300 years.

4.1.2 Non-destructive assay

Although the contents of all waste drums should be carefully documented, such records can be lost, or may be inaccurate. Additionally, legacy waste drums from the early history of the nuclear industry may have no extant documentation of their contents. It is therefore important to develop techniques to assay waste drum contents without opening or destroying the drum, so as to eliminate the risks of personnel radiation exposure and release of hazardous material into the environment. Such methods are collectively described as Non-Destructive Assay (NDA) techniques. For example, X-ray imaging (both radiography and tomography) is often applied to waste drums [65]. This type of NDA specifically is 'active', as additional energy is being artificially introduced into the system. Other NDA techniques are 'passive': measurements are made external to the drum and without using additional generated particles. Gamma spectroscopy of a waste drum is a widely used passive NDA technique; the gamma spectrum can be used to identify and quantify specific radionuclides. In general passive techniques are cheaper and less hazardous than active techniques.

CHANCE [66] was an EU-H2020 funded project to investigate several novel techniques for passive NDA of waste drums. The three methods tested were calorimetry, in which the presence of radionuclides is determined by monitoring the decay heat produced by the drum [67]; cavity ring-down spectroscopy, which characterises drum contents from outgassing; and Muon Scattering Tomography (MST). MST has already been shown to be effective at imaging the interiors of waste drums after several days' exposure [68], allowing the stored objects inside to be viewed. This chapter describes a study into the use of MST data and machine learning techniques for a more detailed assay of drum contents, specifically attempting to identify the material compositions of stored objects.

4.2 Binned clustering algorithm

The PoCA MST algorithm (see section 3.3.2) accounts only for the mean muon scattering angles, and does not take into account the number density of high-angle scatterings. This is useful information to include in an MST algorithm, as more muon scatterings will occur in high-Z material (when normalised by the muon momentum, discussed below) and hence the number density of PoCA scattering vertices is indicative of Z. An algorithm that builds on PoCA by incorporating this additional information is the binned clustering algorithm [50], originally developed for cargo scanning applications and later applied to waste drum imaging.

Specifically, the binned clustering algorithm makes use of metric values m_{ij}, where for two scattering vertices i and j

$$m_{ij} = \frac{|\mathbf{x}_i - \mathbf{x}_j|}{(\theta_i p_i)(\theta_j p_j)} \tag{4.1}$$

where $\mathbf{x}_i$, θ_i and p_i are respectively the position, scattering angle and momentum of scattering vertex i. A higher density of scattering vertices will therefore lead to lower metric values (see figure 4.2). The normalisation by muon momentum is to account for the larger scattering angles of lower energy muons, which could otherwise be misread as indicating the presence of

high-Z materials. The muon momentum is difficult to measure accurately and this should be accounted for when designing an MST simulation; often including even a heavily smeared momentum information can significantly improve results. The momentum approach used in the simulations performed here is discussed in more detail in sections 4.3.2 and 4.4.

The operation of the algorithm with muon detector hit data as input is as follows:

1. Divide the volume of interest into sub-volumes, the size of which will depend on the application (e.g. large volumes for cargo scanning, small voxels for drum imaging). For each sub-volume V prepare empty lists V_θ and $V_\mathbf{x}$.

2. Identify incoming and outgoing tracks for each muon i as straight line fits to triplets of hits in the top and bottom detector modules respectively.

3. Perform a combined fit to obtain the final track slopes $\mathbf{k}_{i,\text{in}}$ and $\mathbf{k}_{i,\text{out}}$ and 'scattering vertex' (the PoCA) position $\mathbf{x}_i$. For each muon, this fit is achieved by using MINUIT [69] to minimise the energy function $E = E_x + E_y$, where

$$E_x = \sum_{j=1}^{3} \frac{(h_{x,j} - (x + k_{x,\text{in}} \cdot t))^2}{\sigma_{h_{x,j}}} + \sum_{j=4}^{6} \frac{(h_{x,j} - (x + k_{x,\text{out}} \cdot t))^2}{\sigma_{h_{x,j}}} \tag{4.2}$$

and E_y is defined similarly. Here $h_{x,j}$ are the measured x hit positions in each layer, $t \equiv h_{z,j} - z$, and $\sigma_{h_{x,j}}$ is the statistical uncertainty on the hit positions.

4. Calculate the scattering angle θ_i from $\mathbf{k}_{i,\text{in}}$ and $\mathbf{k}_{i,\text{out}}$.

5. Append θ_i to V_θ and $\mathbf{x}_i$ to $V_\mathbf{x}$, where V is the sub-volume that contains the scattering vertex.

6. Repeat for all muons.

7. For all V, sort the lists V_θ and $V_\mathbf{x}$ in descending order of scattering angle.

8. Choose a value for the parameter n. Discard sub-volumes containing $< n$ vertices.

9. For each remaining sub-volume, discard all but the n tracks with the largest scattering angles.

10. For each of the $n!/2(n-2)!$ pairs of vertices in each V, calculate the metric value m_{ij} using equation 4.2. Calculate the median of the distribution of $\log m_{ij}$ values; this is the discriminator value for V.

For a cargo scanning application with strict limitations on exposure time, large sub-volumes are used, and a decision (e.g. safe or unsafe) is made based on a cut on the minimum discriminator value. The algorithm was first applied to more granular sub-volumes, allowing full 3D imaging, in [68].

The algorithm's parameter n needs to be treated carefully. A larger n will, in general, improve the contrast between materials, as more tracks and hence more scattering information
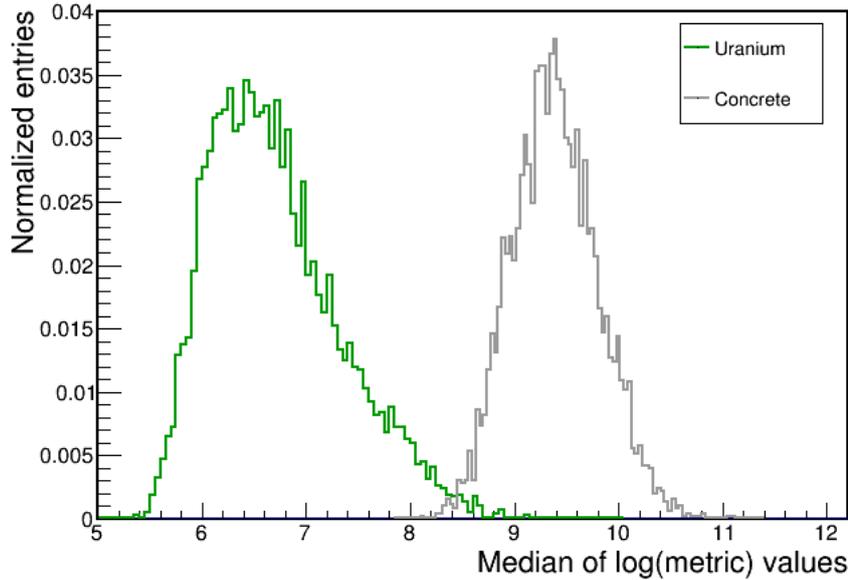
Figure 4.2: Distributions of the binned clustering algorithm's discriminator value for muons passing through 20 cm cubes of uranium and concrete. Muon scatterings occur more often in the high-Z uranium, leading to a higher number density of scattering vertices and hence smaller metric values (see equation 4.2).

is included. But as volumes containing $< n$ vertices are discarded, information from those regions is lost. For a cargo scanning application, in which only a single sub volume is required to make a decision, n can simply be set to an estimate of the number of tracks expected in a sub-volume of that size, for that exposure time, as in [50]. For an imaging application, in which the sub-volumes are cm-scale 'voxels', the choice of n will affect the image, as shown in figure 4.3. A larger n will lead to many image voxels being discarded (coloured white); these voxels generally correspond to lower-Z materials (air and concrete) in the volume of interest where not enough scatters occurred to record n vertices. This leads to less clear images (especially if n is large enough that the borders of the drum itself become indistinct), but the removal of these largely low-Z regions becomes exploitable if the stored objects in the drum are of primary interest.

Figure 4.4 demonstrates that, whilst the algorithm is efficient at imaging metal blocks in a concrete-filled drum, determining the *material* of the blocks 'by eye' is not trivial. For example, the uranium and lead blocks both appear as clear dark regions, but distinguishing the higher-Z material is difficult, and would be more so if the other material was not available for direct comparison. For this reason, it was decided to attempt to use machine learning techniques to build on the binned clustering algorithm by first isolating the voxels corresponding to stored objects and then determining their materials.

39
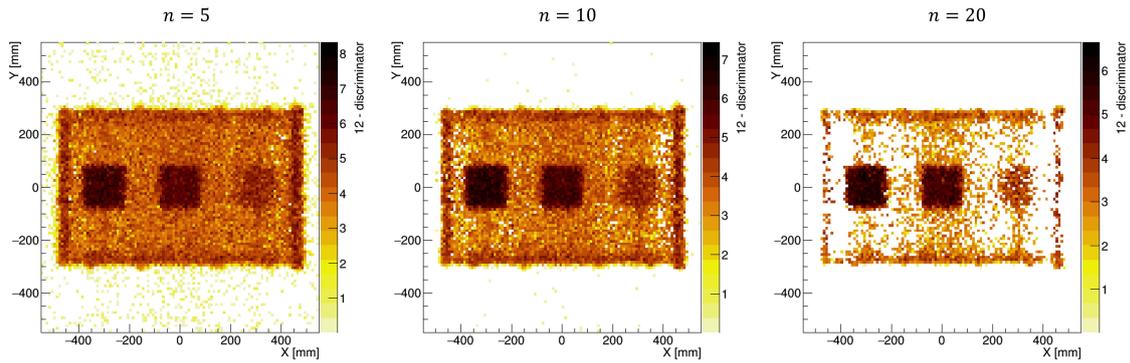
Figure 4.3: Binned clustering algorithm MST images of a drum containing three 15 cm side length cubes of uranium, lead and iron, 10 days' exposure time, for different settings for n (the number of tracks in each voxel). Each 2D image is the $z = 0$ slice through the full 3D output image. The image value of each voxel here is the algorithm's discriminator value (which is lower for higher Z material) subtracted from 12, for visibility.
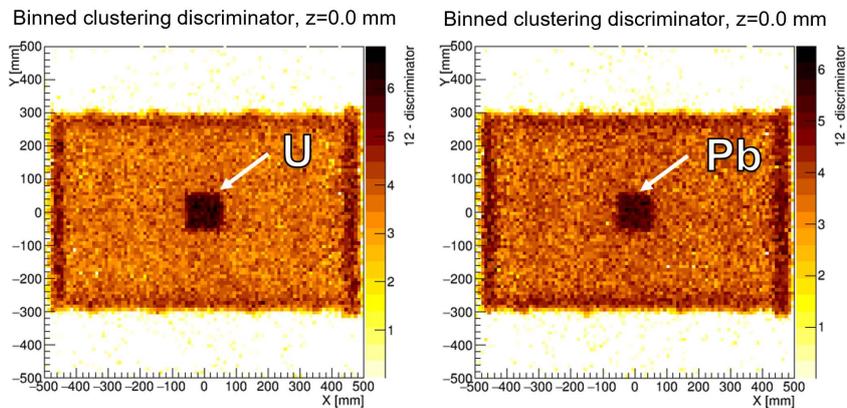
Figure 4.4: Binned clustering algorithm MST images of a drum containing 10 cm cubes of uranium (left) and lead (right). 10 days' exposure time, $n = 5$. In both cases the metal block is clearly visible but determining its material is not trivial.

4.3 MST simulation setup

4.3.1 Software

The study was performed using a set of Monte Carlo simulations (see section 2.5). Simulations of this type are widely used in MST studies; they are extremely useful for testing detector configurations, algorithms and other methods that would be difficult or impractical to perform experimentally. In this case, as the development of the method required MST of a large number of different waste drum configurations with several days' exposure time each, an experimental study was not possible. Simulations also allow for considering cases with large blocks of hazardous material, such as uranium, that would not be obtainable otherwise.

The simulations made use of the CRESTA software [70], which is built on the GEANT4 [13] particle transport toolkit. CRESTA includes an MST detector system, with a concrete-filled waste drum into which objects can be placed. The initial particles were generated using

40

the CRY cosmic ray shower library [28]. The altitude was set to sea level, and the GEANT4 physics used the 'QGSP_BERT' settings as is standard for many HEP applications; this uses the quark gluon string model for particle energies > 20 GeV, and the Bertini cascade model for energies < 10 GeV [71].

4.3.2 Detector specification

As part of the CHANCE project, an MST imaging system was constructed at the University of Bristol. Each detector module consisted of a combination of detector technologies: gaseous drift chambers (see section 2.4) and Resistive Plate Chambers (RPCs, a related form of gaseous detector), along with planes of polystyrene scintillator tiles used as a triggering system. A modelled version of this system had been incorporated into CRESTA at an early stage of the project. For the majority of the time in which the simulations described in this chapter were being conducted, the detector system itself was undergoing commissioning, during which time its design was in flux to a degree with both large-scale (e.g. number and positions of RPC planes) and small-scale (e.g. exact spatial alignments) features subject to change. For this reason, it was decided to use the original detector model (figure 4.5) throughout the simulation studies. This approach meant that new simulations would not need to be continually performed for each detector adjustment, at the cost of the simulation results being less directly applicable to results from the experimental system.
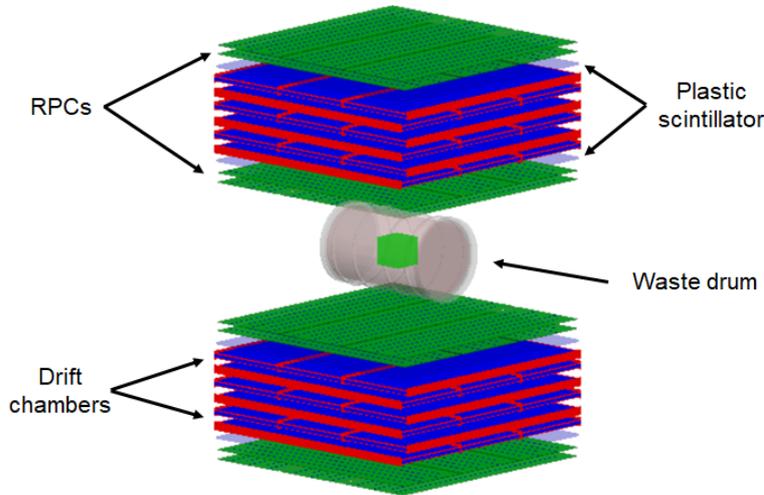
Figure 4.5: The simulated MST detector system, consisting of drift chambers, RPCs and scintillator panels, as viewed in GEANT4 interactive mode. The two detector modules are approximately 200 cm × 200 cm square, the gap between them is ~ 1 m and the waste drum is ~ 1 m in length with radius ~ 30 cm.

The simulated system used two identical detector modules. Each module has two RPC planes consisting of three aligned panels; each panel is $178 \times 56 \times 0.6$ cm, has a spatial resolution of 0.04 cm, and is assumed to be perfectly efficient. The spacing between the panels is 6.5 cm and so the total dimensions of the layers are 178×181 cm. The drift chambers are $180 \times 61.2 \times 6.3$ cm, and primarily filled with argon gas. The modules were placed a distance of 1 m apart, in the gap between them a waste drum model was placed;

the drum was made of steel, filled with homogeneous concrete and approximately 1 m in length. The detailed dimensions of the drum and its features is shown in figure 4.6. Concrete can vary significantly in composition; the material used in the simulation is the standard 'G4_CONCRETE' material, a mixture of silicate minerals with density 2.30 g cm^{-3} [13]. The drum steel is set to be 98% iron, 2% carbon, with density 8.05 g cm^{-3}.
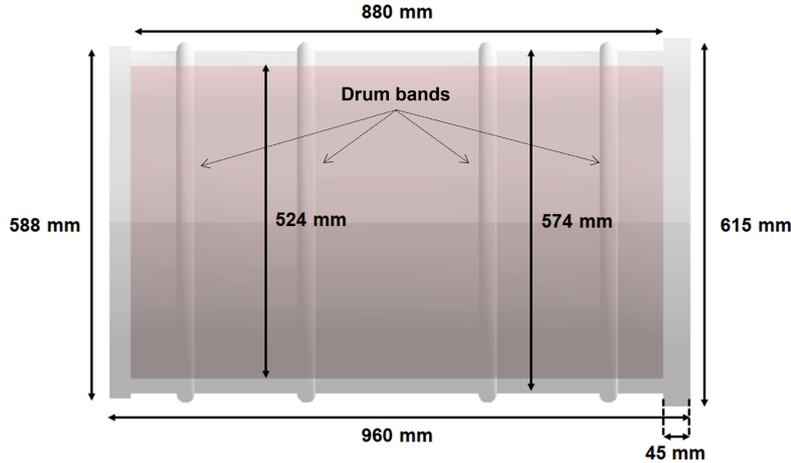
Figure 4.6: The simulated nuclear waste drum used in the CRESTA simulations, with its dimensions and features explicated. The drum is made of steel and filled with concrete; objects can be placed inside to test waste assay and imaging techniques.

The muon momentum is often included in MST algorithms (see equation 4.2); low momentum muons can scatter strongly even in low-Z materials and be mistaken for indicating high-Z material. It thus needs to be accounted for at the simulation phase. In an experimental version of the system it would not be possible to directly measure muon momentum via time-of-flight; this would require alternative and more expensive technologies (although estimating the momentum using gaseous Cherenkov detectors may be a possibility in the future [72]).

An alternative method is to consider the scatterings the muons undergo when passing through a known material, either within an additional module underneath the lower detector module or, ideally, the layers of the lower detector itself. When the muon passes through these layers, it will undergo further scatterings, but as the Z and density of these layers are known the scattering angle distribution will depend only on the muon momentum. Larger intra-detector scattering angles will correspond to a larger residual when fitting a track to the detector hits; thus the residual will give some information about the momentum of the muon.

In [73], it was shown that the residual of a three-layer detector track fit can give a momentum estimate with a precision of $\sim 50\%$. For the simulated system, this is accounted for via a smearing factor applied to the Monte Carlo truth momentum; this factor is drawn from a Gaussian with width 50% of the true momentum. The effect of this method on both imaging and material discrimination, as compared to using either the true momentum or no momentum information, is explored in section 4.4.5.

4.4 Multivariate analysis: separating object voxels

4.4.1 Machine learning and MST

'Machine learning' is a family of computational techniques that use a form of artificial intelligence for data analysis, characterised by algorithms that 'learn' about a problem from some 'training' dataset and can then make decisions or predictions when applied to a fresh 'test' dataset. Applying machine techniques to MST has the potential to enhance the information available from detector data beyond what is possible from algorithms such as PoCA, via detection and incorporation of deeper patterns and relations. As discussed in section 4.2, MST algorithms can be ineffective at identifying the materials of objects in waste drums. Machine learning techniques can be applied to this problem.

One such application was performed in [74], a simulation study. The problem under consideration was making a rapid decision for distinguishing drums containing uranium cubes from drums containing lead or tungsten cubes. The machine learning element was the use of a trained Multivariate Analysis (MVA) classifier: a machine trained to discriminate events between two or more classes based on some characteristic set of variables for each event. Usually one class is designated as 'signal' and the others as 'background'. Each detected muon was used as an event, with the following variables obtained from the fitted tracks: x- and y-projected scattering angles, χ^2 of each track fit, muon momentum, and the offsets between the incoming (outgoing) track extrapolated to the highest (lowest) layer in the other detector module and the measured hit position in that layer. Their trained classifier was able to discriminate very accurately between drums containing central $10 \times 10 \times 10$ cm^3 cubes of each material after a few hours' exposure.

It was decided to reproduce some key results from [74], as a validation exercise and also to build further by comparing different machine learning methods applied to the same problem. The MVA classifiers used were built, trained and operated using the powerful TMVA machine learning platform [75]. TMVA is integrated into ROOT [76], a CERN-developed data analysis framework. TMVA can act on a 'TTree' - a ROOT data structure representing a columnar dataset - with each event (e.g. a muon track) and its variables (e.g. scattering angle) accessible for training the classifier. By default, the training data supplied to a MVA classifier built with TMVA is split randomly into two equal sets, with one dataset used for training, and the remaining data used as a test dataset to assess the classifier's performance, in particular to check for overtraining (see section 4.4.6). TMVA also allows the user to test many machine learning methods - neural networks, boosted decision trees, and so on - simultaneously, by applying each method to the training data and quantifying the performance of each.

The method used in [74] was that of Fisher's linear discriminant [77]. This method works by treating each event as a point in an n-dimensional phase space, where n is the number of input variables, and then attempting to find the optimum hyperplane to divide the signal and background classes; with the plane calculated in the training stage the classifier can then be applied to a new event and classify it based on its position in the phase space. Using the Fisher method, a training set based on 5 days' exposure of muon events from drums containing

10 cm side length cubes of lead and uranium, and applying the classifier to two hour drum exposures produced the results shown in figure 4.7. The left plot shows the classifier output distributions for the lead and uranium datasets: the discrimination between the classes is clearly minimal after 5 days' exposure. However, as was done in [74], if the mean of the classifier output distribution is calculated for a sub-sample of the dataset then the mean is an effective discriminator variable, as shown in the centre plot. The separation of the classes by this method was then quantified via a ROC analysis (see section 4.4.2), as shown in the right plot. The results were in close agreement with those of [74].
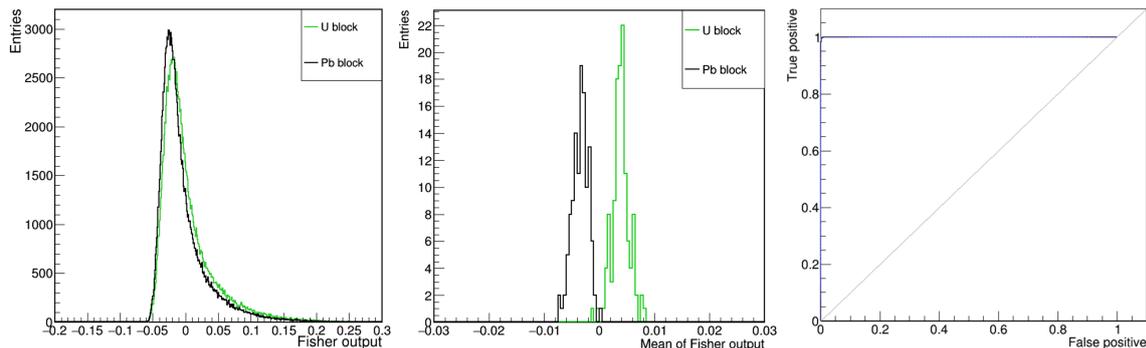
Figure 4.7: Applying Fisher's linear discriminant method for classifying waste drums containing large cubes of lead and uranium, as validation of results in [74]. Left to right: MVA output distributions for the 5 day exposure training datasets, mean of the MVA output for 2 hour exposure samples of muon events, showing good discrimination between the distributions; ROC curve for discriminating the classes from the MVA output mean, with AUC ∼ 1.

4.4.2 ROC analysis

A ROC analysis is a useful and intuitive quantitative measure of the discriminating power of a classifier, involving a comparison of the classifier's rates of true and false positives. For a diagnostic test that aims to classify events - for example, a medical test for a particular disease in patients - the rates of both false positives (type 1 errors) and false negatives (type 2 errors) need to be taken into account. If a test is with maximum sensitivity so as to minimise false negatives, it will likely have a correspondingly high rate of false positives. For the medical example, the former leads to unnecessary patient distress and costly further interventions. For a waste assay application the equivalent could be a high rate of false positive uranium detection necessitating improper, and likely expensive and hazardous, opening or treatment of waste packages.

Most diagnostic tests will involve measuring some continuous variable, such as the concentration of a certain chemical, and defining some cut on the variable for which measurements on one side of the cut are classed as positive and vice versa. A ROC curve is a plot of the true positive rate ('sensitivity', or 'signal efficiency') versus the true negative rate ('specificity', or 'background rejection') for different choices of the cut value. If the cut is chosen so as to miss as few positives as possible, the signal efficiency will be large but the background rejection

will be low, i.e. a high rate of both true and false positives. Conversely a very strict test will lead to a high rate of both true and false negatives. A balance must be achieved for a test to have maximum utility; the optimum cut can be calculated from the ROC curve (see below).

Additionally, the area under a ROC curve (Area Under Curve, AUC) provides a fast, approximate measure of the test's discriminating power. If the test performs no better than randomly assigning events to one class or the other, the background rejection and signal efficiency will be equal no matter the choice of cut. The ROC curve will then be a straight line from $(0, 1)$ to $(1, 0)$ with AUC $= 0.5$. In general the closer the AUC is to unity the better the discriminating power; a classifier with AUC $= 1$ would classify events with perfect accuracy (this would also be true if AUC $= 0$; any AUC < 0.5 indicates that the classifier has discriminating power, but that the signal and background classes are reversed). The AUC is equivalent to the probability that if a pair of random observations, one positive and one negative, are chosen, the positive event is ranked higher than the negative [78]. In practice, AUC is often an optimistic measure of a classifier's performance due to overtraining (see section 4.18).

The optimum cut for selecting the two classes can be determined from the ROC curve. It will broadly correspond to the top right-most point, i.e. the cut that gives the maximum possible combination of signal efficiency and background rejection. During the training phase, TMVA calculates a optimum cut as the point at which the signal efficiency is equal to the background rejection. On the ROC curve, this corresponds to the point with the maximum Youden index [79], defined as signal efficiency $+$ background rejection $- 1$; i.e. the length of the vertical line between the ROC curve and the line between $(0, 1)$ and $(1, 0)$ that connects the curve's ends.

A measure of the effectiveness of applying a machine learning method to the problem of discriminating drums with muon track variables can be therefore found through a ROC analysis of the initial classification output for the two classes, as in figure 4.7 (left). These show only a slight separation of the signal and background output distributions for the Fisher method; it is necessary to use the mean of the distribution (over a sub-sample of tracks, in this case 2 hours' exposure of the full 5 days' exposure) as the discriminator (figure 4.7, middle). That is to say, the method is not effective at distinguishing *tracks* (each event is one muon track) as having passed through uranium or lead blocks, but is effective at distinguishing *drums*, from a large number of tracks.

To determine whether an alternative machine learning method could improve on these results by showing superior ability to distinguish tracks, TMVA was used to apply eight different methods to the same training set, in this case using each method's default hyperparameters for simplicity; details of each method and the hyperparameters used can be found in Appendix B. A ROC curve was calculated for each method to quantify relative performance. The resulting curves are shown in figure 4.8 (left), and indicate that the variation between methods is small but that linear discrimination does not perform quite as well as a Multi-Layer Perceptron (MLP) method. An MLP is a type of neural network in which the neurons are organised in layers and direct connections can only be inter-layer. The output distributions for this

method applied to the same training data are shown in figure 4.8 (right); there is a slightly higher degree of separation than for the Fisher method.
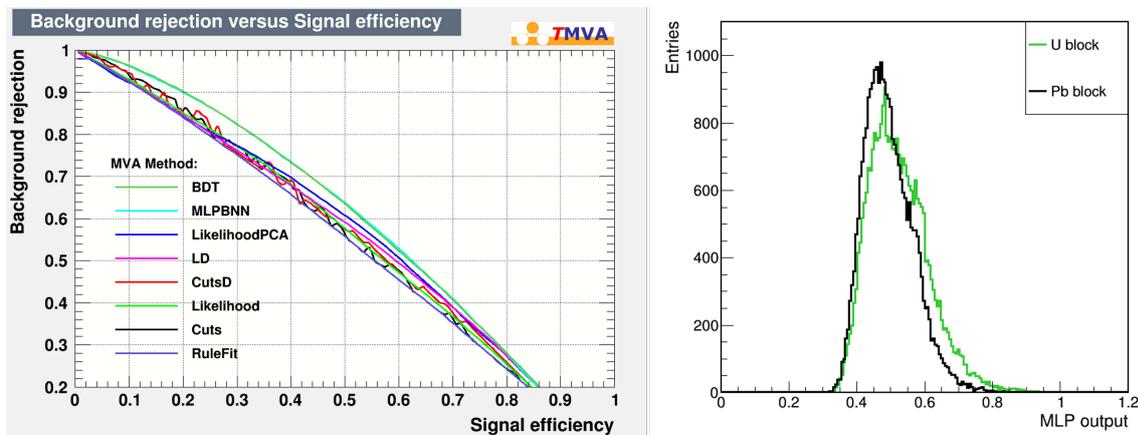

Figure 4.8: Testing possible MVA methods for distinguishing drums containing lead or uranium cubes from muon track data. The left plot shows ROC curves for several MVA methods, including boosted decision trees, multi-layer perceptron, principal components analysis, and linear discrimination. The curves indicate that the MLP has the best performance, with significantly better discriminating power than the linear discriminator method and slightly outperforming the boosted decision tree method. The right plot shows output distributions for 5-day exposure samples from each class. Slightly better separation is achieved than for linear discrimination (figure 4.7, left). More details of the methods tested here can be found in Appendix B.

4.4.3 MVA classification of voxels

The goal of the methods in section 4.4.1 was to distinguish between drums containing large cubes of lead (or tungsten) and uranium, not to give detailed information about the contents of the drums e.g. size, location, number or material of the blocks. This can be thought of as being a 'global' approach to the problem. The remainder of this chapter is focused on the corresponding 'local' approach: using MVA classification to extract granular material information about the drum's contents. The goal is to identify, image and assign a material to each stored object in the drum. This gives much more information at the cost of a longer exposure time, several days rather than hours.

The key transition to a local approach is to change the definition of the 'events', each of which has its associated set of variables and MVA output value when it is input to the classifier. Rather than using muon tracks as events, now each voxel in the voxelised volume of interest comprises a single event. This requires a set of variables to be associated with each voxel. As the 'track' variable set had already shown limited material discriminating power, it was decided to instead use a set of variables based on the binned clustering algorithm (section 4.2).

The algorithm by default, after calculating a voxel's distribution of $\log m_{ij}$ values, outputs only the median of the distribution which becomes the image value for that voxel. Thus information about the metric values and the shape of their distribution is lost, and it was therefore desirable that the variable set used would encode this information. The variable set

chosen were the bin counts of a normalised histogram of $\log m_{ij}$ values. Figure 4.9 shows the distributions of $\log m_{ij}$ values within a sample uranium and concrete voxel (note the lower metric values in the higher-Z material), as projected onto the binning structure used as the final choice to define the MVA variables. As expected the median for uranium is lower than for concrete, and the distributions show differences in width and shape as well.
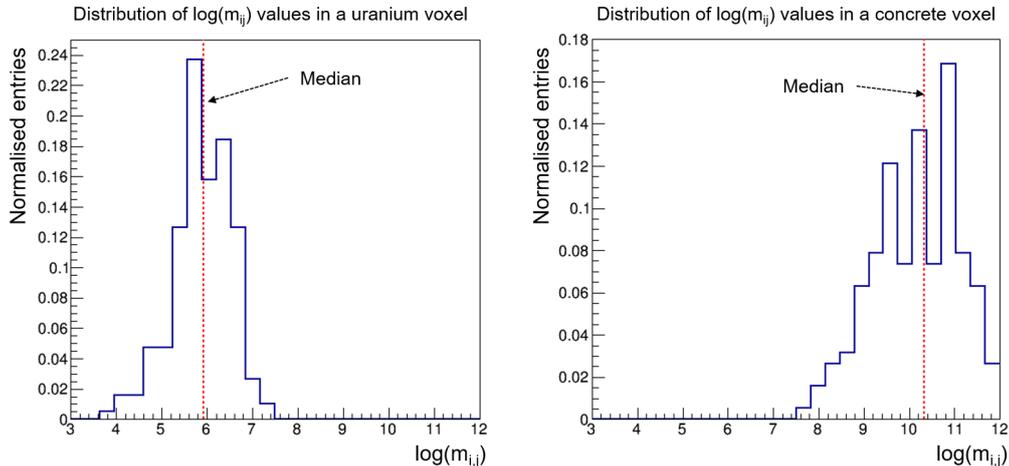
Figure 4.9: Distributions of $\log m_{ij}$ values for a voxel containing uranium (left) and concrete (right). The binned clustering algorithm uses the median of the distribution only. Using the bin counts as variables for training a MVA classifier recovers more information about the distribution.

The final variable set comprised the 28 normalised bin counts, as well as the median itself, for a total of 29 variables. Figure 4.10 shows a sample of the bin count variables, specifically bins $9 - 12$ of the 28 total,in this case for a uranium-lead binary classifier. It is clear that the variables are distinct between the datasets, indicating that they are suitable to test as input to a MVA classifier.

Figure 4.11 shows the performance of an MVA classifier for discriminating lead and uranium voxels using the bin count variables. The calculated ROC curves for multiple machine learning methods show clearly that these variables have discriminating power, as the area under each curve is > 0.5. The best performing methods in this case are Boosted Decision Trees (BDT), the MLP, and the Support Vector Machine (SVM), each of which have ROC AUCs of ~ 0.8. The SVM is a widely used algorithm based on separating vectors of signal and background events in parameter space with an optimised hyperplane, and can operate with nonlinear separating functions [80]. The MLP and SVM have similar performance to the BDT method, but longer computation times.

The BDT method is therefore the superior choice and was chosen as the primary machine learning method for the MVA classifiers. A BDT is a system of chained nodes ('tree'), each of which represents a cut on one of the variables, tuned so as to split the signal and background classes to the greatest extent. Each event progresses through the tree with its variables and the node cuts determining which branches it is sent to. Finally, it is assigned as being either signal or background depending on the final node it enters. Figure 4.12 shows an example
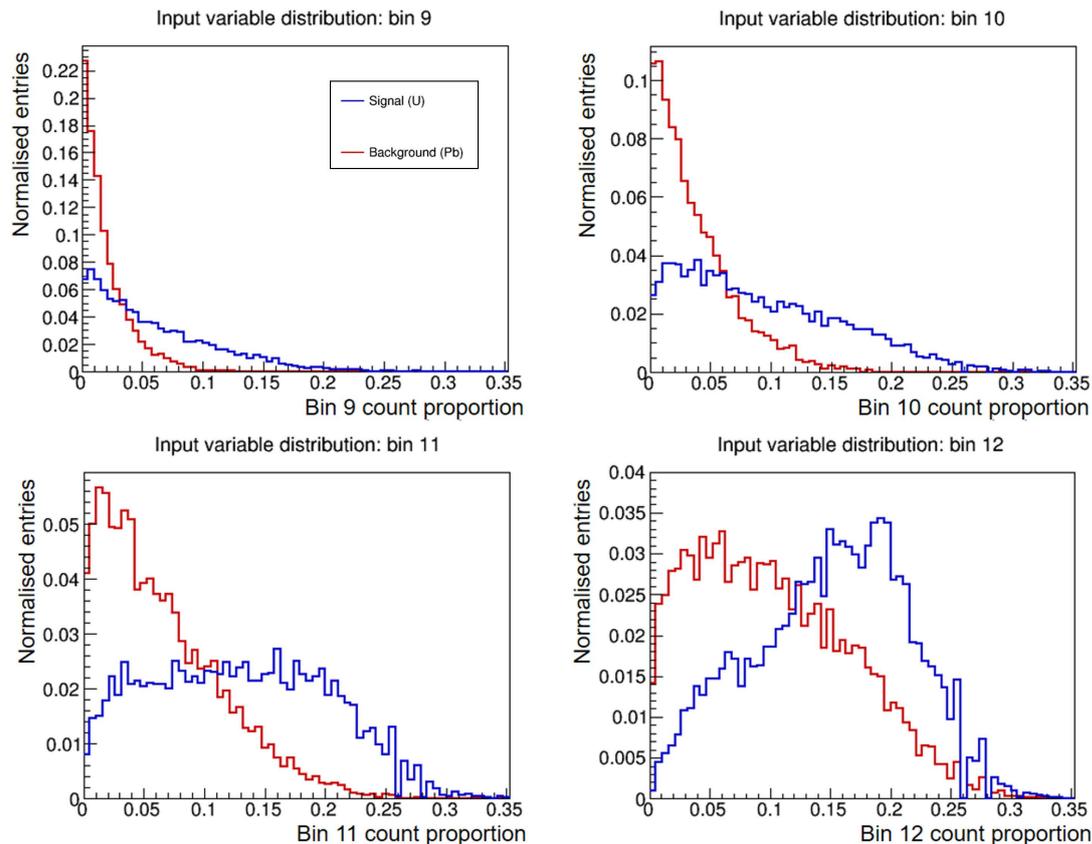
Figure 4.10: Sample of discriminator bin count variables used for MVA classifier training and application. The 'bin' variables are the (normalised) bin counts value of a 28-bin distribution (see figure 4.9) of $\log m_{ij}$ values. Here the 'signal' sample (blue) corresponds to voxels in a 20 cm cube of uranium whilst the 'background' sample (red) is an equivalent cube of lead.

decision tree from a uranium-lead MVA classifier. In practice, using a single tree is very sensitive to statistical fluctuations in the data [80], e.g. a training data fluctuation causing a variable with less discriminating power than another to be chosen for one of the nodes. Instead therefore a *forest* of several hundred decision trees is used. The final output is a weighted average of each of the trees in the forest, with the weights being calculated by a 'boosting' method.

Boosting is used to improve the performance of a decision-tree based classifier by combining the many weak classifiers in the forest into a single more powerful learner. A commonly used method is adaptive boost [81], in which the events that are misclassified by a tree during the training process are given weights α when the dataset is passed to the next tree, via

$$\alpha = \frac{1 - \text{err}}{\text{err}} \tag{4.3}$$

where err is the proportion of misclassified events from the previous tree. It can be shown [80] that this boosting corresponds to the minimisation of a loss function L, which quantifies
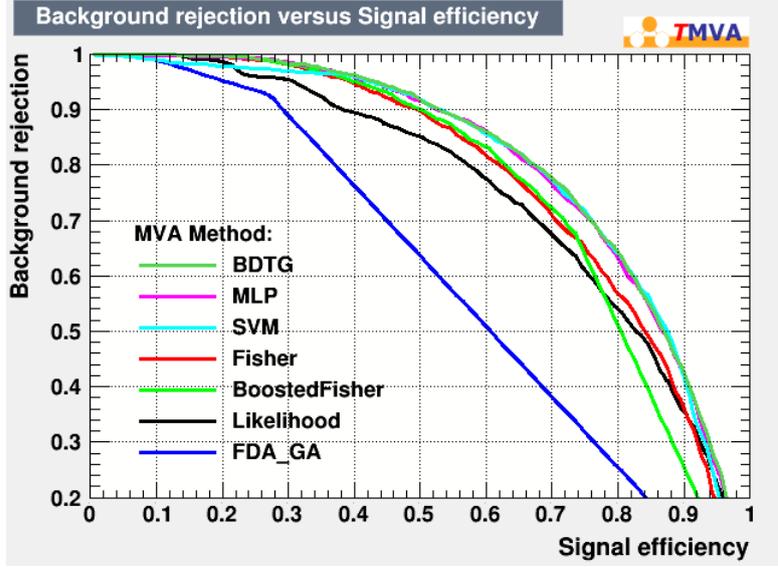
Figure 4.11: TMVA-calculated ROC curves for several machine learning methods, for a binary classifier separating lead and uranium voxels in 20 cm cubes. The exposure time was 10 days, and the algorithm's parameter n (the number of tracks kept for each voxel, see section 4.2) was set to 10. Here, 'SVM'=Support Vector Machine and 'FDA'=Function Discriminant Analysis. The 'G' in 'BDTG' indicates 'gradient-boosting', an optimisation technique. More details of the methods tested here can be found in Appendix B.

the deviation between the model output and the true values and is given by

$$L(F, y) = (F(\mathbf{x}) - y)^2 \tag{4.4}$$

where $F(\mathbf{x})$ is the model output given some event with input variables $\mathbf{x}$ and y is the corresponding true value of the event. The boosting method used in the classifiers presented here, however, is gradient-boosting, which allows for a wider choice of loss functions. The default function used by TMVA in this case is

$$L(F, y) = \log\left(1 + e^{-2F(\mathbf{x})y}\right). \tag{4.5}$$

The corresponding boosting function cannot be found analytically. Instead the gradient of the function is used to perform the minimisation by finding the steepest descent.

4.4.4 MVA classifier training

A useful material classification system will ideally be compatible with a wider range of materials than just uranium or lead/tungsten. The latter approach is suitable for a fast discrimination problem with a global approach to the drum contents, e.g. for cargo assay, in which only a decision about each drum is required. For a local approach, a balance must be struck when choosing the number and type of materials to include as possible categories. If the classifier is a single non-binary classifier (trained on > 2 classes, and optimised for distinguishing one signal class from several background classes), having too many possible

49
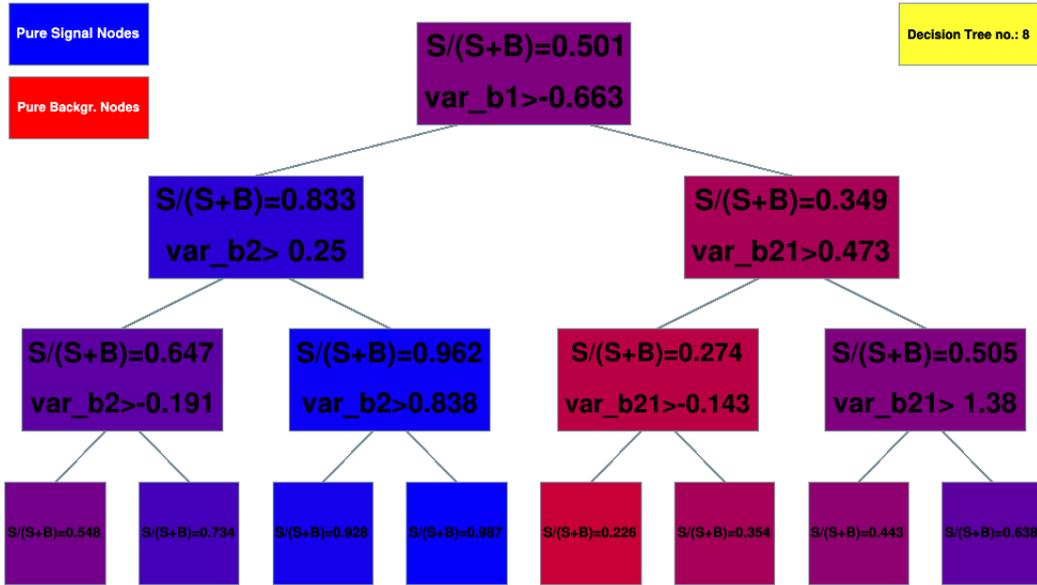
Figure 4.12: Schematic of a decision tree from training a uranium-lead binary MVA voxel classifier. The 'var_b' variables are examples of the discriminator bin count variables used for training the classifier. At each node in the tree, events are sorted based on cuts applied to their variable sets. This is a single tree from a forest of 800 trees, the final classifier incorporates the results of all the trees.

choices will lead to a very low efficiency as materials with similar Z will 'overlap' significantly. If instead a chained series of binary and non-binary classifiers was used (see below), then the required computation time would become impractical with a large number of materials, as each event would need to be input to dozens or more different classifiers.

For these reasons, a conservative choice of four material categories was chosen: concrete, iron, lead and uranium. This set in particular was chosen for three reasons:

1. The simulated drum is made of steel and filled with concrete, using these materials leads naturally to techniques for isolating the outer drum and concrete matrix and removing them.

2. The positions of these materials in the periodic table, with Z values of 11, 26, 82 and 92 for concrete, iron, lead and uranium respectively, covers low-, medium- and high-Z i.e. a broad swathe of existing elements (note 'concrete' is a mixture of silicate minerals with an overall equivalent $Z \sim 11$, see section 4.3.2).

3. Distinguishing these materials involves three separate classification problems of interest: separating metal objects from a surrounding concrete matrix, separating light and heavy metals, and separating two materials of similar Z, with one being benign and one threatening.

With these materials chosen, one approach could be to train four non-binary classifiers, each with one signal class and three background classes, and attempt to discriminate each material from the other three. However, the increasing-Z aspect of these materials lends

itself naturally to an alternative method: a series of chained classifiers, into which each voxel (with its associated metric bin count variables) could be input. Specifically, the chain would consist of first a non-binary classifier with a concrete signal class and iron, lead and uranium background classes, used to separate out the 'concrete' voxels; then an additional non-binary classifier with iron as the signal class and uranium/lead background classes, and finally a binary uranium/lead classifier.

With this structure decided upon, and the MVA classifier machine learning method selected as BDTs, there are still several parameters that need to be studied to understand their effects on the classification efficiency. Most notable are the choice of training dataset, the exposure time, and the choice of the binned clustering algorithm's n parameter (see section 4.2). The training dataset is an example dataset of the signal and each background class used to build the MVA classifier. For the results described in section 4.4.1, the training dataset was muon track information for muons constrained to have passed through a 20 cm cube of lead or uranium, in the centre of a concrete-filled waste drum. This approach was re-implemented for the local voxel-wise material classification problem, using simulation data for iron and concrete cubes in addition to the lead and uranium, and passing to the classifier only the image voxels that fell within the limits of the block dimensions.

However, an alternative was also tested: a large cylinder of each material, filling most of the drum (see figure 4.13). This geometry represents both a greater sample of voxels and a more 'diverse' sample, in that the voxels are located in positions all throughout the drum. This is as opposed to the 'cube' system in which the voxels are all in the centre of the drum. The two possibilities were compared by training a uranium-lead binary classifier (as the two most similar materials in Z, the hardest classification problem involved), and assessing their performance via a ROC analysis. With $n = 20$, and a 10 day exposure time, the AUC for the 'cube'-trained classifier was 0.811 ± 0.008 and for the 'cylinder'-trained classifier it was 0.809 ± 0.003 (the given errors are the 95% confidence intervals, calculated using the standard error on AUC as determined in [78]). The slightly superior (although not statistically significant) performance for the cube geometry training, as well as the shorter computation time, led to this being chosen as the training model for the MVA classifiers. In retrospect, it is likely that the larger sample of training events when using the cylindrical geometry would have reduced the overtraining (see section 4.4.6) of the classifiers, although acceptable results were obtained nonetheless using the cubic geometry.

Next to be assessed was the classifier performance versus the binned clustering algorithm parameter n. Using the cubic training geometry and a fixed exposure time of 10 days, the uranium-lead classifier was tested for values of n between 4 and 25; the results are shown in figure 4.14. Quantifying performance via ROC AUC shows a clear increase in discriminating power for higher values of n. This can be understood as the larger number of tracks increasing the sample of scattering angles and hence giving more information about the material in each voxel. However, as described in section 4.2, a higher value of n will also remove many voxels from the image that contain $< n$ scattering vertices. A too-large n will begin to attenuate smaller or lower-Z objects from the image entirely; the corresponding voxels will have less
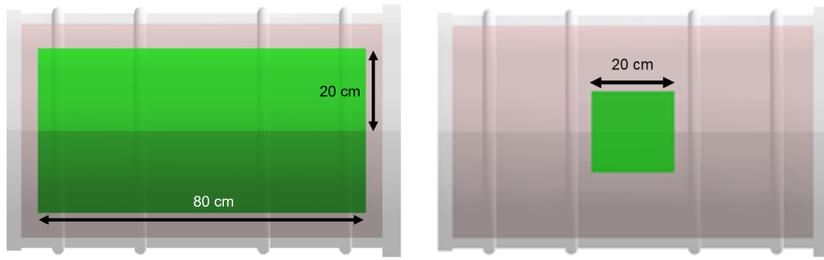
Figure 4.13: Geometries tested as possibilities for the training datasets. Examples shown are the uranium version, however also used were equivalent geometries with lead, iron and concrete objects.

than n tracks and so would be discarded. For the initial study in which the object was to assign a material to each voxel, a lower $n = 5$ value was therefore chosen as a balance of discriminating power and image quality. For later studies in which identifying the materials of stored objects in particular became the priority, $n = 20$ was used, both for the greater discriminating power and to remove some of the concrete matrix at an early stage.

Figure 4.14: The performance of a lead-uranium MVA classifier in terms of ROC AUC (left), as trained on 20 cm cubes of each material with 10 days' exposure time, for different values of the n parameter. Increasing n improves the discrimination between these materials but also attenuates the image, risking removing small iron or lead objects in particular as most of the corresponding voxels would be discarded due to containing less than n tracks. Example classifier training output distributions (right) for $n = 4$ (top) and $n = 25$ (bottom).

More straightforward is the treatment of the exposure time. Intuitively one would expect that a longer exposure would improve material discrimination, as the larger number of muons passing through each voxel will give a larger sample of scattering angles. With a fixed $n = 5$,

the cube-trained U-Pb classifier was applied to several different exposure times, and the performance again quantified via ROC AUC (see figure 4.15). The AUC clearly increases with longer exposure times. In a realistic scenario, in which a waste drum of unknown contents was undergoing assay, exposure times of several days would not be impractical due to the long-term nature of waste storage sites. A fixed value of 10 days' exposure was therefore selected; all subsequent results use this exposure unless stated otherwise.

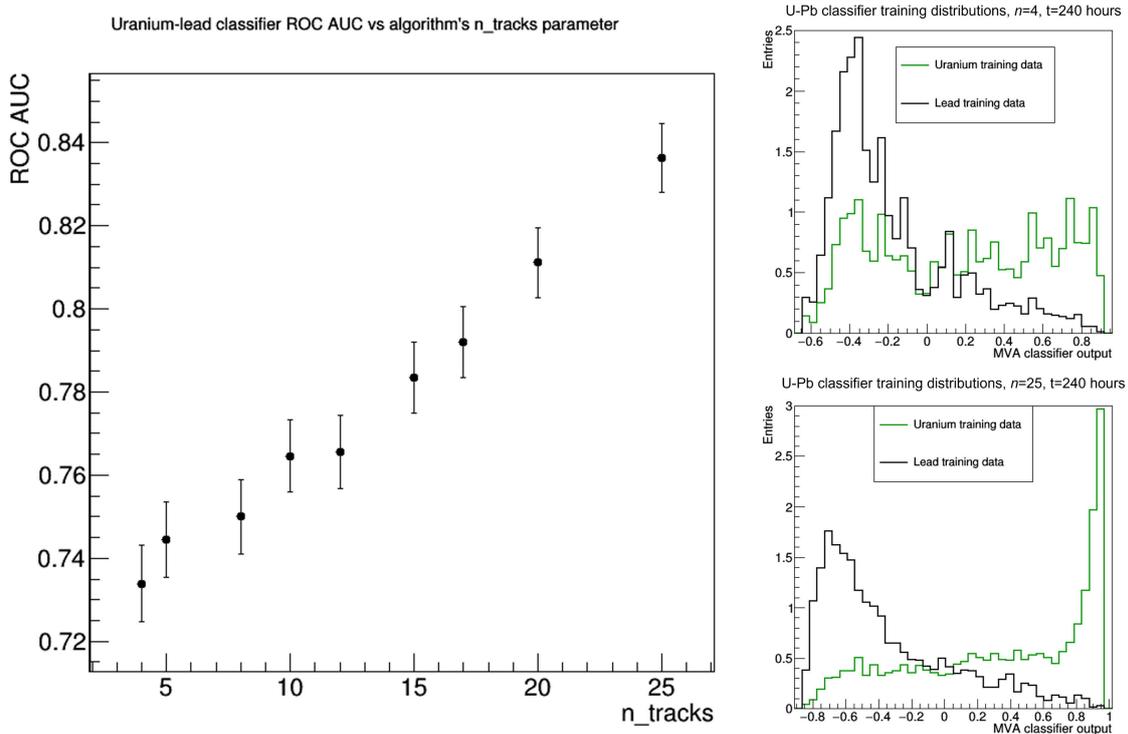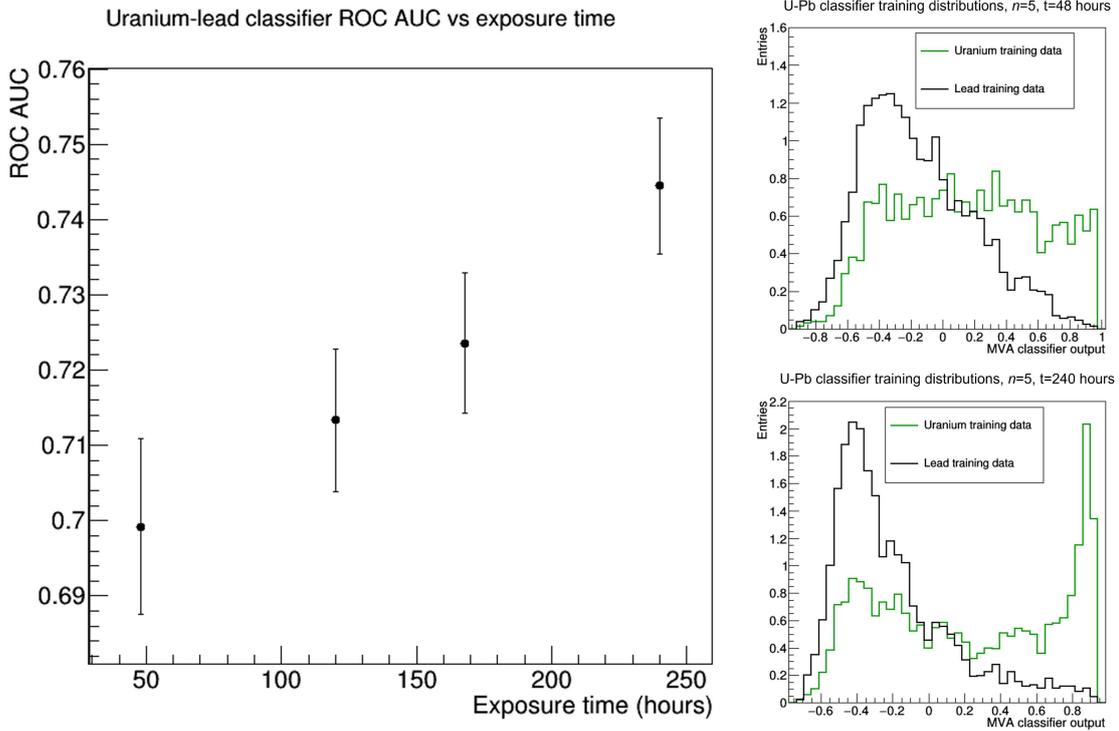Figure 4.15: The performance of a lead-uranium MVA classifier in terms of ROC AUC (left), as trained on 20 cm cubes of each material with $n = 5$, for different exposure times. Increasing the exposure time improves the discrimination between these materials. Example classifier training output distributions (right) for a 2 day exposure (top) and a 10 day exposure (bottom).

4.4.5   Momentum

As discussed in section 4.3.2, the muon momentum should be accounted for when attempting to discriminate materials as low-momentum muons can undergo high-angle scatterings even in low-Z material. The default approach to incorporating muon momentum, used in the results presented thus far, is to apply a 50% Gaussian smear to the Monte Carlo truth momentum, in order to simulate an imprecise momentum measurement estimated from the scatterings of each muon in the lower detector module (see section 4.3.2). To explore the effect of muon momentum in more detail, two other approaches were tested: using the Monte Carlo truth momentum itself, or using a fixed momentum value (set to 3 GeV) for each muon i.e. removing knowledge of momentum entirely. Binned clustering algorithm images of a waste drum containing metal cubes for each of these approaches are shown in figure 4.16. A slight

degradation in the image (particularly in the clarity of the edges of the rightmost iron block) is clear between the true momentum image (left) and the smeared momentum (middle). When momentum information is removed, the image (right) is reduced in quality; the iron block is significantly attenuated. It is concluded that including momentum information improves the algorithm's imaging performance, and that most of the degradation can be rectified even with an imprecise estimate of momentum.
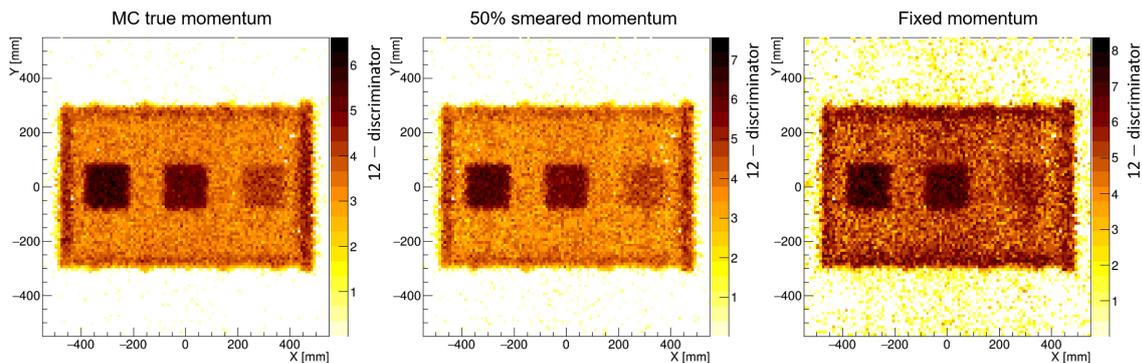
Figure 4.16: Binned clustering algorithm images of a drum containing 15 cm cubes of (left to right) uranium, lead and iron, for three different approaches to the muon momentum: Monte Carlo truth momentum (left), a 50% Gaussian smear applied to the true momentum (middle), no momentum information.

The effect of the different momentum approaches on the MVA classification was also investigated, to assess how important the inclusion of momentum information is to the discriminating power of a classifier. An MVA was trained on simulations of 20 cm cubes of uranium and lead, using the true momentum, smeared momentum or no momentum information. ROC curves were then acquired for each scenario and their AUC calculated for comparison. The results are shown in figure 4.17. As in figure 4.16, using the true momentum ($AUC = 0.852 \pm 0.007$) greatly improves discrimination performance versus using no momentum information ($AUC = 0.631 \pm 0.011$), whilst using even a heavily smeared value recovers much of the lost discriminating power ($AUC = 0.811 \pm 0.008$).

4.4.6 Overtraining

A common problem in classification with machine learning is that of overtraining. This is when a model underperforms when applied to unseen data relative to the test data used to train the model, due to it fitting itself to statistical fluctuations in the training data and treating them as actual characteristic features of the classes. The result is that the classifier's ability to discriminate events is much less than for the training dataset, and hence measures of classifier performance obtained in the training stage will be overly optimistic. This can be adjusted for in general by supplying a larger sample of training data so as to reduce the influence of statistical fluctuations. Additionally, the specific machine learning method used will have its own susceptibility to overtraining and methods for countering it; for example a model based on boosted decision trees will almost always suffer some degree of overtraining but can be
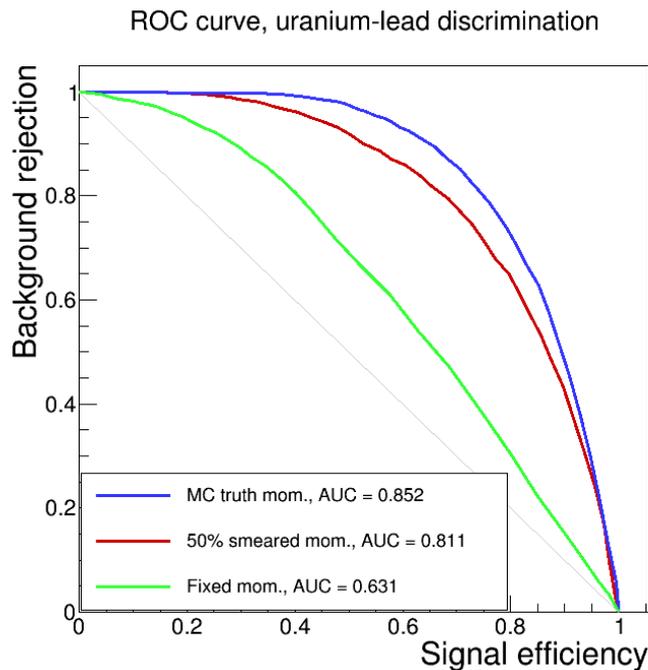
Figure 4.17: ROC curves for a binary uranium-lead MVA classifier, trained on 10-day samples of voxels from 20 cm cubes of each material. Each curve is a different approach to incorporating the muon momentum: using the Monte Carlo truth momentum, using a smeared true momentum, or using a fixed momentum for each muon. The area under each curves demonstrates quantitatively that momentum information improves discriminating power.

improved by removing ('pruning') nodes from the tree that have little discriminating power [80].

TMVA provides a useful overtraining check by splitting the training dataset (randomly) into training and testing sets. The model is trained on the former, then applied to the latter (as if it were an unknown dataset) to assess how well the outputs match; a closer match between the distributions indicates less overtraining. The similarity of the distribution is quantified with a Kolmogorov-Smirnov test. The results of this test for the binary uranium-lead classifier are shown in figure 4.18. It is clear that the distributions of the test and training MVA outputs are a close match visually. The Kolmogorov-Smirnov test values (0.110 for the signal data and 0.114 for the background data) are somewhat low however, indicating some degree of overtraining has taken place. This is likely a result of the model being biased due to the use of the 20 cm cube training geometry, and it is clear in the following results that this has led to the classifiers not performing as well with drum geometries significantly different from the training set. This is adjusted for using an additional calibration step, see section 4.6.3.

4.4.7 Material classification of voxels

Thus three MVA classifiers were designed and trained: a non-binary classifier to distinguish concrete from the metals, a non-binary classifier to distinguish iron from the high-Z metals, and a binary classifier to distinguish lead from uranium. Each classifier returns a single output
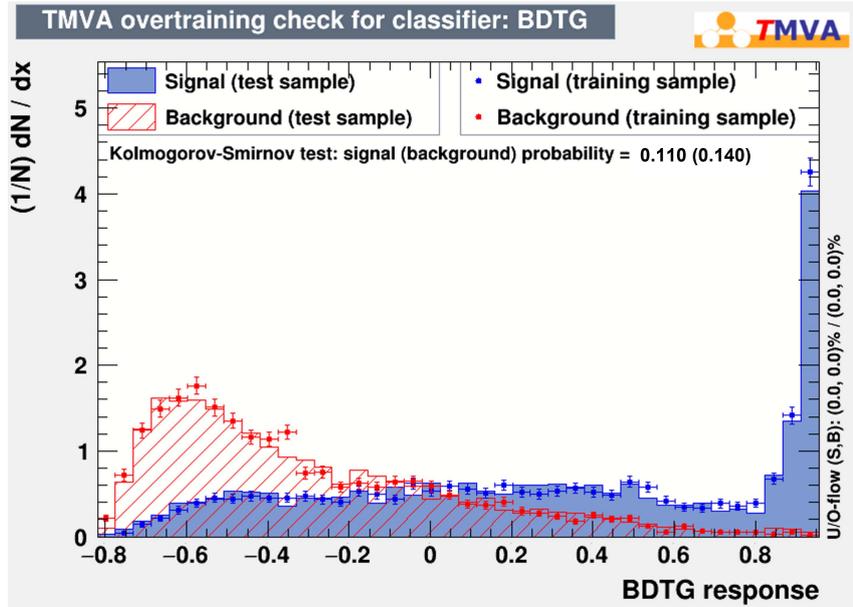
Figure 4.18: TMVA-generated overtraining plot, showing comparison between signal and background classifier outputs for both the training and testing datasets. The distributions match fairly closely, but the Kolmogorov-Smirnov tests indicates a significant degree of overtraining has occurred.

value for an input voxel; it is then classified as being signal- or background-like based on whether this value is above or below an optimum cut. The cut value is calculated in the training stage from a ROC analysis and indicates the point at which the signal efficiency is equal to the background rejection (see section 4.4.2). The chain of classifiers acts as a high-level decision tree (this is distinct from the BDTs used in the classifier itself), in which each voxel is assigned a material based on the final leaf node it is sorted into (see figure 4.19).

This system of classifiers, trained on the 20 cm cube geometries, was then applied to simple test geometries to assess its performance in practice. The input test data was simulated 10-day exposures of waste drums containing metal cubes of various sizes. In this case, n was set to 5; sacrificing some discriminating power (see figure 4.14) but preserving more of the drum image. The goal was to essentially transform a binned clustering algorithm output image (e.g. figure 4.4, left), in which stored objects are clearly visible, into a map of materials in each voxel. Results are shown in figure 4.20. Cubes of 15 cm and 5 cm side length were tested initially. In each case the cubes were positioned along the drum's central axis and aligned with the voxel grid, and one cube of each metal is present. The cubes are separated by several centimetres of concrete. The momentum information used was the true momentum plus a smearing factor, as described in section 4.4.5.

There are several notable features of these results that are immediately apparent. Firstly, the method has in general performed well: the metal cubes are clearly visible against the concrete background, which itself has been reconstructed well; the edges of the steel drum have been identified and classified mostly as iron, and there are clear differences in the voxel material decisions between the metal blocks. The information available is an advance on what

56
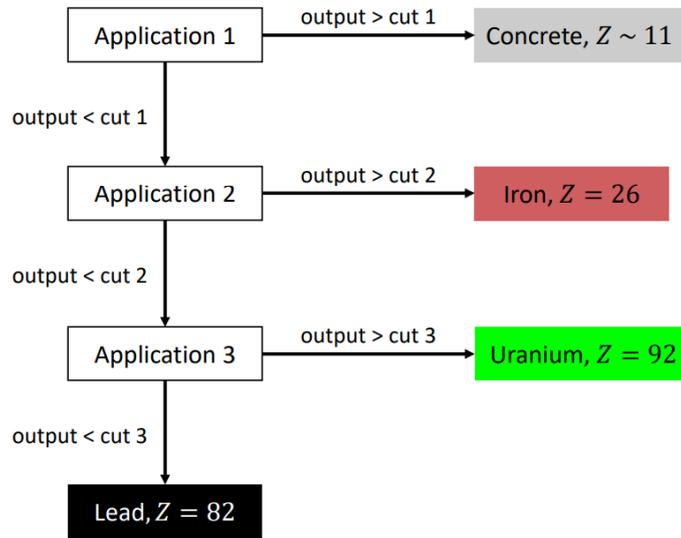
Figure 4.19: The decision tree used to assign a material to each voxel. Application 1 is a non-binary classifier with concrete as the signal class and U/Pb/Fe background classes, application 2 is similar but with Fe signal class and U/Pb background classes, and application 3 is a binary classifier with U signal and Pb background. Each voxel is input into application 1, and moves through the tree based on its classifier output values.

is possible from a binned clustering algorithm output image alone, e.g. comparing figures 4.4 (left) and 4.20 (top right): both are images of a drum containing 15 cm cubes of each metal, but identifying the uranium block is far easier with the MVA classifier system applied.

Indeed the 15 cm cube image (figure 4.20,top) implies that an accurate material decision could be made simply by manually isolating the voxels for each object and determining the most common material. However, the 5 cm example is not so straightforward. Although the objects are visible, the majority of the uranium cube voxels have been assigned as lead; and the other blocks are similarly 'underestimated' in Z. Clearly, the size of the body plays a significant role in the voxel response to the classifiers. The larger cubes are a better match to the training geometry and hence are classified more accurately.

It was therefore decided that additional machinery would need to be implemented to improve the system's results and determine accurate material decisions for objects of different sizes and shapes. Specifically, the goal was to implement a system able to:

1. Filter out the image voxels that correspond to the steel drum and concrete matrix, leaving only those that correspond to stored objects.

2. Determine the number of stored objects in the drum N, and group the voxels accurately into N clusters.

3. Assign a material to each identified cluster and hence determine the materials of all stored objects in the drum.

With the goal now being to classify the materials of objects rather than voxels, the binned clustering algorithm's parameter n was changed from 5 to 20, increasing the discriminating
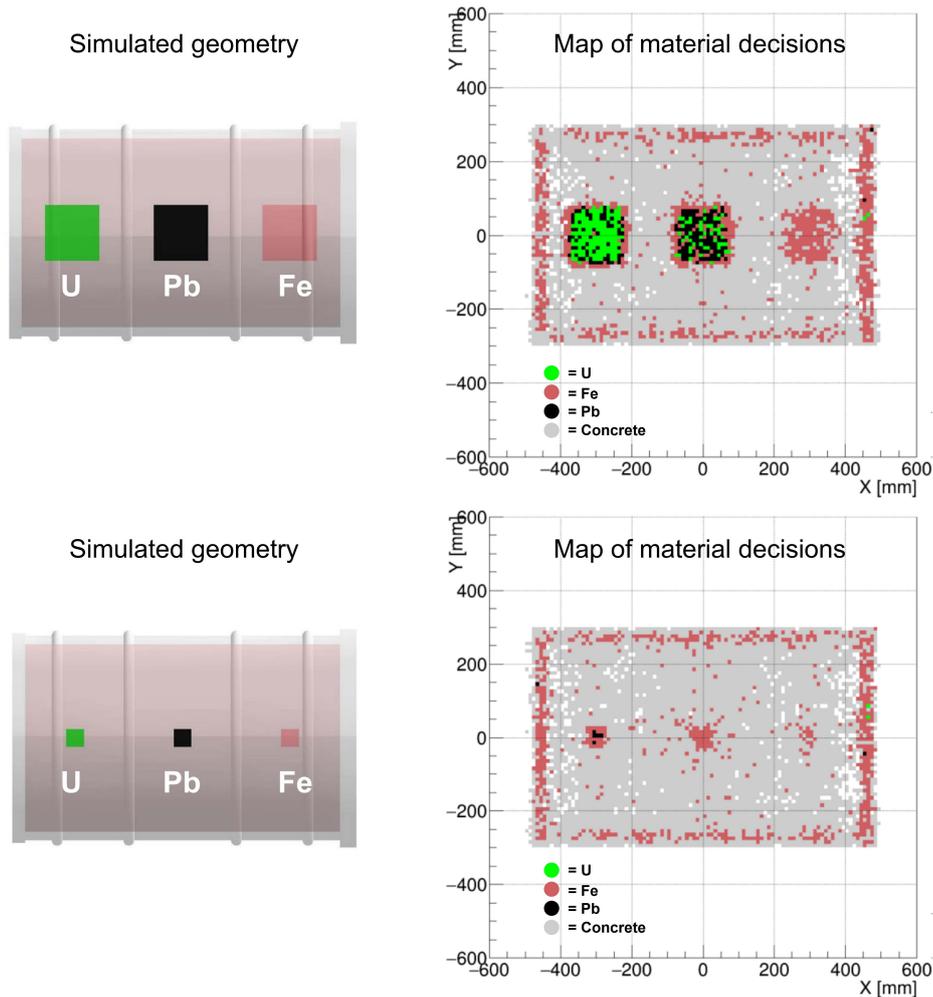
Figure 4.20: The results of applying the the classifier decision tree (figure 4.19) to each of the voxels in binned clustering algorithm output images of drums containing three 15 cm (top) or 5 cm (bottom) cubes of uranium, lead and iron.

power of the classifiers (see figure 4.14). As shown in figure 4.3, a higher n will remove large regions of the image where the voxels contain $< n$ scattering vertices. Almost all of the attenuated voxels will correspond to the concrete matrix, as it is the lowest-Z material present and hence less scattering vertices will be constructed in these voxels. A higher n is therefore doubly advantageous in this framework.

4.4.8   Filtering concrete matrix

To remove the remaining non-object voxels, the first step is to apply a simple cylindrical spatial cut that will remove the outer steel drum voxels. This is possible as the position and thickness of the drum are known; in a realistic scenario this information could be obtained from the details of the storage drum or manually read off a binned clustering output image. Several methods were attempted to remove the remaining concrete voxels. Initially working in the $n = 5$ framework, an algorithm based on locating minima (i.e. the highest-Z voxels,

58

which have the lowest metric values due to a higher density of scattering vertices) and then extrapolating outwards until reaching a boundary was implemented, but could not be made to work reliably and was not effective for non-rectilinear objects. Ultimately, the solution chosen was the simplest: applying the non-binary MVA concrete versus iron/lead/uranium classifier to each image voxel, and removing all voxels classified as concrete. The classifier output distributions and ROC curves for this classifier, now using $n = 20$, are shown in figure 4.21. As expected, very clear separations are achieved for uranium and lead, with the separation of iron being slightly less effective due to the closer Z values.
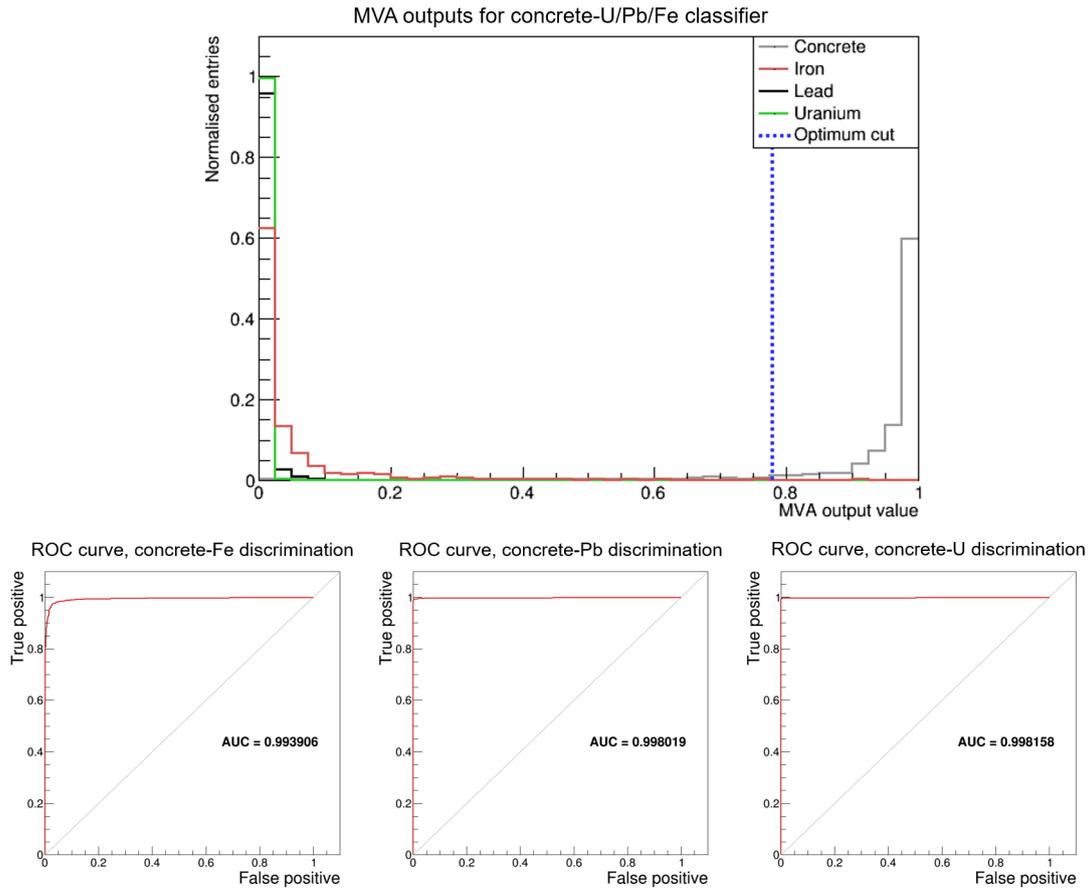
Figure 4.21: Characteristic figures of an MVA classifier trained to distinguish concrete voxels from iron, lead and uranium voxels. Top: classifier outputs in the training phase, showing the clear separation of the concrete class from the three background classes, and the optimum cut. Bottom plots are the ROC curves for separating concrete and iron (left), concrete and lead (middle), and concrete and uranium (right).

A confusion matrix can also be calculated to further interrogate a classifier's performance. This is a table in which the rows show the true classes of the tested events and the columns show how those events were actually classified; from this data the true and false positive and negative rates can be calculated. For the non-binary concrete versus not concrete classifier described above, the confusion matrix contains the number of concrete voxels that were classified as concrete and as not being concrete, and equivalently the numbers of not concrete voxels that were correctly or incorrectly classified. The resulting table is shown in figure

4.22. The low numbers of voxels that were incorrectly classified indicates the classifier has performed well in this case.

		Classifier output material	
	Total: 8768	Concrete: 2040	Other: 6728
True material	Concrete: 2192	2020	172
	Other: 6576	20	6556

Figure 4.22: Confusion matrix showing the true and false positive and negative outputs of the non-binary classifier shown in figure 4.21. It is clear that the classifier has performed well, as the rates of true positives and true negatives are both high, whilst the false positive and false negative rates are low.

When this classifier is applied to simulated drum data, it performs well at removing the vast majority of the remaining 'concrete' voxels, leaving only the voxels corresponding to stored objects. However, there are some single, isolated voxels that correspond to concrete but pass this filtering step, due to the imperfect discriminating power of the classifier. This is partially addressed by applying an additional filtering step: all remaining voxels for which none of their 16 'nearest neighbour' (i.e. adjacent or touching at one corner) voxels remain are also removed. This step removes the majority of the remaining non-object voxels. Figure 4.23 shows the results of applying the filtering system to a drum containing 15 cm side length cubes of the three metals. The remaining voxels clearly correspond well to the stored objects, and can now be grouped via the application of a clustering algorithm.
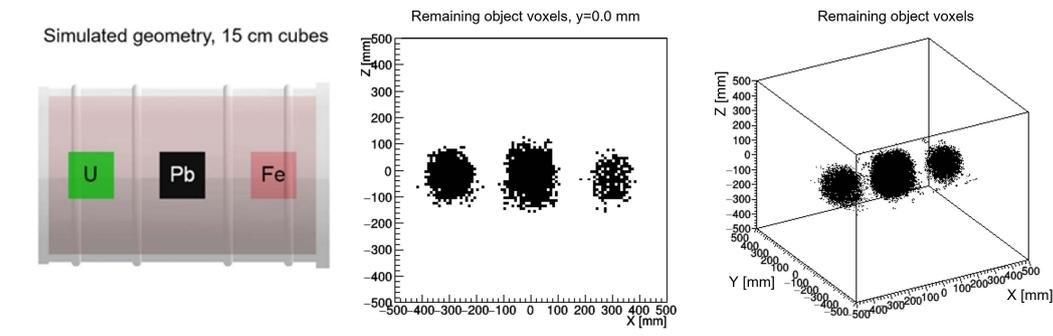
Figure 4.23: Result of removing the steel drum (via a cylindrical cut) and concrete matrix (via a MVA classifier) voxels from a binned clustering algorithm output image of a drum containing 15 cm cubes of iron, lead and uranium (left). The remaining voxels in the image (right) correspond well to the cubes. Also shown for clarity is the $y = 0$ slice (middle) of the 3D image.

4.5 Clustering

With a BDT-based non-binary MVA classifier applied, the image has been reduced to only voxels corresponding approximately to the objects that are stored in the drum. In order to

assign each object a material decision, these voxels must first be grouped into clusters, with each cluster corresponding to a single stored body. For some cases, this is trivial to do 'by eye', for example it is clear that for the example in figure 4.23 there should be three clusters with the voxels assigned to each based on their x coordinate. The correct clustering solution is harder to determine with more complex cases in which the objects are distributed along different axes. Whilst it would be possible for a user to obtain xy and xz slices from the 3D image and determine the clusters manually, it was felt that an automated system based on a clustering algorithm would be preferable in terms of both speed and usability.

The clustering algorithm chosen was the k-means algorithm, which has the advantages of being very simple to implement and having many available adaptations to be made more suitable for the problem. The simplest version of the algorithm operates on a set of points as follows:

1. Choose k, the number of desired clusters to sort the points into.

2. Select k points (uniformly) randomly from the data; these will be the initial cluster 'centroids'.

3. For the remaining points, calculate the Euclidean distance between the point and each of the centroids.

4. Assign each point to the cluster with the closest centroid.

5. Calculate new centroids: the mean position of the points in each cluster.

6. Return to step 3 and iterate until the centroid locations converge.

The result of applying k-means, with $k = 3$, to a toy dataset is shown in figure 4.24.
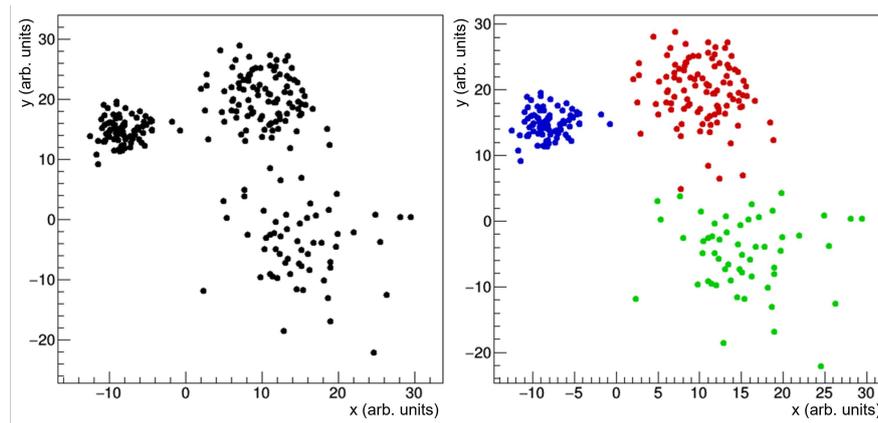
Figure 4.24: k-means clustering algorithm applied to a toy dataset of three Gaussian-distributed point clouds. The algorithm successfully groups the points into clusters corresponding to the sources, converging quickly; however it is necessary to specify that 3 clusters are required.

A limitation of the family of k-means algorithms is that the number of clusters k must be known in advance and used as input. When applied to the problem of clustering voxels into

objects, this cannot be assumed to be known; ideally it will be deduced from the data itself in some way. One approach tested was to project or slice the 3D image of the voxels onto a 2D plane, and then attempt to perform a peak-counting procedure to determine the number of objects. This worked well for some simple examples, but was unreliable for more complex distributions of objects. Similar procedures using aggressive filtering were also unsuccessful. An alternative approach is to simply run the k-means algorithm using a range of reasonable k values as input, and then calculate some figure of merit of the clustering result for each to quantitatively assess the most natural choice for k. If well designed this will be the number of stored objects in the drum for the vast majority of cases.

A commonly used figure of merit is *inertia*: the sum of the squared distances between each point and its cluster centroid. k-means aims to minimise this quantity; a low inertia indicates compact clusters. However, inertia will also decrease as k increases (eventually becoming zero when $k =$number of data points) and so choosing optimum k as the value which minimises inertia is insufficient. The preferred method is to identify an 'elbow' in the inertia-k relation, i.e. the point at which inertia ceases to fall rapidly and begins a slower linear decline. This method was implemented, using a system of calculating the goodness of linear fits to subsets of the inertia-k plots to identify the optimum k. Good performance was observed for some cases, but the method tended to underestimate the number of required clusters (e.g. outputting that 2 clusters was optimum for a drum containing 3 objects), particularly for the case of a small object with two or more larger companions.

The final approach was to use a different figure of merit, that takes into account the distance between the clusters in addition to their internal variances. The figure used was the Dunn index, defined broadly as the ratio of the minimum inter-cluster distance and the maximum intra-cluster distance; a high Dunn index thus corresponds to compact and well-separated clusters [82]. An advantage of this method is that the definitions of both distances are flexible (e.g. the intra-cluster distance could be the mean distance between any pair of points, or the mean distance between each point and the centroid, and so on) and so can be tuned to the problem.

To assess which definitions were most suitable for this problem, the Dunn index method was applied to drum scenarios containing 2, 3, and 4 15 cm cubes of different materials; the measure of success was the value of k with the maximum Dunn index corresponding to the actual number of stored cubes. The following definitions were found to perform best: the inter-cluster distance metric was the distance between any two points in different clusters, and the intra-cluster distance was the distance between the two furthest-apart points in any cluster. The Dunn index versus k relation for the 15 cm cubes example is shown in figure 4.25. The maximum Dunn index is found at $k = 3$ indicating that this is the best choice for k, as expected.

Obtaining k via the maximum Dunn index has been found to be an effective approach for the majority of scenarios. However, it can fail and choose either too many clusters (particularly if a very large object is present that is misinterpreted as two or more objects in close contact) or too few (a small iron object could be incorrectly attenuated out of the image by the filtering
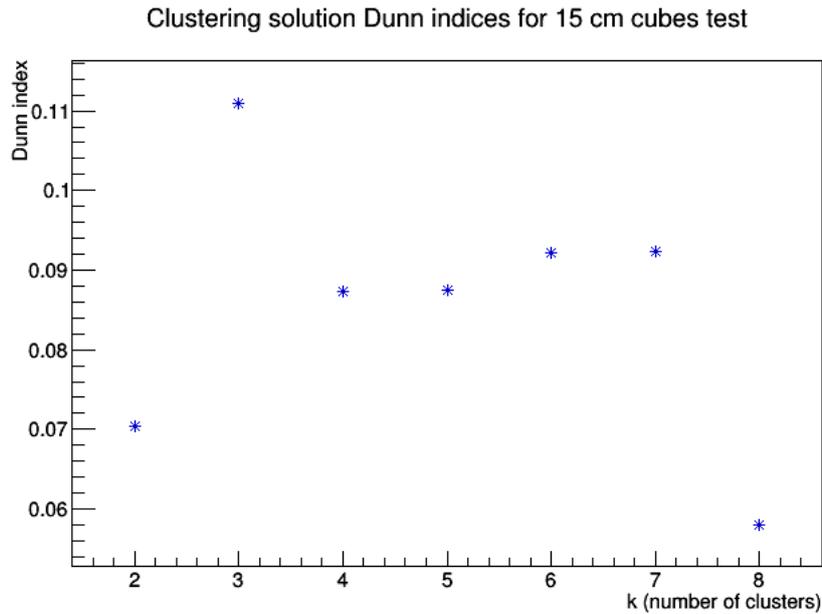
Figure 4.25: Results of applying the k-means clustering algorithm to the voxel map in figure 4.23 (middle), and calculating the Dunn index at each value of k.

algorithm used). Additional contingencies are used to reduce the likelihood of these scenarios. For example, a test is performed to check if two clusters are in close contact along one side to attempt to identify when a large object has been divided into multiple clusters. A final sanity check is also used: the chosen clustering solution is displayed to the user, along with the unclustered and unfiltered images. The user can then reject the solution chosen and set k manually if desired.

Additionally, it was found that the clustering algorithm gave more consistently 'correct' (i.e. the clusters corresponded well to the stored objects in number, size and location) answers when using the 'k-means++' algorithm [83]. The key difference is in the selection of the initial k centroids from the data points; rather than selecting all the points uniformly, in k-means++ the first point only is selected uniformly and the remaining points are chosen from a distribution biased towards the more distant points. Specifically, after the selection of initial centroid 0, the distances d of all other points are calculated and centroid 1 is selected from a distribution weighted by d^2, i.e. the further apart voxels are more likely to be selected. The process is then repeated for the remaining $k - 2$ centroids. In each case d is the distance to the closest already existing centroid. The algorithm then proceeds in the same way as vanilla k-means. The effect is to bias the initial centroids to be far apart; this reduces the risk of a poor choice of initial centroids, increasing the convergence time as well as improving the accuracy of the final result.

The results of applying the full k-means++ algorithm, with k determined via the maximum Dunn index, to the 15 cm example is shown in figure 4.26. An additional filtering step has also been subsequently applied to each cluster: a removal of 'outlier' voxels that are a large

distance from the cluster centroid. Specifically, the standard deviation of the voxel-centroid distance for each cluster is calculated, then voxels that exceed this distance are filtered out. This removes both the few remaining concrete voxels that have passed the previous filters (see figure 4.23), and also approximately the outermost voxel layer of each object, as these voxels are less characteristic of the object's material due to a smearing effect between the object and the concrete at the edges.
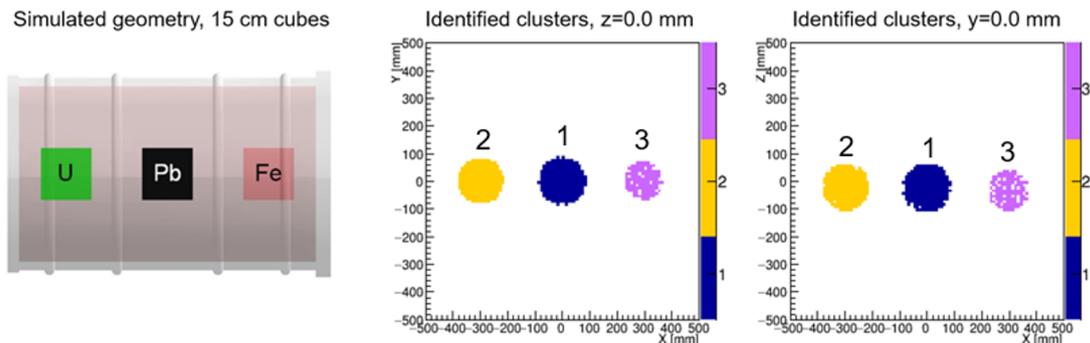
Figure 4.26: Results of applying the k-means++ clustering algorithm to the voxel map in figure 4.23 (middle), with $k = 3$ as determined with the Dunn index figure of merit, and additionally filtering outlier voxels.

4.6 Multivariate analysis: material decisions

4.6.1 Iron and lead classifiers

With the voxels (approximately) corresponding to stored objects now isolated via a concrete versus metal non-binary MVA classifier, and then clustered with the k-means++ algorithm, the next stage is to apply more classifiers to each cluster in order to obtain further material information. These are the aforementioned iron versus uranium/lead non-binary classifier and the binary lead-uranium classifier. The former uses voxels in a 20 cm cube of iron as the signal class, and similar cubes of lead and uranium as two background classes. The MVA output distributions for the training sets, and the two corresponding ROC curves (for distinguishing iron from uranium and iron from lead) are shown in figure 4.27, and the corresponding confusion matrix in figure 4.28. It is clear that, as expected, the classifier is more effective at distinguishing iron from uranium than for lead (although the discrimination between lead and iron is still excellent) due to the greater disparity in Z. This is quantified by the respective ROC AUCs of 0.992 ± 0.002 and 0.998 ± 0.002.

The final classifier applied is the binary uranium versus lead classifier. Its output and ROC curve are shown in figure 4.29, and its confusion matrix in figure 4.30. As expected, the discriminating power of this classifier (with a ROC AUC of 0.811 ± 0.008) is less than for the previous discriminations of materials, due to the closer similarity in Z between uranium and lead. The AUC is however still significantly greater than 0.5, indicating that the classifier is still able to distinguish between the materials to a good degree.

Figure 4.27: Characteristic figures of an MVA classifier trained to distinguish iron voxels from lead and uranium voxels. Top: classifier outputs in the training phase, showing the clear separation of the iron class from the two background classes and the optimum cut. Bottom plots are the ROC curves for separating iron and lead (left) and iron and uranium (right).

4.6.2 Material scores

These classifiers will produce a single output value for each voxel they are applied to. But at this stage, the sub-units of interest are instead clusters of several hundred to several thousand voxels, with each cluster corresponding approximately to a stored object inside the waste drum. The question then becomes how to combine the classifier outputs from both classifiers and for all of the voxels in each cluster to arrive at a single material decision for the cluster.

Both classifiers have their optimum cut (see section 4.4.2), with voxels that have response values above the cut being considered 'signal-like' (i.e. iron for the non-binary classifier, uranium for the binary classifier), and below being 'background-like'. When a classifier acts on a cluster of voxels, the output will be a distribution of output values with some proportion

		Classifier output material	
	Total: 9939	Iron: 3096	Other: 6843
True material	Iron: 3313	3057	256
	Other: 6626	39	6587

Figure 4.28: Confusion matrix showing the true and false positive and negative outputs of the non-binary classifier shown in figure 4.27. It is clear that the classifier has performed well, as the rates of true positives and true negatives are both high, whilst the false positive and false negative rates are low. However, the classifier has not performed as well as the concrete classifier shown in figure 4.21.
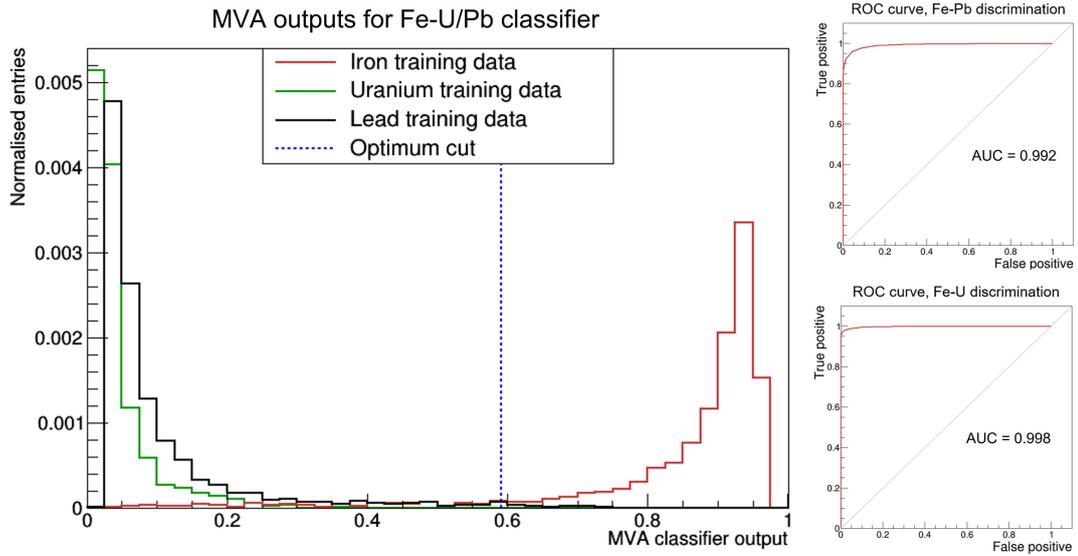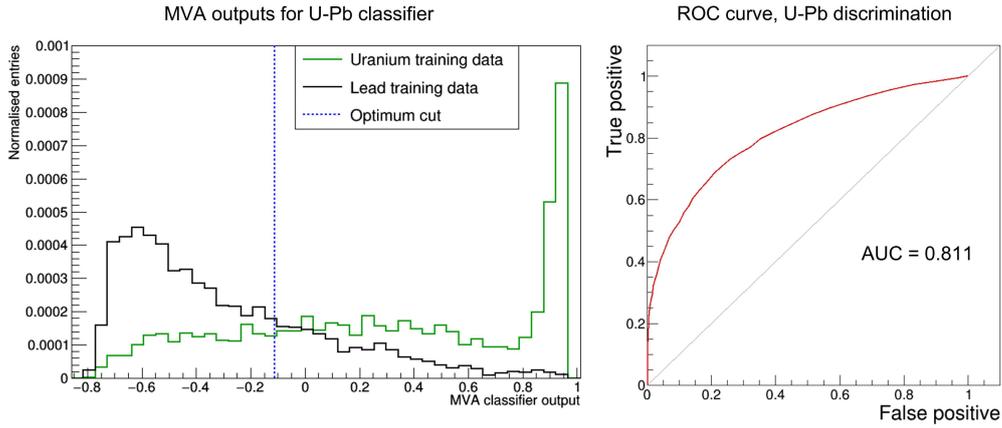
Figure 4.29: Characteristic figures of an MVA classifier trained to distinguish uranium voxels from lead voxels. Left: classifier outputs in the training phase, showing the clear separation of the two classes; right: the ROC curve for separating lead and uranium.

		Classifier output material	
	Total: 7306	Uranium: 3410	Lead: 3896
True material	Uranium: 3616	2557	1059
	Lead: 3690	853	2837

Figure 4.30: Confusion matrix showing the true and false positive and negative outputs of the binary uranium versus lead classifier shown in figure 4.29. It is clear that the classifier has performed fairly well, although it is by far the worst-performing of the three classifiers due to the close similarity in Z between lead and uranium.

above and below the cut. This is illustrated in figure 4.31, which shows the results of applying the full system thus described to a drum containing 15 cm cubes of the three metals. The three identified clusters, found using the concrete filtering process and clustering algorithm described in sections 4.4.8 and 4.5 respectively, are input to the two classifiers and the distributions of the response values obtained.

It is clear that the proportion of values that fall above and below the cut is correlated with the material of the original object. For example, when the iron versus lead and uranium non-binary classifier acts on the cluster corresponding to the lead cube ('object 1'), the vast majority of the response values are below the cut in the 'U/Pb'-like region; i.e. the vast majority of the voxels in the cluster are classified as lead or uranium. When the lead versus uranium classifier is applied to the same cluster, most voxels fall below the cut in the lead-like region, although the proportion is less significant in this case due to the lesser discriminating power of this classifier. Similar patterns can be seen in the response distributions for the iron and uranium blocks as well.
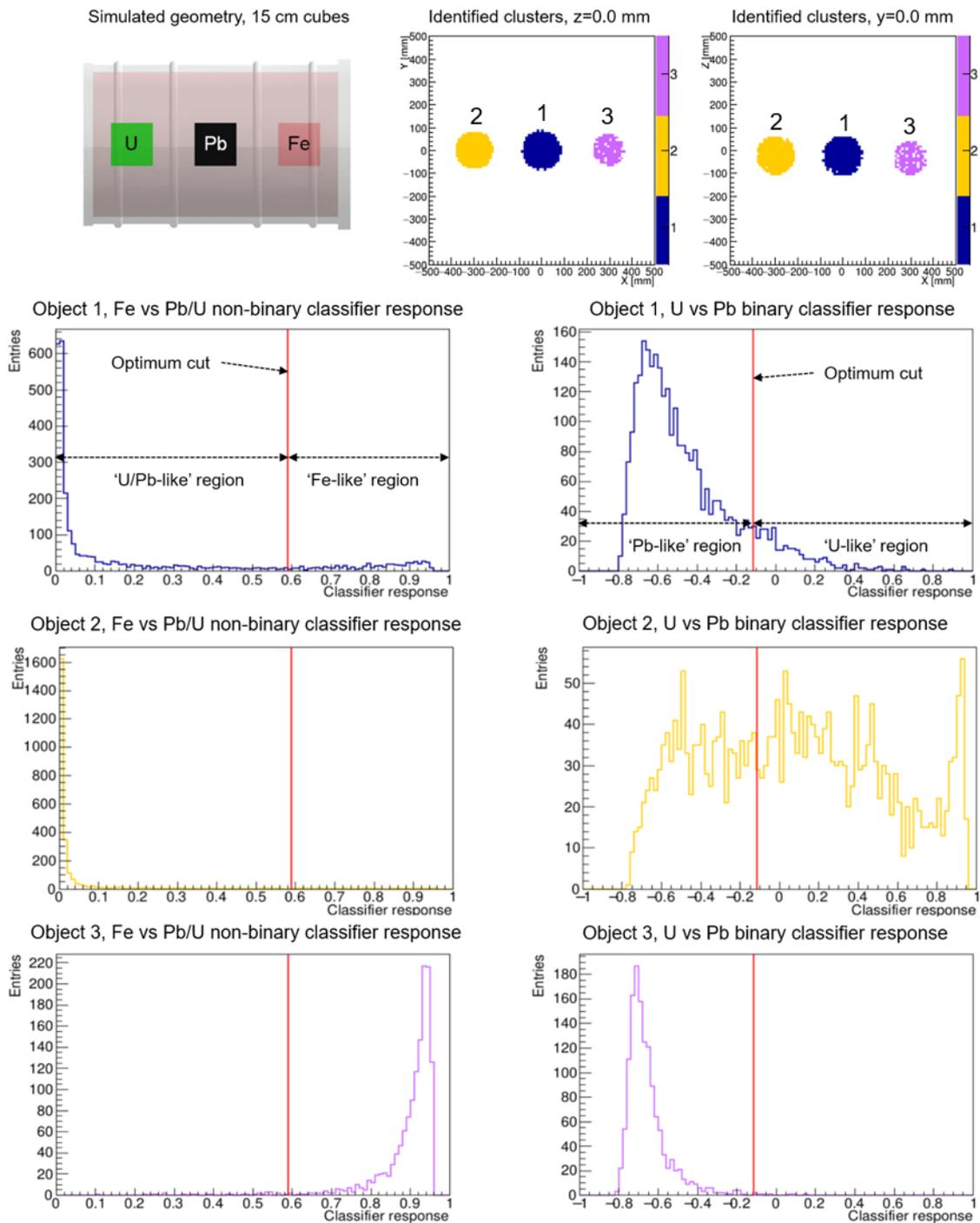
Figure 4.31: Distributions of responses of MVA classifiers applied to found voxel clusters (top right) from a simulated waste drum containing 15 cm cubes of lead (object 1, top), uranium (object 2, middle) and iron (object 3, bottom). The left column shows results from the iron versus lead and uranium non-binary classifier, and the right column results for the lead versus uranium binary classifier.

Therefore, the measure chosen to combine the MVA output values so as to account for the distribution around the optimum cuts was to calculate the proportion of response values above the cut, i.e. to integrate each distribution from the cut to unity, and divide the total by

the total integral. For example, the lead cube non-binary response distribution in figure 4.31 (top left) has 2936 entries, of which 496 fall above the cut. The proportion above the cut, $S_{\mathrm{Fe}} = 496/2936 = 0.169 \pm 0.008$, is then the 'iron score' for that cluster and quantifies how 'iron-like' it is. The error on the score is propagated from the standard errors on the histogram bins i.e. the square roots of the bin contents. For the binary classifier response, 348 voxels have output values above the cut; the equivalent proportion is thus $348/2936 = 0.119 \pm 0.006$ with the corresponding figure for the 'lead-like' region being 0.881 ± 0.017.

The definition of S_{U} and S_{Pb}, the 'uranium score' and 'lead score' respectively, must take into account the results of the non-binary classifier as well. They are therefore defined with an additional multiplicative factor of $1 - S_{\mathrm{Fe}}$, i.e. the 'U/Pb-like' score from the classifier. Hence for the lead cube, $S_{\mathrm{U}} = (1 - 0.196) \times 0.119 = 0.099 \pm 0.006$ and $S_{\mathrm{Pb}} = 0.744 \pm 0.029$. The three scores form a set of material scores, $S_{\mathrm{M}} = \{S_{\mathrm{Fe}}, S_{\mathrm{Pb}}, S_{\mathrm{U}}\}$, that sum to unity and are characteristic of the material of the original object. Figure 4.32 shows the scores for the 15 cm cubes example. The scores are intuitively visualised as a pie chart.
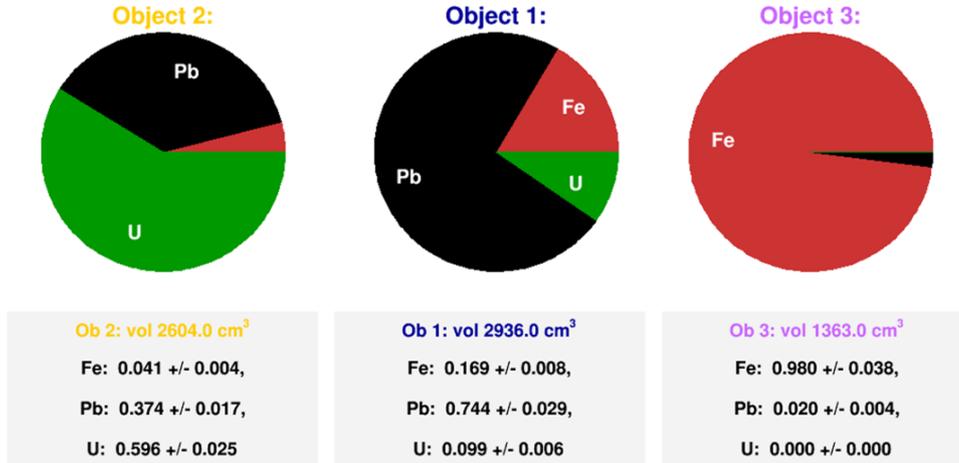
Figure 4.32: The material scores for a test geometry of three 15 cm cubes. Object 1 corresponds approximately to the lead cube, object 2 to the uranium cube, and object 3 to the iron cube. In this case, each object's largest material score corresponds to the actual material of the simulated body.

From this example, it could be assumed that the largest of the three material scores could be taken as the 'decision', i.e. each object could be assigned the material with the largest value. However, in practice this is not effective: the 15 cm example is something of a best case scenario due to the large size of the bodies and their general similarity - large cubes, along the drum axis, aligned with the voxel grid - to the original training geometries (a 20 cm centred cube, see section 4.4.4). A more challenging test case, with smaller 10 cm cubes, is shown in figure 4.33. In this case, although once again the objects have been successfully identified and clustered, the relation between S_{M} and the true material is less clear.

Of note is that, although the largest material score no longer corresponds with the true material, there are still clear and significant differences between the sets S_{M}. For example, the uranium cube has $S_{\mathrm{U}} = 0.221 \pm 0.025$, whereas the equivalent lead cube has $S_{\mathrm{U}} = 0.014 \pm 0.006$;
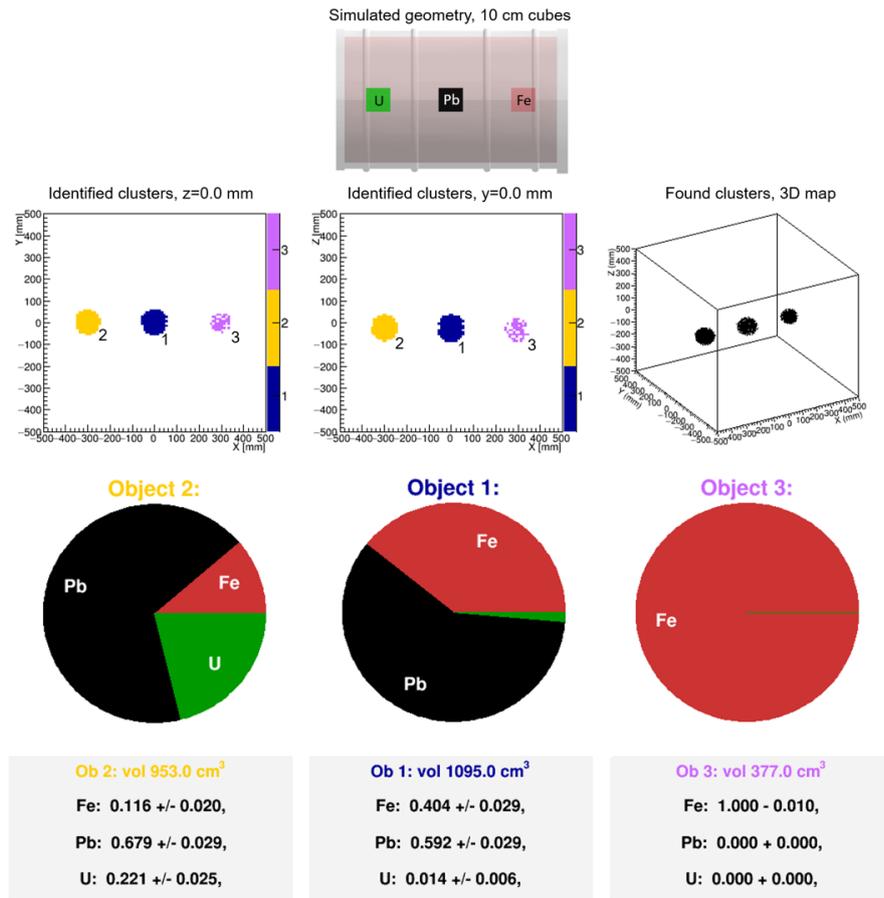
Figure 4.33: The material scores for a test geometry of three 10 cm cubes. In this case, each object's largest material score does not correspond to the actual material of the simulated body, but there are clear differences in the scores between objects of different materials nonetheless.

similarly their respective lead and iron scores are clearly different. This implies that accurate material decisions can still be made from S_M, but an additional calibration step is required to account for the size of the object.

4.6.3 Size calibration

The dependence on the object volume reflects the fact that muons passing through two objects of the same material but different sizes will not necessarily have similar scattering angle distributions: the thickness of the material must also be accounted for also (see equation 3.3). Muons undergoing repeated scatterings in a large high-Z object will lead to larger detected muon scattering angles, and hence a smaller binned clustering metric value (see section 4.2). Hence a large lead object can appear more 'uranium-like' than a smaller lead object.

To account for this, a series of calibration curves were calculated in order to relate the material score set S_M to the object size when assigning material decisions. This was accomplished by running the system on simulated drums containing spheres of different radii

and of each material. The positions of the spheres were generated randomly (but constrained so that all of the sphere was entirely inside the concrete), to reduce bias from the location. Each sphere's material score set was calculated using the methods described above; the volume of a detected object is treated as being equal to the total volume of the 1 cm^3 voxels in its cluster. The results are shown in figure 4.34. Each plot shows the three scores for spheres of different materials (top to bottom: uranium, lead, iron). Spheres were used to remove the effect of the alignment with the voxel grid that affects rectilinear objects.
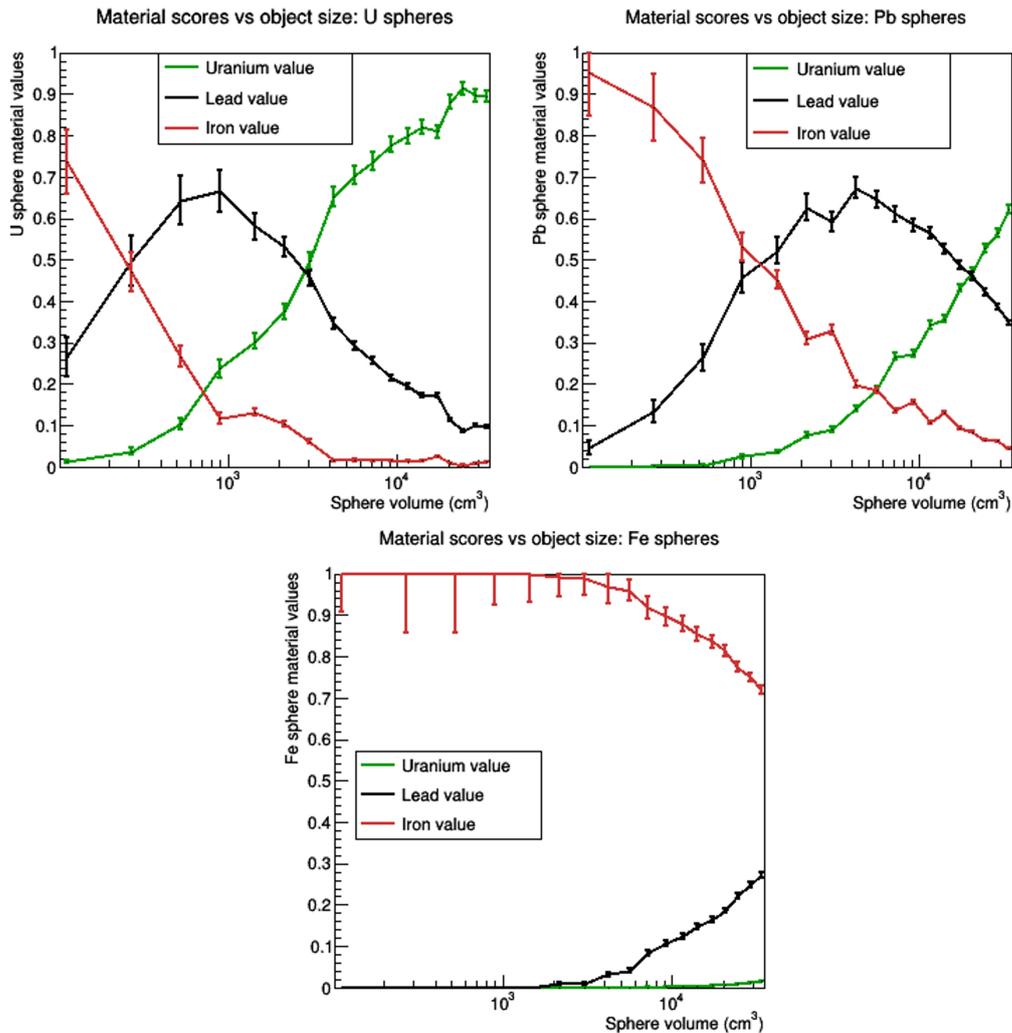
Figure 4.34: The calibration curves used to account for the effect of object volume when assigning materials. Spheres of each material, various radii and random positions were input into the system and their material scores calculated. Top: uranium spheres, middle: lead spheres, bottom: iron spheres.

From these curves, it is clear that there is no simple relation between an object's S_M and its volume. In general, the larger an object is, the more its S_M values are skewed to higher-Z i.e. higher uranium and lead values. Of note is a similar observation as was made when comparing the scores for 15 and 10 cm cubes above: the set of scores as a whole is characteristic of the material, even if the largest score does not correspond to the true material.

For example, it is clear that the uranium sphere with volume around 10^3 cm^3 has a larger S_{Pb} than S_{U}, yet a lead sphere of the same volume has a larger S_{Fe} than S_{Pb}, and a much lower S_{U} than the uranium sphere.

The final material decision for an unknown cluster is then arrived at by calculating its S_{M} and volume as described above. The scores are then compared to the calibration scores for uranium, lead and iron objects of the same volume, linearly interpolating between the data points. The three calibration sets are labelled $S_{\text{M,calib}}$; with $S_{\text{M,U calib}}$ being the calibration scores for the uranium spheres and so on. The final material decision is then made by calculating the Euclidean distances between S_{M} and the three $S_{\text{M,calib}}$ sets and determining the minimum, i.e. calculating each of

$$d_U = \sqrt{\sum_{i=1}^{3}(S_{\text{M}}^i - S_{\text{M,U calib}}^i)^2}, \tag{4.6}$$

$$d_{\text{Pb}} = \sqrt{\sum_{i=1}^{3}(S_{\text{M}}^i - S_{\text{M,Pb calib}}^i)^2}, \tag{4.7}$$

$$d_{\text{Fe}} = \sqrt{\sum_{i=1}^{3}(S_{\text{M}}^i - S_{\text{M,Fe calib}}^i)^2} \tag{4.8}$$

and labelling the object with the material that satisfies $\min\{d_U, d_{\text{Pb}}, d_{\text{Fe}}\}$.

For example, object 1 in figures 4.31 (corresponding to a 15 cm lead cube) and 4.32 has $S_{\text{M}} = \{0.169, 0.744, 0.099\}$, and 'volume' (number of voxels in the cluster) of 2936 cm^3. The three calibration material score sets for an object of this volume are then $S_{\text{M,U calib}} = \{0.066, 0.465, 0.483\}$, $S_{\text{M,Pb calib}} = \{0.328, 0.595, 0.088\}$ and $S_{\text{M,Fe calib}} = \{0.990, 0.010, 0.000\}$. The corresponding Euclidean distances d in this case are then $\{0.486, 0.218, 1.11\}$; the minimum is d_{Pb} and so the object's material decision is lead. For this example, the same method also produces correct material decisions for the uranium and iron cubes.

4.7 Results

4.7.1 Material decisions for various geometries

The results of applying the system to a set of drum geometries, with various number, size, position and materials of objects, are shown in figures 4.35 and 4.36. They demonstrate that the system performs well across diverse geometries. The applicability of the size calibration method described above is shown by the results for 10 cm and 5 cm cubes (figure 4.35 top and bottom respectively); in both cases, the three objects have been correctly identified and their materials classified despite them being much smaller than the training geometries.

For the example geometries shown so far, the objects in question have been cubes, centred along the drum axis and aligned with the voxel grid. Obviously this is an idealised scenario; to test how objects rotated from the grid and displaced from the centre would respond, a
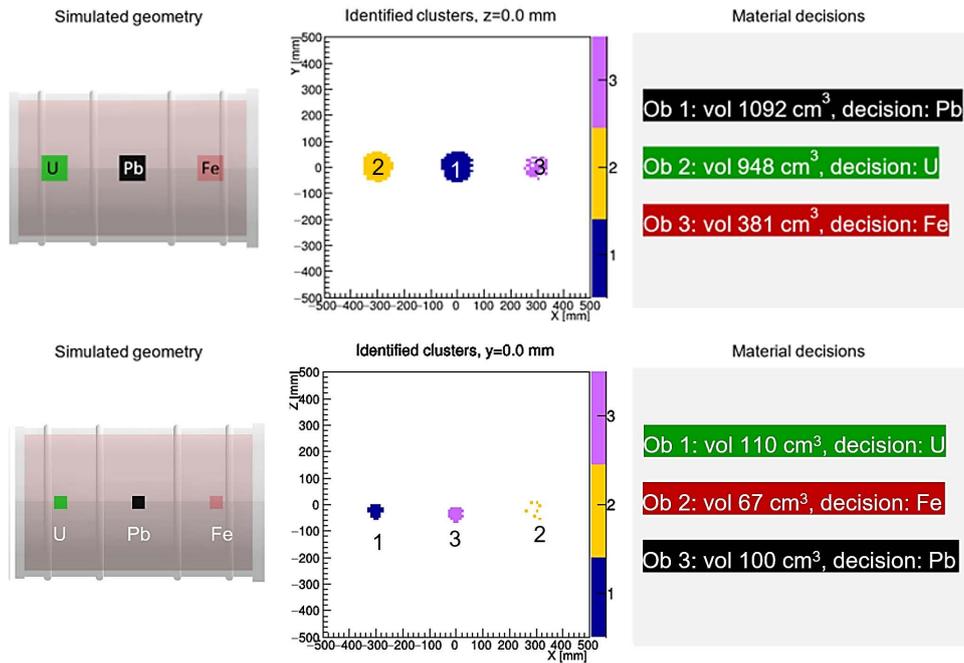
Figure 4.35: Results of applying the full method to drum geometries containing cubes, with the original simulated geometry (left), clustering results (middle), and material decisions (right). Top: 10 cm cubes, bottom:5 cm cubes. All objects have been correctly classified.

geometry containing a rotated cuboid of each metal was input to the system. The results are shown in figure 4.36 (top). The filtering, clustering and material decision systems have been successful in this case, correctly identifying the three bodies, and assigning their true materials.

Finally, a much more complex example was tested, with results in 4.36 (bottom). This drum contained five objects: a uranium cube, an iron sphere, a lead tube, a lead cone, and a uranium tube. The objects were spread throughout the drum away from the central axis. When the system was applied to this drum, once again the filtering and clustering has worked well and the objects have been successfully identified. The material identification results are more complex with four of the objects having been correctly categorised, but the lead tube having been classified as iron. This is likely a result of the size calibration step: the elongated nature of the tube is a poor match to the spherical objects used to generate the calibration curves. An alternative approach might be to implement additional calibration curves with different object shapes, and use a mean of the Euclidean distances for each case to make the final material decision. The uranium objects have been correctly identified however. The correct classification of uranium objects in particular is the most important test of the system; to determine how well it performs at this task in particular, additional tests were performed.
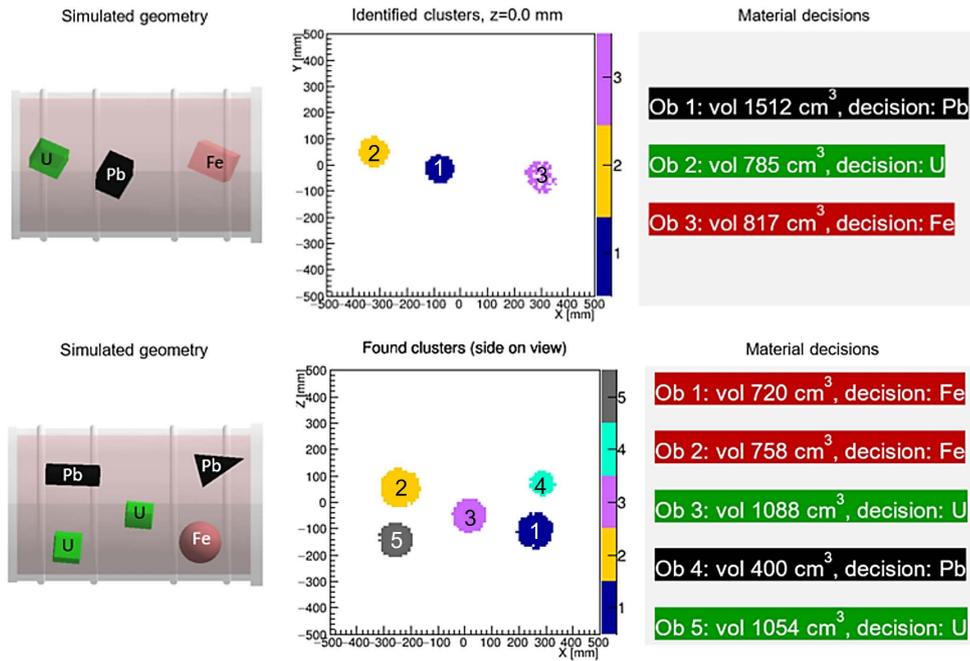
Figure 4.36: Results of applying the full method to drums containing offset and rotated rectilinear objects (top) and five objects of various shapes and positions (bottom). Across the examples shown, all objects have been correctly classified except for object 2 (the lead tube) in the final example, which is misclassified as iron.

4.7.2 Detection of small uranium objects

The application of a system to identify materials in a nuclear waste drum has two main objectives. Firstly, to characterise the contents of the drum in general terms, so as to gain information about its origin or history (e.g. identifying a lead object of a specific size and shape that could be recognised, or determining that a stored object had been broken into pieces from some past violent event). Secondly, to check for the presence of objects of any size of a specific material. In particular, if a piece of uranium were erroneously present in a concrete-filled waste drum, it would represent an enormous hazard to personnel as well as a safeguarding risk.

To assess the system's sensitivity to uranium objects, a set of 100 test geometries was created. Each geometry contains three spheres with radius 6 cm. The positions of the spheres were determined randomly under the constraint that they could not intersect one another or touch the inner drum wall. For 50 of the test drums, the spheres were made of uranium, lead and iron (the 'threat' scenarios); for the other 50 two of the spheres were lead and the other iron ('non-threat'). The full system was applied to 10-day exposures for each drum and the material decisions obtained. A true positive identification of a uranium object was defined as an object with a position close to the true location of a uranium sphere that was classified as uranium by the system. A false positive was accordingly defined as any object in a non-threat drum being classified as uranium. Under these criteria, the achieved true positive rate for small uranium objects was $0.90^{+0.07}_{-0.12}$ and the false positive rate was $0.12^{+0.12}_{-0.07}$

(errors here express 95% Clopper-Pearson confidence intervals).

Additionally, an alternative criterion was tested: a cut on the minimum d_U of all the detected objects in a drum. The closer d_U is to zero the more likely the object is to be classified as uranium. As a continuous measure, this allows for ROC analysis of the system's performance by altering the cut value and calculating the resulting true positive and negative rates. The results are shown in figure 4.37 and indicate a similarly high performance level; the AUC of this ROC curve is 0.91 ± 0.06.
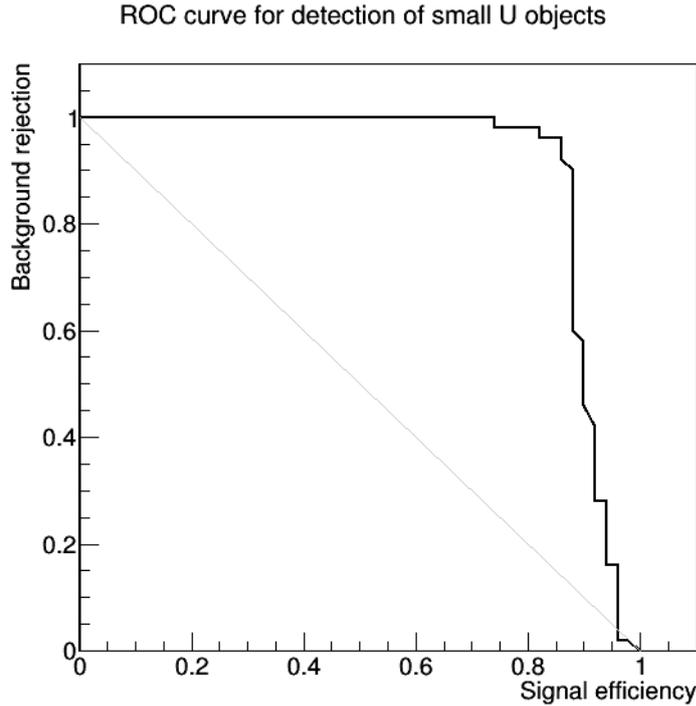
Figure 4.37: ROC curve for the system's ability to detect small uranium objects, as measured by a cut on the Euclidean distance d_U to the calibration uranium score. AUC= 0.91 ± 0.06 indicating the classifier performs well by this measure.

4.8 Conclusions

The primary conclusion of this work is that machine learning techniques, and specifically MVA classifiers, are an effective tool for non-destructive assay of heterogeneous waste packages when combined with muon scattering data. It has been demonstrated additionally that a 'local' approach, in which material information down to a voxel scale as small as 1 cm is targeted, is feasible. Enough usable muon tracks are available for material discrimination at this scale, when using a realistic detection system with a heavily smeared momentum measurement and a practical exposure time of several days.

Specifically, variables obtained from the distribution of binned clustering algorithm metric values are suitable for training MVA classifiers. Using training geometries that are large cubes

of the material classes will produce classifiers that are somewhat overtrained but still suitable for identifying the materials of objects in waste drums, as long as the size of the object is calibrated for. In particular, a classifier trained to recognise concrete from metals is very effective at removing the concrete matrix from a waste drum image. The remaining voxels can then be grouped into objects via simple clustering algorithms. The object clusters can then be input to additional classifiers trained to discriminate between iron, lead and uranium, to produce characteristic sets of material scores S_M that encode information about the object's material.

The material scores are highly dependent on the object's volume. By establishing empirical relations between object size and S_M, a final material decision can be made for each identified body in the waste drum. The full method has been shown to produce accurate results for a wide range of object sizes, shapes and drum locations. Additionally, when tested against a set of simulated drums containing 6 cm radius spheres of different materials with random positions, the system was able to detect the uranium objects effectively with a ROC AUC of 0.91 ± 0.06.

In terms of the method's vulnerabilities, the main identified difficult scenarios are objects of materials with very different Z values (e.g. iron and uranium) being in close contact, and more spatially extended objects being misidentified. Performance in the former case could be improved by applying an edge-finding algorithm to the filtered drum image, in an attempt to identify regions where objects are in direct contact. This information could then be supplied to the clustering algorithm as a bias to force the voxels into separate clusters. The latter could be mitigated by altering the size calibration system to include a wider range of object shapes. In general, using smaller voxels and/or a longer exposure time will improve the system's performance against very small (< 4 cm) or low-Z objects. The training of the classifiers could also be improved, perhaps by supplying a larger training set comprising data from several different drum geometries (e.g. a small cube, a large cube, a group of smaller objects etc.).

Finally, an open question is how sensitive the classifier training is to the specifics of the detector design. For example, if an experimental system of a different design - a larger detection area, or a different gap between the modules - was used, would the classifier weights calculated via using the simulated detector system described in section 4.3.2 be sufficient? Or would a new, dedicated simulation be necessary? The classifier variables are calculated from the binned clustering algorithm's metric values, which depend primarily on the muon scattering angles. This suggests that the classifiers would still be compatible with a detector system of a different design if it did not have a substantially different aperture, which is likely if the system was designed for the same purpose i.e. MST of small waste drums. A system designed instead for e.g. cargo scanning or assaying a much smaller volume would very likely require dedicated simulations to obtain useful classifiers. A change only in detector efficiency, however, could be accounted for by altering the exposure time so as to keep the number of tracks detected consistent.

Chapter 5

Design information verification for geological repositories

5.1 Introduction

This chapter describes a simulation study to assess muon radiography's potential as a design information verification technique for geological nuclear waste repositories. These are final disposal sites for primarily high-level nuclear waste, designed to shield the biosphere from radiation into the far future. The presence of an unknown geological feature, or an attempt to access the stored waste by malicious actors could be catastrophic; it is vital that there are techniques in place to detect these. Muon radiography's sensitivity to density deviations makes it a promising candidate. In this chapter, a software framework for Monte Carlo simulations of large structures is described, and results of applying the software to detect voids in geological repositories are presented and discussed. Finally an attempt to directly image voids in repositories using muon radiography and a powerful iterative reconstruction algorithm is described.

5.1.1 Deep geological waste repositories

Most countries that use nuclear power are planning to ultimately store the resulting high-level waste in deep Geological Repositories (GR) [62]. These structures will be between several hundreds of metres to over a kilometre underground [84] and the chosen sites must possess geological stability over timescales of up to millions of years [85], to ensure that there is no possibility of hazardous waste being released into the environment before it is has decayed enough to become safe for the biosphere or humans in the far future. Additionally the site must be far from major population centres whilst still being easily accessible for waste transport. Although these criteria greatly constrain possible GR sites, the primary challenges in site selection are usually political, with strong opposition often coming from residents local to the site or other stakeholders.

The existing and planned geological repositories around the world vary greatly in design

and progress towards becoming operational. They range from the repository being fully excavated and prepared for imminent waste entry, to having no confirmed site for the GR, or even have decided whether it should be located on land or undersea. For example, the USA selected a site at Yucca Mountain, Nevada in 2002, but funding was withdrawn in 2010 due to strong political opposition [86]. Currently there is no approved site for long-term American waste disposal and waste will continue to be stored in short-term facilities and on-site at nuclear power stations for the foreseeable future. In contrast, the Finnish Onkalo site (see figure 5.1) located in Olkiluoto, Eurajoki has been under construction since 2004 and is expected to begin receiving waste packages in 2023 [87].

Figure 5.1: From [88]. Photograph of waste storage tunnels under construction in the Finnish Onkalo GR site.

In general, the storage process for the waste itself will involve additional barriers to improve the overall reliability of the radiation protection. The waste itself, e.g. spent fuel rods from power stations, will be encapsulated in a copper or steel canister [84], designed to prevent waste leakage for at least 1000 years. This phase is done at an encapsulation plant, often at a facility adjacent to the repository itself. After the waste has been emplaced in a storage tunnel it is surrounded with bentonite clay, which acts as a barrier against radiation and nuclide diffusion; its plastic nature also allows it to self-heal if disturbed. Finally, most repositories have a planned phase of 'backfilling' in which, once a storage tunnel has been filled with waste, the tunnel is filled totally with additional clay. In the long term, the whole GR including access shafts and ramps will be backfilled in this way, surface facilities will be decommissioned and dismantled, and the site will be abandoned.

5.1.2 Design information verification

For a GR to operate safely it is imperative that the geological structure of the sites is well understood. There must be no significant deviations from the Design Information Verification (DIV) during the construction and the operational phase of a GR that could compromise effective safeguarding. In order to ensure that all safeguards-relevant features of the GR design

are as declared by the State, the design information of a GR is verified by using approved techniques such as 3D laser scanning and geophysical monitoring during the construction and operational phases of the facility [89]. However, alternative techniques are required that allow for the detection of unknown chambers or tunnels, which could serve as clandestine access routes to nuclear material, which is subject to international safeguards [90].

Muon radiography has potential as a GR DIV for void detection technique as it is sensitive primarily to density changes. As the density of air is three orders of magnitude smaller than that of rock, muons are attenuated to a far smaller degree; the muon radiation length (see section 3.3) of air is 3.04×10^4 cm whereas for standard rock it is only 10.02 cm. Muon radiography does require long exposure times, particularly in an underground context for which the overburden greatly reduces the muon flux. However this is in general compatible with GR facilities due to their very long operational timescales.

The results presented below represent initial studies into the general feasibility of muon radiography for GR DIV. Some simplifications have therefore been used: the simulated repository model uses homogeneous rock composition rather than a complex stratigraphy, and the detectors used are fairly idealised, with high efficiencies and angular resolutions. The impact of these limitations on results can be partially mitigated in each case; these are discussed with the relevant results. Under these caveats, useful and novel results have nevertheless been obtained.

A variety of simulated detector positions and arrangements were used for the studies presented here. The maximum detection area deployed for any simulation was 48 m^2, in the form of twelve 2 m $\times$ 2 m square planar detectors spread throughout the repository. In terms of feasibility of such a scenario, this is a somewhat larger area than a typical muon radiography experiment, for example the pyramid assay study of [24] used a total area of ~ 10 m^2 for their surveys. However, the relatively inexpensive nature of scintillator-based particle detectors (relative to the usual operating costs of a GR) would allow such a system to be feasibly scaled up.

The detection of narrow unknown shafts dug for clandestine waste access was investigated, determining the effects of void volume, shape and exposure time on the detectability of voids. Further studies tested the effects of using the number and position of detectors used on the detectability of voids. Finally, a system of multiple detectors combined with a powerful imaging algorithm is used to attempt to image voids within the repository.

5.2   Backwards Monte Carlo simulations for muon radiography

A traditional muon tomography simulation uses a 'forward' Monte Carlo particle transport system (see section 2.5). Muons are generated at some 'source' surface, usually a plane, with their initial energy and trajectory sampled from a cosmic ray library such as CRY [28] (see section 4.3.1). The particles are then transported forward with the software simulating the discrete and continuous energy loss processes described in section 2.3; the muons decay, scatter and lose energy until they (or their decay products) reach an edge of the simulated

'world' or another stopping condition is met (e.g. a minimum energy). For Muon Scattering Tomography (MST) simulations, such as those described in chapter 4, the source plane is usually of comparable size to the detectors and placed almost directly above the simulated MST system. This allows for efficient simulations as the majority of the generated particles will encounter a detector surface and hence be usable for the tomography.

In contrast, most muon radiography simulations concern structures on a much larger scale. Simulating common radiography applications, such as volcano imaging or assaying a thick overburden from an underground position, require transporting particles through hundreds or thousands of metres of rock, to encounter a detector of only metre-scale. If the same forward Monte Carlo approach was used the result would be highly inefficient and computationally intensive; one would need to generate initial particles across a very large 'sky' surface, of which only a tiny fraction would reach the detector surface. If the source plane was made of comparable size to the detector and positioned above it, only a small column of material would be assayed, with near-vertical muons. Additionally, a portion of the muons would Coulomb scatter out of the column and be lost, reducing the simulation efficiency.

There are several possible approaches to resolve this problem and allow muon radiography scenarios to be simulated efficiently. The first is to use a forward Monte Carlo simulation, powered by a particle transport toolkit such as GEANT4 [13], but apply extra controls to decrease unnecessary computations. For example, applying a cut to the initial energy of the muons so as to remove low-energy muons that have a very low probability of reaching the detector through the dense medium. A similar cut on the initial muon trajectories can be applied by creating a 'target box' around the detector and only keeping the subset of muons with trajectories that intersect the box. This is an approximation, as it discounts the effect of muons that initially do not point at the box but subsequently scatter into it and contribute to the muography (for further discussion on the target box approximation, see section 6.2.2). A larger target box will therefore be more accurate, but less efficient (as less of the muons that enter the box will encounter the detector). An extension of this method is to restrict the angular distribution that the initial muons are drawn from such that only those pointing at the box are even generated. This increases efficiency further, as the unwanted muons will not be initialised and no trajectory cut needs to be performed. See chapter 6 for further discussion of these methods and analysis of their effects on simulation accuracy.

For some applications, it may be useful to instead use an impossibly big 'detector' (e.g. a 1 km side length square plane) so as to obtain a high simulation efficiency. Whilst this technique is of limited use for modelling an experimental system, it can be convenient for simulation debugging exercises (e.g. obtaining the spectrum of surviving muons passing through a thick layer of material), or for studying the time evolution of a large-scale system in which the objective is to estimate the required exposure time to resolve some feature. In some cases an estimate of the flux that would be detected by an experimental system can be obtained with this method by extrapolating to a smaller detection system. Care must be taken to account for the different geometric aperture of the larger system.

5.2.1 MUSIC

An alternative to GEANT4 is to use a code dedicated to muon transport only, such as MUSIC (MUon SImulation Code) [30]. Unlike a GEANT4 simulation, MUSIC does not transport secondary particles created during muon interactions; this makes it highly efficient for scenarios in which only muons are of interest. For example, when simulating muon flux arriving at a position deep underground, simulating secondaries that cannot penetrate through the rock and reach a detector is not necessary. MUSIC has been shown to give muon spectrum results that agree closely with GEANT4 simulations and experimental data.

5.2.2 PUMAS

A more radical approach is to depart from forward Monte Carlo altogether, and use a method based on 'backward' Monte Carlo. The basic idea is to generate the initial muons on some surface very close to the detector(s), then perform the muon transport backwards i.e. up through the geometry towards the 'sky' (see figure 5.2). The muon-matter interactions are calculated 'in reverse' (e.g. the muons gain energy as they travel). At the end of the transport the muons are assigned weights based on the probability of observing that particular state. The advantage of backwards Monte Carlo is that all of the generated and transported muons will encounter the detector and so be used in the radiographic calculations, vastly increasing the efficiency of the simulation.
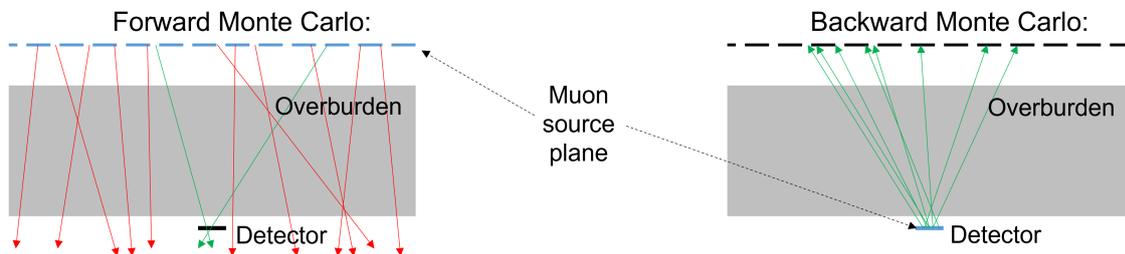
Figure 5.2: Schematic demonstrating the contrasting principles of forward (left) and backward (right) Monte Carlo particle transport. Forward Monte Carlo is inefficient for large scale geometries as only a small fraction of the generated muons will encounter the detector. Backwards Monte Carlo ensures that all of the transported muons will meet the detector and contribute to the muography.

The first practical implementation of backwards Monte Carlo for muon radiography is the PUMAS software ('Semi Analytical MUons -or taus- Propagation, *backwards*') [31], which has been used extensively in the results presented in this chapter. The backwards Monte Carlo of PUMAS aims to estimate the probability density function of the final state $\mathbf{s}_f$ from the initial state $\mathbf{s}_i$; the state is defined as the set comprising a muon's spatial position, kinetic energy and momentum direction. The process is based on the following theorem, here quoted from [31]:

'*Let $\tau_{i,f}(\mathbf{s}_f; \mathbf{s}_i)$ be the probability density function, with respect to $\mathbf{s}_f$, of a transition probability from an initial state $\mathbf{s}_i$ to a final state $\mathbf{s}_f$. Let $\mathbf{s}_f = g(\mathbf{s}_i; X)$ be a Monte-Carlo*

process for generating $\mathbf{s}_f$ distributed as $\tau_{i,f}$, given $\mathbf{s}_i$ and a set of random variables X that do not depend on $\mathbf{s}_i$, e.g. uniform over $(0,1)$. If g^{-1} the inverse of g with respect to $\mathbf{s}_i$ exists, then for any probability density function ρ_i for the initial state, a backward Monte-Carlo estimate of the probability density function ρ_f of the final state is:

$$\rho_f(\mathbf{s}_f) = \int \tau_{i,f}(\mathbf{s}_f; \mathbf{s}_i)\rho_i(\mathbf{s}_i)d\mathbf{s}_i \simeq \frac{1}{N}\sum_{k=1}^{N}\omega_{i,k}\rho_i(\mathbf{s}_{i,k}). \tag{5.1}$$

$$\mathbf{s}_{i,k} = g^{-1}(\mathbf{s}_f; x_k) \tag{5.2}$$

$$\omega_{i,k} = \det(J_{g^{-1},\mathbf{s}_f}(\mathbf{s}_f; x_k)) \tag{5.3}$$

where x_k is a random variate drawn from X and with $J_{g^{-1},\mathbf{s}_f}$, informally $\partial\mathbf{s}_i/\partial\mathbf{s}_f$, the Jacobian matrix of g^{-1} with respect to $\mathbf{s}_f$.'

In other words, an estimate of the final state probability distribution function ρ_f can be obtained by generating final states $\mathbf{s}_f$ uniformly, then inverting the Monte Carlo simulation flow (g^{-1}) to obtain a weighted distribution of initial states $\omega_{i,k}\rho_i(\mathbf{s}_{i,k})$.

In practise, each uniformly-generated final state is assigned an initial weight ω as the inverse of the generating probability distribution function, i.e. $\omega = \frac{1}{\text{PDF}_{\text{gen}}}$. For example, if the final states are generated with a fixed energy and direction but their spatial position is uniformly distributed across a 2D plane of area A m^2, then $\text{PDF}_{\text{gen}} = A^{-1}$ m^{-2} and so each state's initial weight would be $\omega = A$ m^2. The weight is modified during the backwards propagation at every step until the muon reaches the primary flux surface, at which the Jacobian factor $J_{i,f} = \partial\mathbf{s}_i/\partial\mathbf{s}_f$ is calculated, and the weight is multiplied by the primary flux $f(E,\theta)$, in units of GeV^{-1}m^{-2}sr^{-1}s^{-1}. Different flux models are available; the default is that of [5] (see section 2.2). The final weight is therefore given by

$$\omega = J_{i,f}\frac{f}{\text{PDF}_{\text{gen}}}. \tag{5.4}$$

In this case, ω would have units of GeV^{-1}sr^{-1}s^{-1}. A full treatment would be have the final states' momentum directions and energies also be randomised, the former uniformly within some aperture, the latter usually from a reciprocal distribution so that the states are uniform in $\log E$ to account for a wide range of muon energies of different orders of magnitude. In this case the final weight would have units s^{-1} and would represent the estimated rate of events in that final state. The total muon rate is the mean of the weights, with the error on the rate measurement calculated as the standard error on the mean.

Two of the simplest applications of PUMAS are for muon flux measurements in a free-sky scenario and underneath a homogeneous standard rock overburden. Examples for these scenarios are included with the software library; for each, the muon final states are initialised at a single point, originating from a single zenith angle, and with energies sampled from a reciprocal distribution between 1 MeV and 1 PeV. Results from these examples are presented here as a validation exercise. Figure 5.3 shows the free-sky muon differential flux at zenith

angles of 0° and 85° as calculated by PUMAS, for both angles, the results agree closely with the empirical flux parameterisation of [5].
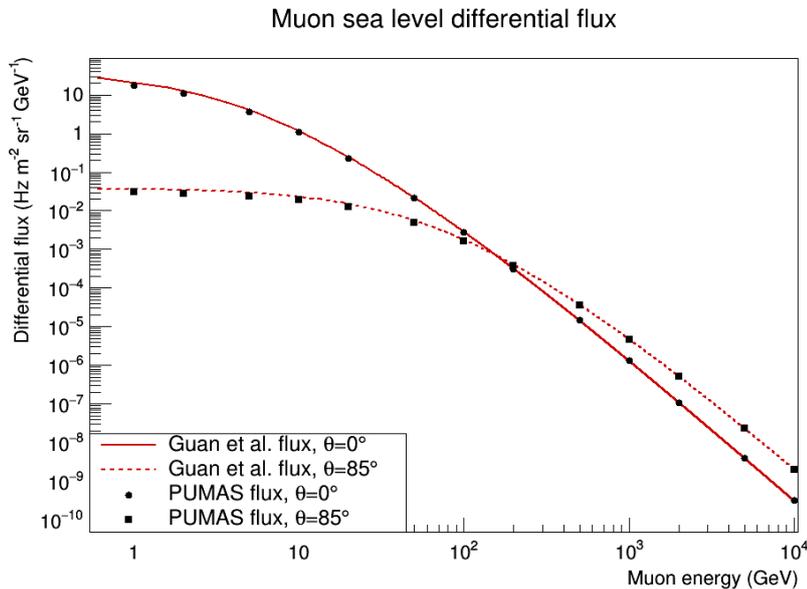

Figure 5.3: Muon differential flux spectrum as calculated by the PUMAS backwards Monte Carlo software [31], compared with the results predicted by the flux model of [5]. For zenith angles of both 0° and 85°, the PUMAS fluxes agree well with the predicted results. Confidence intervals on the PUMAS fluxes are present but smaller than the plot markers.

For validation of muon flux measurements made under standard rock, comparison is made to the empirical model of muon intensity in [7] (see equation 3.2). Results are shown in figure 5.4. Once again the PUMAS-calculated results are a good match to the empirical results, but there is a discrepancy of a factor of ∼ 2 for shallower depths. At the depths considered for the repository scenario (∼ 1000 m.w.e.), there is good agreement.

5.2.3 PUMAS with CRESTA

The default version of PUMAS is not well suited for complex, intricate geometries with many surfaces and boundaries to consider. Geometries are specified with Boolean operators defining the material within each spatial region, and the user must calculate and supply a stepping distance (usual the distance along the muon path to the next material boundary) at each step. Visualisation of constructed geometries and particle trajectories is also not straightforward. Finally, realistic detector effects such as finite apertures, tracking efficiencies and scattering between detector layers are not implemented.

For these reasons, the results presented in this chapter were obtained using a modified software that uses both PUMAS and CRESTA [70], a GEANT4 [13]-based forward Monte Carlo software also used extensively in chapters 4 and 6. The combined software has both backwards and forward Monte Carlo stages. The former is used to obtain a flux weighting that takes account of the overburden and other features of the geometry, and the latter to obtain realistic detector effects.
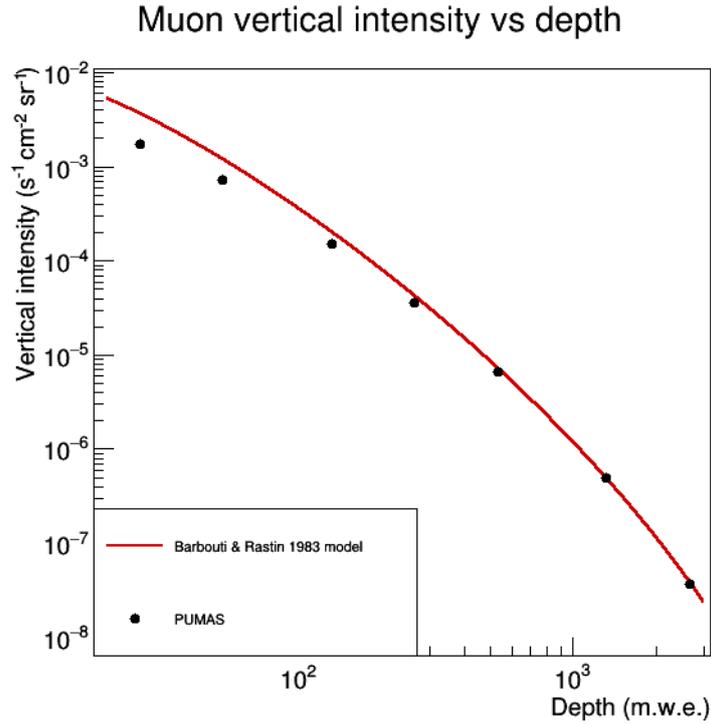
Figure 5.4: Comparison of PUMAS-calculated muon vertical intensity under various depths of standard rock, expressed in metres water equivalent, as compared to the empirical model of [7]. Close agreement is achieved for depths over ~ 100 m.w.e. (≈ 40 m of rock).

As discussed above, a backwards Monte Carlo transport begins with generating initial muons with state vectors drawn from uniform distributions. In our case, each state generated has 5 variables: x and y position, azimuth (ϕ) and zenith (θ) angles, and energy E. The z coordinate is fixed, i.e. the muons are generated on a two-dimensional plane. In order to match the size of the simulated detector systems (see section 5.3), the planes are 2 m × 2 m. ϕ is selected from a uniform distribution between 0 and 2π; θ between 0 and $75° \approx 1.31$ rad. The muon energies are selected from a log-uniform distribution between 1 MeV and 1 TeV. The number of muons generated in each scenario was chosen so as to be approximately equivalent to a particular exposure time, see section 5.4.1. Figure 5.5 demonstrates the effect of the backwards Monte Carlo on the zenith angle θ in a free-sky scenario; the initial values are uniformly generated within the boundaries, when the weights are applied the canonical $\sim \cos^2(\theta)$ relation is produced.

Validation

Validation of the combined software was performed in two stages: first demonstrating that the combined software produced results comparable to equivalent forward Monte Carlo simulations, then testing the full system against experimental data for free-sky and underground scenarios. An example of the first stage is shown in figure 5.6, which compares the free-sky muon spectra as detected with both the combined PUMAS-CRESTA software and with a traditional forward
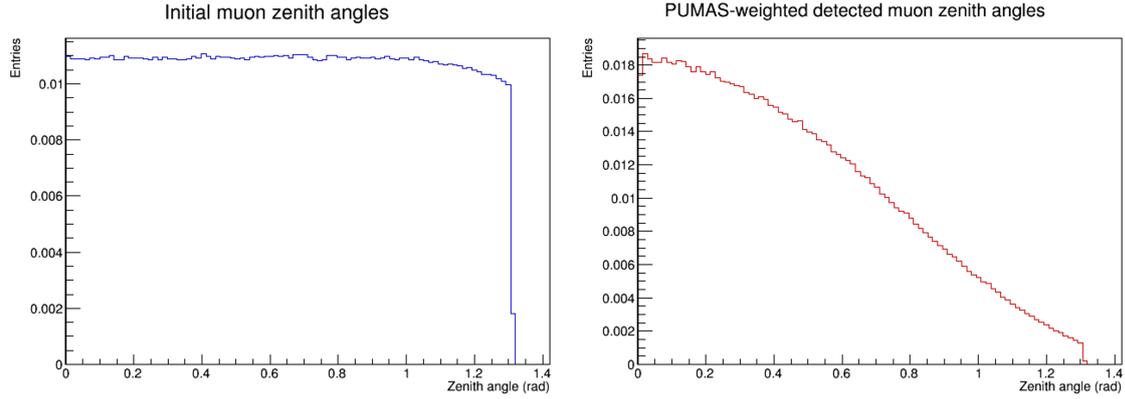
Figure 5.5: Comparison of the zenith angles θ of the uniformly-generated initial muons (left) to the distribution after performing backwards Monte Carlo using the PUMAS-CRESTA software (right). The initial angles are approximately uniform between 0 and 75° (the slight dip at higher angles is due to muons that encounter the walls of the world volume being removed). The weightings applied by the backwards Monte Carlo produce the expected $\cos^2(\theta)$ shape.

Monte Carlo. There is a close match between the muon energy distributions.
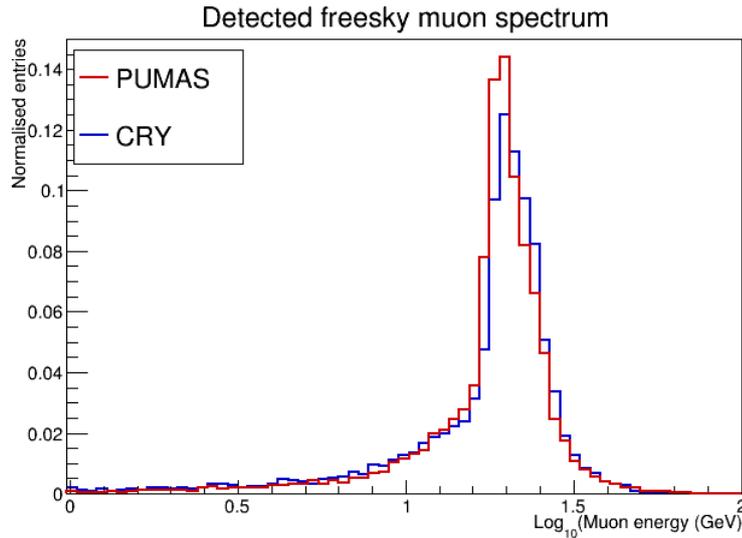
Figure 5.6: Distributions of detected muon energies for two simulated free-sky geometries. One (red) uses the combined PUMAS-CRESTA software to perform the transport and the other (blue) is a forward Monte Carlo simulation with the initial muon energy and angular distributions sampled from the CRY library [28]. The distributions agree closely.

The validation comparisons to experimental data are shown in figure 5.7. Tests of the free sky energy spectrum (left) and the vertical flux underground (right) were performed. For the former, the results from the software are in close agreement with those from the data; however for the latter a deficit of a factor of ~ 2 is once again observed at low depths. This is likely a result of the underlying difference in flux model used in PUMAS, as shown in figure 5.4. The affected depths are much less than those used in the GR simulations.
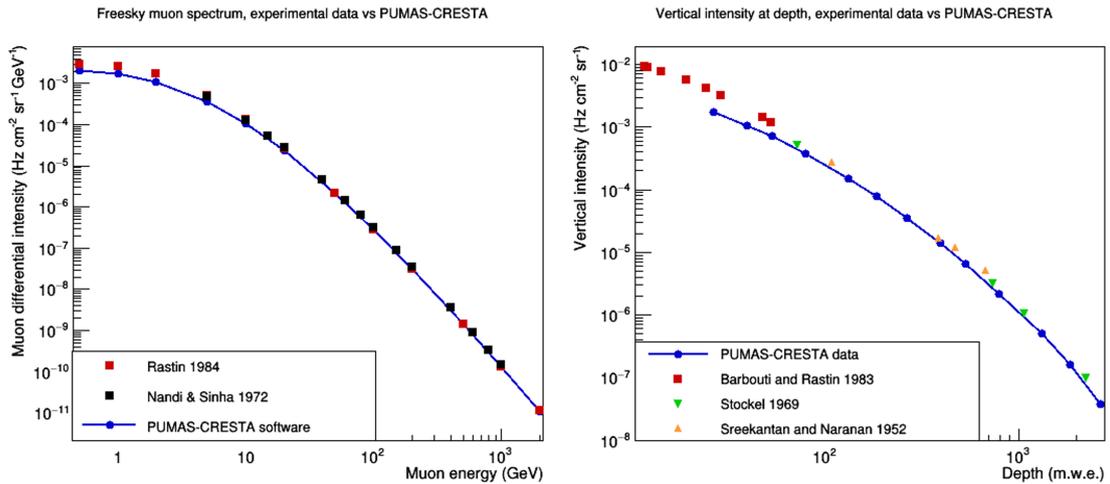
Figure 5.7: Validation tests for the PUMAS-CRESTA combined software. Left: comparison of the free-sky differential muon spectrum generated by the simulation (blue) versus experimental data from [91] (red) and [92] (black). Right: comparison of simulated muon vertical flux measurements (blue) under various overburdens of standard rock, expressed in metres water equivalent, versus experimental data [7] (red), [93] (green) and [44] (orange). For both plots, error bars on points are smaller than data markers.

5.3 Repository model

Despite their greatly varying designs, all GR will include a rock overburden several hundreds of metres thick, a network of branching waste storage tunnels, and, for access and waste transport, either a system of ramps or a shaft (see figure 5.8). It was decided therefore to design the simulated repository model so as to contain each of these features, but not to attempt to replicate an existing or planned repository in detail; in other words to simulate a generic repository and obtain results that would be broadly applicable to existing or planned GR.
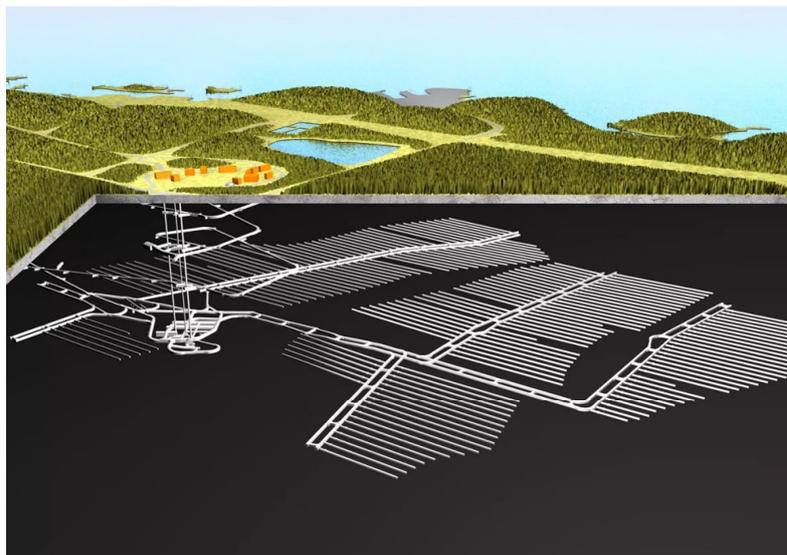
Figure 5.8: From [94], diagram of the storage tunnel network at the Finnish Onkalo GR site.

5.3.1 Simulated model design

The rock overburden thickness was set at 440 m, a similar depth to that at the Finnish Onkalo GR [95]. A system of four storage tunnel networks was included, each consisting of a 600 m long cylindrical main tunnel with a radius of 3 m and a set of ten intersecting shorter tunnels, representing the 'waste storage' areas of the repository. No actual waste containers were included in the simulation, as they would not be visible to the detectors positioned in the same plane as the tunnels. Assaying waste containers with muon radiography (or MST, see section 3.3) at this depth would be very challenging but would be possible at sea level (i.e. pre-emplacement). Each tunnel is made of air only, the tunnel wall materials were not considered as they would not deviate significantly in density from the surrounding rock. The tunnel networks branched off a 'gallery', an air-filled cuboid with dimensions $10 \times 10 \times 610$ m. Also included was a cylindrical access shaft between the surface and the storage tunnel depth. The detailed dimensions of the repository features are shown in figure 5.9.
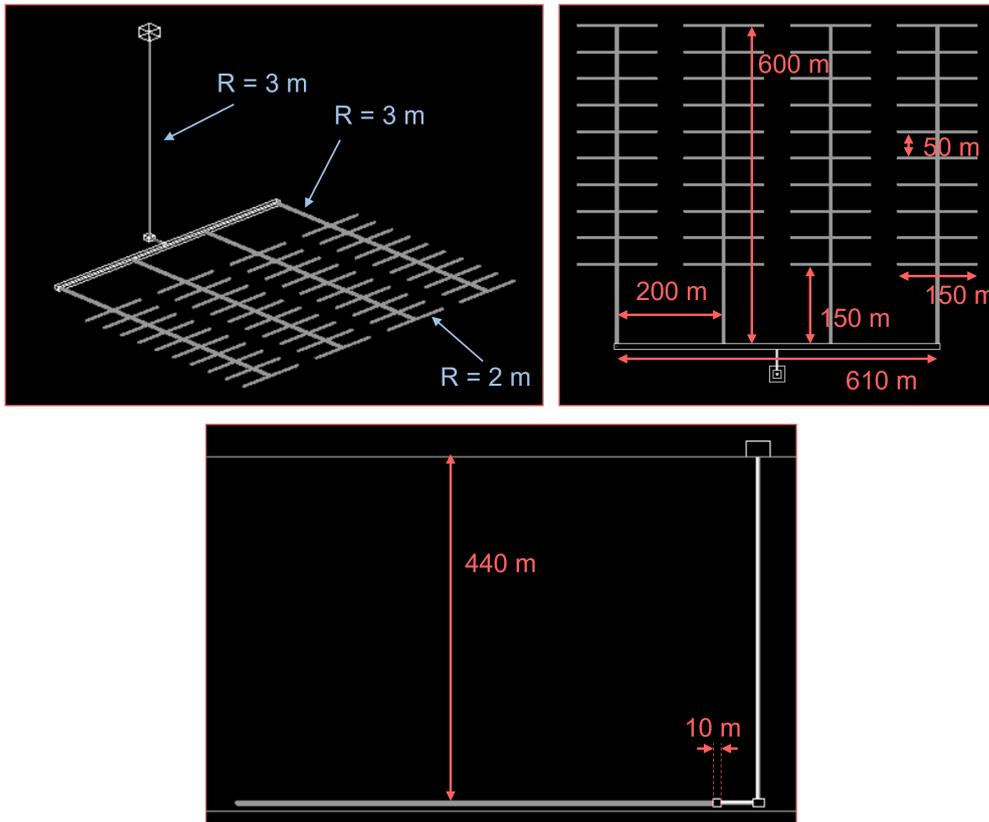
Figure 5.9: The simulated geological repository model, from three views, with key dimensions and distances labelled. The body of the repository consists of a set of branching waste storage tunnels underneath a rock overburden 440 m thick.

5.3.2 Overburden material

For the overburden material, it was decided that, since the overburden's bulk density would have by far the biggest impact on the muon flux at the scales considered, to use a homogeneous

material rather than a complex stratigraphy of several minerals. The mineral used was granite, an igneous rock composed primarily of quartz i.e. silicon dioxide. Granite was chosen due to its presence at the Onkalo site. At the time of implementation, PUMAS did not have a granite material included by default; it was necessary to create it manually. The required information is bulk density and elemental proportions. Granite from different sites can vary in density from around $2.6 - 2.75$ g cm^{-3} and so a density of 2.75 g cm^{-3} was chosen for the simulation in order to study a less idealised scenario (a denser overburden will lead to a reduced muon flux and hence a longer required exposure time for useful data). For the elemental proportions, average worldwide granite chemical composition data from [96] (e.g. 72.04% Si O$_2$, 14.42% Al$_2$O$_3$ etc.), was used to calculate the element ratios by mass, shown in table 5.1.

Element	Granite proportion by mass (%)
O	48.1
Si	33.7
Al	7.63
K	3.42
Na	2.74
Fe	2.16
Ca	1.30

Table 5.1: Proportion by mass of elements in the granite overburden material.

Using a homogeneous granite overburden simplifies the simulations whilst also allowing the results to remain generalised. However, this is not a realistic scenario and does introduce caveats into the results. The assumption being made is, in essence, that the geological structure of the repository site is sufficiently well understood that a significant deviation in the expected muon flux due to the presence of a new or previously unknown void would be detectable. Many of the results presented in this chapter rely on a comparison of detected muon fluxes between repositories containing voids or not, the latter being in an experimental scenario a result obtained from a simulation of the known geological structure. The homogeneous overburden used here means that the case in which there is a density deviation due to unknown geological features (e.g. a layer of less dense rock), which could produce a false 'void' feature, is not explored. The problem is partially mitigated by the fact that any GR site will by necessity have a relatively stable and homogeneous geology that is very well characterised. Additionally, later studies (see section 5.5) account for the specific directions of density deviations. Results of this type can distinguish a highly localised density anomaly such as a void from a larger feature.

For the detector design, as the objective was to determine the general feasibility of the technique rather than to simulate a specific experimental scenario, simple single planes of scintillator were used. Each plane was $2 \times 2 \times 0.1$ m, made of polystyrene with a density of 1.06 g cm^{-3}. Each detector was considered to have perfect angular and spatial resolution, but a finite efficiency set to 85% (independent of muon direction and energy), to increase the

difficulty of the void detection studies. The maximum detectable zenith angle was set to 75°. A set of 12 detectors was created and positioned throughout the repository model, at the ends and midpoints of the waste storage tunnels (see figure 5.10). Not all of the detectors were used in all of the simulations. For a particular detector to operate, a PUMAS muon initialisation plane was placed at its surface and the backward muon transport performed as described in section 5.2.2. A separate simulation was required for each detector in use, and so for efficiency most simulations use only a subset of the twelve detectors.
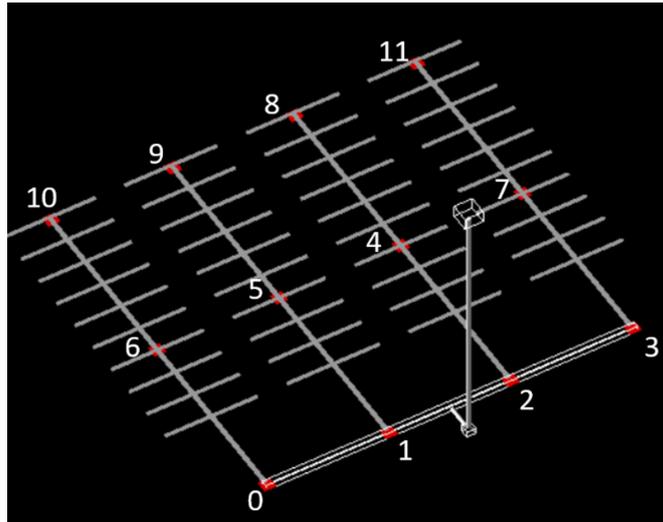
Figure 5.10: The positions of the 12 planar scintillator detectors (red) in the repository model. Each detector is a 2×2 m plane of polystyrene; they have been enlarged in the figure for clarity.

5.4 Void detectability

The first set of simulations were designed to study the variation in detected muon rate due to the presence of voids in the repository model. The goal was to determine how the detectability of the voids would vary as a function of the void's volume, shape and location; number and position of detectors, and the exposure time. To this end, voids of various types were simulated under different scenarios, with accompanying equivalent simulations without voids present also being performed. The detected muon rate for the case with and without the voids were both determined and compared. For each scenario, the detectability was then calculated in terms of the number of standard deviations separating the two measurements.

5.4.1 Exposure time

The void and detector characteristics can be easily modified in the JSON tables describing the geometry. The treatment of exposure time requires more care. For a traditional forward Monte Carlo simulation, exposure time is a key parameter of the particle initialisation and transport: the clock advances with each particle state sampled from the initial flux distribution

88

and as it is transported through the geometry. This allows the user to study how the quality of the muon radiography of the volume of interest will change with longer exposures, a key parameter for a practical experimental system. Exposure time for a backwards Monte Carlo is less clearly defined. The results of a backwards Monte Carlo are point estimates for the rates at which different final states will be detected. The total rate is obtained as the mean of the calculated weights; the error on the rate is the standard error on the mean and so decreases as the number of generated initial particles increases. There is therefore not an equivalent 'clock' that accounts for the time necessary to obtain a particular muography.

A natural choice for defining an exposure time equivalent for a particular backwards Monte Carlo simulated muography is to consider the error on the muon rate measurement. Consider a simple free-sky scenario with a 1m × 1m planar detector placed perpendicular to the zenith. For an experimental or forward Monte Carlo simulated measurement, the detector will record some number of hits N within an exposure time Δt, with the measured muon rate then $r = N/\delta t$. The detection of the hits is a Poisson process and hence the error on this rate measurement $\delta r = \sqrt{N}/\Delta t$, with the relative error

$$\frac{\delta r}{r} = \frac{1}{\sqrt{N}}. \tag{5.5}$$

As the exposure time and hence N increases, the relative error will therefore fall proportional to the square root of the time. If a precise estimate of the rate from a long exposure is available, the relative error on a shorter measurement can be estimated. For example, if after a long exposure a total muon rate of 1469 Hz is measured in a particular detector, then the number that would be detected after 2 s can be drawn from a Poisson distribution with $\lambda = 2$ s $\times$ 1469 Hz = 2938, giving e.g. $N = 3006$ and an estimated rate of $r = 1503$ Hz, and hence (from equation 5.4.1) $\frac{\delta r}{r} \simeq 1.82\%$. If this process is repeated for many values of Δt, the convergence of the estimated rate on the 'true' rate and the corresponding reduction in relative statistical error (inversely proportional to the square root of the exposure time, see equation 5.4.1) on the rate become clear, as shown in figure 5.11.

A similar exercise was then performed using the combined PUMAS-CRESTA backwards Monte Carlo software. Free-sky muon rate measurements for different numbers of initialised muons were taken, and the relative rate on each measurement calculated. The errors on PUMAS rate measurements originate from the standard error on the mean of each sample (see section 5.2.2). Figure 5.12 shows the results; as expected, the relative errors fall with the square root of N.

With a relation between exposure time and relative error on the rate measurement established, it is now possible to directly estimate a relation between equivalent exposure time and number of initial muons generated in a backwards Monte Carlo, for a particular geometry. For the free-sky scenario discussed above, combining the best fits in figures 5.11 and 5.12 gives the relation Δt [s] $\approx 4750N$. One could therefore use backwards Monte Carlo to approximate 1 hour of exposure time by generating 1.7×10^6 initial muons.

The calculations above apply only to a free-sky scenario and the specific detector configu-
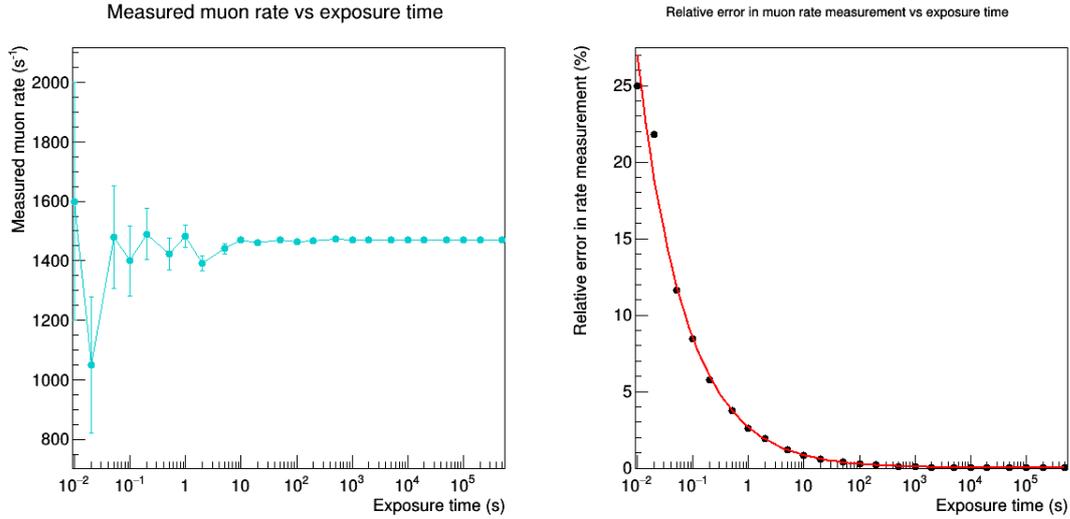
Figure 5.11: Demonstration of the increase in the precision of a muon rate measurement, which is subject to Poisson statistics. Left: estimates of the rate that could be calculated for different exposure times, between 0.01s and ~ 5 days, if the 'true' muon rate was 1469 Hz, calculated by drawing from the appropriate Poisson distributions. Right: resulting relative error on each rate measurement. The fitted curve (red) is $2.68/\sqrt{\Delta t}$.

ration, and must be repeated for alternative detector characteristics and locations. This is because a muon rate measurement taken e.g. underground would require a longer exposure time to obtain an equally precise muon rate measurement due to the reduced muon flux. For the geological repository model (see section 5.3), with its thick overburden of rock, the exposure time-N relation will be significantly different. When the procedure above was performed for this geometry, it was found that the relative error-exposure time curve was well fitted by $331/\sqrt{\Delta t}$, and the relative error-N curve by $266/\sqrt{N}$. Hence the relation between the desired exposure time and the number of muons to transport via backwards Monte Carlo within the repository model is given by

$$\Delta t \text{ [s]} \approx 1.55N \tag{5.6}$$

Figure 5.13 demonstrates this relation by plotting the relative rate errors for N and Δt as evaluated from the fits calculated as described above for test values, allowing the resulting straight line fit to become clear. In practise, this fit was used to read off an appropriate value for N for different desired exposure times, so as to choose how many PUMAS muons to initialise for each case and so study the effect of varying exposure times.

5.4.2 Large void - single detector

With an efficient backwards Monte Carlo-based muon radiography simulation software, a GR model with detectors, and a method to estimate equivalent exposure time for N generated muons, it was now possible to begin adding voids to the model and studying their detectability. For the initial study, a large cylindrical void, filled with air of constant density 1.2×10^{-3} g cm^{-3}
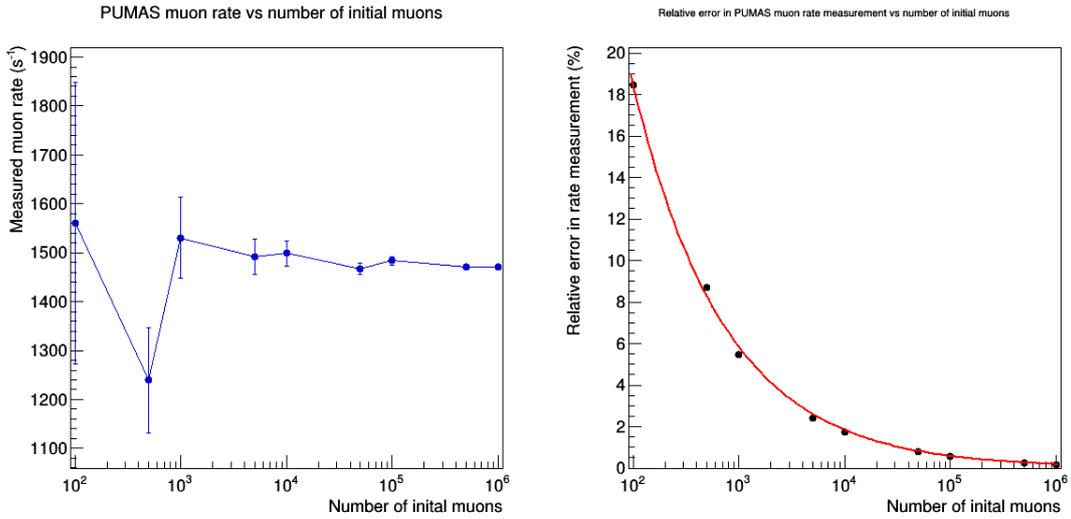
90

Figure 5.12: The relation between a backwards Monte Carlo muon rate calculation, performed using the combined PUMAS-CRESTA software, and the number of initial muons generated N. The left plot shows the muon rate measurement versus number generated, and the right shows the consequent relative error on the measurement for each value of N. The fitted curve (red) is $185/\sqrt{N}$.

was placed at a position close to a single scintillator detector. This cylindrical geometry has been chosen to represent a possible 'second shaft' drilled by malicious actors to access the waste storage tunnels.

This test void had length 400 m and radius 20 m and was orientated parallel to the model's waste access shaft (see figure 5.14). In the xy plane, the void's horizontal centre-to-centre distance from the single 2 m $\times$ 2 m detector was 147.5 m. The void is positioned vertically such that the vertical distance from the plane of the detector to the void base is 15 m, and the vertical distance between the surface and the void top is 25 m. The void therefore subtends a solid angle of ~ 0.15 sr from the detector position, approximately 3% of the total solid angle visible to the detector, which has an aperture of around 75°.

To study the relationship between the detectability of this void and the exposure time, muon flux measurements were taken with and without the void present for a range of exposure times between 1 day and 1 year (see figure 5.15). A larger exposure time corresponds to more muons encountering the detector and hence a more accurate muon flux measurement. A clear separation between the flux measurements taken with and without the void present becomes apparent after around 5 days' exposure, with the flux with the void higher as expected. At longer exposure times the two cases are increasingly clearly distinguished as the excess in muon flux due to the presence of the void is detectable at higher confidence.

The detectability d of a particular void is expressed in terms of the number of standard deviations separating a muon flux measurement taken with a void present (i.e. the 'signal' flux measurement S) from a measurement with no void present ('background', B), with the latter case representing the null hypothesis of no large unknown voids being present in the

91
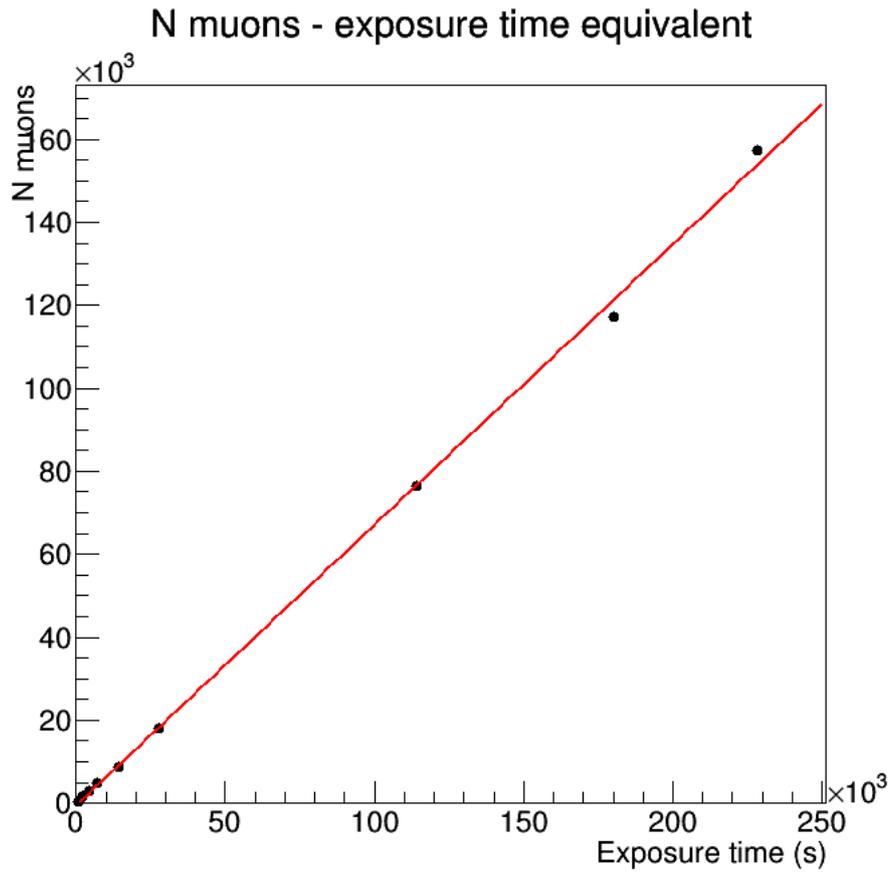
Figure 5.13: Relation between the number of muons initialised in a backwards Monte Carlo of a simulated geological repository versus the equivalent exposure time, as calculated by equating the relative errors on muon rate measurements obtained in each case.

repository:

$$d = \frac{S - B}{\delta S} \tag{5.7}$$

The equivalent of the latter measurement for an experimental study could be obtained with a detailed Monte Carlo simulation making use of the geological and structural knowledge available.

It is clear that by this measure of detectability, the large void above is much more detectable after longer exposure times. To quantify this relationship, d as defined in equation 5.4.2 was calculated for each exposure time in figure 5.15 and plotted against time (see figure 5.16). For each d, the corresponding error δd was propagated as

$$\delta d = \frac{\sqrt{\delta S^2 + \delta B^2}}{\delta S}. \tag{5.8}$$

It is clear that the detectability increases in proportion to the square root of the exposure time, due to the errors on the flux measurements decreasing in proportion to the same. The best fit curve suggests that a detection of this void to 3σ could be achieved (with a single
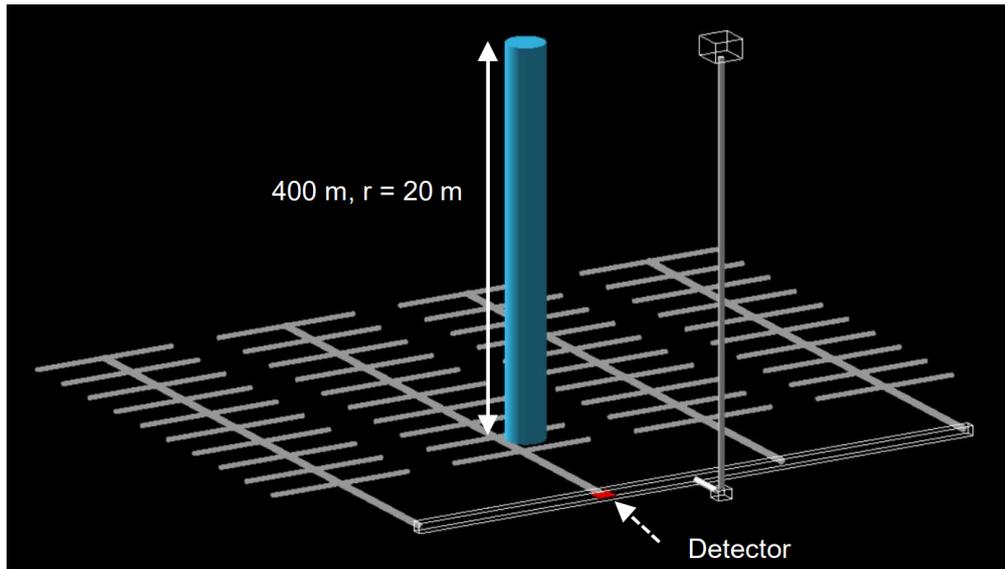
Figure 5.14: View of the simulation setup for the first scenario tested. A large cylindrical void placed into the repository model close to a single detector, as an initial bench-marking exercise for the detectability of voids. The position of the detector is is shown in red, its size is exaggerated for visibility.

detector) after a few days, with a 5σ detection requiring closer to 2-3 weeks.

Another relevant factor determining the detectability of a void is its volume. Muon tracks passing through a larger void will experience a greater difference in opacity and hence the difference in detected muon flux will be greater; there will also be more total muon tracks that pass through the void. To study the relation between a void's volume and its detectability, cylindrical voids of increasing radii were assayed for a fixed exposure time of 4 weeks. Each void was positioned a horizontal distance of 147.5 m from a single 2m $\times$ 2m detector as in the example in figure 5.14. The recorded flux measurements were compared to an equivalent exposure with no void present, then the detectability of each void calculated via equation 5.4.2. The resulting relation between detectability and void volume, expressed in terms of the solid angle that is subtended at the detector position, is shown in figure 5.17. The relation is well fitted by a linear fit. It was found that a void that subtends approximately 0.1 sr, equivalent in this case to a radius of $\sim$ 17 m, is detectable to 3σ after a 4-week exposure time. By comparison, the 20 m radius void described above which subtends $\sim$ 0.12 sr has reached a detectability of 6.5σ after 4 weeks.

5.4.3 Large void - multiple detectors

The results in section 5.4.2 suggest that a single detector with an area of a few square metres may not be adequate for practical void detection, as several weeks' exposure are required to detect even an unrealistically large void to a high level of significance. It was decided therefore to determine how using instead a system of multiple detectors would affect void detectability results. The tested alternative configuration contains 12 detectors, each a 2 m $\times$ 2 m plane as before, positioned within the four long storage tunnels in the repository model (see figure
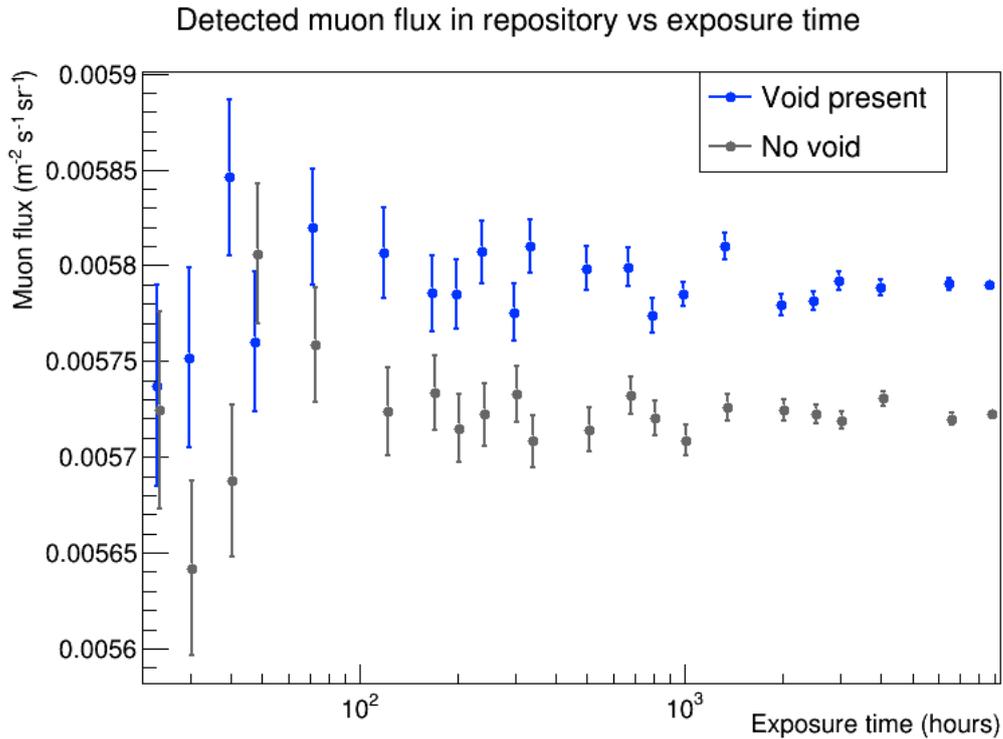
Figure 5.15: Muon flux measurements taken in the repository model at a single detector, with without a large void present (see figure 5.14), for a range of exposure times. For shorter exposures the two cases are difficult to distinguish, but as the exposure increases it becomes clear that the flux with the void present is higher, by around 1%

5.10). Measurements of the muon rate are made by combining muon hit data from all 12 detector surfaces. There are two clear advantages to using several coupled detectors spread out through the repository: firstly the increased detector area, which will increase the number of detected muons and hence allow more precise estimates of the muon flux to be calculated; secondly more of the repository can be assayed as lines of sight from the multiple detectors will pass through a greater proportion of the overburden. Using multiple detectors provides an immediate increase in the precision of the muon rate measurement, reducing the relative error after a 2-week exposure in the GR with no void present to 0.07% from 0.24% with a single detector.

To assess how the detectability of a cylindrical void would be affected by using multiple detector, muon flux measurements were taken in both single and multiple configurations, for geometries with the void present and not present. The void was 400 m in length and had radius 10 m. As with the results in section 5.4.2, the void's vertical position was fixed 15 m above the detector plane, however its x and y position was randomly determined (but constrained to be within the area above the waste storage tunnels). The process was repeated for five configurations, each with a 4-week exposure time. With the single detector, the mean detectability across the sample was $0.6 \pm 0.4\sigma$, whereas with multiple detectors the detectability improved to $3.0 \pm 0.7\sigma$ for the same period of time.
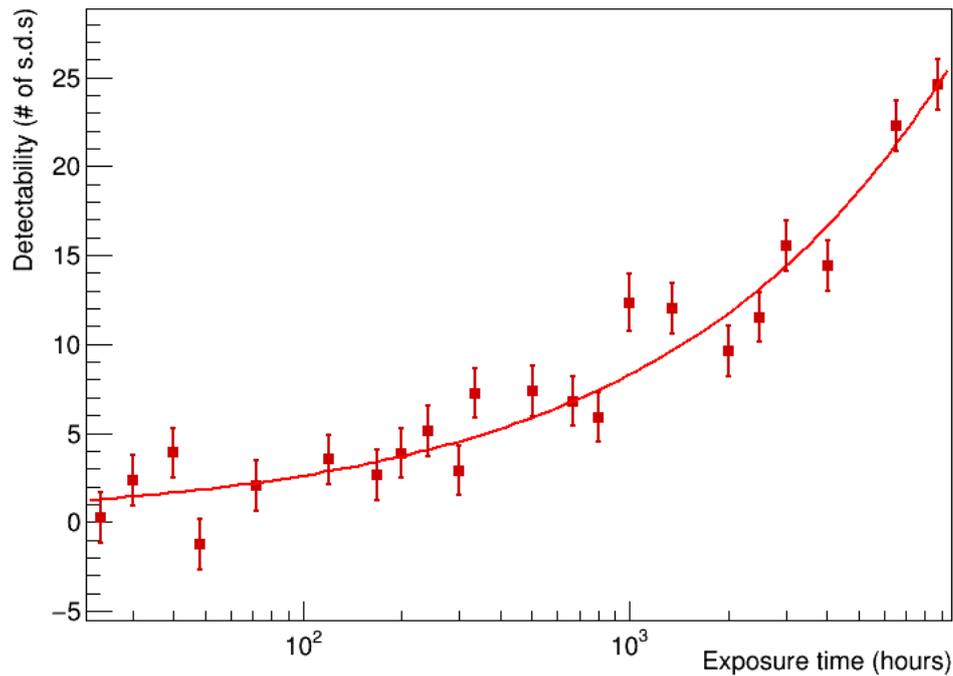
Figure 5.16: Detectability of a large void (figure 5.14) in a GR with a single detector for different exposure times. The fitted curve is proportional to the square root of the exposure time.

An alternative scenario considered was the presence of an unknown large air-filled chamber or cavity in the repository that could threaten the repository's safety, as it could become a potential path for water into the GR during the construction or operational phase. After back-filling and decommissioning of a GR it is important that the long-term safety is ensured by the retention of the radionuclides of the stored nuclear waste from the biosphere for 1 million years. This is the time until the radioactive dose emitted from the long-lived radionuclides has decayed to a non-hazardous limit for the biosphere. Therefore, unknown large air-filled chambers or cavities have the potential to accelerate the migration of the radionuclides of the stored nuclear waste to the biosphere. The feasibility of detecting such a cavity was investigated by adding a spherical void to the repository model with a radius of 30 m, positioned such that its centre is 165 m above the storage tunnel plane and a horizontal distance of 313.9 m from the detector (see figure 5.18). 4-week detectability measurements were taken with both the single and multiple detector configurations. For the former case, the measured detectability was $0.4 \pm 1.3\sigma$, and for the latter it was 2.6 ± 1.3 once again demonstrating a significant improvement in detectability when using a multiple detector configuration.
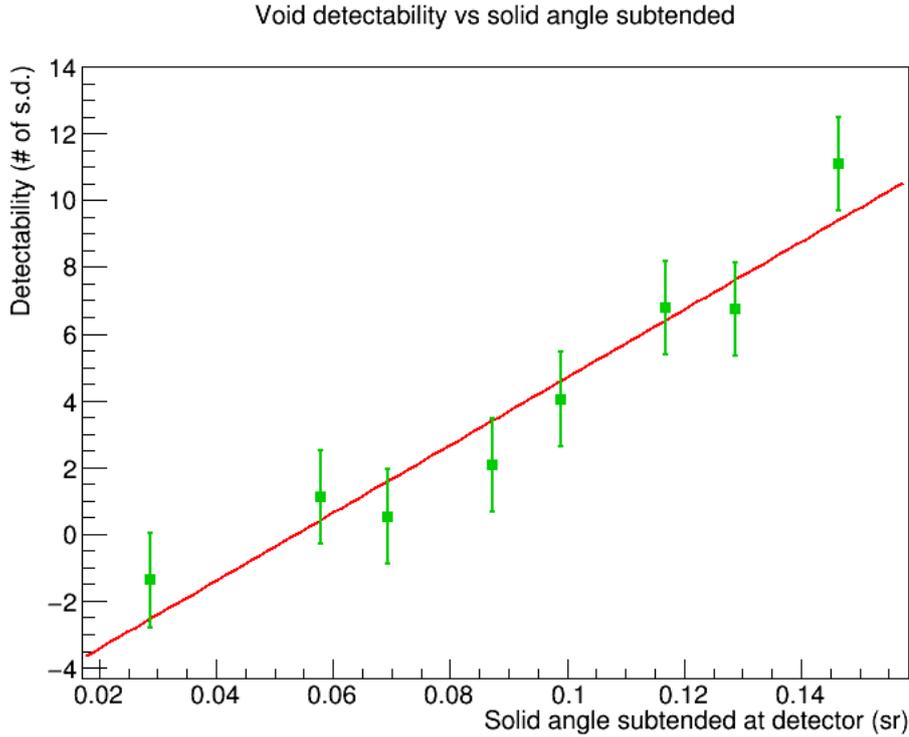
Figure 5.17: Detectability of cylindrical voids after a 4-week exposure time. Each void is positioned close to the 2m × 2m detector, as in figure 5.14. The radii of the cylinder is varied and the detectability calculated in each case from flux measurements with and without the void present. For generality, the void size is expressed in the solid angle subtended.

5.4.4 Small voids

Void shape analysis

The cavity detectability example above also demonstrates that the solid angle that a void subtends at the detector position does not fully characterise the void, in terms of generalising detectability to other scenarios. The 30 m radius sphere tested is 355 m (centre-to-centre distance) from the detector and so subtends a solid angle of $\Omega = A/r^2 \approx 0.02$ sr, where A is the cross sectional area of the sphere and r is the distance. The single detector was able to identify this void to $\sim 0.6\sigma$; a cylindrical void subtending the same solid angle would not be detectable to this level according to the best fit line in figure 5.17. It is clear that the shape of the void must be taken into account.

To compare how the detectability of spherical and cylindrical voids change with void volume, a large set of test geometries with voids was prepared and tested for detectability across a range of exposure times. Each geometry contained a single void, cylindrical or spherical, with the cylindrical voids positioned close to the detector (as in figure 5.14) and the spherical voids centred in the repository (as in figure 5.18). The radii of the voids were varied and the detectability of each resulting volume was calculated, for a particular exposure time. This was then repeated for various exposure times between a few days and several months,

Figure 5.18: Simulated geometry used to test the detectability of a spherical cavity-type void, comparing single and multiple detector configurations. The void (blue) is filled with air, has a radius of 30 m and is positioned 165 m above the storage tunnel plane.

resulting in a set of detectability-volume relations. For each relation, a straight line was fitted and its gradient determined; this then allowed the relation between the volume gradient and the exposure time to be plotted (see figure 5.19).

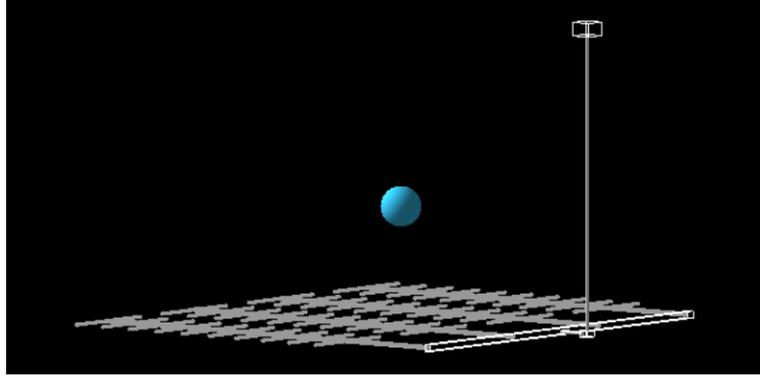This process was repeated for both the cylindrical and spherical void shapes. The obtained relations were then extrapolated to estimate the required exposure time for a void subtending a particular solid angle to be detected to 3σ (with a single detector), for both cases. The results are shown in figure 5.20; it is clear that for solid angles $\lesssim 0.05$ sr much shorter exposure times are required to detect spherical voids than cylindrical 'shaft' voids to a high significance. Therefore detectability results obtained using a particular void shape cannot be naively applied to other voids with different or unknown shapes.

This result indicates a challenge for further void detectability studies, namely that the parameter space under investigation is very large, including as it does the shape and volume of a void in addition to the exposure time, detection area and detector specification. Indeed even if the void's shape, volume and solid angle subtended at the detector are known, the description is still incomplete due to the angular dependence of the muon flux: two identical spherical voids, both positioned the same distance away from a detector, will not have the same detectability unless they are additionally at the same depth in the repository. If, for example, one void is directly above the detector whilst the other is close to the horizon, the former will be detectable in a time orders of magnitude shorter than the latter due to the much greater muon flux at smaller zenith angles. It is necessary therefore to apply constraints to reduce the size of the parameter space and explore fewer scenarios in more detail.

The first chosen constraint was to fix the void shape as being cylindrical, i.e. a shaft-type void, extending from close to the surface to close to the waste storage tunnels. From a safeguarding perspective, a pre-existing or illicitly drilled shaft would represent a potential pathway for waste to be removed from the repository, and detecting the presence of such a shaft would be vital. Consequently the length and vertical position of the shaft-type voids were also fixed, at 400 m and 15 m above the storage tunnels (25 m below the surface) respectively. The areas and details of the individual detectors were also fixed as 2 m × 2 m planes of
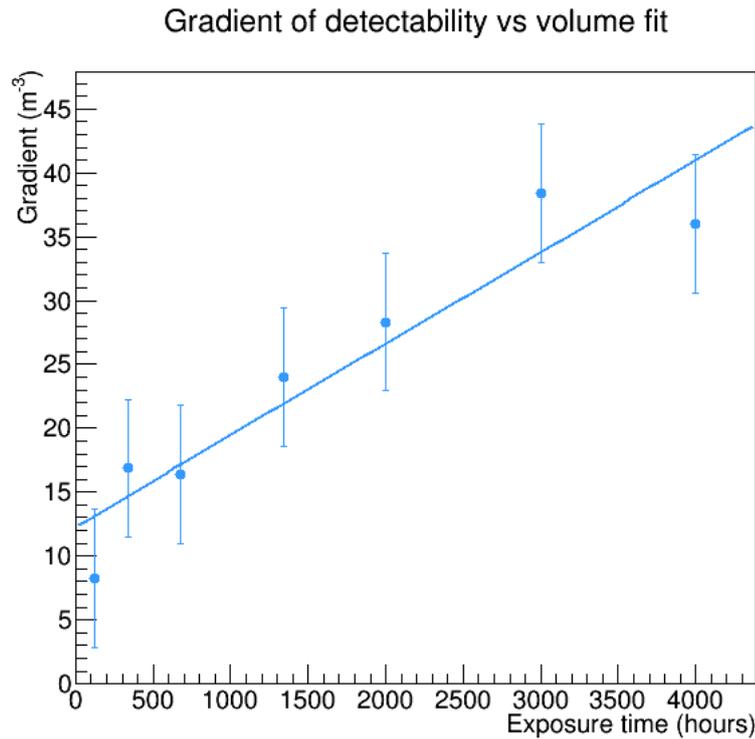
Figure 5.19: Relationship between the gradients of linear fits to plots of void detectability versus void volume, evaluated at a range of exposure times. The voids in this example were spherical and positioned centrally in the repository (see figure 5.18).

polystyrene scintillator orientated perpendicular to the zenith. Using these constraints left three key remaining variables: the void radius, the void position (in the xy plane), and the number of detectors used. These are each varied and the resulting effects on void detectability and imaging are explored in the following sections.

Extrapolation to realistic void sizes

The large void detectability results in sections 5.4.2 and 5.4.3 are interesting and useful for what they suggest about the relations between a void's detectability in a repository, its size, and the detection area and exposure time used. However, the voids concerned are mostly unrealistically large. The illicit tunnelling of a shaft-type void with a radius of tens of metres would be virtually impossible to conceal; a pre-existing void of that size in the repository would be easily detectable by seismic imaging or other assay methods. It is therefore necessary to consider the ability for muon radiography to detect smaller voids in order to get practically applicable results.

However, simulating the long exposure times required to detect such voids is not trivial. Even using a backwards Monte Carlo-based software (see section 5.2.2), simulating 1 hour of muon transport with a single detector requires a computation time (on a standard desktop computer) of approximately 2 minutes. The use of a powerful computing cluster reduces this
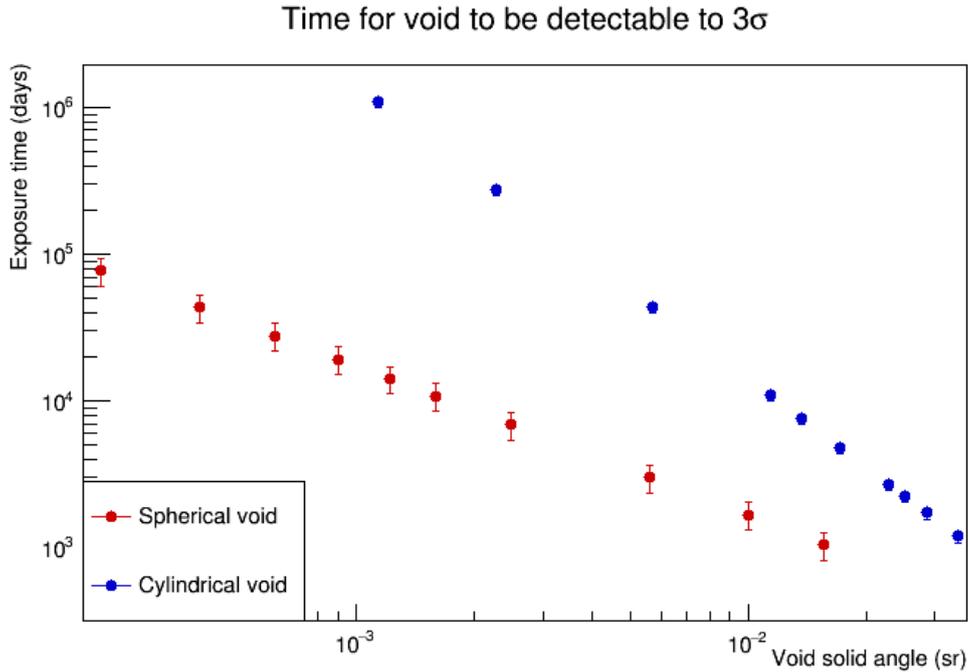
Figure 5.20: Estimated exposure times required to detect voids of various volumes to 3σ, using a single detector. The void volume is expressed as solid angle subtended at the detector position to account for the distance between detector and void. Results are shown for both spherical (red) and cylindrical (blue) (with 400 m length) voids; spherical voids require shorter exposure times to detect.

by around two orders of magnitude, but simulating several months of exposure for several detectors (each requiring a separate simulation, see section 5.3) will still require around a day of computing time. Testing a large number of scenarios e.g. different void positions and radii with long exposures would therefore be quite impractical.

An alternative approach is to extrapolate results obtained from shorter exposures of larger voids to estimate the exposure times required to detect smaller voids. This has the advantage that many different combinations of void position and radius can be tested relatively quickly, allowing useful results to be obtained within practical times. The primary relation investigated was the detectability of voids with varying radii and a fixed number and arrangement of detectors; this study was repeated for several exposure times, giving relations that could then be extrapolated to smaller void radii than were directly simulated.

For this study the void position must also be accounted for. Previously described results in sections 5.4.2 and 5.4.3 used either a fixed void position (varying the void's size and/or the exposure time) or a small number of random positions (in order to demonstrate the improvement in detection power with more detectors). It was decided for the extrapolation study to extend the latter idea and use randomised void positions. Allowing the shaft-type voids to be positioned anywhere within the repository generalises the detectability results; as previously discussed, the solid angle that a void subtends is not sufficient to fully characterise it.

To implement this study in practice, scripts were written to automatically print JSON

table-based geometry description files that would contain a cylindrical void of a chosen radius and vertical position but have its x and y coordinates set randomly, constrained to be within the area covered by the waste storage tunnels. This was repeated for void radii in 5 m increments between 10 and 40 m. The chosen detector arrangement was to use four detectors only, positioned on the 'corners' of the repository (i.e. detectors 0, 3, 10 and 11 in figure 5.10). Using four detectors instead of the full twelve reduces the required computation time for each scenario and so allows more throws (i.e. more void positions) to be tested.

Ultimately a sample of 50 throws was used for each detector radius; in each case muon flux measurements were taken for simulated exposure times of 1, 2, 5 and 10 days. Each flux was compared to a flux measured for a long (2 month) exposure of a repository containing no void. The voids' detectability was then calculated using equation 5.4.2 in the same way as was done for previous results in sections 5.4.2 and 5.4.3. Results are shown in figure 5.21. Each line represents the mean detectability across all 50 throws for each void radius and one exposure time. In each case, the void detectability increases approximately linearly with void radius. The gradient of the increase is steeper for longer exposure times. The displayed errors are the standard error on the mean of the 50 samples.
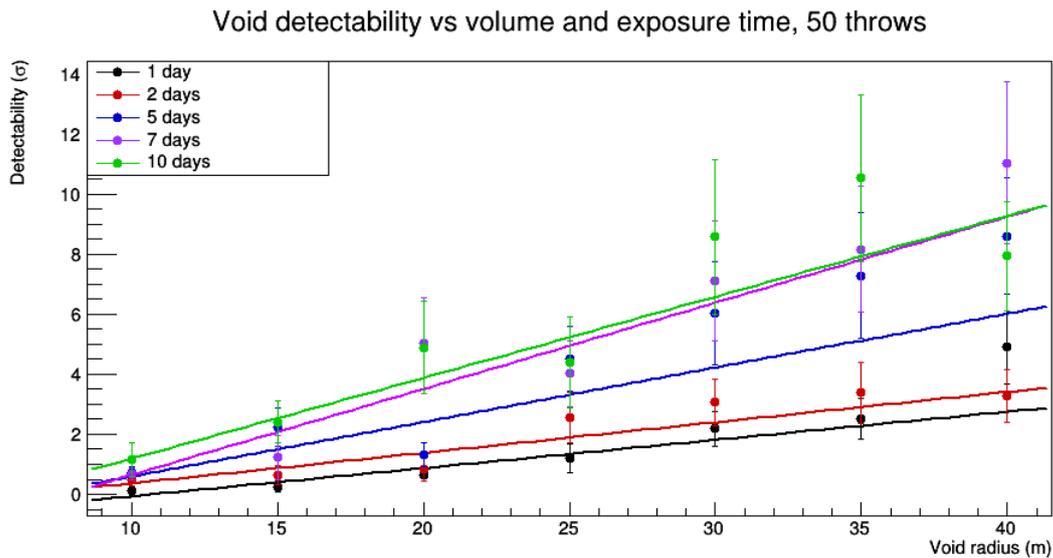
Figure 5.21: Mean detectability of samples of 50 cylindrical, 400 m-long voids with randomly determined positions and four detectors, as tested with voids of various radii between 10 and 40 metres. The process has been repeated for exposure times between 1 and 10 days. In each case a straight line has been fitted to the data to show the trend. The voids become more detectable as the radius increases and for longer exposure times; the determined relations can now be extrapolated to smaller voids.

The next stage of this study is to determine the relation between the gradients of the best-fit lines and the exposure time. The results are shown in figure 5.22. The gradient increases monotonically and approximately linearly with exposure time; the errors are obtained from the uncertainty on the gradient of each best-fit line. This propagates into an uncertainty on the gradient of the best-fit line to the gradient-time relation itself, which can be used to provide a coarse estimate of the exposure time required to detect a void of a given radius to a

given level of significance.
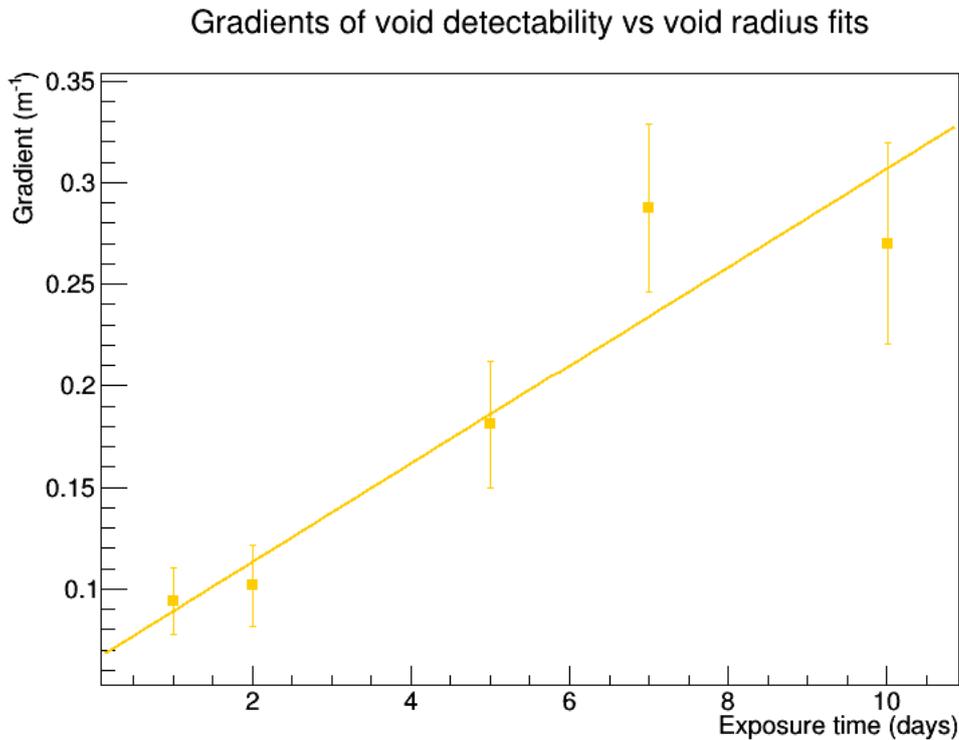
Figure 5.22: Relation between the gradients of the detectability versus radius best-fit lines (figure 5.21) and the exposure times tested.

Finally, the extrapolation itself is done by choosing a significance level (e.g. 3σ) and a void radius, then calculating a corresponding detectability versus radius gradient (in m^{-1}) as significance/radius (i.e. $3\sigma/r$), and inserting this value into the gradient-time relation of figure 5.22. The resulting estimated exposure times to detect voids with radii 1, 2 and 5 m to 3 and 5σ level are shown in table 5.2; they indicate that an average of several months would be required to detect a 1 m radius shaft-type void, at a random position and with four detectors present, to a 3σ level of significance. As geological repositories operate on timescales of many years, an exposure time of $\sim$ 6 months is not overtly impractical (and could be further improved with more detectors), suggesting that muon radiography is a feasible technique for detecting shafts of realistic size. However, this assumes that there is sufficient time between a shaft of this type being illicitly tunnelled, and a major safeguarding breach (e.g. theft of stored waste) occurring. If several months are required to confidently detect the shaft's presence then this may be too slow from a safeguarding perspective.

101

Void radius (m)	Time to detect to 3σ (days)	Time to detect to 5σ (days)
1.0	118 ± 22	201 ± 37
2.0	58 ± 11	99 ± 18
5.0	22 ± 4	38 ± 7

Table 5.2: Estimated required exposure times to detect cylindrical voids with radii of 1, 2 and 5 m, using four detectors positioned at the corners of the repository.

5.5 Opacity

The results described above are primarily concerned with identifying the presence of voids in the repository globally, i.e. distinguishing one case from another without specific data as to the location of the void. To progress further, it is necessary to incorporate directional information. The goal is to calculate $\varrho(\theta, \phi)$, the *opacity* - the integral of density along a line of sight - as a function of angle; a large void in an overburden would be visible in such an image as a region of lower opacity due to the lower density material along the corresponding lines of sight. With a single detector (with an area much smaller than the features of interest) it would not be possible to distinguish the position of the void along the line of sight, or its extent (although this could be estimated by assuming a particular average density value in the void, e.g. that of air). This problem can be ameliorated by combing opacity data from multiple detectors. The fullest evolution of this approach is to use the opacity data as input to an imaging algorithm to produce full 3D density maps (see section 5.6).

If a certain number N muons are detected within some angular bin, at a detector viewing some region of interest (e.g. placed underground), an estimate of the opacity along the corresponding line of sight can be obtained by considering *muon transmission* (T): the ratio of N to N_{free}, where N_{free} is the equivalent number of muons measured in the free-sky case. The exposure time and detector characteristics must also be accounted for. Specifically, N can be expressed as [97]

$$N(\theta, \phi) = \Delta t \cdot S_{\text{eff}}(\theta, \phi) \int_{E_{\min}}^{\infty} f(\theta, \phi, E) dE. \tag{5.9}$$

Δt is the elapsed exposure time and $E_{\min}$ is the minimum muon energy necessary to pass through the opacity along that line of sight and reach the detector. $S_{\text{eff}}(\theta, \phi)$ is the 'effective detector area', which takes account of a detector's physical area, geometrical acceptance and detection efficiency for a particular line of sight and muon energy. For the simplified simulated detectors used for the studies described in this chapter (see section 5.3), S_{eff} can be assumed to be the same for both underground and free-sky measurements and thus will cancel out during the opacity calculation described below.

$E_{\min}$ is a key factor. The number of muons reaching a detector underground from a particular direction will be attenuated by the overburden; only muons with sufficient energy to cross the opacity along that line of sight will be detectable. Therefore, the ratio of the number of detected muons underground to the number detected in the free-sky case will be

equal to the corresponding ratio of the integral of the muon differential flux, i.e.

$$\frac{N(\theta,\phi)}{N_{\text{free}}(\theta,\phi)} = \frac{\int_{E_{\min}}^{\infty} f(\theta,\phi,E)dE}{\int_{E_0}^{\infty} f(\theta,\phi,E)dE} \tag{5.10}$$

where E_0 is the smallest detectable muon energy. Hence, for a particular line of sight (θ,ϕ), if N and N_{free} are measured and $f(\theta,E)$ is known, it is possible to calculate $E_{\min}$.

The final step is to use $E_{\min}$ to obtain an estimate of the opacity. This is done by considering the muon energy loss in matter. As discussed in section 2.3, the mean stopping power can be expressed in terms of functions a and b, which encode electronic and radiative processes respectively, via

$$\left\langle -\frac{dE}{dx} \right\rangle = a(E) + b(E)E. \tag{5.11}$$

If a and b for a particular material are known, an estimate of the maximum range $R(E_{\min}$ of muons with energy $E_{\min}$ can be found by inverting and integrating this expression between E_0 and $E_{\min}$:

$$R(E_{\min}) = \int_{E_0}^{E_{\min}} [a(E) + b(E)E]^{-1} dE \tag{5.12}$$

This can then be converted to an opacity with the material density ρ, $\varrho_{\text{CSDA}}(E_{\min}) = R(E_{\min})\rho$. This estimate of opacity is called the Continuous Slowing Down Approximation (CSDA), as it assumes that muon energy loss is deterministic and depends only on the stopping power of the material; in reality, muon energy loss is a stochastic process, leading to energy fluctuations and range straggling. This is a variation in the range of muons of a particular energy due to the random nature of the energy loss processes they undergo. Nevertheless ϱ_{CSDA} is a useful approximation to the true opacity in some situations.

Tabulated values of ϱ_{CSDA} as a function of energy for many materials can be accessed in databases e.g. [8]. For a custom material, such as the granite used in our repository model (section 5.3.2), it is possible to calculate equivalent CSDA tables using a Monte Carlo transport code; PUMAS includes functions for generating stopping power tables for new materials. This method was used to calculate the CSDA range for muons in granite between 1 MeV and 1 EeV, shown in figure 5.23 (right). The muon differential flux function $f(\theta,E)$ used was the empirical relation from [5], see figure 5.23 (left).

Thus, the full process for calculating an opacity estimate along a particular line of sight was as follows:

1. Use the PUMAS-CRESTA software (section 5.2.3) to obtain muon rate measurements as a function of angle, for both a measurement underground and a free-sky measurement, using the same angular binning choices.

2. Calculate the ratio of the two rates.

3. Integrate $f(\theta,E)$ between $\theta_{\min}$ and $\theta_{\max}$ (the zenith angle boundaries of that angular bin), then between test values of E and ∞, until the ratio of the integral to the total
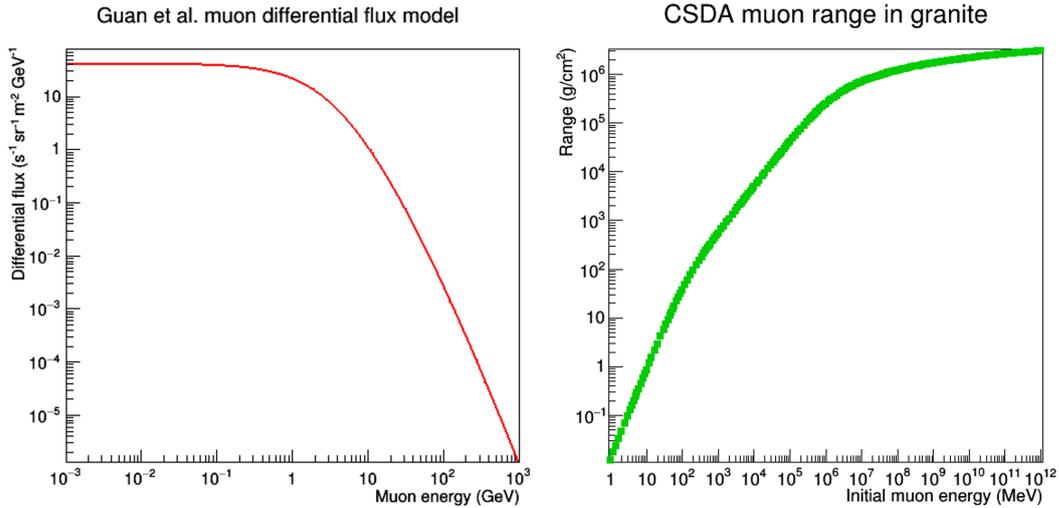
Figure 5.23: Key functions in calculating an opacity estimate using the CSDA. Left: muon differential flux function, from [5]. Right: CSDA muon range in the custom granite material used as the repository model overburden, calculated using PUMAS [31].

integral is approximately equal to the ratio of the rates, i.e. until equation 5.10 is satisfied. Label the corresponding value of E as $E_{\min}$.

4. Use the tabulated ϱ_{CSDA} data for that material to obtain the CSDA opacity that corresponds to $E_{\min}$, using linear interpolation between the tabulated values.

To test this method, simulated vertical muon rate measurements were taken under various depths of granite overburden between 10 and 500 m, with exposure times of 2 weeks, alongside equivalent free-sky measurements. The two rates were then used to calculate the CSDA opacity along the vertical direction, and subsequently a reconstructed depth by dividing the opacity by the known material density. A comparison of the reconstructed and true depths is shown in figure 5.24. It is clear that the reconstructed depth is a close match to the true depth but is also a slight underestimate; approximately 80% of the true value.

This underestimation of the opacity is likely a result of the reliance on the CSDA. The approximation loses accuracy at greater depths of rock due to the greater change in the muon intensity from the free-sky case [8]. Thus, the opacities used in the final section of this chapter suffer from an inaccuracy which is exacerbated at greater depths and at higher zenith angles. Nevertheless, useful data concerning the use of imaging algorithms with muon radiography data was still obtained. For example, figure 5.25 shows angular opacity data for a detector a horizontal distance of 78.6 m from a 10 m radius cylindrical void; despite the underestimated opacity values, the void location can be seen (at approximately $-2.9 < \phi < -2.7$), data of this type can then be used for a 3D reconstruction (see section 5.6). A more rigorous approach would be to use a dedicated Monte Carlo simulation to relate muon transmission factors to opacity directly; this has not been implemented but is a clear possibility for extension. Alternatively, the empirical formula obtain in [7] (equation 3.2) also provides a reasonably accurate relation between a flux measurement and an opacity; figure 5.28 shows a comparison
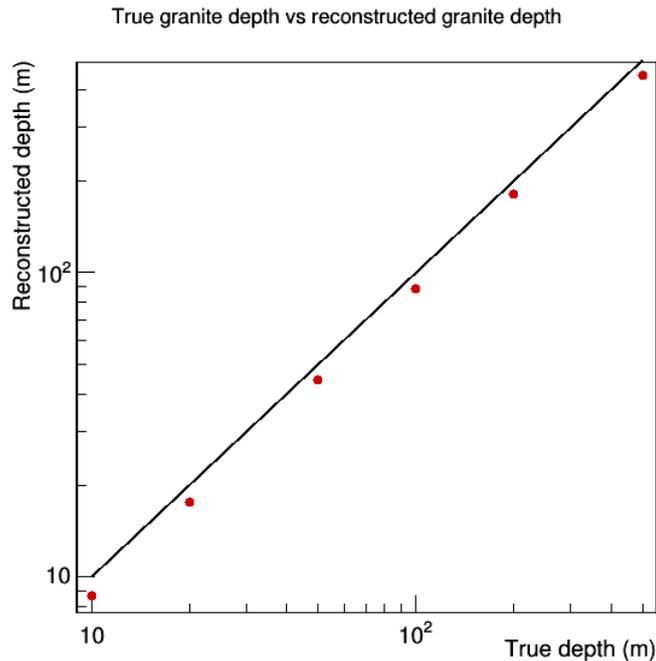
Figure 5.24: The thickness of a granite overburden as reconstructed from the method described above to calculate the CSDA opacity, compared to the true thickness (red). The black line is $y = x$. The reconstructed depth is a close match to the truth depth but tends to underestimate; confidence intervals are present, calculated by propagating the uncertainty on the flux measurements, but are smaller than the markers.

of obtained opacities with the empirical formula and the CSDA method described above.

5.6 Void imaging via SART

To produce a 3D image from multiple 2D opacity projections, it is necessary to implement an imaging algorithm. There are several options available, often algorithms adapted from the highly analogous medical imaging field; X-ray radiography is in many aspects the same technique as muon radiography on a smaller scale and with artificially introduced particles. In general all imaging algorithms take opacity data as input, and attempt to determine the density distribution in the volume of interest that would produce that opacity. For example, [98] use a linear regularisation method on opacity data from four muon radiographs to image the density above an underground tunnel at Los Alamos.

5.6.1 SART algorithm overview

Another approach that has seen great success in recent years is to use algebraic reconstruction techniques to solve the density problem via an iterative process. The key algorithm is the Simultaneous Algebraic Reconstruction Technique (SART), an evolution of the Algebraic Reconstruction Technique (ART) algorithm [99] commonly used in medical tomography applications. This powerful algorithm attempts to iteratively solve a set of matrix equations

Figure 5.25: Opacity as a function of angle calculated from a PUMAS-CRESTA simulation, with detector close to a large cylindrical void. The void is in an azimuthal direction $-2.9 \lessapprox \phi \lessapprox -2.7$.

that relate a vector of density values ρ (i.e. the density distribution within a set of 3D voxels) to opacity measurements $\mathbf{o}$ via a 'distance matrix', $[a]$, i.e. solve $\mathbf{o} = [a]\rho$.

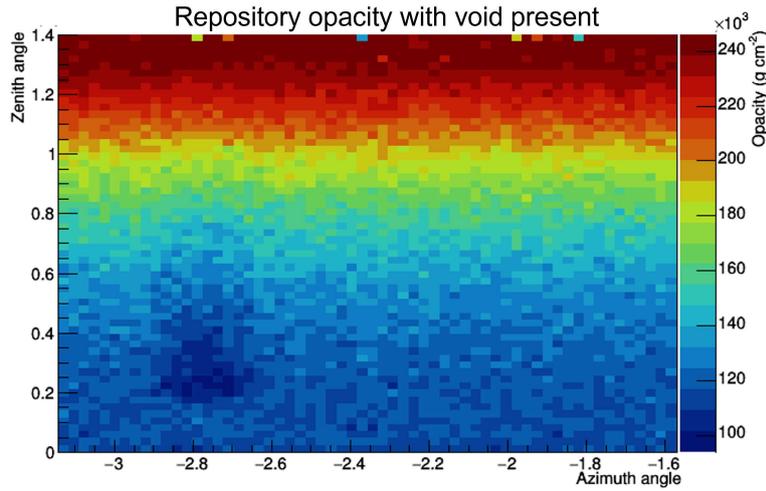Previous implementations of SART for muon radiography include [100] (imaging of a box containing lead bricks, within the muon telescope itself) and [101] (imaging of a ~ 50 m scale concrete cube with internal voids, with several external detectors). They have showed great success at reconstructing densities with muon radiography data and the SART algorithm. For the results presented in this section, the goal was to implement a version of SART applied to the much larger repository volume, and study its efficacy and how it varied with different voxelisations and numbers of projections used. The methods used draw heavily from those of [101] due to their very thorough explanation of the technique and the necessary computations.

The key equation at the heart of SART is $\mathbf{O} = [A]\rho$. Here $\mathbf{O}$ is the 'opacity vector': a column vector made by vertically concatenating the vectors containing the measured opacities along each line of sight from each detector. In other words, if there are M detectors, each producing opacity data (calculated from muon rate measurements following the methods described in section 5.5) binned into n_m angular bins and expressed as a column vector $\mathbf{o}_m$, then $\mathbf{O}$ is the vertical concatenation of the $\mathbf{o}_m$ and has $N_m = M \times n_m$ elements. The density vector ρ contains the density in each voxel in the image volume; it has N_v (number of voxels) elements. The distance matrix $[A]$ is thus a $N_m \times N_v$ matrix. The calculation and use of $[A]$ is described in section 5.6.2.

SART's *a priori* steps are to calculate O and $[A]$, and to create an initial density vector ρ; this can vary but the simplest method is to have this represent the null hypothesis e.g. setting all the density values to the rock density when attempting to image voids in an overburden.

The iterative step then updates the density value in each voxel v via $\rho_v^j = \rho_v^{j-1} + \Delta\rho_v^j$, where

$$\Delta\rho_v^j = \frac{1}{\sum_{m=0}^{N_m} w_m} \sum_{m=0}^{N_m} w_m \frac{O_m - O_m^{j-1}}{\mathbf{A}_m \cdot \mathbf{A}_m} A_{m,v}. \tag{5.13}$$

Here w_m is present to weight the values by their statistical accuracy. For an experimental measurement, this could be set by propagating the standard error on the muon rate measurement into the opacity values. The goal of the iteration is to minimise the difference between the measured opacity vector O_m and that calculated from the current version of the density vector, O_m^{j-1}. This is given by

$$O_m^{j-1} = \sum_{v=0}^{N_v} A_{m,v} \rho_v^{j-1}. \tag{5.14}$$

ROOT's 2D and 3D histogram objects are suitable for storing the opacity and density data respectively, for example the density vector ρ can be made by 'unfolding' a 3D histogram with bin values representing the density in each voxel, making use of ROOT's global bin indexing system. The algorithm's output can then be reformed into a 3D histogram for viewing.

5.6.2 Distance matrix

The distance matrix is a very large and very sparse matrix that encodes the relation between the angular binning of the opacity data and the world volume. It has as many rows as there are angular bins (across all the detectors) and as many columns as there are voxels in the world volume i.e. it has dimensions $N_m \times N_v$. Each row of the matrix corresponds to a particular angular bin. The N_v elements of that row are then the mean distance that muons in that bin pass through each of the voxels in the world, under the approximations that the muons are purely ballistic and that their angular distribution is flat within each bin. It is an extremely sparse (generally $> 99\%$ of the elements are zero) matrix as, unless the binning is extremely coarse, muons in any particular angular bin will only pass through a small fraction of the world voxels.

A key feature of $[A]$ is that its size and elements are a function only of the number of detectors, the chosen angular binning of each detector, and the world characteristics i.e. the number and size of the voxels and the positions of the detectors. It is thus independent of the material distribution in the voxels of the world, and so the distance matrix can often be calculated in a single *a priori* step and then applied for imaging many scenarios, e.g. a distance matrix calculated for the repository geometry with 12 detectors would not need to be recalculated for the addition or transformation of voids.

Voxel traversal

To calculate the matrix, one must choose an angular bin (i.e. one of the N_m rows of the distance matrix), generate some number of throws ($10,000$ was used as default) each with a zenith and azimuth angle selected (uniformly) from the ranges defined by the bin. The multiple throws are necessary to sample the possible set of trajectories for that angular bin.

Each thrown trajectory is then projected through the voxel grid through the world until an edge is reached. The voxel traversal algorithm implemented was the efficient method of [102]; in essence, one moves along the trajectory until a voxel boundary is reached, then determines which new voxel has been entered, until an edge of the world is encountered. This algorithm can be easily adapted to work with the bin-indexing system of a ROOT 2- or 3-dimensional histogram. The distance passed through each encountered voxel is logged; when all the traversals have terminated, the mean distance in each voxel is calculated, and stored in the N_v columns of the matrix. This is then repeated for the remaining $N_m - 1$ bins.

The process of voxel traversal and distance calculation is visualised in two dimensions in figure 5.26. The left plot shows the voxel traversal of a single throw, and the right plot shows the mean distance traversed in each voxel. For each throw, a temporary voxel map (initialised with null entries in all voxels) is updated to increment the total distances traversed in each voxel for that angular bin. A separate map keeps count of the number of throws that encounter each voxel. After all throws for an angular bin, the former map is divided by the latter to give the whole row of the distance matrix.
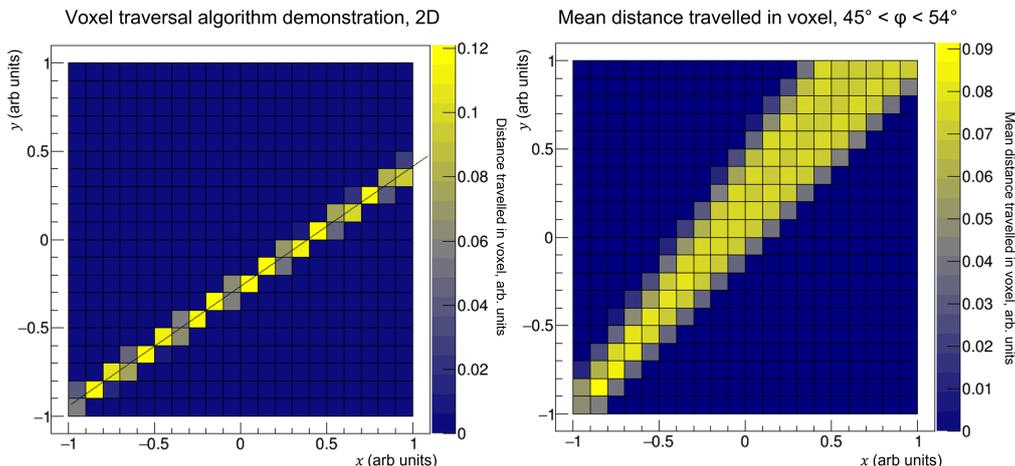
Figure 5.26: Visualisation of the distance matrix calculation, in two dimensions. Left: a single thrown 'muon' traverses the voxel grid, with the distances that it passes through each voxel displayed. Right: after repeating this process over many throws, the mean distance in all the voxels for that angular bin is calculated.

Sparse matrix formulations

In practice, the distance matrices are too large to calculate and store in their entirety. Due to their extreme sparsity it is appropriate to implement some form of sparse matrix formulation instead. This is in general a set of vectors that encode all the information stored in the sparse matrix but at a fraction of the memory; one simple example would be a set of three vectors with one containing the non-zero matrix elements, one their row indices and one their column indices. Often, a sensible choice of the formulation allows also for very efficient operations with the matrix.

A very useful pair of sparse formulations are 'Compressed Column Storage' (CCS) and

'Compressed Row Storage' (CRS). These are both sets of three vectors, one vector being a list of the matrix's non-zero elements, but for the former in order of column and the latter in order of row. These will be denoted $\mathbf{V}_{\mathrm{CCS}}$ and $\mathbf{V}_{\mathrm{CRS}}$ respectively. For CCS, the other two vectors are the row indices of each non-zero element ($\mathbf{I}_{\mathrm{CCS}}$), and a final vector ($\mathbf{S}_{\mathrm{CCS}}$) that contains the cumulative sum of the number of non-zero elements at the end of each column, with an additional 0th element of 0 by convention. For CRS, $\mathbf{I}_{\mathrm{CRS}}$ contains the column indices of each non-zero element, and $\mathbf{S}_{\mathrm{CRS}}$ the cumulative sum of the number of non-zero elements at the end of each row. An example small sparse matrix and its corresponding CCS and CRS vectors is

$$\begin{bmatrix} 0 & 0 & 2 & 1 \\ 5 & 0 & 0 & 0 \\ 0 & 1 & 6 & 0 \end{bmatrix} \tag{5.15}$$

$$\mathbf{V}_{\mathrm{CRS}} = (2,1,5,1,6), \ \mathbf{I}_{\mathrm{CRS}} = (2,3,0,1,2), \ \mathbf{S}_{\mathrm{CRS}} = (0,2,3,5) \tag{5.16}$$

$$\mathbf{V}_{\mathrm{CCS}} = (5,1,2,6,1), \ \mathbf{I}_{\mathrm{CCS}} = (1,2,0,2,0), \ \mathbf{S}_{\mathrm{CCS}} = (0,1,2,4,5) \tag{5.17}$$

It is clear that the V and I vectors will have size NNZ (number of non-zero elements), whereas $\mathbf{S}_{\mathrm{CRS}}$ will have size $m+1$ and $\mathbf{S}_{\mathrm{CCS}}$ size $n+1$.

These formulations are particularly suitable when matrix-vector multiplication is necessary. In the iterative step of SART, as discussed above (equation 5.14), it is necessary to calculate e.g. $O_m \sum_{v=0}^{N_v} A_{m,v} \rho_v$. In a CRS formulation, this can be formulated so only the non-zero elements of $[A]$ are considered, i.e.

$$O_m = \sum_{v'=S_m}^{S_{m+1}} V_{v'} \times \rho_{I_{v'}} \tag{5.18}$$

For the distance matrix calculation and storage, it was ultimately found that using a combination of CRS and CCS sparse matrix formulations was necessary to efficiently iterate the algorithm. In practice, the generation of the distance matrix itself was designed so as to automatically generate the data in the form of a set of CRS vectors; in fact M separate distance matrices are calculated for the M muographies, which are then vertically concatenated by appending the V and I vectors and adding the final element of the previous S vector onto all the elements of the next. A separate process was then used to generate the corresponding CCS vectors directly from the CRS vectors. Both of these steps were slow, taking up to several hours for the voxelisation and angular binning choices used in the results presented below. However they need only be done once for each particular combination of detector and world size.

5.6.3 SART imaging results

Blocks geometry imaging

Before attempting to image the repository geometry with the SART algorithm, a simpler test geometry was used to debug the algorithm and test its performance. This was a large air-filled volume containing several blocks of granite as imaging targets, and a system of (initially) five detectors on the base beneath the blocks. Figure 5.27 shows the geometry, its dimensions and the positions of the detectors. The world volume is $3 \times 3 \times 0.5$ km and divided into cubic voxels of side length 10 m; however to reduce the computations required only a smaller volume centred on the blocks was actually imaged i.e. only these voxels were part of the distance matrix. This smaller volume was $1 \times 1 \times 0.4$ km; there is a 100 m gap between the vertical position of the detectors and the base of the imaged volume. N_v is therefore 4×10^5.
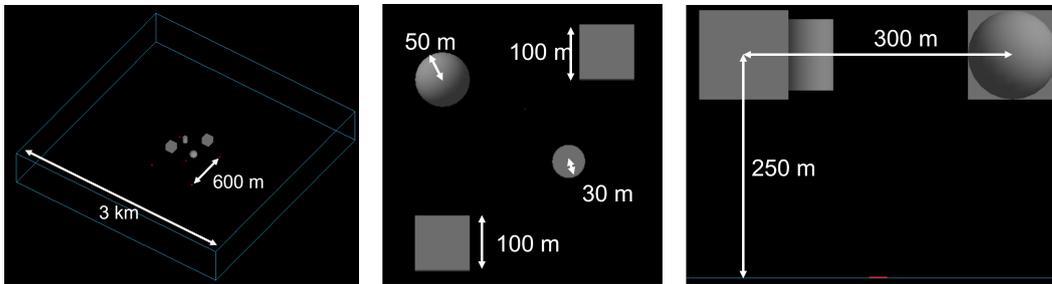

Figure 5.27: 'Blocks' geometry used for SART development and testing. The four solid volumes - two cubes, a sphere, and a cylinder - are made of granite (see section 5.3.2), the surrounding volume is filled with air.

For each detector, a PUMAS-CRESTA backwards Monte Carlo simulation was performed to calculate the muon rates along different lines of sight. This data was then used to estimate the opacity in each direction, using the method of section 5.5. The chosen angular binning was 50×50, with the azimuth angle ϕ between 0 and 2π and the zenith angle θ between 0 and $\pi/3$. N_m was therefore $5 \times 50 \times 50 = 12,500$; hence the total size of the distance matrix was $N_m \times N_v = 5 \times 10^9$. However $> 99.9\%$ of these elements are null, and hence the sparse matrix formulation described above hugely reduces the necessary storage and allows for efficient computation with the matrix elements.

The opacity as a function of angle for detector 0, located directly underneath the blocks, is shown in figure 5.28 (left). It is clear that the blocks have been clearly identified and appear as sharp regions of increased opacity. The density values themselves are of the right order of magnitude (the blocks have density 2.75 g cm^{-3} and are of ~ 100 m scale) but suffer from an underestimation due to the use of the CSDA, as described in section 5.5. This can be seen by comparison to the opacity as calculated from the flux directly, making use of the empirical formula of equation 3.2.

Before using this opacity data in the SART formulation, an additional step was implemented to better understand the output images. As the distance matrix $[A]$ directly relates opacity to density via $\mathbf{O} = [A]\rho$ (see section 5.6.2), if $[A]$ has been calculated, a custom density vector ρ_{true} can be inserted into this equation to obtain $\mathbf{O}_{\text{true}}$, the 'true' opacity.
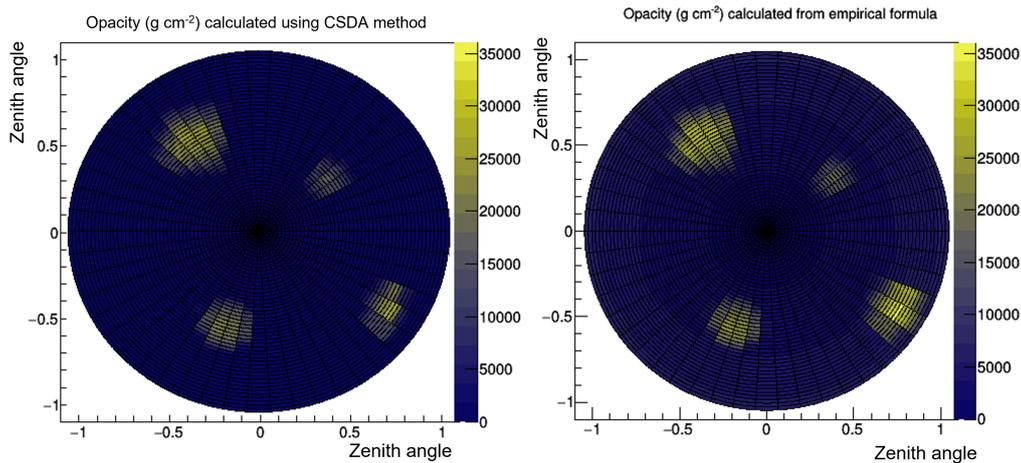
Figure 5.28: Opacity as a function of angle, on polar axes, for a detector positioned underneath the blocks (figure 5.27). The angular binning is 50×50. The opacities are calculated by using a backwards Monte Carlo simulation to obtain the muon flux as a function of angle. Two methods are compared: using a free-sky measurement to obtain the CSDA opacity as described in section 5.5 (left), and using an empirical formula (equation 3.2) (right) followed by then the method described in section 5.5, which introduces an inaccuracy due to the use of the CSDA.

This is an approximation to the true mean density length along each line of sight from that position; it is free of the inaccuracies introduced from approximations such as the CSDA and from the inherent Poisson noise on the muon rate measurement. However, it is still affected by the limitations of the distance matrix calculation itself: a finite number ($10,000$ by default) of trajectories arriving in each angular bin are tested to obtain the rows of the matrix. Nevertheless $\mathbf{O}_{\text{true}}$ can be used as input data to SART to study the algorithm's performance when free of the error on the rate measurement and perform relatively fast tests of e.g. different detector positions without requiring a new simulation.

An example of this process is shown in figure 5.29. The true density ρ_{true} for the 'blocks' geometry has been calculated by manually setting the bin values in the 3D map corresponding to the granite bodies to the granite density and the remaining bins to the air density; the map is then converted into a vector of densities, which is acted on with the distance matrix $[A]$ to calculate $\mathbf{O}_{\text{true}}$. The opacity calculated by this method agrees closely, in terms of the positions and boundaries of the blocks, with the previously shown (figure 5.28) method from simulation. The values however are free of the error introduced by the CSDA. It should also be noted that the opacities of curved objects such as the sphere and tube granite bodies cannot be represented perfectly by this method, as the density map covering these regions is voxelised.

At this stage, with $[A]$ and $\mathbf{O}$ ('true' or simulated) available, it is possible to run the SART algorithm and obtain a 3D image. The chosen initial density map for the 'blocks' scenario had air density in every voxel. With the distance matrix expressed in the sparse formats, the algorithm iterations were fast, taking < 1 s per iteration allowing several thousand iterations to be performed efficiently.
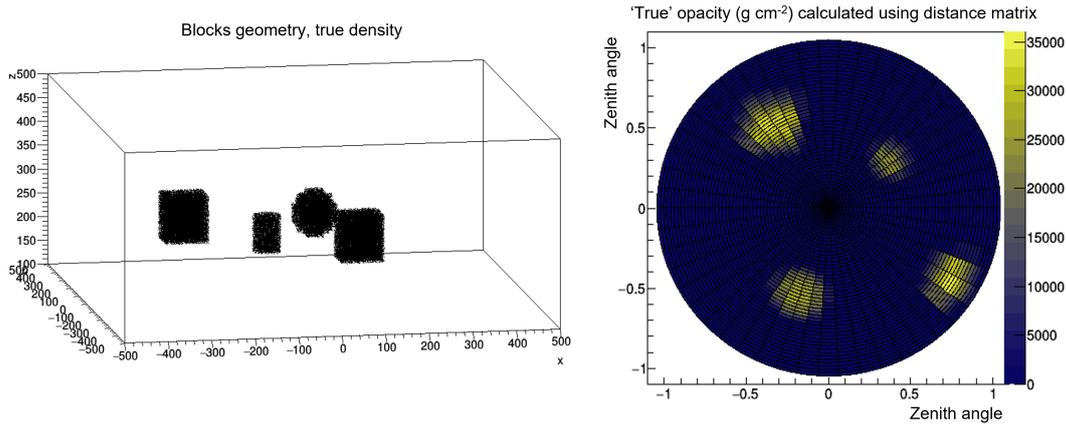
Figure 5.29: Demonstration of the 'true density' method. A 3D map (left) of the imaging volume with the voxel density values set to those of the materials (via boolean expressions) is converted to a density vector that can be acted on with the distance matrix, producing a 'true opacity' vector. This can then be viewed as a polar projection (right) and input into the SART algorithm for 3D imaging.

Results of applying SART to 'true' opacities for the blocks geometry, with different numbers of iterations, are shown in figure 5.30. The plots shown are the central 2D slice through the full 3D image in each case. Some features of interest are immediately apparent: even after ~ 10 iterations, the positions of the blocks, and their cross-sections, are clearly identified and are a very good match to the actual size and position of the objects. However, with small numbers of iterations the density values themselves are not accurate: being around $1 - 3$ kg m^{-3} rather than the true value of 2750 kg m^{-3}. Increasing the number of iterations by several orders of magnitude shows that the blocks remain identified but the density values become more accurate, reaching the right order of magnitude of 10^3 kg m^{-3}.

An additional problem with the output images was that of vertical smearing i.e. the regions above and below the blocks not being well imaged. These regions contain no material; the density values should be that of air but instead have higher densities from the block densities being smeared up and down. This is an artifact of the detector position and orientation: the detectors are orientated perpendicular to the zenith, with a maximum aperture of 60°; they therefore do not cover the regions above and below the blocks well. A better approach would be to tilt the outer detectors so as to 'point' them at the sides of the block region, allowing the regions above and below to be better viewed. However this would require significant modification to the software used and was not implemented.

It was initially suspected that the increase in density values over iterations was due to an error, and that the density values would simply continue to increase further until they were much higher than the true values. However, in practice it was found that this does not happen, and that the density map in fact converges. This can be seen clearly by introducing a figure of merit to assess the quality of a SART output image; this also allows quantitative comparison between different SART settings e.g. number of detectors.

The output of the SART algorithm is a density vector ρ_{SART}, which contains the density in each voxel in the imaging volume, and can be converted into a viewable 3D map using
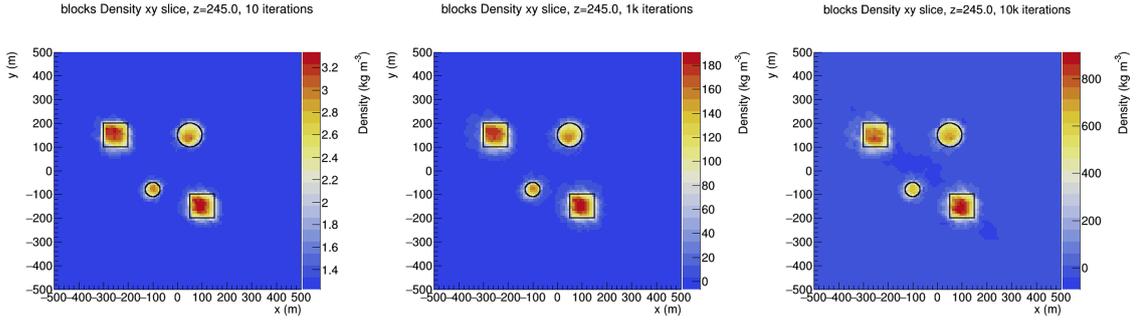
112

Figure 5.30: Comparison of SART images of the 'blocks' geometry (figure 5.27), using the true opacity, for different numbers of iterations.

a ROOT 3D histogram. As the true density map ρ_{true} can also be expressed in this form, a useful figure of merit would be a method that quantifies the similarity of two vectors. A simple method of this kind is 'cosine similarity': the cosine (C_s) of the angle between two vectors with N_d elements in an N_d-dimensional inner product space. For comparisons between density vectors, C_s is bounded between 0 (for orthogonal vectors) and 1 (for parallel vectors). It can be calculated simply as

$$C_s = \frac{\rho_{\text{SART}} \cdot \rho_{\text{true}}}{|\rho_{\text{SART}}||\rho_{\text{true}}|} \tag{5.19}$$

The result of calculating this figure of merit for the SART output images obtained after different numbers of iterations is shown in figure 5.31. It demonstrates that the SART output is a poor match for the true density after few iterations, but rapidly increases in quality after many iterations and eventually reaches a plateau where the density values do not significantly increase further. Also shown is a comparison between the quality of output images obtained using the 'true' opacity and the opacity calculated via Monte Carlo simulation. The true opacity performs slightly better, but the two methods have approximately the same relationship with the number of iterations.

A final test of SART was to compare how the algorithm was affected by using different numbers of detectors. Using more detectors requires a larger distance matrix (due to N_m being larger) and slower iterations, but could offer an improvement in image quality due to the larger number of 'views' of the geometry able to contribute to the calculation. The initial test was to add an additional four detectors to the five present, positioned on the midpoints of the edges of the square defined by the outer detectors (see figure 5.27). The 'true' opacity method, described above, was used for efficiency.

Comparison of the algorithm's performance with five and nine detectors was quantified using the cosine similarity figure of merit. The results are shown in figure 5.32, and indicate that the larger number of detectors has improved the image quality by around 20%. A similar pattern of a convergence after $\sim 10^4$ iterations is observed in both cases.

Finally, the effect of reducing the number of detectors was tested. If the method achieved comparable results with even less detectors then the overall computation time could be
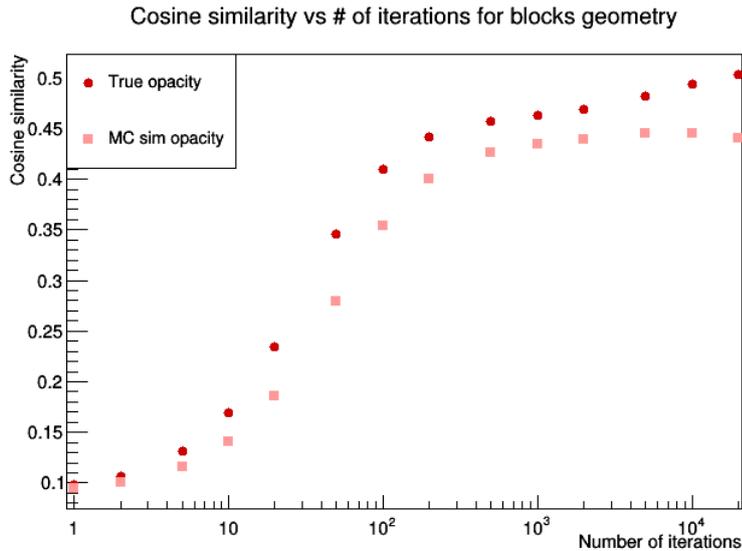
113

Figure 5.31: Cosine similarity figure of merit for SART output images of the 'blocks' geometry, versus the number of SART iterations. Results obtained using true and simulated opacity are compared, with the true opacity performing better; both methods converge after $\sim 10^4$ iterations.

reduced; an experimental system could use less detectors and hence be cheaper. A system with only three detectors was tested for the same geometry and with 10^4 iterations. The results are shown in figure 5.33, which also includes a schematic showing the detector positions in each case. The relatively close agreement between using five and nine detectors is observed, as expected from the figure of merit results; using nine detectors gives a better match to the true densities. However, the three detector arrangement performs poorly with very significant smearing of the bodies in many directions. This is likely a result of the three-detector arrangement not including the 'central' detector present in the other cases; this detector can see all four bodies and so contributes greatly to the output image.

Repository imaging

Finally, the SART algorithm was applied to the repository geometry itself. The imaging volume chosen chosen was a $2 \times 2 \times 0.6$ km cuboid, covering the storage tunnel region and some of the surrounding rock volume. Initially 20 m voxels were used, giving $N_v = 3 \times 10^5$. The angular binning chosen was once again 50×50, however for this geometry there are 12 detectors available, in the arrangement shown in figure 5.10. This gives $N_m = 12 \times 50 \times 50 = 3 \times 10^4$. The larger resulting distance matrix increases the computation time required; a few hours are needed to compute the matrix and store it in sparse vector form, and algorithm iterations now take $1 - 2$ s. This still allows many iterations (up to order $\sim 10^4$) to be performed. The initial density vector was set to have the granite density, 2.75 g cm^{-3}, in all voxels.

Initial tests were performed using a test geometry containing two (extremely) large voids, a cube (side length 200 m) and a sphere (radius 100 m). These were positioned above the storage tunnel plane, equidistant between the tunnels and the surface. Muon rate measurements were
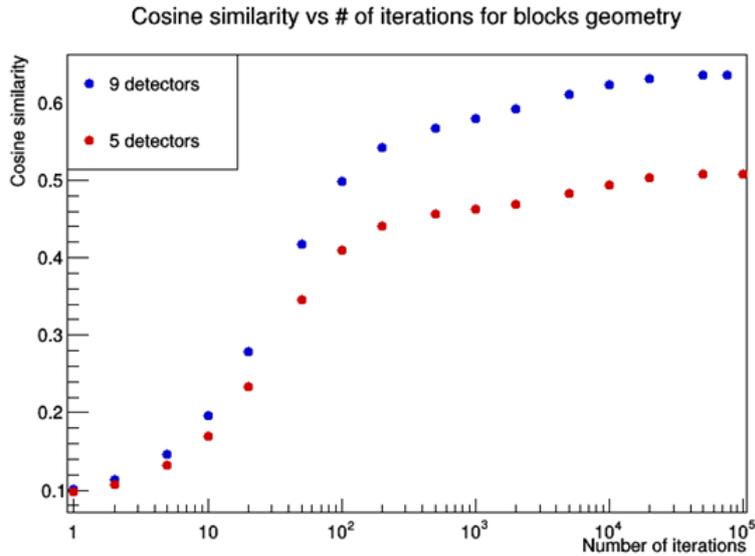
Figure 5.32: Cosine similarity figure of merit for SART output images of the 'blocks' geometry, versus the number of SART iterations, when using both five and nine detectors. The increased number of detectors offers an improvement in image quality at the cost of a longer computation time.

taken with all 12 detectors, for an exposure time of ~ 20 days, then the opacity calculated following the method of section 5.5. A visualisation of the geometry, and an example opacity plot, are shown in figure 5.34.

The results of applying the SART algorithm to the opacity data obtained from this geometry are shown in figure 5.35; this SART image slice was obtained using $20,000$ iterations. The algorithm has clearly performed well at locating the voids, with both identified very clearly, in the correct locations. The edges of the voids in particular are very well defined. However, the density values are not accurate. The voids are clearly of less density than the surrounding material but the true density of 0.001 g cm^{-3} (i.e. air density) is not reconstructed well. This may be due to the number of iterations performed as it bears some similarity to previously seen cases with the blocks examples described above, in which the features are identified clearly but the density values are inaccurate if insufficient iterations have been performed. However due to the much greater computation times required, testing significantly higher numbers of iterations was not possible.

Using 20 m side length voxels as above is not suitable for imaging voids of more realistic sizes. Smaller voxels are necessary in order to resolve features on the $\lesssim 10$ m scale such as shaft-type voids with radii of a few metres, however, using smaller voxels will significantly increase the computation time. Nevertheless, a test case with a 10 m radius shaft type void was imaged using an alternative, finer voxelisation: a map with 10 m side length voxels. The imaging volume was also reduced in size in order to make the computation practical. The new volume is $1 \times 1 \times 0.6$ km, and so $N_v = 6 \times 10^5$. The angular binning remained at 50×50 bins. The result of attempting to image the shaft void is shown in figure 5.36; the shaft location is marked in black. This represents $40,000$ algorithm iterations. The exposure time was also
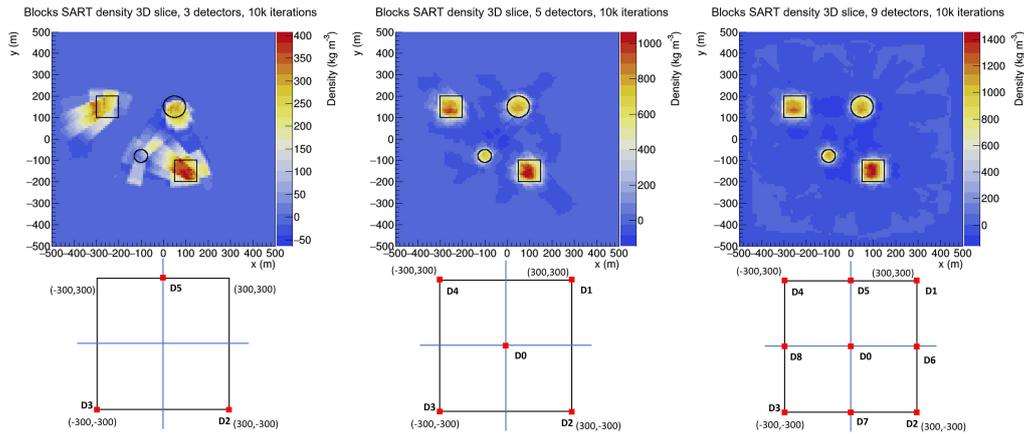
Figure 5.33: Comparison of SART output images obtained using (left to right, top) 3, 5 and 9 detectors, using 'true' opacity and 10,000 iterations in each case. The corresponding detector positions are shown also (bottom). The 5 and 9 detector arrangements give comparable results; using only 3 performs poorly.
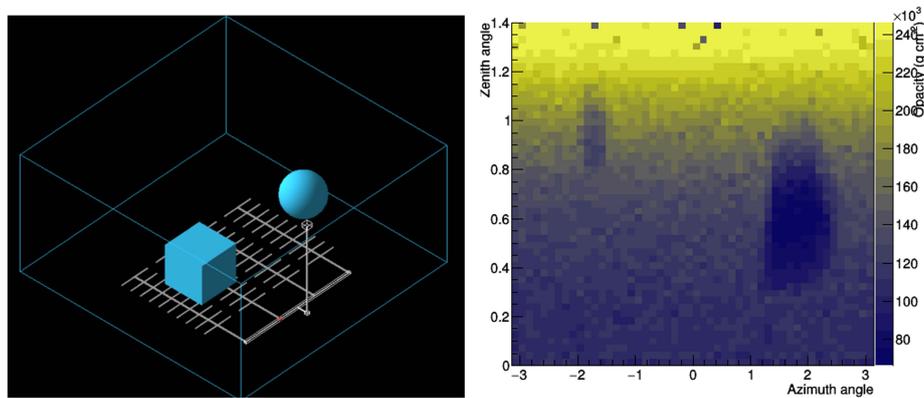
Figure 5.34: Test version of repository geometry for imaging (left), containing 2 large voids. Opacity measurements were taken with all 12 detectors, the opacity for detector 5 (right), clearly shows both voids as regions of lower opacity.

increased, to ~ 100 days.

The algorithm has not performed well in this scenario. Although a small region of lower density is visible in the correct location, it cannot be distinguished from the density variations in the image from general noise. It is possible that a longer exposure time would allow the signal to be more clearly distinguished, but it is clear that it would need to be substantially longer and very likely beyond what is practical in a geological repository operational scenario.

An alternative method to improve results would be to view the void region exclusively to a higher resolution; even with the smaller voxels, in this setup the void will be only 2 voxels wide and smaller voids will be unresolvable. Attempts were made to address this by diving the repository into sectors, each visible by a subset of the 12 detector array, in the hope that these could then be viewed with $2-$ or $5-$metre voxels and that the voids would then become clearly distinguishable. However, this requires sacrificing most of the detectors available in the repository and attempting to image with only four detectors. The reduced number of

Figure 5.35: 2D slice of 3D SART output image of the repository containing two large voids. 12 detectors were used to obtain the opacity data used as input to the algorithm. Both voids are sharply identified as density deficits, however the density values in the void regions are not accurate.

detectors and hence opacity information prevented successful imaging in this scenario. A more radical approach is necessary; some ideas are discussed in section 5.7.

5.7 Conclusions

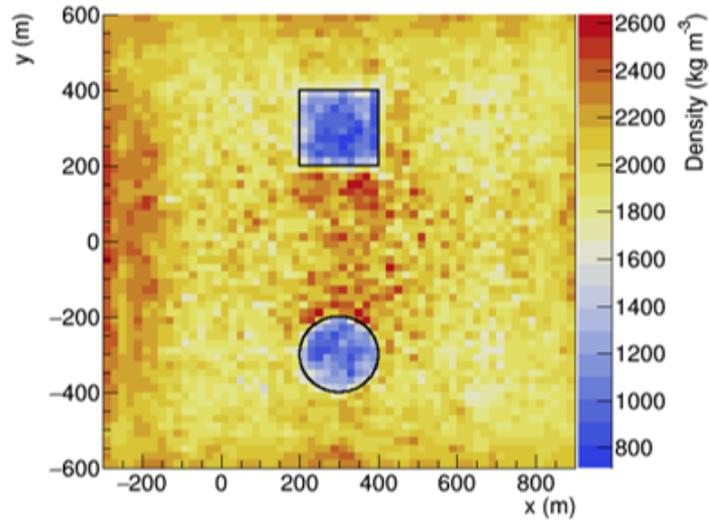The primary conclusions of the studies presented in this chapter concern the feasibility of using muon radiography to detect voids of realistic sizes in a geological repository context. A simulation framework based on backwards Monte Carlo was very efficient for the problem, and was able to perform simulations equivalent to many days' exposure time in a few hours: many orders of magnitude faster than a traditional forward Monte Carlo-based GEANT4 simulation.

Subject to the caveat of a single, semi-idealised particle detector, the results demonstrate that cylindrical voids that subtend solid angles of ~ 0.1 sr are detectable (3σ) under a crystalline rock overburden with a thickness of several hundred metres, within a practical exposure time of ~ 4 weeks. Larger voids require shorter exposure times to be detectable; a cylindrical void subtending 0.12 sr was found to be detectable in ~ 10 days' exposure time.

Additionally, it was found that using a network of multiple detectors enhances the detectability of present voids, with a 10 m radius shaft type void being undetectable in 4 weeks when positioned far from a single detector, whereas a system of 12 fixed detectors was able to detect the same void to $3.0 \pm 0.7\sigma$. Similar improvements were observed for detecting a single large spherical void centred in the repository.

By determining quantitative relations between the detectability of shaft type voids and their radius, the exposure time, and their position relative to a detector, it was possible to
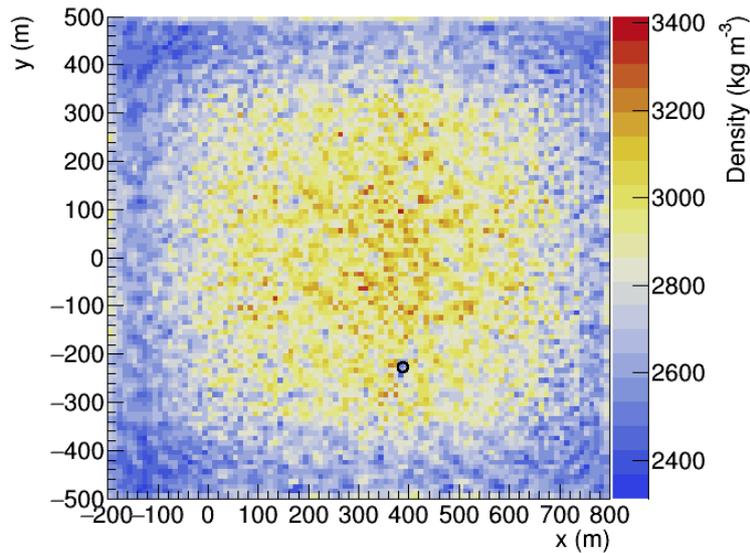
Figure 5.36: 2D slice of 3D SART output image of the repository containing a 10 m radius shaft-type void, location marked with a black circle. Despite reducing the voxel size to 10 m and increasing the exposure time, the void has not been successfully imaged by the algorithm.

estimate the required exposure time to detect smaller voids to 3σ significance. The results suggest that voids with a radius of 1 m could be detected with just four 2×2 m detectors in around 5 months, with $2 - 3$ months needed for 2 m radius voids. These timescales are broadly compatible with the operational timeframes of a geological repository, indicating that muon radiography does indeed have potential for practical detection of small-radius shafts.

An algorithm to image a full-scale repository was implemented. It was found that imaging to sufficient resolution to view voids of a few metres' radius was very challenging without substantial computing resources. A better approach would be to use an iterative approach towards deciding the region to image: an initially very coarse scan to identify density deviations of interest that can then be viewed at a progressively higher resolution; the process repeating until the resolution is sufficient that the feature can either be identified or determined to be benign. This could be tested computationally by modelling the detecting surface as being very (impossibly) large, e.g. the size of the repository itself, before reducing the size to a realistic level as the resolution increases and the viewing region decreases. It would also be advantageous to modify the software to allow for more flexibility in detector arrangement, in particular to have angled detectors to improve the input data available to the imaging algorithm.

A final caveat to the results presented in this chapter concerns the composition of the overburden above the repository. A general limitation of void detection with muon radiography is the accuracy of the geological model used to represent a 'no void' scenario to which experimental data could be compared. For example, if along a particular line of sight through the overburden of length 500 m the modelled mineral has density 4 g cm^{-3} whereas in reality the average density is 3.9 g cm^{-3}, the difference in opacity would be 50 m.w.e., which

118

could be misinterpreted as indicating the presence of a ~ 10 m-scale void. Therefore, the conclusions presented above assume that the geological model is very accurate in order for new voids to be clearly identified. However, any GR site will necessarily have a particularly well-characterised geology. Additionally, the angular information provided by muon radiography can distinguish between a large-scale density deviation such as an unknown stratigraphy layer and a highly localised deviation such as a void. Thus whilst a raw total muon rate would be highly affected by an inaccurate model, incorporating angular rate information (particular if using a system of multiple detectors) would substantially mitigate this limitation.

To take this work further, an experimental study is now being prepared as a partnership between the University of Sheffield, Forschungzentrum Jülich [103], and BGE [104]. This project, if approved, will be make use of the Grimsel test site [105], a Swiss laboratory situated underneath a glacier at a depth of ~ 500 m. The Grimsel laboratory is frequently used for projects on geological repositories as the overburden thickness is similar to that of most repository designs. The planned project will begin with a simulation phase to test the performance of different detector systems and placements, followed by a long experimental phase most likely using a combination of planar detectors and cylindrical 'borehole' detectors. The known shafts and tunnels in the Grimsel overburden can be used as targets, mimicking the effect of unknown voids on the muon flux. Using the work presented here as a foundation, the goal of the study will be to ascertain in detail the suitability of muon radiography for GR assay.

Chapter 6

Assay of a real CCS site with muon radiography

6.1 Introduction

This chapter describes a simulation study into the feasibility of muon radiography as a long-term monitoring technique for a Carbon Capture and Storage (CCS) site. CCS has the potential to significantly mitigate anthropogenic climate change by reducing the quantity of greenhouse gases released into the atmosphere, but the storage sites will require continuous monitoring to ensure that the carbon dioxide is not leaking. Movement of CO_2 will cause density changes in the storage body; the sensitivity of muon radiography to density changes, combined with its well-known efficacy at great depth (> 1 km of rock), suggest it is a candidate for CCS site monitoring.

The study makes use of real experimental data from the Sleipner CCS site in the North Sea. The data, obtained from seismic measurements, has been adapted by the geophysical company CGG [106] into maps of the site at several timesteps between 1994 and 2010. These have subsequently been converted into a form compatible with the GEANT4-based [13] simulation software CRESTA [70] by Geoptic [20]. First, testing and calibrations have been performed to prepare the software to simulate the scenario efficiently and accurately. Next, the model was studied using ray-tracing to better understand the dependence of opacity measurements of the model on the detector position, the timestep, and other adjustables. Finally full muon radiography simulations are performed; these are used to determine the significance of the measured opacity changes over time, and estimate the required exposure time to detect and model the CO_2 plume with various detector setups.

6.1.1 Carbon Capture and Storage (CCS)

The burning of fossil fuels for energy releases huge quantities of carbon dioxide gas into the atmosphere. CO_2 is a greenhouse gas: its absorbance of infrared radiation from the Earth's surface prevents heat from the Sun from fully escaping into space [107]. The greenhouse effect is vital for the survival of complex life, as the Earth's average surface temperature would be around 30°C colder without it. However, following the Industrial Revolution the concentration of CO_2 in the atmosphere has increased by over 50% due to human activity, leading directly to an increase in mean temperature of over 1°C [108]. Further increases are predicted to lead to sea level rises, widespread failure of agriculture, increased frequency of extreme weather events and ecological devastation over the next few decades; huge areas of the tropics may become effectively uninhabitable [109]. Additionally, higher concentrations of atmospheric CO_2 cause increased ocean acidification, which is devastating to marine life. Fully preventing catastrophe may have become impossible at this very late stage, but any reduction in the amount of warming could still mitigate the worst effects of climate change. This will require a rapid transition to green energy and transport. If warming is to be limited in the short term, the continued use of fossil fuel-derived energy must be coupled with measures to greatly reduce the quantity of CO_2 released, alongside active removal of existing gas from the atmosphere.

CCS is a blanket term for various methods of capturing and storing carbon dioxide gas released from fossil fuel burning. The underlying principle is to pump liquified CO_2 deep underground, to depths $\gtrsim 1$ km, at which the CO_2 is in a supercritical liquid phase [110] and can be contained in various geological formations. The exact physical method of containment can vary. For some cases the goal is the CO_2 to react with brine to form solid carbonate minerals, whereas at other sites the CO_2 is expected to remain chemically unaltered (at least over short timescales) but is prevented from escaping by an impermeable 'caprock' layer above the main storage body [111]. For this latter case, the storage body is some form of aquifer such as a sandstone formation, saturated with brine. The injected carbon dioxide displaces the brine and is then held within the porous rock. Figure 6.1 shows the principle of the aquifer storage method; an offshore fossil fuel extraction platform diverts captured CO_2 to the shallower aquifer via a second well.

For a CCS site to operate safely, it must be monitored to ensure that the stored CO_2's behaviour remains as predicted and within acceptable safety parameters. If the CO_2 is able to escape the aquifer, there is a great risk it will eventually reach the seafloor and enter the ocean. This will increase environmentally devastating ocean acidification and ultimately lead to much of the CO_2 entering the atmosphere and contributing to climate change. Muon radiography has potential as a passive monitoring method for CCS sites due to its sensitivity to the density changes caused by CO_2 movement and, in particular, its ability to operate continuously. This is in contrast to other commonly used methods, such as seismic surveys or monitoring pressure, which are applied only on an episodic (often annual) basis [112]. This risks CO_2 movement in the intervening periods not being observed. Additional drawbacks of these methods are that they require highly skilled personnel, analysis of the resulting data
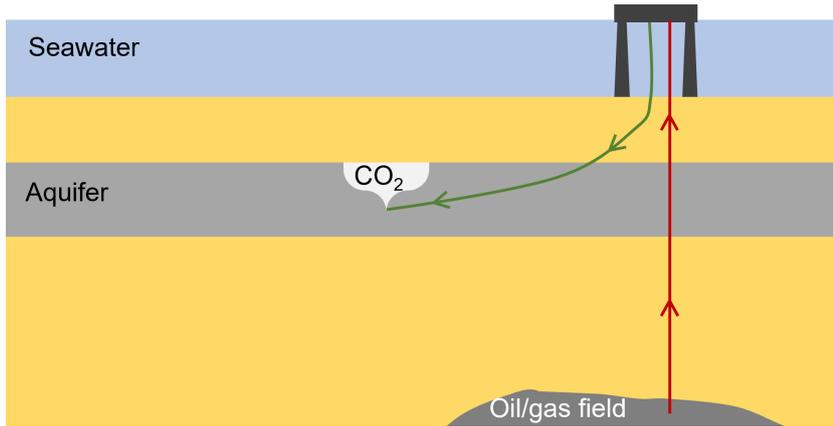
Figure 6.1: (Schematic (not to scale) of an undersea CCS site. An offshore platform extracts oil or natural gas from a deep field, whilst pumping liquefied carbon dioxide into a shallower aquifer layer for storage.

can take several months, and the local weather during surveys can introduce noise [110]; these effects are relatively minimal for muon radiography.

6.2 Muon radiography simulation software

As described in sections 4.3.1 and 5.2.3, CRESTA [70] is a powerful, GEANT4-based Monte Carlo particle simulation software primarily intended for muon transport simulations. Both 1 m-scale scattering tomography and $100 \, \text{m} - 1$ km-scale radiography simulations are possible. To apply CRESTA to the CCS scenario, it was necessary to make use of detector models compatible with placement in a narrow borehole with a diameter of ~ 10 cm. The overburden would be rock and at least 1 km thick; the simulated 'world' in x and y would be several km in order to have a field of view covering the region of interest. This challenging scenario would require a heavily optimised simulation in order to achieve useful results in a practical computation time.

6.2.1 Helical detector for borehole emplacement

To match the symmetry of a long, narrow borehole, a useful detector geometry is some form of cylinder. Multiple cylinders can then be coupled together to increase the overall detecting area. Geoptic have developed a borehole-compatible detector based on strands of solid organic scintillator, arranged in a helical shape; tracking is provided by using three nested layers of strands (see figure 6.2). The outermost and second layer are helices of opposite axial chirality and have 16 and 10 strands respectively. The innermost layer has 6 parallel strands. Each strand has a radius of 4.0 mm, the radii of the cross-section of the outermost, second and central layers are 30 mm, 20 mm and 10 mm respectively; and the overall length of the system is 500 mm. Each helical detector system presents an area of approximately $3 \times 10^3 \, \text{cm}^2$ to the sky and has an angular acceptance of ~ 4 sr, giving a geometrical aperture of $\sim 1.2 \times 10^4 \, \text{sr cm}^2$.
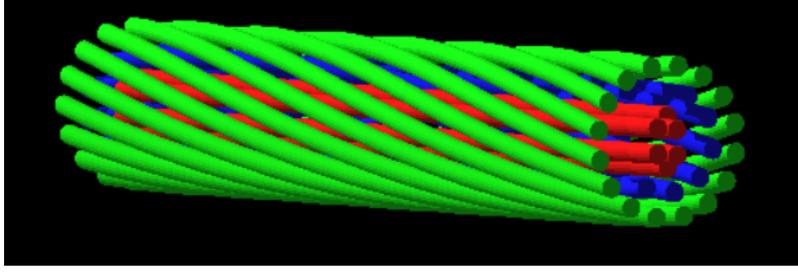
Figure 6.2: Simulation view of the 'helix' detector model designed for borehole applications. The outermost (green) and second (blue) layers of strands are helical whilst the central layer (red) is parallel.

A rock overburden of order 1 km thickness, corresponding to a CCS site, attenuates the muon flux significantly. To estimate the expected flux at this depth, the empirical relation of [7] is used; this equation relating a vertical flux I to an opacity ϱ (in hg cm^{-2} $\equiv$ m.w.e.) is used extensively in this chapter and so is restated here:

$$I(\varrho) \; \left[\mathrm{cm}^{-2} \; \mathrm{sr}^{-1} \; \mathrm{s}^{-1}\right] = \frac{Ke^{-\eta\varrho}}{(\varrho^\alpha + a)(\varrho + H)} \tag{6.1}$$

where $K = 270.7$ hg cm^{-2}, $\eta = 5.5 \times 10^{-4}$, $\alpha = 1.68$, $a = 75$ and $H = 200$ hg cm^{-2}. Equation 6.1 suggests the vertical intensity at the borehole depth is $\sim 3.0 \times 10^{-8}$ Hz cm^{-2} sr^{-1}. Therefore a single helical detector would be expected to measure a muon rate of order 4×10^{-4} Hz. Even coupling together many detectors to form an effective ~ 100 m-long system, with an area of 6×10^5 cm^2, would be expected to detect only around one muon every ~ 10 s. It is clear that therefore that long exposure times of many days will be required to detect small opacity changes in the CCS site, and hence the required computation time of the simulations is likely to be large. Measures to improve the efficiency of the simulation are described below.

6.2.2   Target box size calibration

As discussed in chapter 5 (see section 5.2), simulating a muon radiography of a large structure on the scale of 100 m - 1 km is computationally challenging, as a large proportion of the initialised and transported particles will never encounter a detector and so are wasted. This leads to inefficient simulations. One solution is to use a backwards Monte Carlo system, for which uniformly-distributed muon states are initialised at a detector surface and the transport is performed backwards to calculate the appropriate weighting of each muon. If however a forward Monte Carlo-based transport is used, for most cases it will be necessary to implement measures to improve simulation efficiency in order to obtain useful results in a practical time.

One strategy to improve simulation efficiency is the use of a 'target box'. This is a structure with no material that is added to a geometry and surrounds a detector system. When muon states are generated at the source plane, the software performs an intersection test to determine whether or not their trajectories are pointing at the target box i.e. whether they will enter a volume close to the detector assuming that they are not scattered or attenuated.

The goal is to prevent wasting computation time on transporting muons that have a very low chance of encountering the detector. Figure 6.3 demonstrates the principle by comparing visualisations of 100 generated muon trajectories for a simple scenario of a planar detector under several metres of concrete overburden. With no target box present (top), only a small fraction of the muons encounter the detector. With a target box (with its dimensions set to match the detector in this case) almost all of the generated particles are detected and thus usable for the radiography.
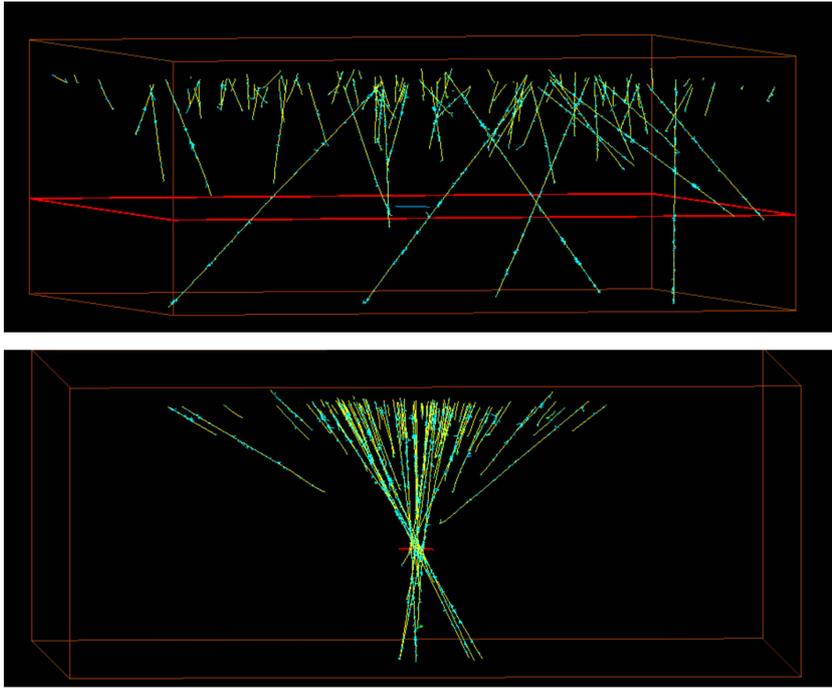
Figure 6.3: Visual comparison of muon radiography simulations with (bottom) and without (top) the use of a target box. The box method improves simulation efficiency by preventing transport of particles that will not reach the detector. If the box is too small however the simulation accuracy is degraded.

Care should be taken when using a target box, as the increased simulation efficiency comes with a cost of reduced simulation accuracy. The artificial cut on the simulated muon trajectories discounts the effect of scattering, as muons that initially point away from the detector can be deflected so as to encounter it; additionally a box that is too small will remove high-angle muons that could still be detectable. A balance must therefore be found when choosing the target box size. The goal is to use the smallest box possible that still maintains an acceptable simulation accuracy for that scenario. For example, if the simulation involved a thin layer of a high-Z material such as lead, muon scatterings would have a significant effect on the detected rate; a small target box around the detector would filter out muons that would scatter into it.

To understand how to size the target box positioned around the helical detector sized above, the hit rate (i.e. proportion of initialised muons that are detected) when using different side lengths of (cubic) target boxes was obtained. The geometry used a concrete overburden of thickness 100 m; the results are shown in figure 6.4 (left). As the target box size increases, the

hit rate falls rapidly. This represents the decreasing simulation efficiency with a larger target box: the acceptable range of initial trajectories is wider and more muons are transported without encountering the detector.

The effect of the box size on simulation accuracy is determined by measuring the source rate (i.e. detector hits per unit exposure time) with varying box side lengths, see figure 6.4 (right). The source rate initially increases rapidly as muons that were excluded by the cut on initial trajectory now pass the cut and are detected. The rate plateaus once the target box size is large enough that the probability of excluding a muon that would be detected is negligible. The plateau region therefore represent an accurate simulation; target boxes smaller than ~ 3 m exclude detectable muons and so reduce simulation accuracy.
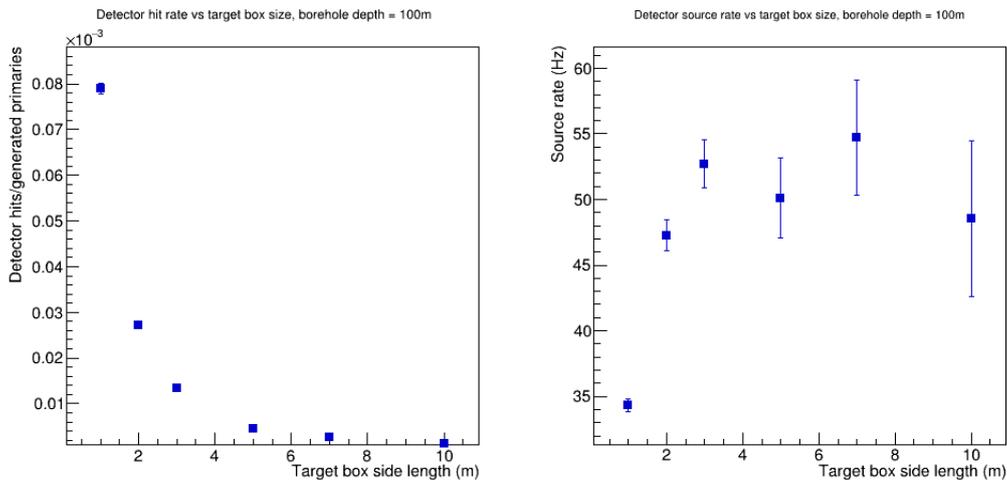
Figure 6.4: Comparison of simulation results for a helical detector underneath a 100 m-thick concrete overburden, with varying target box size. The hit rate (left) encodes the simulation efficiency; the simulation is more efficient with smaller target boxes. The source rate (right) shows the simulation accuracy is lower with a small target box.

6.2.3 Minimum muon energy

Another method to improve simulation efficiency is a cut on the minimum muon energy. As with the target box implementation described above, the goal is to reduce the computation time wasted on transporting particles that will never reach the detector, in this case due to their energies being sufficiently low that they have a negligible probability of passing through the overburden without being stopped. For example, a muon with an initial energy of 10 GeV passing through standard rock has a range of ~ 20 m under the Continuous Slowing Down Approximation (CSDA) [8] (some muons of this energy will travel further due to range straggling). For a radiography scenario with an e.g. 100 m-thick rock overburden, simulating the transport of 10 GeV muons is therefore unnecessary, and so rejecting initialised muons with energies below an appropriate cut can significantly reduce computation time.

As with the target box method, a balance must be struck between between simulation accuracy and efficiency in the choice of cut. A too-high cut will remove muons that had a

non-negligible chance of reaching the detector and contributing to the rate measurement; additionally if the cut is much greater than the optimum choice then efficiency will decrease as so many throws will be necessary to initialise particles that pass the cut. A too-low cut will be accurate but very inefficient.

To determine a general principle for choosing the cut value, a set of simulations using the helical detector underneath concrete overburdens was performed with different choices of the minimum energy cut. In each case the hit rate (ratio of the number of detected particles to the number initialised) and source rate (ratio of detector hits to exposure time) were calculated; the results for a 10 m thick overburden are shown in figure 6.5. The hit rate (left) increases with the energy cut, as a higher proportion of the initialised muons are able to reach the detector. It is clear that the hit rate initially increases rapidly then reaches a plateau at ~ 5 GeV; this is due to the muon energies being such that their chance of being stopped before reaching the detector is negligible (although only a small fraction actually do due to their trajectories not intersecting the narrow helix detector). The source rate (right) is initially approximately constant then falls rapidly as muons that have sufficient energy to be detected are increasingly cut out.
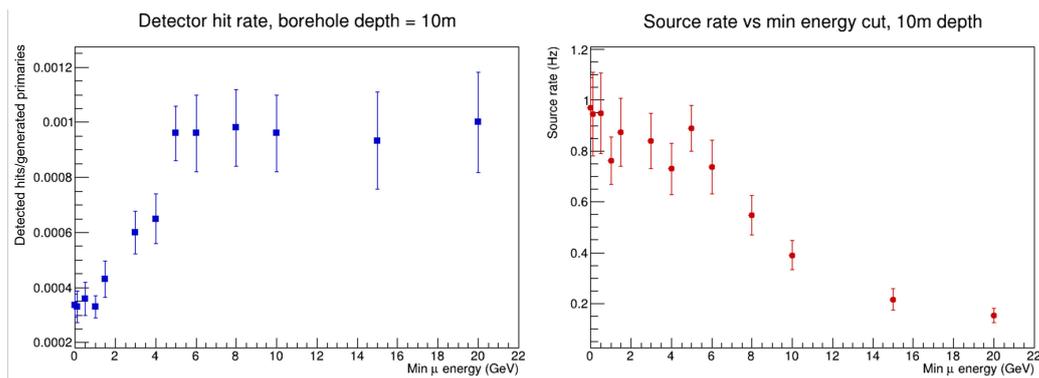
Figure 6.5: The effect of imposing a minimum cut on the initial muon energies, for a helical detector under a 10 m thick concrete overburden. The hit rate (left) is the ratio of detector hits to initial particles generated; it increases rapidly then plateaus. The source rate (right) is the detected muon rate. It is constant at a low energy cut then falls rapidly.

The best choice for the energy cut, then, is the largest value for which the source rate has not reduced significantly from the 'no-cut' value. It is clear from figure 6.5 that, with a 10 m thick concrete overburden, the best choice is ~ 5 GeV: the source rate at this value is compatible to that with no energy cut, and the hit rate has begun to plateau. Under the CSDA, the expected range for 5 GeV muons is approximately 11 m [8]). The implication is that the muon energy corresponding to the CSDA range for the overburden (in the vertical direction) is a useful indicator for the cut value to choose.

To test this, the muon hit rate was also obtained for a thicker concrete overburden with the detector at a depth of 100 m. Applying the CSDA estimation would suggest an appropriate choice for the energy cut of around 55 GeV. In practise, the results indicate (see figure 6.6) that this choice of cut would be slightly conservative; the optimum choice would be ~ 100 GeV

as this is approximately the energy at which the hit rate begins to plateau. Nevertheless the CSDA method provides a useful order of magnitude estimate of the ideal energy cut.

Figure 6.6: Hit rate of the helix detector under a 100 m thick concrete overburden. As for the 10 m depth test, the hit rate increases rapidly then plateaus, indicating that the maximum increase in efficiency has been reached.

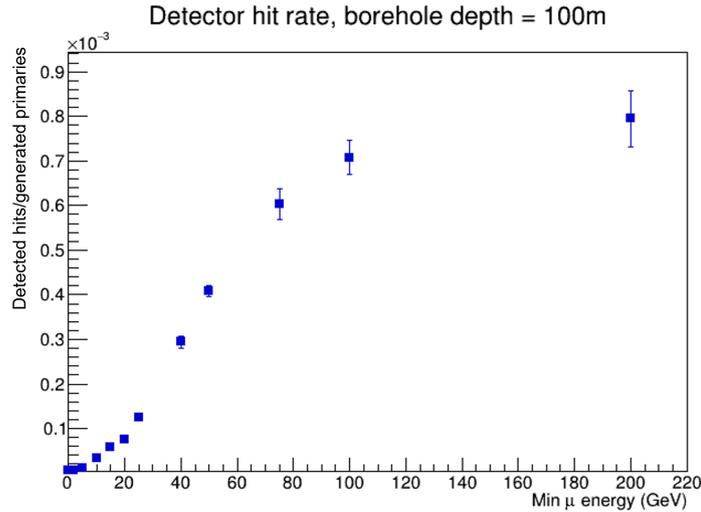In practise, to choose the energy cut for a large simulation it is best to use the CSDA-suggested, conservative cut as a starting point only. If a smaller simulation is run using this cut, the energies of the detected particles can be directly accessed from the Monte Carlo truth; hence the smallest energies of detectable muons can be determined empirically. The full simulation can then use this cut value, boosting efficiency at a negligible cost to the simulation's accuracy.

As described, cuts on the initial muon states can boost transport simulation efficiency. However, it was found that the increase was not sufficient for detector systems of single or coupled helical configurations to obtain useful results within practical computation times; the detection area of even a km-length borehole system is $< 100 \text{ m}^2$. An alternative approach is to simulate a much larger (but still small compared to the 'world' size) detector, and scale results appropriately to different lengths of borehole helical detectors. Therefore, it was decided to use a 100 m × 100 m square planar detector to obtain initial results.

6.3 Sleipner model

6.3.1 Sleipner CCS site

'Sleipner', named after the Sleipnir eight-legged horse of Norse mythology, is a fossil fuel extraction field in the North Sea, approximately 250 km from the Norwegian coast [113]. Its main product is natural gas. Since 1996, the Sleipner site has hosted the world's first offshore CCS operation. This is possible due to the presence of the Utsira formation, which is a very large, very well characterised aquifer body composed primarily of sandstone saturated with

brine. It is an ideal CO_2 injection site due to its depth ($800 - 1000$ m below the seabed, deep enough for CO_2 to remain in a liquid phase), ease of access and material composition. Figure 6.7 shows a natural gas extraction platform at work at the Sleipner site; similar platforms are subsequently used to pump the CO_2 to the aquifer depth.

Figure 6.7: From [114]. An extraction platform in the off-shore Sleipner oil and gas field.

6.3.2 CGG

CGG SA [106] is a geoscience company based in Massy, France. Its areas of expertise include mapping, subsurface imaging, data analysis and geology; it has a wide range of clients across the environmental, energy and infrastructure sectors. The CCS modelling is from CGG's GeoVerse Carbon Storage study [115]. The seismic and geological model used are available under the Sleipner CO_2 reference dataset license [116].

The model is obtained from seismic assay of the Sleipner site. A set of six timesteps are available, corresponding to the years 1994, 2001, 2004, 2006, 2008 and 2010. The 1994 set represents the site before any injection of CO_2 and so is vital for comparison to the other datasets. CO_2 has been injected at an approximately constant rate of 1 Mt/year since the year 1996.

6.3.3 Simulated model

Data from CGG has been adapted into a GEANT4-compatible simulated model by Geoptic [20] (see figure 6.8). The model is large: its dimensions are $3.19 \times 5.80 \times 1.22$ km. Its component materials are a complex static stratigraphy of many kaolinite minerals, with density varying between $1.5 - 2.6$ g cm^{-3} (see figure 6.8, top) but the same chemical composition. The specific mineral used has the empirical formula $Al_2Si_2OH_4$; the elemental compositions by mass fraction are shown in table 6.1.

Element	Atomic number Z	Atomic mass A (g/mole)	Kaolinite mass fraction (%)
Al	13	26.982	11.76
Si	14	28.085	11.76
O	8	15.999	52.94
H	1	1.008	23.53

Table 6.1: Basic properties and mass fraction of the elements in the kaolinite mineral used in the Sleipner model overburden.

This leads to the mineral having a mean atomic number of $Z = 7.646$ and a mean atomic mass of $A = 15.182$. The aquifer itself is made of sandstone; the man-made features include a vertical borehole and the deviated well, both narrow cylindrical shafts.
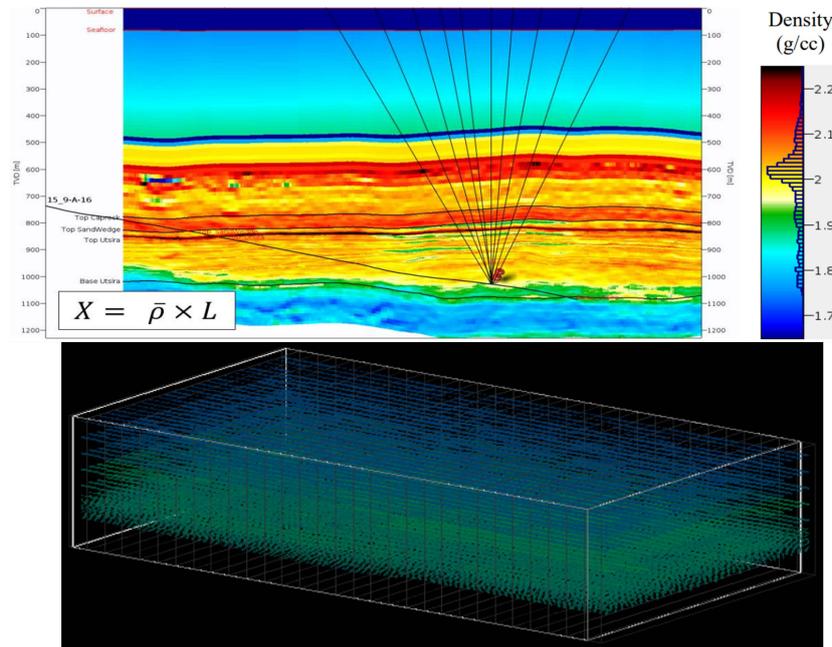
Figure 6.8: Two views of the Sleipner geological model. Top: stratigraphy of the model, showing the densities of the different layers and the deviated well [117]. Bottom: CRESTA interactive view of the Sleipner CCS site simulated model. Note that the 'grid' structure is in place for easier visualisation.

To be compatible with simulation, the computational memory required to store the full-resolution model must be reduced. The CRESTA version of the model is therefore divided into cells in the xy plane (see figure 6.8, bottom), whilst maintaining the full resolution in the z direction. The larger the resolution settings, the less memory the model occupies and hence the more simultaneous simulations can be run with the use of a multi-machine computing cluster. However, accuracy is lost with a coarser resolution and so a balance must be struck. Figure 6.9 shows the memory occupied by the model with different resolution settings between 100 m and 500 m. From this data, it was decided to proceed with the resolution set to 200 m in x and y, to allow for a reasonably finely resolved model with which simulations could be run efficiently.

Figure 6.9: Relationship between the model x and y resolution setting and the memory occupied by the full model at simulation time.

6.3.4 CO$_2$ plume

Since 1996, ~ 1 Mt of CO$_2$ per year has been injected into the Sleipner aquifer. This is reflected in the six available model vintages, covering the year 1994 (before any injected carbon dioxide is present) up to 2010. The gas is released from the end of the deviated well, and migrates upward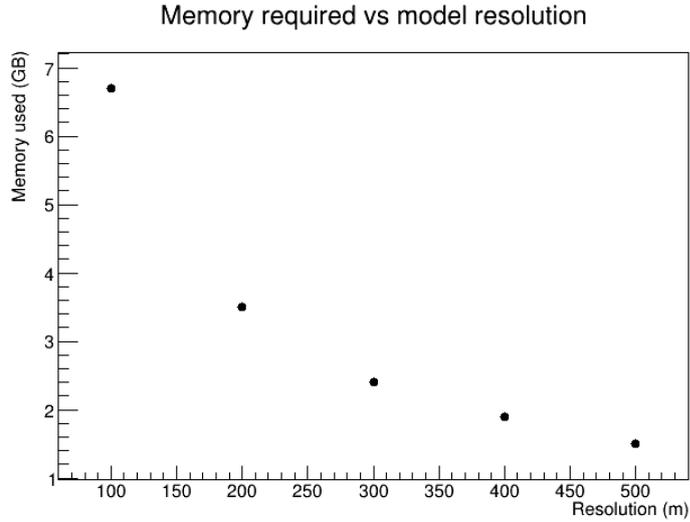s towards the caprock, spreading out in the process. The shape that the CO$_2$ plume forms over time is known from seismic measurements and can be used for comparison to data obtained from simulated muon radiography for validation. Figure 6.10 shows a top-down view of the edges of the CO$_2$ plume for each of the five dataset years for which any CO$_2$ is present. The increase in size of the plume over time is clear. Also shown are the locations of the deviated well and the vertical borehole, and the boundaries of the simulated 'world'.

To model how this region, corresponding approximately to the CO$_2$ plume, would appear to a detector, a simple test was conducted whereby the plume outlines of figure 6.10 were converted to angular data collected by a detector. The chosen detector position was inside the deviated well, at a depth of 1030 m. Specifically, for each azimuth angle ϕ (to a resolution of 1°) the zenith angle θ that, from that detector position, would point to the edge of the plume was calculated. This was repeated for each of the model timesteps and the data was then plotted on polar axes; the results are shown in figure 6.11. This modelling also shows how the edges of the 'world' region appear from this detector position, which leads to regions along the east-west axis being cut out of the field of view.

130
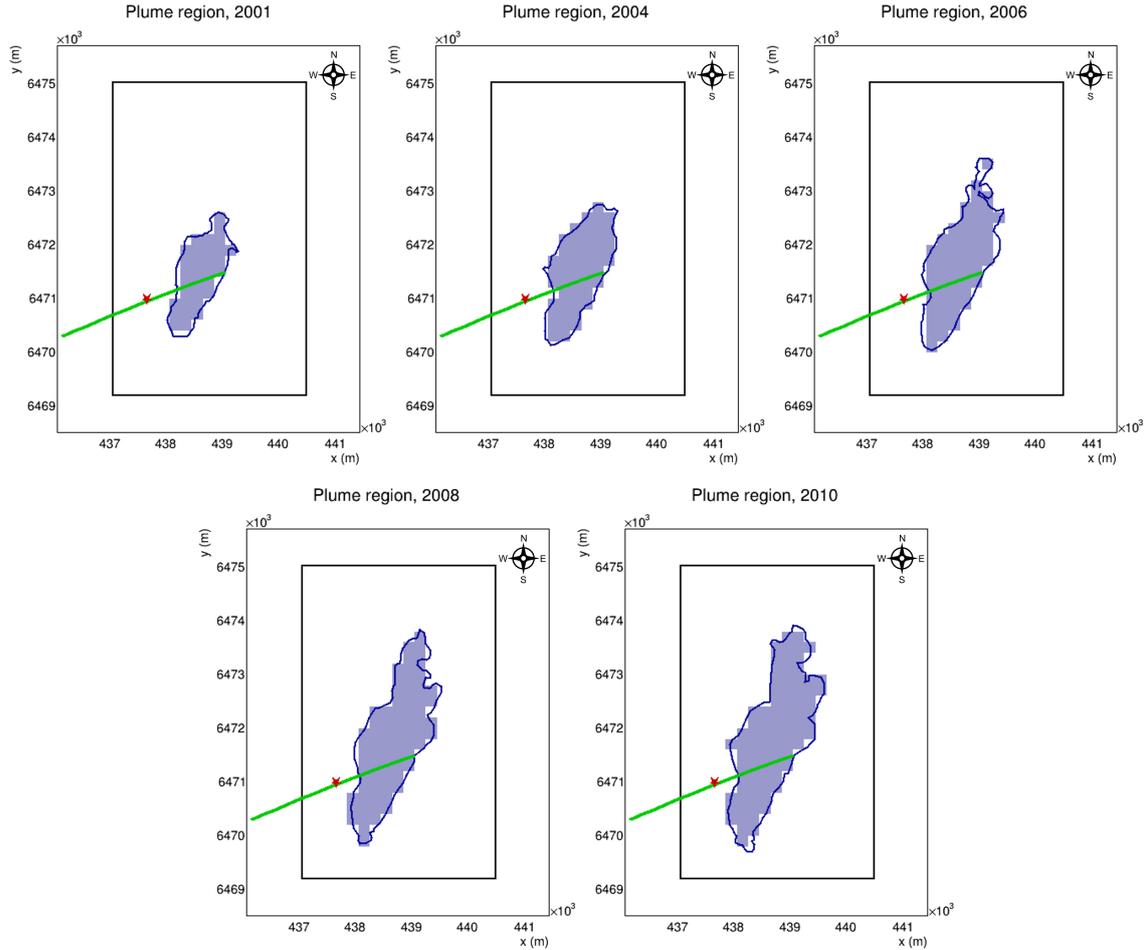
Figure 6.10: Top-down view of the carbon dioxide plume edges in the simulated model (blue). Shown for comparison (light blue) is the 200 m × 200 m xy grid applied to lower the model resolution and hence required memory. Other shown features are the deviated well (green), vertical borehole (red), and edges of the world volume (black).

6.4   Geantino analysis

A useful aspect of GEANT4-based simulation software is the ability to generate and transport 'geantino' particles. These are fictitious particles that have no charge and no mass, and do not undergo any physical interactions with materials. They are intended primarily for debugging purposes, and are particularly useful when considering particle transport through a complex geometry. For example, when considering a scenario of a simulated muon radiography study of a volcano with a complicated geometry model, a long computation time may be required to transport sufficient muons to verify that the geometry is correct, or that a particular detector placement is suitable to view the structure as desired. Conversely geantinos can usually be transported very quickly, as they traverse on purely ballistic trajectories. It is therefore possible to quickly verify these aspects of the geometry and modify the system if necessary before performing the full muon transport exercise.

A particularly useful application of geantinos is to combine them with a facility for direct

Figure 6.11: Projection of the appearance of the CO_2 plume outlines to a 'detector' at a depth of 1030 m, positioned inside the deviated well. The region of the field of view covered by the plume is in yellow, and the remaining portion is in blue. Note the plots are rotated by π from those of figure 6.10.
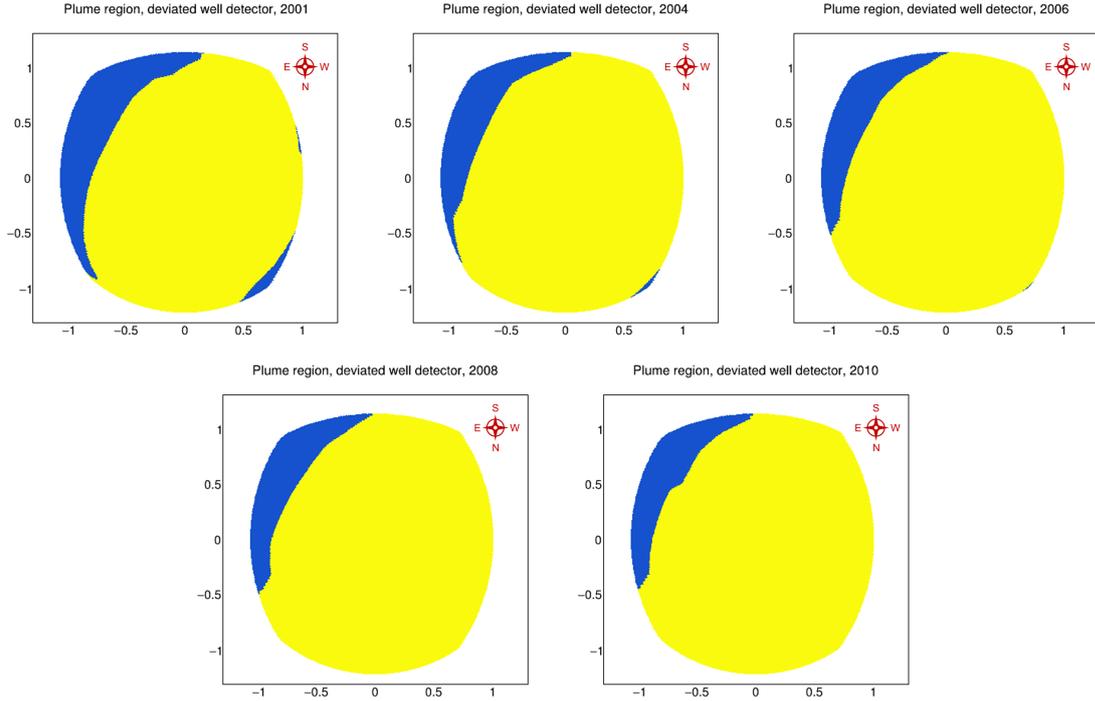
opacity measurements within the GEANT4-based software. When used with generated muons, rather than calculate an opacity measurement from an empirical relation (e.g. equation 6.1) or by considering energetics (see section 5.5), instead the density and distance travelled through each geometry cell by each muon is recorded from the Monte Carlo truth, and the total opacity is then the sum of each product of these. This method therefore cannot represent a physical measurement. However, when geantinos are generated instead, the result is a measurement of the 'true' opacity of the model as seen along a particular line of sight, obtained in an efficient computation time. The uncertainty of the measurement propagates from the spread of the geantinos arriving in the corresponding angular bin; for these studies the geantinos are initialised with an angular distribution drawn from the CRY library so as to match the distribution of cosmic-ray muons.

The ballistic trajectories taken by geantinos are analogous to those that would be taken by very high-energy muons, i.e. muons sufficiently energetic that they will not be noticeably scattered even after traversing hundreds of metres of rock. This can be demonstrated by comparing the distributions of the Monte Carlo truth opacity values as obtained with simulations of both particles. Figure 6.12 (left) shows geantino and muon opacity distributions after simulated transport through a geometry consisting of a 100 m-thick layer of homogeneous concrete with density 2.3 g cm^{-3}; both distributions therefore have their minimum Monte Carlo truth opacity as 2.3×10^4 g cm^{-2} (i.e. the opacity recorded by vertical particles). The muon-measured opacities have a clear distinct population separate from the main peak. This

132

is due to lower-energy muons being scattered during the transport, and hence travelling along much longer (up to an order of magnitude longer) paths; this extra distance is integrated into the total opacity recorded. If a cut is applied to remove muons with initial energies below 100 GeV (figure 6.12, right) this second population is removed and the muon and geantino distributions become consistent.
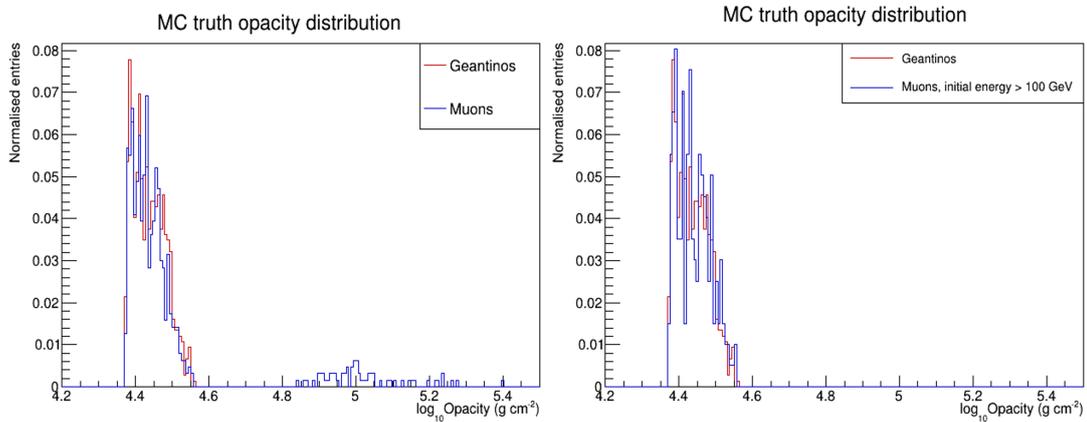

Figure 6.12: Comparison of the distributions of MC truth opacity as measured with geantinos and muons. The muon distribution has a tail of larger opacity values (left) due to scattering of low-energy muons. If a cut is applied to the minimum initial energy of the muons (right), the tail disappears and the distributions match closely.

6.4.1 Assay of Sleipner model with geantinos

The Sleipner CCS site geometry described in section 6.3 has a complex density distribution, containing both a stratigraphy with many materials and an aquifer layer with both time- and spatially-varying concentrations of brine and carbon dioxide. It is therefore difficult to verify that opacity measurements obtained with muon radiography are correct, or to predict the position, shape and intensity of observed features. Geantino analysis provides an efficient method of assaying the Sleipner geometry, as an intermediate step to plan the 'physical' muon simulations and improve understanding of their results.

For example, geantino data was used to choose the position of the test detector in the simulations. In the Sleipner geometry, there are two obvious possible choices for the detector position: the vertical borehole (see figure 6.10, marked in red) and the deviated well (green). The former is a cylindrical shaft positioned around 1 km to the west of the centre of the aquifer. The latter is an angled shaft that runs beneath the aquifer; this is the pipe from which the CO_2 is released from. Either of these volumes could hold up to several hundreds of metres of helix-type cylindrical particle detectors, but the borehole would represent a more straightforward deployment as the coupled detectors could be held suspended whereas the well may require a more complex method of detector insertion. Therefore the initial plan was to run muon simulations with a detector placed at the position (in x and y) of the vertical borehole. The results of testing this approach with an initial geantino simulation are in figure

6.13 (left), which shows the opacity as a function of angle as obtained with a 100 m $\times$ 100 m planar detector placed in the borehole at a depth of 1030 m (approximately 200 m below the base of the CO_2 plume). It is clear that a large section of the field of view, towards the west, is not visible from this position. This is due to the position of the borehole being offset towards the edge of the simulated world volume (see figure 6.10), and consequently some lines of sight passing outside of the simulated world volume.

Choosing to instead place the detector in the deviated well allows the xy position to be tuned so as to eliminate this problem. If the detector depth is chosen again to be 1030 m, the corresponding x and y coordinates in the deviated well are $x = 438659$ m and $y = 6471333$ m. This position is close to the center of the world volume, removing the large invisible region that affects results obtained at the borehole position, and is also directly underneath the plume and so gives the best possible view of the resulting change in opacity. Therefore, it was decided to choose this detector position as the default for future measurements and the computationally intensive muon simulations in particular. An opacity measurement taken at this location is shown in figure 6.13 (right).
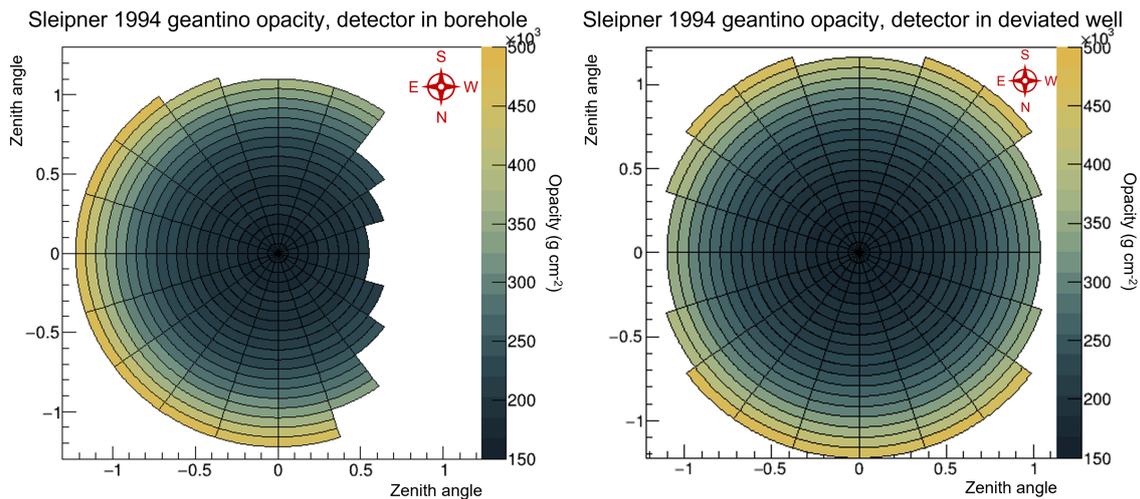
Figure 6.13: Comparison of geantino-measured opacities as a function of angle as measured by a 10^4 m^2 square detector, placed in either the vertical borehole (left) or the deviated well (right). The position of the borehole relative to the simulated 'world' volume leads to a large gap in the field of view towards the west in the former case. For the latter, the detector can be placed so as to give the best possible view of the world volume and the CO_2 plume itself.

As liquefied CO_2 is released from the well into the aquifer, the displacement of the denser brine lowers the overall density of the overburden above the detector position. Hence opacity measurements along lines of sight that pass through the CO_2 plume will be reduced over time. Geantino measurements are effective for confirming and visualising this opacity change in the Sleipner model; for example, taking polar opacity measurements in the '1994' and '2010' vintages and calculating the relative change in opacity along each line of sight shows a large region of reduced opacity, see figure 6.14. The location of the region as viewed from the deviated well position corresponds well to that predicted from the known shape of the plume,

see figure 6.11 (left). As expected, the smallest change in opacity is towards the south-east, with reductions in opacity in all other directions. The opacity reductions are small, around a 0.2% reduction along most lines of sight.
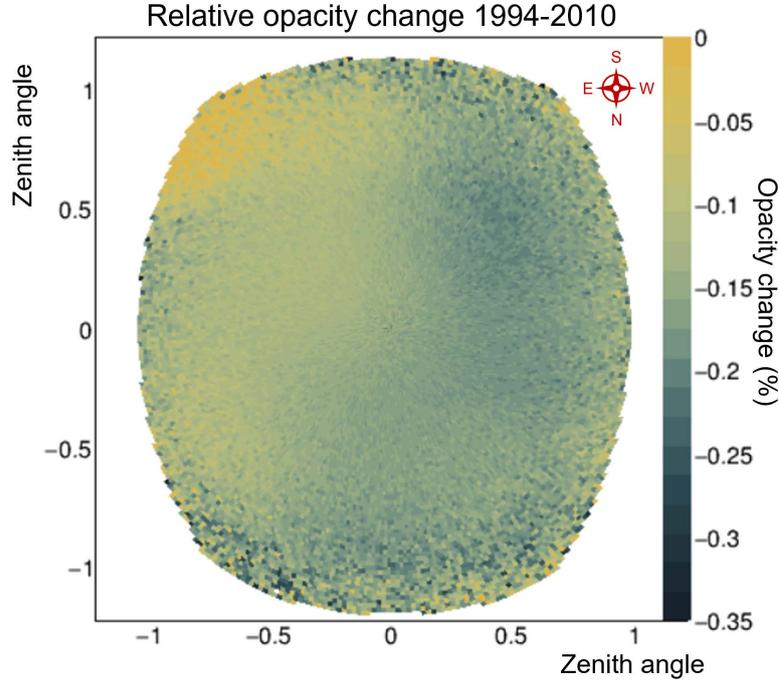
Figure 6.14: Relative change in opacity, viewed on polar axes under a fine $1°$ binning, between the 1994 and 2010 model timesteps, as measured with samples of 4×10^7 geantinos. A large region of reduced opacity is clearly visible, indicating the presence of the CO_2 plume.

6.4.2 Estimation of carbon dioxide injection rate

This data can be converted into an estimate of the amount of carbon dioxide that has been released into the storage site. As the CO_2 is displacing the brine from the aquifer, using the known densities of both materials in combination with the calculated opacity change gives first a CO_2 'length', then incorporating additionally the known properties of the rock a percentage of saturation along each line of sight.

The total opacity ϱ along a line of sight is the sum of the product of the densities and lengths of each material i.e. $\varrho = \sum_{i=0}^{n} \rho_i x_i$. The materials used in the Sleipner model geometries (as discussed in section 6.3) are several different kaolinite minerals, brine, and CO_2. The distribution of the minerals is constant over time, so any change in opacity is due to the movements of brine and CO_2. If these materials are approximated as having constant density, then the change in opacity $\Delta\varrho$ is given by

$$\Delta\varrho = \rho_c \Delta x_c + \rho_b \Delta x_b \tag{6.2}$$

where ρ_c (ρ_b) and Δx_c (Δx_b) are respectively the density and change in 'length' of the carbon

dioxide (brine). Physically, x_c and x_b correspond to the length of the line that would be formed if all of that fluid were removed from the pores within the rock along that line of sight and lined up. Both fluids occupy the same porous volume within the rock, and so it can be approximated that $\Delta x_c = -\Delta x_b$. Hence, from equation 6.2, it follows that $\Delta x_c = \Delta \varrho / (\rho_c - \rho_b)$. The density of the liquid CO_2 varies with its temperature, but on average is approximately $\rho_c \sim 0.7$ g cm^{-3}. Brine is slightly denser than pure water, $\rho_b \sim 1.1$ g cm^{-3}. Therefore, an approximate relation between opacity change and change in CO_2 length is Δx_c [cm] $= -2.5 \cdot \Delta \varrho$.

Δx_c can now be converted into an estimate of the rock saturation along each line of sight. This is done by considering the porosity of the aquifer rock, which is, on average, $30 - 40\%$. Therefore, if the length of aquifer rock along a line of sight is known, it can be approximated that $\sim 35\%$ of this length is available to be filled with CO_2. As data from the 1994 timestep geometry represents the aquifer containing no carbon dioxide, taking the ratio of Δx_c in a particular year to this 'fillable' length then gives an estimate of the saturation in that year. The total length can be approximated as $l_0 / \cos(\theta)$, where l_0 is the thickness of the aquifer (≈ 100 m) and θ is the zenith angle. For example, if in one angular bin at a zenith angle of $\pi/4$ a measurement of $\Delta x_c = 10$ m is made using the 2010 dataset, then the saturation along that line of sight is estimated to be $\frac{10[\text{m}]}{0.35 \times 100[\text{m}] / \cos(\pi/4)} \approx 20\%$. Figure 6.15 shows the estimated saturation along each line of sight between 1994 and 2010, taking as input the data shown in figure 6.14.
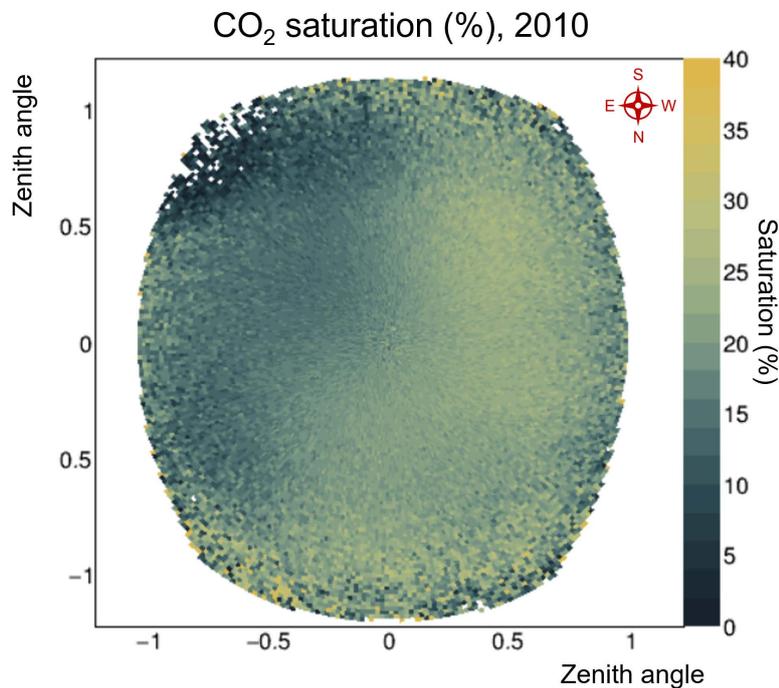

Figure 6.15: Estimation of the CO_2 saturation along each line of sight on a $1°$ binning. The saturation is calculated by deducing the change in carbon dioxide 'length' from the relative change in opacity, then dividing by the available aquifer length.

Obtaining this data for each of the five post-1994 model timesteps now allows the total

volume of injected CO_2 in the portion of the aquifer within the field of view to be estimated. The visible region of the aquifer is a truncated cone; with a 70° maximum zenith, a detector to base distance of 200 m and a height of 100 m (see figure 6.16), the total volume of the cone is 1.5×10^8 m^3. With an average porosity of 35%, the available volume for carbon dioxide injection in this region is 5.3×10^7 m^3.
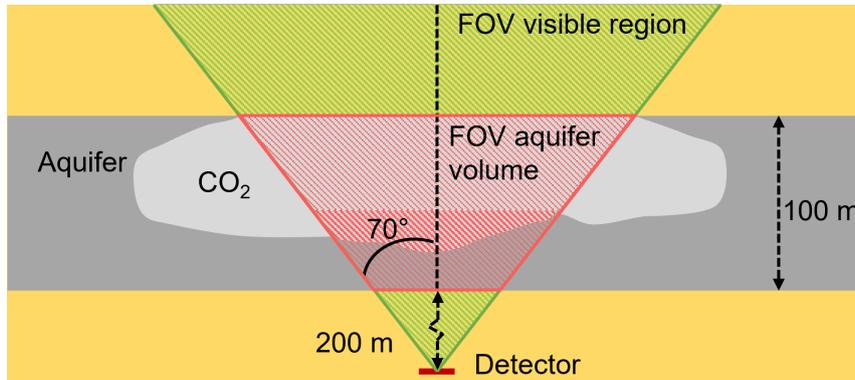

Figure 6.16: Schematic (not to scale) of the region of the aquifer within the field of view of the detector. The aquifer is modelled as a 100 m thick uniform layer at a distance of 200 m from the detector; with a maximum zenith angle of 70° the volume of the truncated cone formed is 1.5×10^8 m^3.

Taking the mean of the CO_2 saturation proportions across all bins and multiplying by this available volume then gives the total volume of carbon dioxide injected into the visible region of the aquifer for that timestep. The calculated total volume in each year is shown in figure 6.17 (left); the volume initially increases, then plateaus out between 2006 and 2008. The total volume injected per year is approximately linear. This plateau therefore represents the visible region becoming fully saturated: further CO_2 injected enters the aquifer in regions outside the detector field of view.

The results can be verified by directly comparing the deduced CO_2 mass injection rate to the known rate. The Sleipner site has been injected with approximately 1Mt of carbon dioxide per year since 1996 [118]. The results of converting the estimated total volume into total mass via $\rho_c \approx 0.7$ g cm^{-3} are shown in figure 6.17 (right). When compared to the true injection rate (black), the estimated injected mass shows close agreement for the first few timesteps before plateauing.

6.4.3 Reconstruction of plume shape

The results above are obtained largely via considering the mean opacity changes across many lines of sight. However, detailed angular information is of more use for most applications, e.g. identifying the current shape of the CO_2 plume in a particular timestep. The required statistics for analysis of this kind are high; the expected opacity changes are on the order of $\lesssim 0.3\%$, at a depth of 1 km, this corresponds (using the empirical formula in equation 6.1) approximately to muon vertical intensity changes of just $\sim 1\%$. Geantino analysis therefore once again represents a useful tool for understanding the plume evolution and how it could be
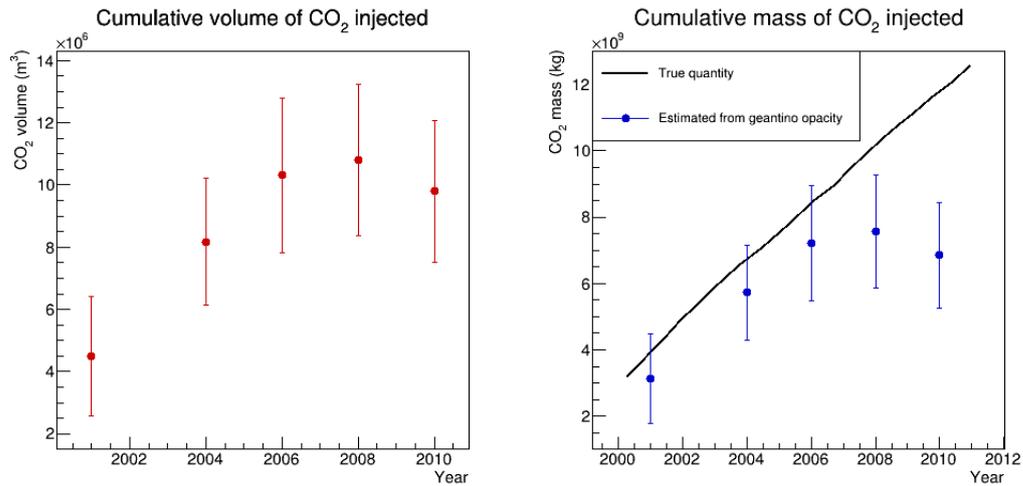
Figure 6.17: The estimated cumulative volume (left) and mass (right) of liquefied carbon dioxide injected into the visible region of the aquifer by year, as calculated from the relative opacity change. The CO_2 quantity initially increases linearly then plateaus as the visible region becomes saturated. Comparing the known mass injected to the estimated shows close agreement up to 2006, at which point further injected CO_2 enters regions outside the detector field of view.

assayed in much shorter computation times than a full muon simulation.

Once example of this is to project obtained angular relative opacity change data, such as the 1994-2010 change shown in figure 6.14, onto a plane; this can be repeated for each timestep to show the plume evolution over time, and then compared to the known empirical data in the model. The relative opacity change on polar axes, with a fine 1° binning, in each timestep is shown in figure 6.18. A 3×3 Gaussian smoothing has been applied to each bin value to reduce noise and clarify the plume region. It is clear that the plume has increased in volume over time, hence the greater reductions in opacity in the later timesteps, and that it has physically moved, from east to west.

This data can then be projected to directly compare with the known plume location and movement. Section 6.3.4 describes the integrated density maps of the carbon dioxide plume shown in figure 6.10. The results of projecting the relative opacity change in each angular bin (see figure 6.18) onto a plane 250 m above the detector (i.e. the approximate height of the centre of the plume), in each of the five timesteps, and overlaying onto the integrated density change outline are shown in figure 6.19. The movement of the plume from east to west, and its final north east-south west orientation, are clear in the projected data.

6.5 Muon radiography of Sleipner site

The geantino studies above are of interest for understanding the model in greater detail, visualising the progress of the carbon dioxide plume over time, and studying the effects of different detector positions and sizes. However, full muon simulations are necessary for the final studies, in order to match physical measurements as closely as possible and obtain results
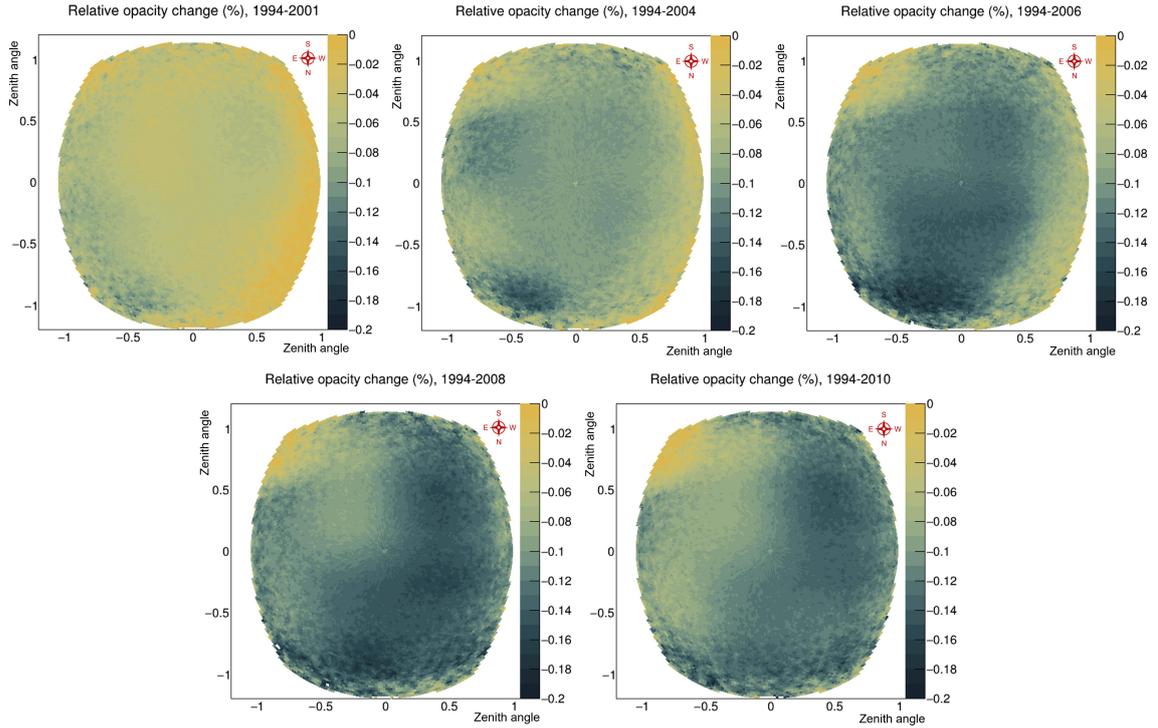
Figure 6.18: Relative change in opacity as a function of angle, from 1994 to each of the five available timesteps. A Gaussian smoothing has been applied to each image for clarity; figure 6.14 shows the equivalent data for the 2010 timestep without smoothing applied. It is clear that the opacity reduction increases over time, and that the plume moves from east to west.

usable by a site operator. Having developed methods of optimising the simulation software, muon exposures up to several hours were run for the 1994 and 2010 timesteps in order to calculate the statistical significance of the resulting opacity change and estimate the required exposure time to detect the plume, and determine its shape, with muon radiography.

As discussed in section 6.2, a large detector area is required to obtain sufficiently large muon fluxes in a practical computation time. The detector used is a 100 m × 100 m plane, positioned at a depth of 1030 m; it is placed in the deviated well which at this depth gives xy coordinates of $(438659 \text{ m}, 6471333 \text{ m})$. The CRY source plane size is 3 km × 3 km. The CSDA-based estimate of the minimum muon energy that will reach the detector (assuming 1 km of standard rock overburden) is ~ 1 TeV, hence a conservative choice of $E_{\min} = 500$ GeV is used. A check on the distribution of the initial energies of muons that reached the detector after a test exposure showed no muons with initial energy< 500 GeV were detected, and so this choice of cut is valid. A square target box was also used with a side length of 1.2 km. Additionally, in test simulations it was noticed that no secondaries (i.e. non-muon particles produced via muon interactions with the overburden) were reaching the detector; the produced particles are far less penetrating than muons and so are attenuated after short distances ($\ll 1$ m) through the rock. An additional cut was therefore applied to restrict the computational transport of secondary particles to further increase efficiency. With these settings, the relation between the single-thread computation time and the simulated exposure time is shown in figure 6.20.

139
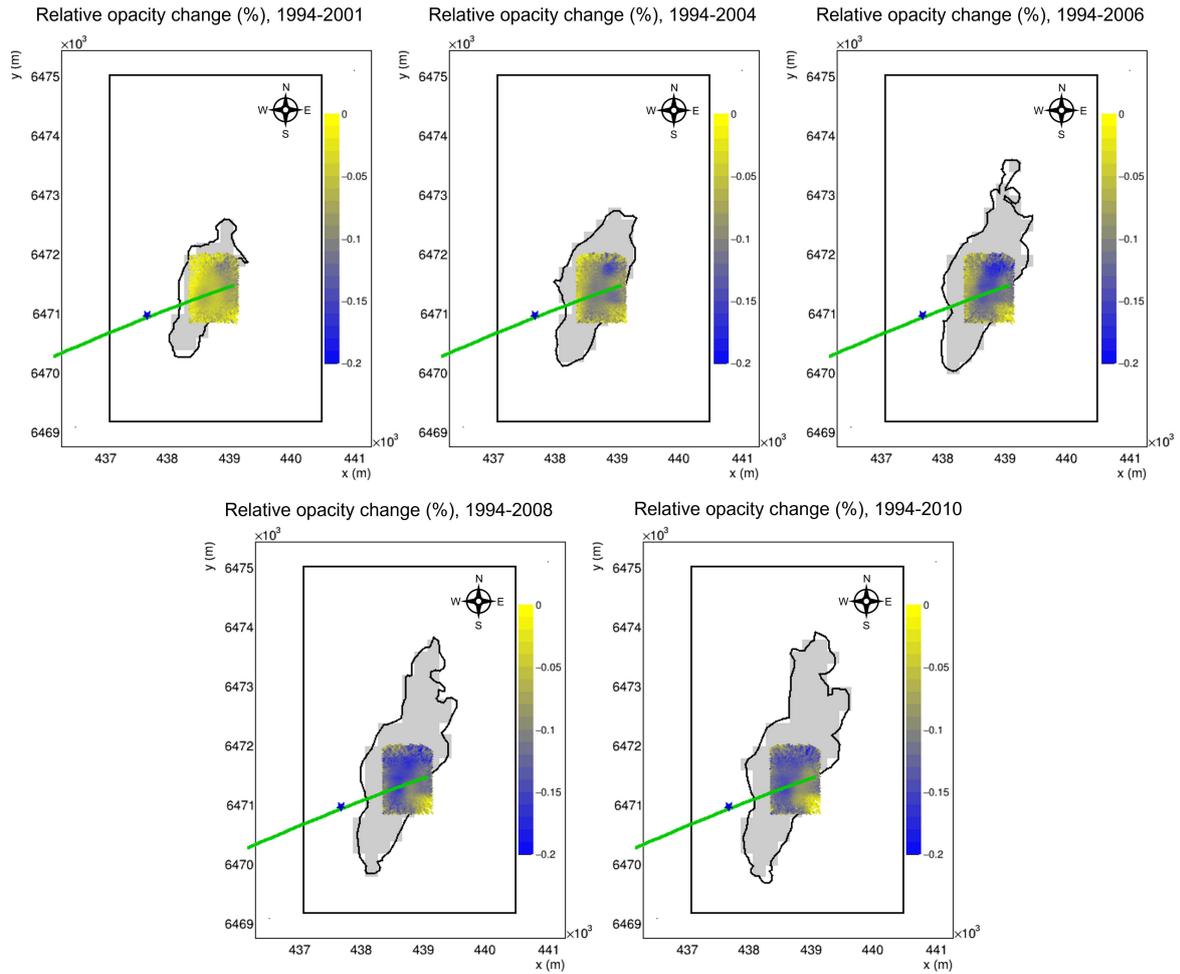
Figure 6.19: Relative opacity change as a function of angle, projected onto a plane 250 m above the detector and shown in comparison to the plume integrated density (see section 6.3.4).

The obtained relation shows that even with the described methods to improve simulation efficiency, long computation times are still required; e.g. around 1×10^7 s (~ 4 months) for 1 hour of exposure time. However, the use of a powerful computing cluster allows simulations to be divided and ran concurrently. The required computation time is thus reduced by a factor of ~ 50, allowing 1 hour of exposure time to be simulated in just $2-3$ days. The total exposure obtained for the 1994 and 2010 timesteps was 6.8 hours, requiring 2.1×10^6 initialised muons.

6.5.1 Significance of muon flux changes

With this long exposure available, it is now possible to directly compare the detected muon flux in both 1994 and 2010. The expected result is an increase in flux in the later timestep due to the increased quantity of CO_2 present in the aquifer; the resulting reduced opacity attenuates the muons to a lesser degree. Geantino data (e.g. figure 6.18) indicates that the expected opacity changes are of the order of $0.1-0.2\%$, which corresponds to a flux increase of $\sim 1\%$. Figure 6.21 shows both the calculated flux (right) of muons in the 1994 and 2010 timesteps and

140
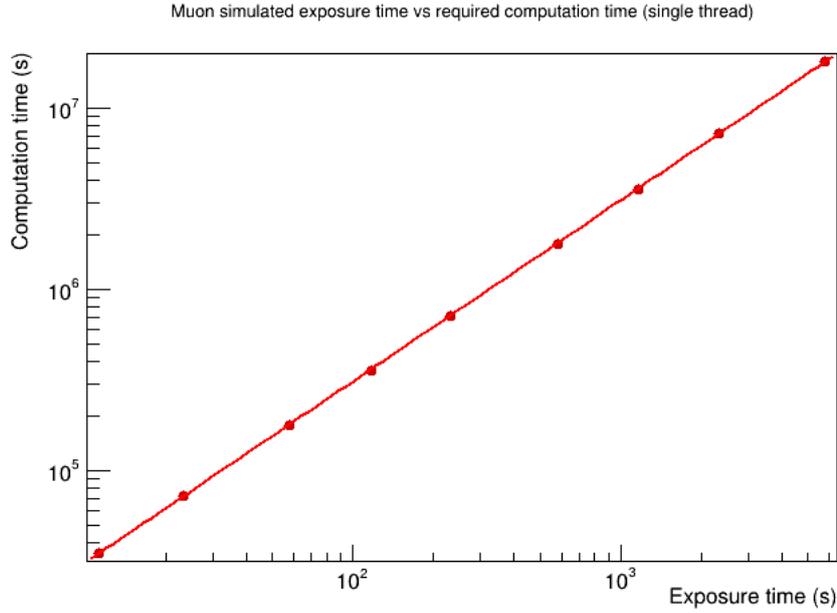
Figure 6.20: Single-thread computation time for muon transport through the Sleipner geometry with CRESTA, using a large 100 m × 100 m planar detector. The results indicate that e.g. 1 minute of exposure requires approximately 5 hours of computation time; this is mitigated via the use of a computing cluster.

the corresponding ratio of the muon counts (left). As the exposure time increases, it becomes clear that there is a significant separation between the recorded fluxes. After the longest exposure time (6.8 hours), the recorded flux in 1994 is $3.42 \pm 0.01 \times 10^{-8}$ Hz sr^{-1} cm^{-2}, and in 2010 is $3.45 \pm 0.01 \times 10^{-8}$ Hz sr^{-1} cm^{-2}. The increase is thus $0.9 \pm 0.5\%$, a close match to the predicted $\sim 1\%$. A first-principles analysis using the empirical formula of equation 6.1 would suggest the expected vertical intensity under 1 km of standard rock (density 2.65 g cm^{-3}) is 3.0×10^{-8} Hz sr^{-1} cm^{-2}; the Sleipner overburden is on average less dense than standard rock, and so the simulated detected flux is in good agreement with that expected.

This data can now be used to determine the statistical significance of the detected change in muon flux. The chosen equation for the significance derives originally from astrophysics, in the context of detecting a gamma-ray source by a flux comparison to a known region with no source [119]. One defines the number of muons detected from the source as N_{on} and the number detected from the 'off' region as N_{off}; when applied to this problem, N_{on} (N_{off}) is the muon count in 2010 (1994). Combined with a factor α that encodes the ratio of the size of the regions, the full expression for significance S is

$$S = \sqrt{2}\left[N_{\mathrm{on}} \cdot \log\left(\frac{(1+\alpha)N_{\mathrm{on}}}{\alpha(N_{\mathrm{on}} + N_{\mathrm{off}})}\right) + N_{\mathrm{off}} \cdot \log\left(\frac{(1+\alpha)N_{\mathrm{off}}}{(N_{\mathrm{on}} + N_{\mathrm{off}})}\right)\right]^{1/2}, \qquad (6.3)$$

$$\alpha = \frac{\kappa_{\mathrm{on}} \cdot t_{\mathrm{on}} \cdot A_{\mathrm{on}}}{\kappa_{\mathrm{off}} \cdot t_{\mathrm{off}} \cdot A_{\mathrm{off}}} \qquad (6.4)$$

where κ, A and t are respectively the angular size, acceptance, and exposure time in the
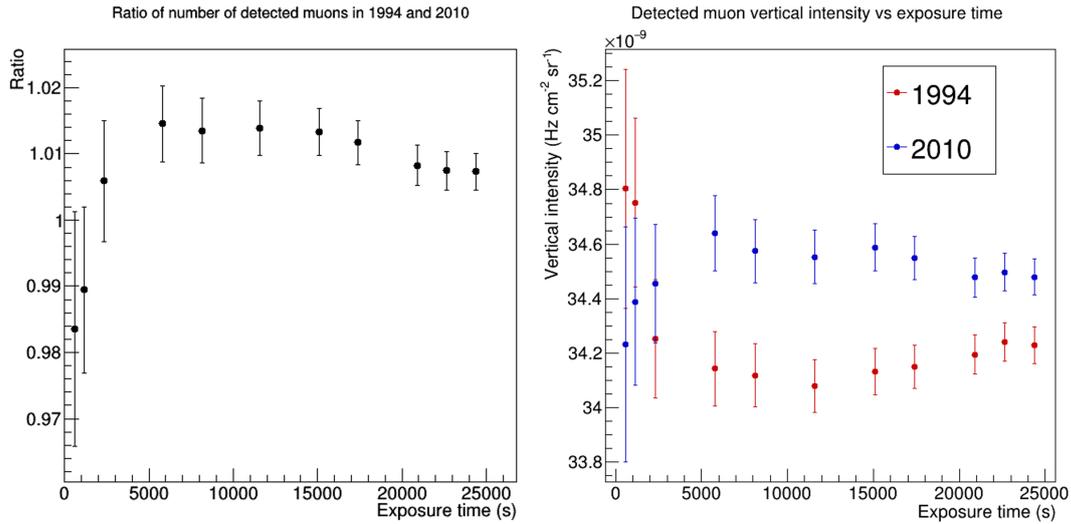
141

Figure 6.21: Comparison of muon flux data obtained after 6.8 hour exposures of the 1994 and 2010 model timesteps. Left: ratio of the cumulative number of muons detected in 1994 and 2010, with increasing exposure time (error bars propagated from $\sqrt{N}$). Right: the calculated muon fluxes for each exposure time, with errors propagated from the Poisson noise on the rate measurements.

on and off regions. As $\kappa_{\text{on}} = \kappa_{\text{off}}$ and $A_{\text{on}} = A_{\text{off}}$ if the detector and angular binning are the same for both measurements, α reduces to just the ratio of exposure times, $\alpha = t_{\text{on}}/t_{\text{off}}$. The formula as originally derived was presented as valid for large values of N only. It was demonstrated via a Bayesian analysis in [120] that the method is valid for all N.

For the longest exposure tested, ~ 6.8 hours, the number of muons detected in 1994 was $N_{\text{off}} = 262164$, and in 2010 was $N_{\text{on}} = 264099$. The exposure times for both differed by < 60 s and hence $\alpha = 1.00$. Therefore, via equation 6.5.1, the significance $S = 2.7 \pm 1.0\sigma$. The confidence intervals here are derived from a Monte Carlo analysis: N_{on} and N_{off} are drawn repeatedly from Poisson distributions with $\lambda = N$, S is calculated for each pair, and the confidence intervals are the width of the distribution of S. A significance of almost 3σ indicates that the presence of the CO_2 plume has been detected with confidence after a few hours' exposure, when using a 10^4 m^2 detecting area.

A detecting area of this size is not feasible in reality. To relate this obtained significance to other detector configurations, the detected 1994 and 2010 muon rates, and their uncertainties, are scaled by the ratio of the detector areas. The scaled rate is then multiplied by a time to obtain an estimate of the number of muons one would detect on that area in that time, and finally the corresponding statistical significance is calculated using equation 6.5.1. For example, the detected muon count rates after the 6.8 hour exposure were 10.75 ± 0.02 Hz in 1994 and 10.83 ± 0.02 Hz in 2010. Scaling these rates to a 1000 m^2 detector, i.e. by a factor of 0.1, and using a time of 1 day$= 8.64 \times 10^4$ s, would give $N_{\text{on}} = 935899 \pm 1821$ and $N_{\text{off}} = 929075 \pm 1815$. The significance would then be $S = 5.0 \pm 2.7\sigma$, propagating the errors from the original rate measurements. As a test of the feasibility of muon radiography at a CCS site, the study in [112] simulated a CCS scenario using the Boulby mine site as a model,

142

and a detecting area of 1000 m^2. The depth of the overburden in both cases is comparable, and so a useful validation of this method is to compare the detected significances of the carbon dioxide plume as a function of exposure time in both cases. The results of this are shown in figure 6.22 (left). The results from [112] are larger than the significances extrapolated from this study (black). This is likely due to the thinner overburden used, 776 m thick as opposed to a detector depth of 1030 m for the present results.
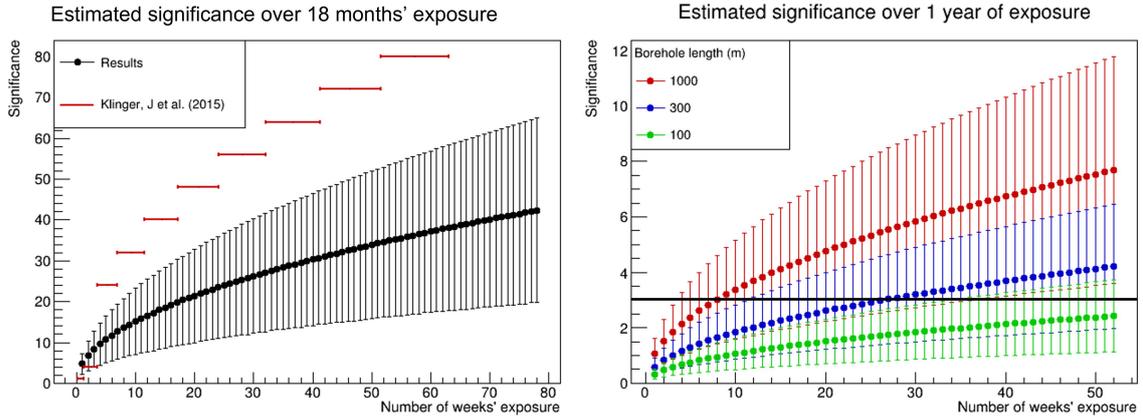
Figure 6.22: Estimated plume detection significances for other detecting areas vs exposure times up to a year, as extrapolated from a 6.8 hour exposure (2.1×10^6 initialised muons) using a 10^4 m^2 detector. The rates in 1994 and 2010 are scaled by the ratio of the detecting apertures and then used to estimate N and S. Left: extrapolation to a 10^3 m^2 detector with comparison to results from [112]. Right: extrapolation to different lengths of helical borehole detectors by the same method, black line marks significance= 3σ.

This same process was also performed for much smaller detecting areas, specifically the areas available to a system of coupled helical detectors, of the type described in section 6.2.1, in a borehole. Each individual system is treated as having a diameter of 65 mm and it is assumed that the whole length of each system is sensitive (i.e. disregarding electronics or scaffolding in between the scintillating materials). Using borehole lengths of 100 m, 300 m and 1 km, the significance was estimated at 1-week intervals over 1 year of total exposure. The results are shown in figure 6.22 (right). The longer borehole lengths give higher significances in less time. For each case, the significance-time relation follows a $\sqrt{t}$ relation; this is due to the uncertainty of the rate measurements deriving from Poisson noise i.e. $\sqrt{N}$. For each curve, the required time to reach a 3σ level of significance has been calculated:

Borehole length (m)	Time to detect to 3σ (weeks)
100	79^{+284}_{-45}
300	26^{+95}_{-15}
1000	8^{+28}_{-5}

6.5.2 Angular opacity measurements

Detecting the presence of the plume via bulk muon rate measurements is of interest, as it would allow e.g. a scenario in which a large fraction of the stored carbon dioxide has leaked to

be ascertained. However, angular measurements that, ideally, show the shape and movement of the plume over time are more useful. As was shown in section 6.4.3 via geantino analysis, maps of the relative change in opacity along different lines of sight show the plume region clearly if sufficient statistics are available. This is more challenging for a muon simulation as much longer exposure and hence computation times are required to discern the small changes in opacity.

Obtaining the muon flux as a function of angle is achieved simply by binning the number of muons detected by their zenith (θ) and azimuth (ϕ) angles, then dividing the bin counts by the product of the exposure time, the detector area (10^4 m^2), and the solid angle subtended by the bin. The results of this calculation, using a coarse, $10°$ bin size, is shown in figure 6.23, for both the 1994 (left) and 2010 (right) vintages of the geometry. As expected the flux is most intense at smaller zenith angles towards the vertical, due to the shorter path lengths and hence reduced attenuation along those directions. When calculated using the 3.2 hour exposure time, the errors on these flux values are approximately $5 - 7\%$.
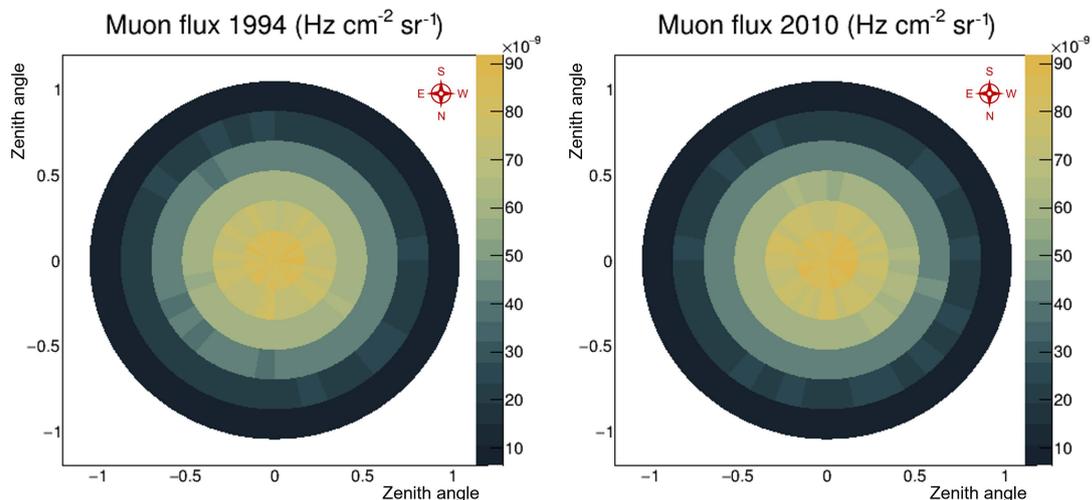
Figure 6.23: Muon flux as a function of angle, as calculated from a 3.2 hour exposure for the CCS site in 1994 (left) and 2010 (right). With a coarse binning of $10°$ angular bins, the errors on the flux values ($\sim 6\%$) values are larger than the changes caused by the injection of the CO$_2$ plume ($\sim 1\%$) and so are not apparent in this data.

To convert these flux measurements to an estimate of the opacity along each line of sight, the empirical formula derived in [7] (equation 6.1) has been used. This formula is obtained from many experimental measurements, with the varying overburdens normalised to an opacity assuming homogeneous standard rock. When applied to the 1994 flux as a function of angle in figure 6.23 (left), the results are shown in figure 6.24 (left), in comparison to the equivalent opacity measurement conducted using geantinos. The opacity values are expressed in 'metres water equivalent' (m.w.e.): the thickness of pure water that would result in the equivalent opacity. 1 m.w.e. $\equiv 1$ hg cm^{-2}. The latter result, as described in section 6.4, represents the ray-traced product of density and length and so is the 'true' opacity of the model.
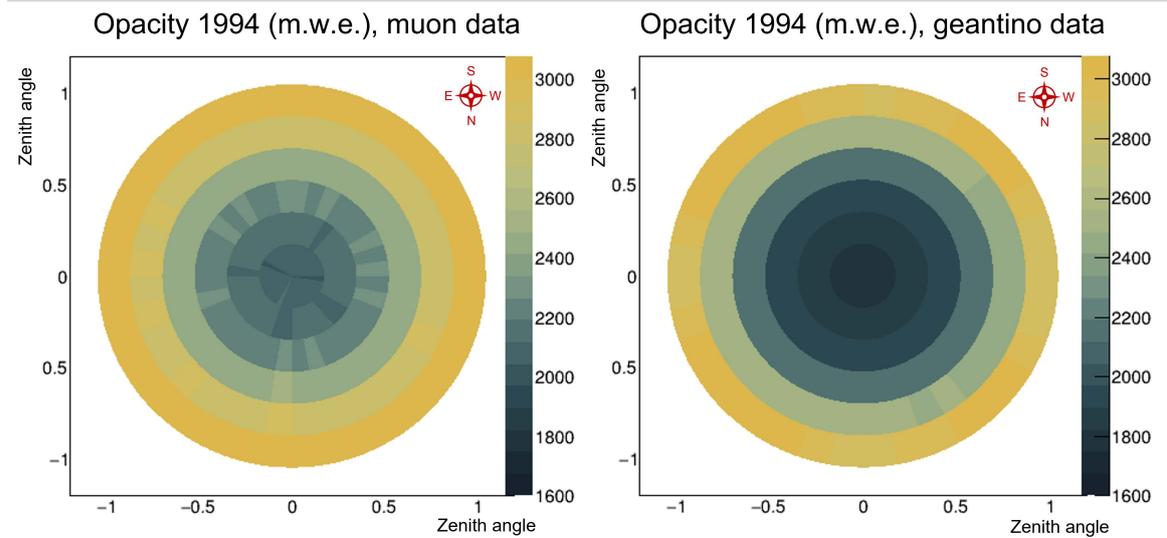
Figure 6.24: Comparison of the measured opacity as a function of angle, via a muon simulation (left) or a geantino measurement (right). The muon opacity is obtained by taking the flux measurements (figure 6.23) as input to the empirical flux-opacity relation of equation 6.1. The two measurements match well, but the muon opacity values are clearly lower by a factor of $10 - 20\%$, indicating a higher flux by a factor ~ 2.

The results indicate that the opacity values as calculated by applying the empirical formula of [7] (equation 6.1) to the muon flux measurements are larger than the true values, by a factor of ~ 1.25 in the near-vertical direction. This demonstrates the limitations of the empirical formula: it is useful for determining an approximation of the opacity corresponding to a measured flux, but performs somewhat poorly when compared to the ray-traced (i.e. Monte Carlo truth) opacity available in a simulation context. It is clear in the derivation of the relation that factors of $2 - 3$ between the underlying experimental data and the curve are not uncommon, particularly at depth. One explanation for this effect is that the formula accounts only for the opacity of the overburden materials, and not their chemical nature. It has been demonstrated ([45], [121]) that at depth, assumptions based on an overburden of density-corrected standard rock break down. This topic is explored in more detail in section 6.5.4. In this case the excess opacity factor will be very similar between the 1994 and 2010 datasets, as the bulk mineral structure is the same in both cases. Hence, whilst the absolute opacities calculated are inaccurate as described, the relative changes in opacity between the two are still valid.

6.5.3 Plume angular opacity change detection

The muon opacity measurements made in 1994 and 2010 were also used to estimate the required exposure time to determine the edge of the plume itself, in order to monitor its evolution over time. The relative opacity changes as calculated from muon data are shown in figure 6.25 in comparison to the equivalent data as obtained with geantinos. It is clear that the muon measurements are much larger in both magnitude and spread than the geantino measurements; this is a result of the relatively high uncertainties on the underlying opacity

145

measurements. As described above, the uncertainties on the original flux measurements in each bin are $5 - 7\%$, arising from the Poisson noise on the muon counts. When propagated (via a Monte Carlo method) onto the opacity measurements, the uncertainties are $1 - 2\%$, leading to a high uncertainty on the relative opacity change. For example, if in a particular angular bin the opacity in 1994 is $O_{1994} = 2.13 \pm 0.02 \times 10^3$ m.w.e. and in 2010 is $O_{2010} = 2.11 \pm 0.02 \times 10^3$ m.w.e., then the relative change in opacity $O_r = (O_{2010} - O_{1994})/O_{1994} = -0.94 \pm 1.35\%$. In other words, the error on the opacity change is comparable to the change itself.
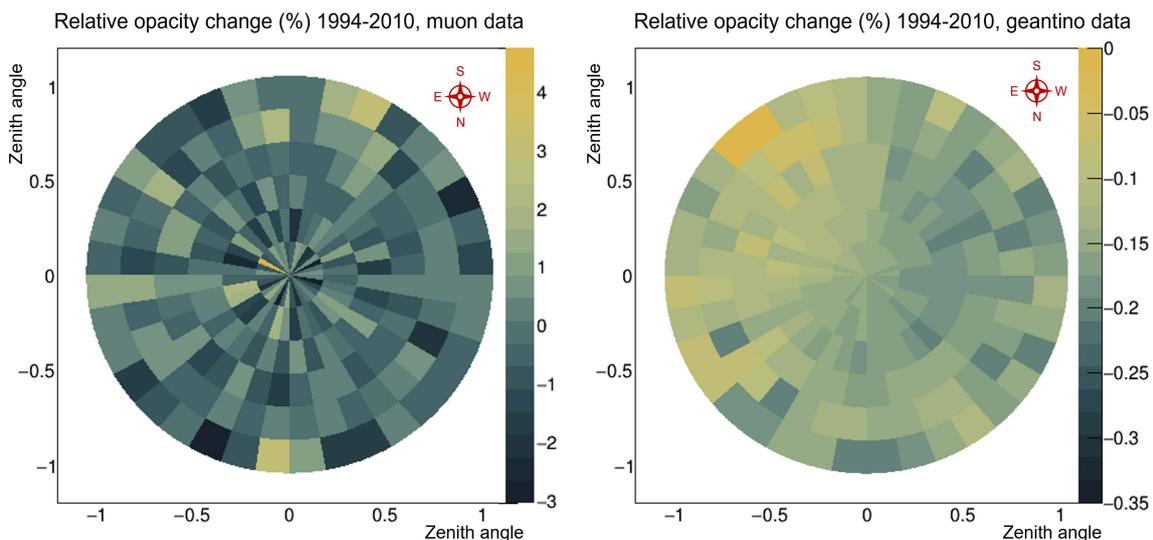
Figure 6.25: Comparison of angular measurements of the relative opacity changes between 1994 and 2010, as measured via a full muon simulation (left) or via geantino analysis (right). The exposure time for the muon simulation is 6.8 hours. The angular binning is coarse in both cases with $10° \times 10°$ bins. The muon results indicate much larger and noisier relative opacity changes due to the high uncertainties on the opacity measurements.

In order for the muon exposure to identify the plume region, the exposure time must be sufficient that the uncertainties on the opacity changes are smaller than the true changes due to the plume movement. The geantino data (figure 6.26, right) indicate that this requires the opacity change uncertainty to be $\lesssim 0.2\%$. To estimate the required exposure time to achieve this, the mean error on the opacity change for various exposure times was calculated; the results are shown in figure 6.26).

As expected, the error falls with increasing exposure time t, following a $1/\sqrt{t}$ shape due to the underlying Poisson statistics of the flux measurements. This fit can then be extrapolated to estimate the required exposure time to reach a desired level of uncertainty. The resulting estimated required times, having used a 10^4 m^2 detecting area, for select values of the mean error are shown in table 6.2.
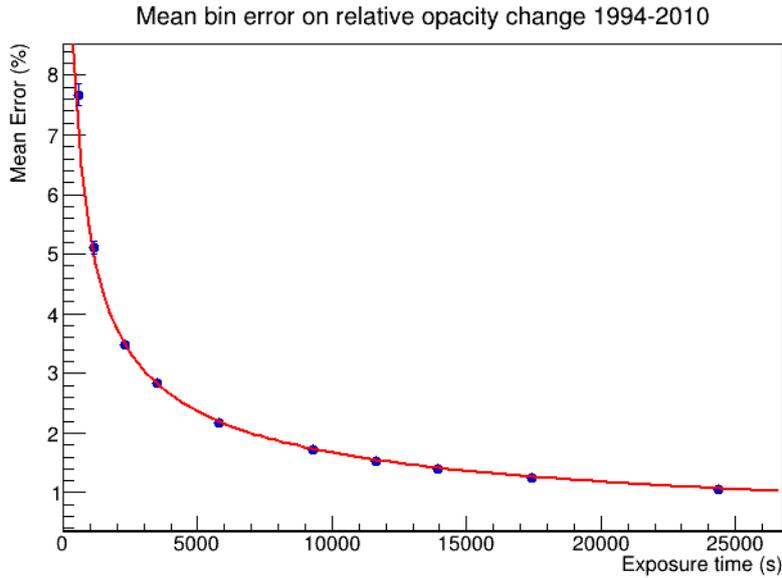
Figure 6.26: Mean error on the relative opacity changes in each angular bin between 1994 and 2010, as measured with muon radiography, for various exposure times. The data is well fitted by a $1/\sqrt{t}$ curve, where t is exposure time.

Mean uncertainty on opacity change (%)	Required exposure time (hours)
1	7.8 ± 0.1
0.5	31.1 ± 0.4
0.1	777 ± 9

Table 6.2: Extrapolated exposure times required to reach selected mean bin errors for the opacity change between 1994 and 2010, on a $10° \times 10°$ binning.

6.5.4 Generalising to other overburden materials

Finally, it is interesting to consider how these results would change at an alternative CCS site with a different overburden. For a location for which the overburden opacity is known, e.g. via seismic measurements, the corresponding muon flux can be estimated from the empirical relation equation 6.1. However, this relation only accounts for the opacity and not the additional effect of the chemical nature of the overburden minerals: materials with the same density but different chemical makeup will not attenuate muons to exactly the same degree. This was shown clearly in [45], which demonstrated via Monte Carlo simulation that muon flux through an overburden of a particular mineral was different than through a 'standard rock' (which has, by definition, $Z = 11$ and $A = 22$) with its density set to match the mineral. It was also demonstrated that the magnitude of the relative change in flux increases with depth, and that the key variable of the mineral is its $\langle Z^2/A \rangle$ i.e. the mass-weighted mean of the Z^2/A of each of its constituent elements.

Figure 6.27 shows this clearly by comparing the mineral to density-corrected standard rock flux ratios to each mineral's $\langle Z^2/A \rangle$, for a 1 km-thick overburden. For example, limestone

has density $\rho = 2.711$ g cm^{-3} and $\langle Z^2/A \rangle = 6.275$; the flux under 1 km of limestone was found to be $\sim 86\%$ of that under 1 km of standard rock with density 2.711 g cm^{-3}. With a linear fit applied to the data, the equivalent flux ratios of other materials can be estimated. As described in section 6.3 the bulk composition of the model used in the results above is the mineral kaolinite, for which $\langle Z^2/A \rangle = 3.908$. The corresponding flux ratio is then calculated to be 1.30 ± 0.03. The uncertainty on this value is propagated from the uncertainties on the fit parameters, which are relatively large with this dataset.
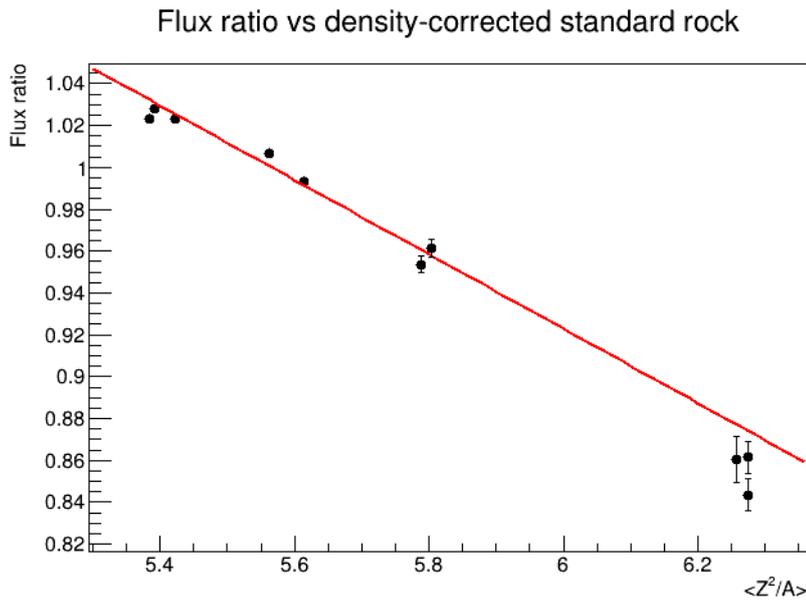
Figure 6.27: The ratio of the muon flux at 1km depth under various minerals to the flux at 1 km depth of density-corrected standard rock, data taken from [45]. The higher the mineral's $\langle Z^2/A \rangle$, the greater the deficit in flux.

This relation, in combination with the empirical relation in equation 6.1, can be used to estimate the muon vertical flux that would be measured under overburdens of different materials at a particular depth. For example, a 1 km depth of basalt, with density $\rho = 3.156$ g cm^{-3}, would have an opacity in the vertical direction of 3156 m.w.e.. Using equation 6.1, an equivalent opacity of standard rock would lead to a vertical flux of 1.37×10^{-8} Hz cm^{-2} sr^{-1}. However, basalt has $\langle Z^2/A \rangle = 6.258$, and hence the relation shown in figure 6.27 indicates that the flux should be adjusted by a factor of 0.877 ± 0.025, giving a final estimated measured flux of $1.20 \pm 0.03 \times 10^{-8}$ Hz cm^{-2} sr^{-1}. The equivalent fluxes for a range of minerals are shown in table 6.3.

Mineral	Density (g cm^{-3})	$\langle Z^2/A \rangle$	Estimated vertical flux (10^{-8} Hz cm^{-2} s^{-1})
Aragonite	2.939	6.275	1.66 ± 0.05
Limestone	2.711	6.275	2.38 ± 0.07
Arkose	2.347	5.563	5.03 ± 0.13
Dolomite	2.859	5.423	2.21 ± 0.06
Shale	2.512	5.384	3.91 ± 0.10
Arenite	2.357	5.392	5.03 ± 0.13
Granite	2.650	5.615	2.98 ± 0.08
Peridotite	3.340	5.788	1.01 ± 0.03
Andesite	2.812	5.803	2.22 ± 0.06
Basalt	3.156	6.258	1.20 ± 0.03
Kaolinite	2.630	3.908	4.23 ± 0.08

Table 6.3: The estimated vertical muon fluxes at a depth of 1 km, calculated using equation 6.1 and the flux ratios from [45] (shown in figure 6.27).

The $\sim 2-3\%$ errors on these values are larger than the $< 1\%$ flux changes between the 1994 and 2010 vintages of the Sleipner dataset. Therefore, it is not currently possible to predict e.g. the exposure time necessary for a 3σ detection of the CO_2 plume under different overburden materials. The results are still of qualitative interest however. A larger flux than that at the Sleipner site would be expected at sites with, for example, an overburden of arenite (a mineral similar to sandstone); around 150% of the Sleipner flux. This will reduce the necessary exposure time for plume detection, as less time will be required to obtain sufficient hits to detect the effect of the plume on the flux. Conversely, a site with a predominant denser overburden mineral, such as basalt, will require longer exposure times.

6.6 Conclusions

To conclude, the results presented above represent the first full simulation of muon radiography of a real, currently active CCS site: the Sleipner gas field in the North Sea. Previous simulated [112] and experimental [122] have demonstrated the potential of the technique via study of the Boulby mine site, which has a similar overburden to a typical CCS site. This work builds on these results by extending the simulation phase to a dedicated CCS site.

A 'ray-tracing' assay of the model using fictional geantino particles has demonstrated how the CO_2 plume and its motion appear to a detector placed in the deviated well at a depth of 1030 m. The resulting opacity changes are found to be of the order of $0.1-0.2\%$. Comparing the evolution of the opacity deficits over time clear shows the increasing volume of CO_2 from the larger region of reduced opacity; also visible is finer structure in the plume itself. The estimate of CO_2 mass injection rate derived from these opacities is a good match to the known injected mass, and clearly demonstrates the increased spreading out of the plume over time as the region within the detector field of view becomes saturated, with further CO_2 entering other regions of the aquifer.

To progress to full muon simulations, methods were tested to boost the efficiency of simulations via cuts on the allowable initial muon states. It was shown that whilst these

techniques can boost efficiency, performing an efficient muon simulation through a km-scale overburden to a small detector remains computationally challenging. Instead approximations are necessary in which a large detector is used to obtain muon rate measurements that are then scaled to test other, much smaller detector sizes.

Using a large 10^4 m^2 detector, a detection of the CO_2 plume in 2010 i.e. after injection of ~ 14 Mt of CO_2 is possible to a high ($\sim 5\sigma$) significance after a few hours of exposure time. Using these results to extrapolate to a smaller 10^3 m^2 system, as was used in [112], shows somewhat lower expected significances, most likely due to the deeper detector location used in these results. When adjusted to a smaller detecting area corresponding to coupled cylindrical detectors of diameter 65 mm placed in a borehole, it was found that a 100m length of borehole would require exposure times on the order of 1-3 years to establish the presence of the plume to 3σ. However, longer systems require shorter exposures, $2-4$ months' exposure for 1 km length.

Finally, it was shown that with a large detecting area (10^4 m^2), long exposures of several weeks are required for opacity measurements to be precise enough for the small changes in opacity caused by the plume movement over time to be identified clearly. This suggests that a more realistic detecting area, on the order of 10^2 m^2, would require an impractically long exposure time ($\gtrsim$ several years). The obvious conclusion is that the technique is unlikely to be suitable for detecting the detailed movement of the CO_2 plume in real time at this depth of rock.

The most significant limitation of the results presented above is computational. The ~ 7 hour exposure time of the muon simulations is sufficient to acquire order of magnitude estimates of the times required for detection of the plume, but there are considerable uncertainties on the values obtained that would be reduced with a longer simulated exposure time. A clear direction for future study therefore would be to use an alternative muon transport code (e.g. MUSIC [30] or PUMAS [31]) to obtain longer exposure times in practical computation times. The ~ 7 hour exposure times yield a total flux measurement with a relative uncertainty of $\sim 0.2\%$, as compared to a flux change between 1994 and 2010 of $\sim 1\%$; a reasonable target would be to reduce this to $\sim 0.05\%$ requiring a simulated exposure of $2-3$ days. This would also allow the smaller flux changes between the other available model timesteps, e.g. between 2001 and 2004, to become resolvable.

Chapter 7

Summary and conclusions

In this thesis, a number of interesting and novel studies into the use of muons for imaging and assay have been presented. Here some key results from the above are summarised, followed by a look into the future of the field.

A study into the use of machine learning techniques for muon scattering tomography showed that a system based on multivariate analysis classifiers could accurately identify the materials of objects inside concrete-filled waste drums after 10 days' exposure. The classifiers were based on exploiting variables generated using the binned clustering algorithm. The completed system was able to separate out image voxels corresponding to stored objects, group them into clusters, and, when calibrated by the object volume, accurately assign them their materials. When tested for its ability to identify 6 cm radius uranium spheres, the system performed well with a ROC AUC of 0.91 ± 0.06.

In a larger context, the obtained results, alongside the discussed previous applications of machine learning, illustrate that the information content accessible to MST is very rich and is not fully accessed by the described imaging algorithms in their usual operation. Machine learning is an ideal tool for extracting this additional information. In the near future, it is likely that practical MST systems will make use of machine learning as an inherent part of their data interpretation, allowing for, for example, faster 'threat/non-threat' decisions when assaying cargo or more accurate detection of boundaries between materials inside waste drums.

A test of muon radiography for void detection in long-term waste repositories gave promising results. Even with a fairly small total detection area (< 20 m^2), narrow (but long) cylindrical voids were found to detectable within exposure times < 1 year. As repositories have operational lifetimes of many years to decades, this is (broadly speaking) an acceptable level of performance. However, this is dependent on the required void detection being a 'stable' concern, as would be the case for a pre-existing void in the overburden, but not for a clandestine shaft for malicious waste access. A counter to this problem would be to increase the detection area by adding additional detection systems. This will reduce the necessary exposure time for void detection in addition to providing a more holistic view of the whole repository and overburden volume.

In the near future, it is intended that this work will be extended via an experimental demonstration. This would involve the Grimsel test site [105], a research laboratory located underneath a Swiss glacier, at a depth of ~ 500 m. The goal of this study would be to demonstrate the efficacy of muon radiography for void detection at this depth, using a combination of planar and cylindrical borehole detection systems. If approved, the programme of research will begin with a simulation phase based on an accurate digital model of the Grimsel overburden, in order to obtain a 'null hypothesis' radiography and to test different detector positions. It will then progress to an experimental study at the site.

Results from the first muon radiography simulation study of a modelled real, active carbon capture and storage site were presented. Using a very large detection area (10^4 m^2) and extrapolating to different lengths of cylindrical borehole detectors, the results indicated that, to detect the CO_2 plume within a few months, the required detector length would be $0.5 - 1$ km. This would be challenging to implement for a CCS site due to the considerable cost of these advanced detectors. However, shorter lengths of borehole systems could still be used for continuous monitoring of a plume if the detection requirement is relaxed and the system is instead used in combination with other episodic methods. For example, a low-significance ($1 - 2\sigma$) indication of possible dangerous plume movement detected by a borehole muography system could be taken as a signal to perform a seismic survey, allowing the situation to be ascertained; the episodic nature of seismic surveys could lead to such movement being missed if they are e.g. only performed at fixed intervals.

The field of muography is entering a potentially seismic moment. Increasing awareness of the technique's potential, coupled with the widespread availability of both powerful computational simulation software and practical particle detection systems, is leading to a rapid increase in both muon radiography and muon scattering tomography studies across many fields. Figure 7.1 shows the trajectory of muography over the last decade by comparing the number of Google Scholar results for the two key techniques by year. Although this is a somewhat crude measure of the number of publications, the trend is very clear; the number of results having more than doubled between 2010 and 2021. This is also reflected in the initiation in 2021 of the first conference devoted exclusively to muography, the 'International Workshop on Cosmic-Ray Muography' [123]. This first edition had 142 participants and it is hoped that the conference will become an annual event, bringing together the whole muography community.

As the field continues to grow, it is hoped that muography expertise will become widely utilised across many fields, with 'muographer' eventually being a recognised career in its own right. As the necessary detector technology continues to improve and develop, cosmic-ray muons will be used increasingly routinely across the energy sector and beyond. In the long term, it is hoped that the the technique will reach its maximum potential, and contribute to the fullest extent to science and to the collective good of humanity.
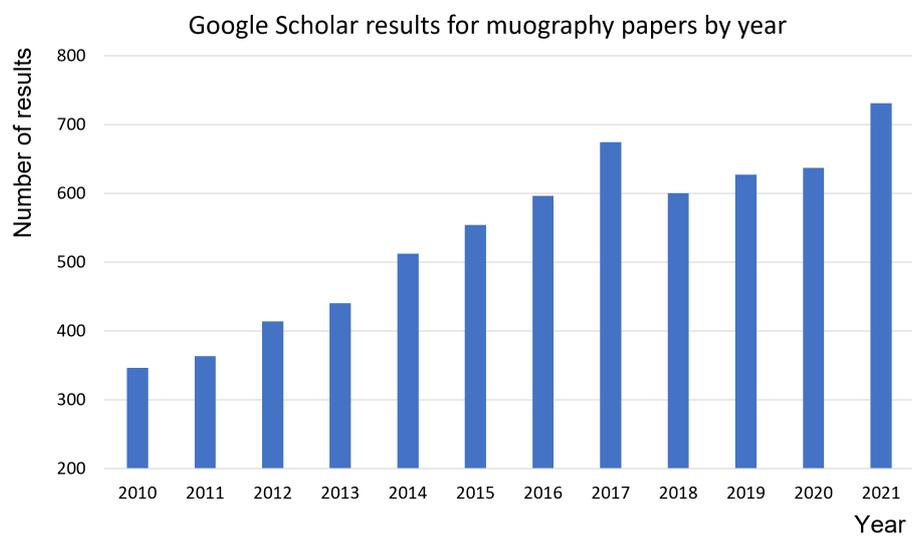
Figure 7.1: Number of Google Scholar results for 'muon radiography' and 'muon scattering tomography' per year. This is a crude measure, but the growth of the field over the last decade is clear.

Acknowledgements

First and foremost, I would like to express huge gratitude to my supervisor, Professor Lee Thompson. These have been very difficult years for everyone in the world, and I could not have completed this work without Lee's support, understanding and kindness. I am extremely grateful also to Dr Patrick Stowell for years of help and guidance. I would like to thank additionally Dr Chris Steer for his invaluable expertise, patience and advice over the last year. I would like to extend special thanks to Dr Sam Fargher and Dr Celeste Pidcott in particular, for putting up with years of daft questions with fortitude and coffee.

My family have been a source of constant support throughout my PhD. During the difficult days at the height of the pandemic, I moved back home for a time, and through that period I knew I could rely on my parents and sister to help me continue and ultimately complete my work. I am forever grateful for their belief and love.

Finally, I would like to express my gratitude and love to my friends Helen, Chris, Will, Ruth, Esmé, Karen, and Max, without whom I could have transported nary a single muon.

Appendix A: Machine learning methods

Here are described the various machine learning methods tested for the material discrimination studies of chapter 4, along with the hyperparameter sets used for each.

Boosted Decision Trees (BDT)
A BDT classifies an event by passing its variables down through a chained series of nodes, called a 'tree'. At each node, the next branch is determined by a cut on one of the variables, until the final nodes at the end of the tree ('leaf' nodes) are reached. The phase space is thus split into signal and background regions, with the goal being for each event to end in a leaf node that will characterise it as clearly signal or background. An example decision tree from one of the classifiers used in shown in figure 4.12.

A single BDT is vulnerable to statistical fluctuations in the training sample, which can be mitigated by using a large number of trees (a 'forest'), with the final classification arrived at as a weighted average of the set of trees. Performance of the classifier can also be improved using a boosting method to combine the weak classifiers of the trees into a single stronger classifier. Here gradient boosting has been used.

The following hyperparameters were used for the BDT classifiers:

- Number of trees: 800

- Maximum depth of each tree: 3

- Minimum proportion of training events in one leaf node: 5%

- Gradient boost learning rate: 0.01

- Number of grid points used for optimal variable splitting: 20

Multi-Layer Perceptron (MLP)
The MLP method is a type of artificial neural network. It consists in general of input and output layers and at least one hidden layer, with only inter-layer connections between adjacent layers permitted. The following hyperparameters were used for the MLP classifier test:

- Neuron activation function: tanh

- Number of training cycles: 600

- Number of epochs before overtraining check performed: 5

- Training method: back-propagation

- Number of hidden layers: 1

- Number of neurons in hidden layer: $n_{\mathrm{vars}} + 5$

Friedman-Popescu's method (RuleFit)

This method implements the RuleFit technique described in [124]. A series of rules $r(\mathbf{x})$ are defined that act on the set of variables x_i for an event. Each rule is the product of a series of boolean conditions on the variables; the rule will only return 1 if all conditions are met. The rules can be extracted from a forest of decision trees. A weighted sum of the set of rules is calculated for each event to provide the final response for that event.

The following hyperparameters were used for the RuleFit classifier test:

- Minimum rule importance accepted: 0.001

- Minimum distance between rules: 0.001

- Number of trees in forest: 20

- Minimum fraction of events in a node: 0.01

- Maximum fraction of events in a node: 0.5

- Number of steps along gradient-directed path: 10000

Support Vector Machine (SVM)

The SVM method treats training events as vectors with length n_{vars} then attempts to find an optimal hyperplane by maximising the distance between the plane and a subset of the vectors. If the data is not linearly separable, SVM is able to map the input variables onto a higher-dimensional space in which linear separation becomes possible. Import parameters of the method include the cost parameter, which controls the scale of the effect of misclassified events during the training process; and the tolerance parameter, which controls the precision of the minimisation process used when selecting the support vectors. The following hyperparameters were used for the SVM classifier test:

- Cost parameter: 1

- Tolerance parameter: 0.01

- Maximum number of training loops: 1000

Function Discriminant Analysis (FDA)

The FDA method attempts to fit the parameters of a provided simple function such that, when data in input, the output of signal (background) events is as close as possible to 1 (0). It is most suited for relatively simple problems with few variables. The fitting method used can be chosen by the user, but by default the MINUIT [69] software library is used.

The following hyperparameters were used for the FDA classifier test:

- Fitting method: MINUIT

- Discrimination formula: $(0) + (1)x_0 + (2)x_1 + (3)x_2 + (4)x_3$

- Parameter ranges: $(-1, 1); (-10, 10); (-10, 10); (-10, 10); (-10, 10)$

Appendix B: List of publications

Journal papers

- *Material identification in nuclear waste drums using muon scattering tomography and multivariate analysis*, M. Weekes, A. Alrheli, D. Barker, D. Kikoła, A. K. Kopp, M. Mhaidra, J. Stowell, L. Thompson, and J. J. Velthuis. Journal of Instrumentation, vol. 16, no. 05, 2021

Conference papers and presentations

- *Material identification of bodies stored in nuclear waste drums using muon scattering tomography and machine learning*, M. Weekes, A. Alrheli, D. Barker, D. Kikoła, A. K. Kopp, M. Mhaidra, J. Stowell, L. Thompson, and J. J. Velthuis. Presented at WM2021 Radwaste Management Virtual Conference, March 2021

- *Identifying materials in nuclear waste drums using muon scattering tomography and machine learning*, M. Weekes, A. Alrheli, D. Barker, C. De Sio, D. Kikoła, A. K. Kopp, M. Mhaidra, J. Stowell, L. Thompson, and J. J. Velthuis. Presented at IOP Joint APP, HEPP and NP Conference, April 2021

- *Design information verification of a geological repository for high level waste using muon radiography*, M. Weekes, K. Aymanns, I. Niemeyer, P. Stowell, L. Thompson, C. Vieh. Presented at INMM & ESARDA Joint Virtual Annual Meeting, August 2021

Posters

- *Applications of muon scattering tomography to image and characterise materials in nuclear waste drums*, M. Weekes. Presented at IOP Joint APP and HEPP Conference, April 2019

- *Identifying materials in nuclear waste drums with muon scattering tomography*, M. Weekes. Presented at International Workshop on Cosmic-Ray Muography, November 2021

Bibliography

[1] V. Hess, "On the observations of the penetrating radiation during seven balloon flights," *Phys. Zeitschr.*, vol. 13, pp. 1084–1091, 1912.

[2] D. Pacini, "Penetrating radiation at the surface and in water," *Nuovo Cimento VI*, vol. 3, pp. 93–100, 1912.

[3] S. H. Neddermeyer and C. D. Anderson, "Note on the nature of cosmic-ray particles," *Physical Review*, vol. 51, no. 10, p. 884, 1937.

[4] T. K. Gaisser, R. Engel, and E. Resconi, *Atmospheric muons and neutrinos*, p. 126–148. Cambridge University Press, 2 ed., 2016.

[5] M. Guan, M.-C. Chu, J. Cao, K.-B. Luk, and C. Yang, "A parametrization of the cosmic-ray muon flux at sea-level," *arXiv preprint arXiv:1509.06176*, 2015.

[6] D. Chirkin, "Fluxes of atmospheric leptons at 600 GeV-60 TeV," *arXiv preprint hep-ph/0407078*, 2004.

[7] A. Barbouti and B. Rastin, "A study of the absolute intensity of muons at sea level and under various thicknesses of absorber," *Journal of Physics G: Nuclear Physics*, vol. 9, no. 12, p. 1577, 1983.

[8] D. E. Groom, N. V. Mokhov, and S. I. Striganov, "Muon stopping power and range tables 10 MeV–100 TeV," *Atomic Data and Nuclear Data Tables*, vol. 78, no. 2, pp. 183–356, 2001.

[9] A. Bogdanov, H. Burkhardt, V. Ivanchenko, S. Kelner, R. Kokoulin, M. Maire, A. Rybin, and L. Urban, "Geant4 simulation of production and interaction of muons," *IEEE Transactions on nuclear science*, vol. 53, no. 2, pp. 513–519, 2006.

[10] P. NS *et al.*, "Range straggling of muons in the energy range 200-100 000 GeV," *Proceedings of the Physical Society*, 1962.

[11] H. Bethe, "Zur theorie des durchgangs schneller korpuskularstrahlen durch materie," *Annalen der Physik*, vol. 397, no. 3, pp. 325–400, 1930.

[12] R. Sternheimer and R. Peierls, "General expression for the density effect for the ionization loss of charged particles," *Physical Review B*, vol. 3, no. 11, p. 3681, 1971.

[13] G. Collaboration, S. Agostinelli, *et al.*, "GEANT4: a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, no. 25, p. 0, 2003.

[14] S. R. Kelner, R. P. Kokoulin, and A. A. Petrukhin, "About cross section for high-energy muon bremsstrahlung," tech. rep., Aug 1995.

[15] H. Bethe and W. Heitler, "On the stopping of fast particles and on the creation of positive electrons," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 146, no. 856, pp. 83–112, 1934.

[16] A. Bulmahn and M. H. Reno, "Cross sections and energy loss for lepton pair production in muon transport," *Physical Review D*, vol. 79, no. 5, p. 053008, 2009.

[17] L. Bezrukov and E. Bugaev, "Nucleon shadowing effects in photonuclear interactions," *Sov. J. Nucl. Phys.(Engl. Transl.);(United States)*, vol. 33, no. 5, 1981.

[18] W. R. Leo, *Techniques for nuclear and particle physics experiments: a how-to approach.* Springer Science & Business Media, 2012.

[19] V. Anghel, J. Armitage, F. Baig, K. Boniface, K. Boudjemline, J. Bueno, E. Charles, P. Drouin, A. Erlandson, G. Gallant, *et al.*, "A plastic scintillator-based muon tomography system with an integrated muon spectrometer," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 798, pp. 12–23, 2015.

[20] "Geoptic infrastructure investigations." https://www.geoptic.co.uk/.

[21] N. Ampilogov, I. Astapov, N. Barbashina, V. Borog, D. Chernov, A. Dmitrieva, K. Kompaniets, A. Petrukhin, V. Shutenko, A. Teregulov, *et al.*, "Large area scintillation muon hodoscope for monitoring of atmospheric and heliospheric processes," *Astrophysics and Space Sciences Transactions*, vol. 7, no. 3, pp. 435–438, 2011.

[22] A. Ariga, T. Ariga, A. Ereditato, S. Käser, A. Lechmann, D. Mair, R. Nishiyama, C. Pistillo, P. Scampoli, F. Schlunegger, *et al.*, "A nuclear emulsion detector for the muon radiography of a glacier structure," *Instruments*, vol. 2, no. 2, p. 7, 2018.

[23] T. Fukuda, "The nuclear emulsion technology and the analysis of the OPERA experiment data," *arXiv preprint arXiv:0910.3274*, 2009.

[24] K. Morishima, M. Kuno, A. Nishio, N. Kitagawa, Y. Manabe, M. Moto, F. Takasaki, H. Fujii, K. Satoh, H. Kodama, *et al.*, "Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons," *Nature*, vol. 552, no. 7685, pp. 386–390, 2017.

[25] G. Charpak and F. Sauli, "Multiwire proportional chambers and drift chambers," *Nuclear Instruments and Methods*, vol. 162, no. 1-3, pp. 405–428, 1979.

[26] Wiso, "Equipotential line and field line in a MWPC," 2007. This work is licensed under the CC BY-SA 3.0 License. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/3.0/deed.en.

[27] L. Oláh, S. J. Balogh, Á. L. Gera, G. Hamar, G. Nyitrai, H. K. Tanaka, and D. Varga, "MWPC-based Muographic Observation System for remote monitoring of active volcanoes," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 57–58, 2019.

[28] C. Hagmann, D. Lange, and D. Wright, "Cosmic-ray shower generator (CRY) for Monte Carlo transport codes," in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 2, pp. 1143–1146, IEEE, 2007.

[29] P. Mitra and P. Shukla, "Energy and angular distribution of cosmic muons," in *Proceedings of the DAE-BRNS Symp. on Nucl. Phys*, vol. 60, p. 890, 2015.

[30] V. Kudryavtsev, "Muon simulation codes MUSIC and MUSUN for underground physics," *Computer Physics Communications*, vol. 180, no. 3, pp. 339–346, 2009.

[31] V. Niess, A. Barnoud, C. Carloganu, and E. Le Ménédeu, "Backward Monte-Carlo applied to muon transport," *Computer Physics Communications*, vol. 229, pp. 54–67, 2018.

[32] Z.-X. Zhang, T. Enqvist, M. Holma, and P. Kuusiniemi, "Muography and its potential applications to mining and rock engineering," *Rock Mechanics and Rock Engineering*, vol. 53, no. 11, pp. 4893–4907, 2020.

[33] E. George, "Cosmic rays measure overburden of tunnel," *Commonwealth Engineer*, vol. 455, 1955.

[34] L. Thompson, J. Stowell, S. Fargher, C. Steer, K. Loughney, E. O'Sullivan, J. Gluyas, S. Blaney, and R. Pidcock, "Muon tomography for railway tunnel imaging," *Physical Review Research*, vol. 2, no. 2, p. 023017, 2020.

[35] H. Tanaka, T. Nakano, S. Takahashi, J. Yoshida, and K. Niwa, "Development of an emulsion imaging system for cosmic-ray muon radiography to explore the internal structure of a volcano, Mt. Asama," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 575, no. 3, pp. 489–497, 2007.

[36] H. K. Tanaka, T. Kusagaya, and H. Shinohara, "Radiographic visualization of magma dynamics in an erupting volcano," *Nature Communications*, vol. 5, no. 1, pp. 1–9, 2014.

[37] H. K. Tanaka, T. Uchida, M. Tanaka, H. Shinohara, and H. Taira, "Cosmic-ray muon imaging of magma in a conduit: Degassing process of satsuma-iwojima volcano, japan," *Geophysical Research Letters*, vol. 36, no. 1, 2009.

[38] A. Anastasio, F. Ambrosino, D. Basta, L. Bonechi, M. Brianzi, A. Bross, S. Callier, A. Caputo, R. Ciaranfi, L. Cimmino, *et al.*, "The mu-ray detector for muon radiography of volcanoes," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 732, pp. 423–426, 2013.

[39] A. Nishio, K. Ishiguro, N. Kitagawa, M. Kuno, Y. Manabe, and K. Morishima, "Development of a nondestructive inspection method for trees using cosmic rays," *Journal of Advanced Instrumentation in Science*, vol. 2022, Apr. 2022.

[40] L. W. Alvarez, J. A. Anderson, F. E. Bedwei, J. Burkhard, A. Fakhry, A. Girgis, A. Goneid, F. Hassan, D. Iverson, G. Lynch, *et al.*, "Search for hidden chambers in the pyramids: The structure of the Second Pyramid of Giza is determined by cosmic-ray absorption.," *Science*, vol. 167, no. 3919, pp. 832–839, 1970.

[41] P. Y. Lipscy, K. E. Kushida, and T. Incerti, "The Fukushima disaster and Japan's nuclear plant vulnerability in comparative perspective," *Environmental science & technology*, vol. 47, no. 12, pp. 6082–6088, 2013.

[42] J. Percy and I. Rogers, "Cosmic ray intensities at shallow depths," in *International Cosmic Ray Conference*, vol. 6, pp. 99–103, 1977.

[43] A. Wright, "A study of muons underground and their energy spectrum at sea level," *Journal of Physics A: Mathematical, Nuclear and General*, vol. 7, no. 16, p. 2085, 1974.

[44] B. Srikantan and S. Naranan, "Cosmic rays underground," in *Proceedings of the Indian Academy of Sciences-Section A*, vol. 36, pp. 97–117, Springer India, 1952.

[45] A. Lechmann, D. Mair, A. Ariga, T. Ariga, A. Ereditato, R. Nishiyama, C. Pistillo, P. Scampoli, F. Schlunegger, and M. Vladymyrov, "The effect of rock composition on muon tomography measurements," *Solid Earth*, vol. 9, no. 6, pp. 1517–1533, 2018.

[46] K. Nagamine, M. Iwasaki, K. Shimomura, and K. Ishida, "Method of probing inner-structure of geophysical substance with the horizontal cosmic-ray muons and possible application to volcanic eruption prediction," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 356, no. 2-3, pp. 585–595, 1995.

[47] V. L. Highland, "Some practical remarks on multiple scattering," *Nuclear Instruments and Methods*, vol. 129, no. 2, pp. 497–499, 1975.

[48] D. E. Groom and S. Klein, "Passage of particles through matter," *The European Physical Journal C-Particles and Fields*, vol. 15, no. 1, pp. 163–173, 2000.

[49] S. Riggi, V. Antonuccio-Delogu, M. Bandieramonte, U. Becciani, A. Costa, P. La Rocca, P. Massimino, C. Petta, C. Pistagna, F. Riggi, *et al.*, "Muon tomography imaging

algorithms for nuclear threat detection inside large volume containers with the Muon Portal detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 728, pp. 59–68, 2013.

[50] C. Thomay, J. J. Velthuis, P. Baesso, D. Cussans, P. Morris, C. Steer, J. Burns, S. Quillin, and M. Stapleton, "A binned clustering algorithm to detect high-Z material using cosmic muons," *Journal of Instrumentation*, vol. 8, no. 10, p. P10013, 2013.

[51] M. Stapleton, J. Burns, S. Quillin, and C. Steer, "Angle statistics reconstruction: a robust reconstruction algorithm for muon scattering tomography," *Journal of instrumentation*, vol. 9, no. 11, p. P11019, 2014.

[52] U.S. Congress, "Atomic Energy Act Title 1," 1954.

[53] W. Bank, "The Great East Japan earthquake, learning from megadisasters: knowledge notes executive summary," *International Bank for Reconstruction and Development; The World Bank*, 2012.

[54] K. Borozdin, S. Greene, Z. Lukić, E. Milner, H. Miyadera, C. Morris, and J. Perry, "Cosmic ray radiography of the damaged cores of the Fukushima reactors," *Physical review letters*, vol. 109, no. 15, p. 152501, 2012.

[55] H. Miyadera, K. N. Borozdin, S. J. Greene, Z. Lukić, K. Masuda, E. C. Milner, C. L. Morris, and J. O. Perry, "Imaging Fukushima Daiichi reactors with muons," *Aip Advances*, vol. 3, no. 5, p. 052133, 2013.

[56] H. Fujii, K. Hara, S. Hashimoto, F. Ito, H. Kakuno, S. Kim, M. Kochiyama, K. Nagamine, A. Suzuki, Y. Takada, *et al.*, "Performance of a remotely located muon radiography system to identify the inner structure of a nuclear plant," *Progress of Theoretical and Experimental Physics*, vol. 2013, no. 7, p. 073C01, 2013.

[57] H. Fujii, K. Hara, K. Hayashi, H. Kakuno, H. Kodama, K. Nagamine, K. Sato, S.-H. Kim, A. Suzuki, T. Sumiyoshi, *et al.*, "Investigation of the Unit-1 nuclear reactor of Fukushima Daiichi by cosmic muon radiography," *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 4, p. 043C02, 2020.

[58] H. Fujii, M. Gi, K. Hara, S. Hashimoto, K. Hayashi, H. Kakuno, H. Kodama, M. Mizokami, S. Mizokami, K. Nagamine, *et al.*, "Investigation of the status of Unit 2 nuclear reactor of the Fukushima Daiichi by cosmic muon radiography," *Progress of Theoretical and Experimental Physics*, vol. 2021, no. 2, p. 023C01, 2021.

[59] H. Fujii, K. Hara, K. Hayashi, H. Kakuno, H. Kodama, K. Nagamine, K. Sato, S.-H. Kim, A. Suzuki, T. Sumiyoshi, *et al.*, "Study of the Unit-3 nuclear reactor of Fukushima Daiichi by cosmic muon radiography," *Progress of Theoretical and Experimental Physics*, 2022.

[60] Eurostat, "Complete energy balances." Online data code: NRG_BAL_C, 03 2022.

[61] J. Donev, B. Afework, J. Hanania, and K. Stenhouse, "Energy education - RBMK." Online, 2018. Available: https://energyeducation.ca/encyclopedia/RBMK, Accessed: August 2, 2022.

[62] Environment Agency and Office for Nuclear Regulation, "Scrutiny of RWM's work on geological disposal - annual report 2020 to 2021," tech. rep., May 2022.

[63] R. E. Berlin and C. C. Stanton, *Radioactive waste management.* John Wiley and Sons Inc., New York, NY, 1988.

[64] Low-Level Radioactive Waste Policy Act. 42 U.S.C. § 2021b. 1980.

[65] B. Pérot, F. Jallu, C. Passard, O. Gueton, P.-G. Allinei, L. Loubet, N. Estre, E. Simon, C. Carasco, C. Roure, *et al.*, "The characterization of radioactive waste: a critical review of techniques implemented or under development at CEA, France," *EPJ Nuclear Sciences & Technologies*, vol. 4, p. 3, 2018.

[66] "Chance: Characterization of conditioned radioactive waste." www.chance-h2020.eu/.

[67] W. R. Kubinski, D. Kikola, D. Tefelski, C. Carasco, C. Mathonat, B. Rogiers, and H. Tietze-Jaensch, "Radiation leakage and the precision of characterizing nuclear material using calorimetric assay," WM Symposia, Inc., PO Box 27646, 85285-7646 Tempe, AZ (United States), 2019.

[68] C. Thomay, J. Velthuis, T. Poffley, P. Baesso, D. Cussans, and L. Frazão, "Passive 3d imaging of nuclear waste containers with muon scattering tomography," *Journal of Instrumentation*, vol. 11, no. 03, p. P03008, 2016.

[69] F. James and M. Roos, "Minuit: a system for function minimization and analysis of the parameter errors and corrections," *Comput. Phys. Commun.*, vol. 10, no. CERN-DD-75-20, pp. 343–367, 1975.

[70] C. Steer, P. Stowell, and L. Thompson, "CRESTA: Cosmic rays for engineering, scientific, and technology applications." https://gitlab.com/cosmicraysim/cresta.

[71] "CERN accelerating science." geant4.web.cern.ch/node/155.

[72] J. Bae and S. Chatzidakis, "Fieldable muon momentum measurement using coupled pressurized gaseous Cherenkov detectors," *arXiv preprint arXiv:2201.02591*, 2022.

[73] C. Morris, C. Alexander, J. Bacon, K. Borozdin, D. Clark, R. Chartrand, C. Espinoza, A. Fraser, M. Galassi, J. Green, *et al.*, "Tomographic imaging with cosmic ray muons," *Science & Global Security*, vol. 16, no. 1-2, pp. 37–53, 2008.

[74] L. Frazão, J. Velthuis, C. Thomay, and C. Steer, "Discrimination of high-z materials in concrete-filled containers using muon scattering tomography," *Journal of Instrumentation*, vol. 11, no. 07, p. P07020, 2016.

[75] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, *et al.*, "Tmva-toolkit for multivariate data analysis," *arXiv preprint physics/0703039*, 2007.

[76] R. Brun and F. Rademakers, "ROOT—an object oriented data analysis framework," *Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment*, vol. 389, no. 1-2, pp. 81–86, 1997.

[77] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[78] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[79] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[80] A. Hocker, X. Prudent, J. Therhaag, Y. Mahalalel, M. Backes, R. Ospanov, M. Kruk, M. Jachowski, A. Voight, A. Robert, *et al.*, "TMVA-toolkit for multivariate data analysis with ROOT: users guide," tech. rep., 2007.

[81] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory: Second European Conference, EuroCOLT'95 Barcelona, Spain, March 13–15, 1995 Proceedings 2*, pp. 23–37, Springer, 1995.

[82] J. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[83] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," tech. rep., Stanford, 2006.

[84] W. R. Alexander and L. McKinley, *Deep geological disposal of radioactive waste.* Elsevier, 2011.

[85] R. L. Clark, "Background on 40 CFR part 197 environmental radiation protection standards for Yucca Mountain," in *Proceedings of the 1997 Waste Management Conference*, 1997.

[86] M. Adams, "Yucca Mountain-Nevada's perspective," *Idaho L. Rev.*, vol. 46, p. 423, 2009.

[87] "Posiva annual report 2021," tech. rep., Posiva Oy, april 2022.

[88] P. Oy, "Onkalo demo," oct 2020.

[89] Z. Bukovska, I. Soejono, L. Vondrovic, M. Vavro, K. Souček, D. Burianek, P. Dobeš, O. Švagera, P. Waclawik, J. Řihošek, *et al.*, "Characterization and 3d visualization of underground research facility for deep geological repository experiments: A case study of underground research facility bukov, czech republic," *Engineering Geology*, vol. 259, p. 105186, 2019.

[90] IAEA, *Safeguards Implementation Practices Guide on Establishing and Maintaining State Safeguards Infrastructure.* No. 31 in Services Series, 2018.

[91] B. Rastin, "An accurate measurement of the sea-level muon spectrum within the range 4 to 3000 GeV/c," *Journal of Physics G: Nuclear Physics*, vol. 10, no. 11, p. 1609, 1984.

[92] B. Nandi and M. Sinha, "The momentum spectrum of muons at sea level in the range 5-1200 GeV/c," *Journal of Physics A: General Physics*, vol. 5, no. 9, p. 1384, 1972.

[93] C. Stockel, "A study of muons deep underground. I. angular distribution and vertical intensity," *Journal of Physics A: General Physics*, vol. 2, no. 6, p. 639, 1969.

[94] P. Oy, "Havainnekuva loppusijoituslaitoksesta," oct 2020.

[95] J. Mattila, J. Suikkanen, R. Read, J. Valli, M. Hakala, J. Sjöberg, B. Figueiredo, R. Kiuru, and S. Haapalehto, "Rock mechanics of Olkiluoto," *Posiva Report 2021-18*, 2022.

[96] H. Blatt and R. Tracy, *Petrology: Igneous, Sedimentary, and Metamorphic.* W. H. Freeman & Company, 1995.

[97] G. Saracino, F. Ambrosino, L. Bonechi, L. Cimmino, R. D'Alessandro, M. D'Errico, P. Noli, L. Scognamiglio, and P. Strolin, "Applications of muon absorption radiography to the fields of archaeology and civil engineering," *Philosophical Transactions of the Royal Society A*, vol. 377, no. 2137, p. 20180057, 2019.

[98] E. Guardincerri, C. Rowe, E. Schultz-Fellenz, M. Roy, N. George, C. Morris, J. Bacon, M. Durham, D. Morley, K. Plaud-Ramos, *et al.*, "3D cosmic ray muon tomography from an underground tunnel," *Pure and Applied Geophysics*, vol. 174, no. 5, pp. 2133–2141, 2017.

[99] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *Journal of theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.

[100] G. Baccani, "Muon absorption tomography of a lead structure through the use of iterative algorithms," *Journal of Instrumentation*, vol. 15, no. 12, p. P12024, 2020.

[101] S. Procureur, "Muon tomography of large structures with 2D projections," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1013, p. 165665, 2021.

[102] J. Amanatides and A. Woo, "A fast voxel traversal algorithm for ray tracing," in *In Eurographics '87*, pp. 3–10, 1987.

[103] "Forschungszentrum Jülich." https://www.fz-juelich.de/en.

[104] "Bundesgesellschaft Für Endlagerung." https://www.bge.de/de/abfaelle/.

[105] "The Grimsel Test Site." https://www.grimsel.com/.

[106] "CGG SA (CGG)," July 2022. https://www.cgg.com.

[107] E. Wolff, "Climate change: Evidence and causes.," *School Science Review*, vol. 96, no. 354, pp. 17–23, 2014.

[108] J. Hansen, R. Ruedy, M. Sato, and K. Lo, "Global surface temperature change," *Reviews of Geophysics*, vol. 48, no. 4, 2010.

[109] H. Pörtner, D. Roberts, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama, "Climate change 2022: Impacts, adaptation, and vulnerability," tech. rep., IPCC, 2022.

[110] V. A. Kudryavtsev, N. J. Spooner, J. Gluyas, C. Fung, and M. Coleman, "Monitoring subsurface CO_2 emplacement and security of storage using muon tomography," *International journal of greenhouse gas control*, vol. 11, pp. 21–24, 2012.

[111] S. Iglauer, *Dissolution trapping of carbon dioxide in reservoir formation brine-a carbon storage mechanism.* INTECH Open Access Publisher, 2011.

[112] J. Klinger, S. Clark, M. Coleman, J. Gluyas, V. Kudryavtsev, D. Lincoln, S. Pal, S. Paling, N. Spooner, S. Telfer, *et al.*, "Simulation of muon radiography for monitoring CO_2 stored in a geological reservoir," *International Journal of Greenhouse Gas Control*, vol. 42, pp. 644–654, 2015.

[113] S. Solomon, "Carbon dioxide storage: Geological security and environmental issues, case study on the Sleipner gas field in Norway," *Bellona report*, vol. 128, 2007.

[114] Bair175, "Sleipner oil field," 2013. This work is licensed under the CC BY-SA 3.0 License. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/3.0/deed.en.

[115] C. Barnini, "CGG delivers new multi-client screening study for CO_2 storage sites in UK and Norway northern North Sea," Nov 2021. https://www.cgg.com/newsroom/press-release/cgg-delivers-new-multi-client-screening-study-co2-storage-sites-uk-and.

[116] The Sleipner Group, "Sleipner CO_2 reference dataset license." https://co2datashare.org/landing_pages/sleipner-2019-benchmark-model/static/license.pdf.

[117] T. Crozat, C. Steer, L. Thompson, and M. Weekes, "Monitoring changes in CO_2 reservoirs using cosmic ray muons." Private communication.

[118] P. Kelemen, S. M. Benson, H. Pilorgé, P. Psarras, and J. Wilcox, "An overview of the status and challenges of CO_2 storage in minerals and geological formations," *Frontiers in Climate*, vol. 1, p. 9, 2019.

[119] T.-P. Li and Y.-Q. Ma, "Analysis methods for results in gamma-ray astronomy," *The Astrophysical Journal*, vol. 272, pp. 317–324, 1983.

[120] S. Gillessen and H. Harney, "Significance in gamma-ray astronomy–the Li & Ma problem in bayesian statistics," *Astronomy & Astrophysics*, vol. 430, no. 1, pp. 355–362, 2005.

[121] P. Kuusiniemi, M. Holma, K. Loo, and T. Enqvist, "Muography and geology: Does it matter which continent you stand on?," *Journal for Advanced Instrumentation in Science*, vol. 2022, no. 1, 2022.

[122] J. Gluyas, L. Thompson, D. Allen, C. Benton, P. Chadwick, S. Clark, J. Klinger, V. Kudryavtsev, D. Lincoln, B. Maunder, *et al.*, "Passive, continuous monitoring of carbon dioxide geostorage using muon tomography," *Philosophical Transactions of the Royal Society A*, vol. 377, no. 2137, p. 20180059, 2019.

[123] "International workshop on cosmic-ray muography (Muography2021)." https://indico.cern.ch/event/1033631/.

[124] J. H. Friedman and B. E. Popescu, "Predictive learning via rule ensembles," *The annals of applied statistics*, pp. 916–954, 2008.