



**LUNDS**  
UNIVERSITET

# Radiation Transport Calculations Around the ESS Accelerator Using Kernel Density Estimation

Rasmus Sjöo

---

Master Thesis

Project duration: One semester, 30 credits

Supervisors: Joakim Cederkäll, Douglas Di Julio,  
Jose Ignacio Marquez Damian

Department of Physics  
Division of Particle and Nuclear Physics  
May 2024



## Acknowledgements

Firstly, I want to give thanks to my supervisors at ESS, Douglas Di Julio and José Ignacio Marquez Damian. They guided me through the whole project and were always willing to help whenever I needed it. In addition, I want to thank Joakim Cederkäll for getting me in contact with ESS and helping make this project possible. I also want to acknowledge Gunther Muhrer, Shuqi Xu, Zoe Prieto, Michael Wohlmuther and Linus Persson, who all helped out with the project in different ways. Finally, I would like to acknowledge ESS for providing funding for this project.

## **Abstract**

In order to safely design accelerator-based research facilities, such as the European Spallation Source, shielding calculations are carried out using Monte-Carlo particle simulations. However, these calculations gain high uncertainty when large amounts of shielding are needed, due to decreasing particle statistics over distance and mass. To circumvent this issue, variance reduction methods are used to reduce errors in calculations. A possible method for this purpose is Kernel Density Estimation (KDE). In this thesis, KDE methods has been applied to shielding calculations done around the ESS linear accelerator. A model of the accelerator was constructed in which the calculations were ran using a Monte Carlo particle transport code. It was found that using KDE can be used to reduce uncertainty in shielding calculations. However, the parameters of the bandwidth selection method, such as the property importance, needs to be worked with in order to get a good estimate of the distributions of the particle properties.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	European Spallation Source . . . . .	1
1.2	Accelerator . . . . .	1
1.3	Goal . . . . .	2
<b>2</b>	<b>Theory</b>	<b>3</b>
2.1	Monte Carlo Methods . . . . .	3
2.2	Kernel density estimation . . . . .	4
2.3	Multivariate KDE . . . . .	4
2.4	Kernel Function Selection . . . . .	5
2.5	Bandwidth Selection . . . . .	5
2.5.1	Silverman’s Rule . . . . .	5
2.5.2	Non-Adaptive Maximum-Likelihood Cross-Validation . . . . .	6
2.5.3	Adaptive Maximum-Likelihood Cross-Validation . . . . .	6
<b>3</b>	<b>Software</b>	<b>6</b>
3.1	Comblayer . . . . .	7
3.2	PHITS . . . . .	7
3.3	KDSOURCE . . . . .	8
<b>4</b>	<b>Results and Discussion</b>	<b>8</b>
4.1	Model . . . . .	8
4.2	Source and Tallies . . . . .	11
4.3	KDSOURCE Resampling . . . . .	13
4.4	KDE Source Dose Rates . . . . .	14
4.4.1	Surface 1 . . . . .	14
4.4.2	Surface 2 . . . . .	15
<b>5</b>	<b>Summary</b>	<b>21</b>
<b>A</b>	<b>Test Problem</b>	<b>24</b>
A.1	Test Problem . . . . .	24
<b>B</b>	<b>Bandwidth Test</b>	<b>26</b>

# 1 Introduction

Radiation shielding calculations are an important aspect when building a particle accelerator. This is no exception for the accelerator at the European Spallation Source (ESS) [1]. Workers need to perform duties in rooms connected to the accelerator, even when it is active. In order to reduce dose rates during operation in these sections, large amounts of shielding are built around the accelerator tunnel. This does, however, create computational challenges for Monte Carlo shielding calculations. Uncertainty in calculations increases with distance from the source when large amounts of shielding are included, due to a decreasing amount of particles caused by probabilistic interactions happening in the shielding. This makes it difficult to get an accurate estimate of the dose rate far from the particle source. There are several variance reduction methods that can be used to get around this problem, however. One of them is to re-sample the source term at a distance away from the primary source. By estimating the probability distribution of particles in a region, a new source can be created by drawing from said distribution. A prospecting method for this end is called Kernel Density Estimation (KDE) [2]. This method is explained in section 2.2. It has the potential to be used for variance reduction in shielding calculations.

## 1.1 European Spallation Source

In 1998, an OECD report declared the need for new neutron sources in the world [3]. Up to that point, nuclear reactors had been used as the primary neutron sources for scattering experiments [4]. However, increasing scientific interest combined with the existing sources using outdated technology led to demand for next generation neutron sources. The European Strategy Forum for Research Infrastructures concluded in 2003 that a 5 MW neutron source with a single target station and 22 instruments would meet the scientific needs in Europe [4]. This led to the construction of The European Spallation Source in Lund, Sweden, which is still underway. The facility is expected to be fully operational by the year 2027.

## 1.2 Accelerator

One of the components of the facility is the linear accelerator, which provides protons for inducing spallation reactions. In this process, high-energy protons hit a metal target, exciting nuclei, which then release neutrons upon being de-excited. Protons are produced from a Microwave Discharge Ion Source in the start of the tunnel. 62.5 mA worth of protons are produced and subsequently accelerated to 2 GeV through a series of conducting and superconducting magnets placed in different sections of the tunnel [5]. Figure 1 shows the sections that the accelerator is split into. The high-

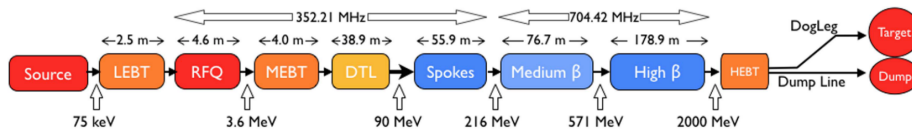


Figure 1: The different sections of the linear accelerator at ESS, including their lengths and the energy levels protons are accelerated to at different steps. Image credit: [4]

energy protons are then fired at a rotating tungsten target, which produces neutron beams with an average power of 5 MW.

The high-energy particles emit radiation in the tunnel during operation. Therefore, heavy shielding has been built around the accelerator in order to protect workers and the public. The shielding does however contain some penetrations, for access to utilities such as water and cables, as well as allowing workers to perform accelerator maintenance. There are also chicanes in the shielding which connect the tunnel to other sections of the facility, such as the Klystron gallery. This creates potential weak spots, which might require additional shielding. There is also the issue of deep-penetration. Both neutrons and photons can travel deep into even heavy materials before stopping due to interactions, especially when the primary energy is high.

### 1.3 Goal

The aim of this project was to investigate if Kernel Density Estimation is a viable method for variance reduction in shielding calculations for high-energy accelerators. To do this, Monte Carlo particle simulations were used in a model of the ESS linear accelerator, using a section of the beamline as a source. A previously constructed model of the accelerator was extended to include additional sections of the tunnel before running the simulations. Simulated particles in the model were tracked, and their properties were recorded at certain locations. This data was then used to resample new particles with KDE software, which were used as the source in new simulations. The properties of the resampled particles were compared to the original ones to evaluate how well KDE captured the beamline source. In addition, the respective dose rates were used to evaluate if the resampled source produced doses similar to the ones from the beamline. The following section explains how Monte Carlo simulations work, and also goes into the relevant theory needed to understand Kernel Density Estimation. Section 3 explains the different software used in the project, such as KDSOURCE, which uses KDE to generate Monte Carlo particle sources. Part of this work has been featured in the SATIF-16 conference [6].

## 2 Theory

In this section, Monte Carlo methods and KDE are explained.

### 2.1 Monte Carlo Methods

Monte Carlo methods are a group of algorithms that solve deterministic problems by generating random objects, such as sampling from probability distributions [7]. One application of the Monte Carlo method is modeling particle transport. This works differently from deterministic models, which commonly solve the transport equation to approximate particle behavior. The Monte Carlo method instead simulates individual particles and tallies some properties of their average behavior. The central limit theorem is then used to infer how the particles behave in physical space.

When using Monte Carlo simulations, it is important to determine the uncertainties in order to know how reliable any results are. A main cause of errors is limitations in sampling data [8]. If there isn't sufficient material to draw from, Monte Carlo methods will give unreliable results. The amount of particles in a Monte Carlo simulation will decrease with distance from the source if there is significant shielding, due to particles being stopped along the way by interactions. This causes high uncertainty far from the source, which is why variance reduction is needed. To evaluate how accurate a Monte Carlo simulation is, the Relative error,  $R$  is used, which is defined as

$$R = \frac{S_{\bar{x}}}{\bar{x}} \quad (1)$$

where  $S_{\bar{x}}$  is the estimated standard deviation of the mean and  $\bar{x}$  is the estimated mean [9]. The estimated standard deviation is determined using the standard deviation of the population of scores,  $S$ , and the number of data points,  $N$ ,  $S_{\bar{x}} = S/\sqrt{N}$ . This means that the relative error is inversely proportional to the square root of  $N$ . Due to this, reducing the relative error can require a lot more data, which becomes computationally demanding at a certain point. This is an additional reason to why variance reduction is needed.

Once the relative error has been determined, it can be used to infer the statistical significance of a result. A general rule of thumb for interpreting results is that anything above 0.1 error will mostly be unreliable. Under 0.1 the results start being relevant and can be considered useful (see table 1).



Table 1: Table showing how to interpret the relative error in a Monte Carlo simulation [9].

Relative Error	Result Quality
0.5-1.0	Not meaningful
0.2-0.5	Factor of a few
0.1-0.2	Questionable
<0.1	Generally Reliable

## 2.2 Kernel density estimation

The main use of Kernel Density Estimations is to estimate an unknown Probability Density Function (PDF). While there are many other methods for this, KDE has the upside of always producing a smooth PDF. An unknown PDF, denoted by  $f(x)$ , can be estimated by drawing  $n$  samples,  $x_1, x_2, \dots, x_n$ , from it. The estimator  $\hat{f}$  is then defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \quad (2)$$

where  $x$  is the position of the estimate,  $K$  is the kernel function,  $K_h$  is the scaled kernel function,  $K_h(x) = h^{-1}K(h^{-1}x)$  and  $h$  is the bandwidth, a smoothing parameter [10]. The kernel function must satisfy  $\int K(x) dx = 1$  and is often selected to be a unimodal probability density function which is also symmetric around zero (although other selections are also possible) [11]. The selection of bandwidth will be explained in section 2.5.

## 2.3 Multivariate KDE

Radiated particles have multiple properties that need to be looked at for shielding calculations, namely position, momentum, and energy. One therefore has to find their joint PDF in order to resample the source term. This requires extending KDE to higher dimensions, which can be done using multivariate KDE. Samples are then drawn as  $D$ -dimensional vectors  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$ , where  $\mathbf{p}_i = ((p_i)_1, (p_i)_2, \dots, (p_i)_D)$ . The joint density is then determined as so

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^N w_i \left\{ \prod_{j=1}^D \frac{1}{h} K\left(\frac{x_j - (\tilde{p}_i)_j}{h}\right) \right\}. \quad (3)$$

Here,  $w_i$  are normalized weights of the sample vectors, and  $\tilde{p}$  are the rescaled sample vectors. The samples are divided by their respective standard distributions in order to

make them dimensionless, as the different particle properties are measured in different units [2].

## 2.4 Kernel Function Selection

There are several options when selecting the kernel function. However, it is most commonly selected to be a normal density [10]. This choice of kernel function was used for all estimates done in this project. Note that for multivariate KDE, the kernel function has to be extended to higher dimensions in order for the method to work.

## 2.5 Bandwidth Selection

The bandwidth,  $h$ , has a large impact on the outcome of an KDE estimate [12]. As such, it must be selected carefully. If it is too small, the model will not correlate with the unknown probability density. If it is too large on the other hand, the model will become overtrained, meaning it becomes too sensitive to local trends. The software used to perform KDE in this project supports three different algorithms for selecting bandwidth: Silverman’s rule [13], non-adaptive Maximum-Likelihood Cross-Validation (MLCV) [2], and adaptive  $k^{\text{th}}$  nearest neighbor [14].

### 2.5.1 Silverman’s Rule

Silverman’s rule is a non-adaptive method for selecting bandwidth for an unknown PDF, which is assumed to be normal. It is determined using a simple equation, rather than testing with algorithms. It is derived from a rule of thumb, which states that for a normal distribution with mean  $\mu$  and variance  $\sigma^2$ , where  $\sigma$  is the standard deviation, the bandwidth that minimizes the mean integrated squared error is

$$h = \left( \frac{4\sigma^5}{3n} \right)^{\frac{1}{5}} \approx 1.06\sigma n^{-1/5}. \quad (4)$$

Here,  $n$  is the number of samples collected. The  $\sigma$  in equation 4 is then replaced with a new term  $A$ , which is defined as

$$A = \min \left( \sigma, \frac{\text{IQR}}{1.34} \right) \quad (5)$$

where IQR is the interquartile range, or the middle 50 % of a normal distribution. The factor 1.06 is also reduced to 0.9. This gives the bandwidth choice as

$$h = 0.9An^{-1/5} = 0.9 \min \left( \sigma, \frac{\text{IQR}}{1.34} \right) n^{-1/5}. \quad (6)$$

The advantage of this method is that it is very simple to use, and uses very little computational power compared to more advanced methods. Its main disadvantage is that it is very rigid, and only works for normal distributions. Applying Silverman’s rule to non-normal distributions will give highly inaccurate results, and as such, its use cases are very limited.

### 2.5.2 Non-Adaptive Maximum-Likelihood Cross-Validation

Non-adaptive MLCV determine bandwidths using  $k$ -fold cross validation. A set of samples are collected from the unknown distribution, which are then split into  $k$  equally large folds. One is used for evaluation, while the rest are used for training [15]. The performance is evaluated using a property known as Figure-of-Merit (FoM), which is calculated using the distribution estimate found by applying KDE to the training data,  $\hat{f}_{\text{train}}$

$$\text{FoM}_{\text{MLCV}} = \sum_{i=1}^{N_{\text{test}}} w_i \log(\hat{f}_{\text{train}}(\mathbf{x}_i)) \quad (7)$$

where the weights  $w_i$  are the same as in equation 3. This process is repeated  $k$  times, using a new fold for evaluation each time, and the average FoM is calculated afterward. This whole process is done for several bandwidths, and the one that maximizes the average FoM is selected to be used in the estimation [2]. This method is applicable to all types of PDFs, not just normal ones.

### 2.5.3 Adaptive Maximum-Likelihood Cross-Validation

Unlike the other methods, adaptive KDE selects a set of non-constant bandwidths, one for each data point. This makes it a more flexible method than the previously presented ones. The bandwidths are selected using the kNN algorithm. This is an unsupervised learning method that sets bandwidths equal to the distance from each corresponding data point to its  $k^{\text{th}}$  nearest neighbor, after normalizing the properties to be dimensionless [16]. Several values for  $k$  are tested and evaluated using the MCLV method presented in the previous section [2].

## 3 Software

This section goes over the software used in this project

### 3.1 Comblayer

Comblayer is a C++ based software that has been used to model the ESS accelerator geometry [17]. It was designed to improve geometry construction and variance reduction using popular Monte Carlo transport codes, such as Monte Carlo N-Particle code (MCNP), which is an internationally recognized particle transport code that is primarily used to model neutron and gamma ray particles [18]. Models in these codes are constructed using different types of surfaces, such as planes, cylinders and spheres, that the user defines. Cells can be defined in the intersections and unions of surfaces using boolean operators, and the user can create materials that the cells consist of. The issue is that this process becomes cumbersome when constructing complex geometries [19]. Also, each object must be completely defined, and other objects cannot overlap or cut into new ones. User-created geometries can suffer from unoptimized geometrical construction, leading to longer computation times, especially for intricate geometries. Comblayer optimizes this process by removing any non-unique surfaces and opposite plane surfaces, which increases performance. Another advantage of Comblayer is that it can build input files for multiple different particle codes, such as MCNP, PHITS [20] and FLUKA [21]. The same geometry is used for all types of input files, making it a very convenient option when working with several codes. Comblayer also has the possibility to provide methods for variance reductions.

### 3.2 PHITS

PHITS (Particle and Heavy Ion Transport code System) is a Monte Carlo transport code that can model most particle types, with energies up to 1 TeV [20]. PHITS uses the same type of geometry as MCNP. It is possible to put particle tallies inside a model used in PHITS calculations. These will track any particles, and create a file containing the properties recorded in MCPL format (Monte-Carlo Particle List, a binary file format used in physics simulations) [22]. PHITS supports a wide array of tallies that have different applications. In this project specifically, the T-track and T-cross tallies have been used for particle tracking. T-track is used to track particle fluence in a region. Any particles crossing a specified volume will have its track length recorded. PHITS then calculates the fluence by dividing the sum of all track lengths with the region volume and the number of source particles [23]. This can then be converted to dose rates. The T-cross tally can instead be used to record particle flux and current through a surface. Any particle crossing the surface will add 1 to the current, which is evaluated by dividing by the surface area  $S$ . Both types of tallies will take particle weight into account if provided.

### 3.3 KDSource

KDSource is a tool that uses multivariate Kernel Density Estimation to approximate joint probability distributions of phase space variables recorded using Monte Carlo Simulations [2]. The program has a Python API that can be used to model, plot, and analyze source terms. It also has a C API that allows resampling sources on the fly.

KDSource can read MCPL files containing particle energy ( $E$ ), 3D-position ( $x, y, z$ ), and momentum ( $d_x, d_y, d_z$ ). This requires the sample phase vector to be formatted like so

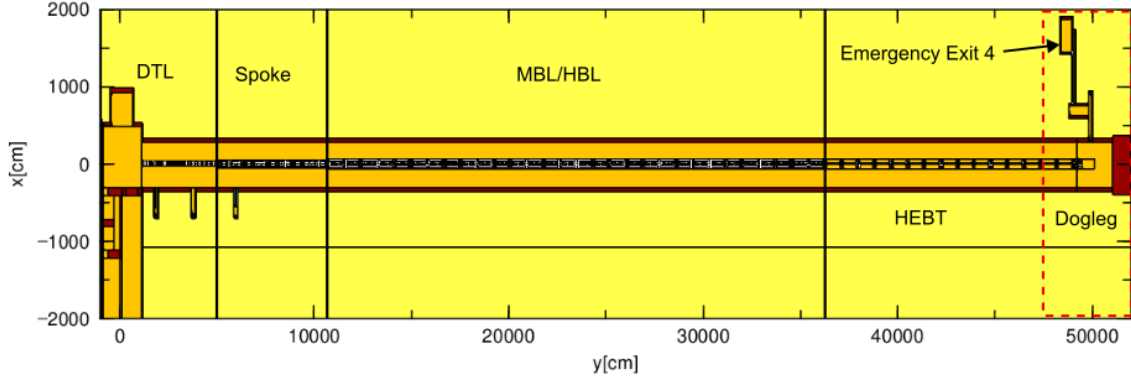
$$\mathbf{p} = (E, x, y, z, d_x, d_y, d_z). \quad (8)$$

These parameters can be given relative importance to each other, so that KDSource weights a parameter more when resampling. Note that KDSource does support other variables, such as time or polarization, but including them is not needed for this project. KDSource is also able to perform various variable transformations, in cases where the source term and/or problem geometry does not work with the vector format in equation 8.

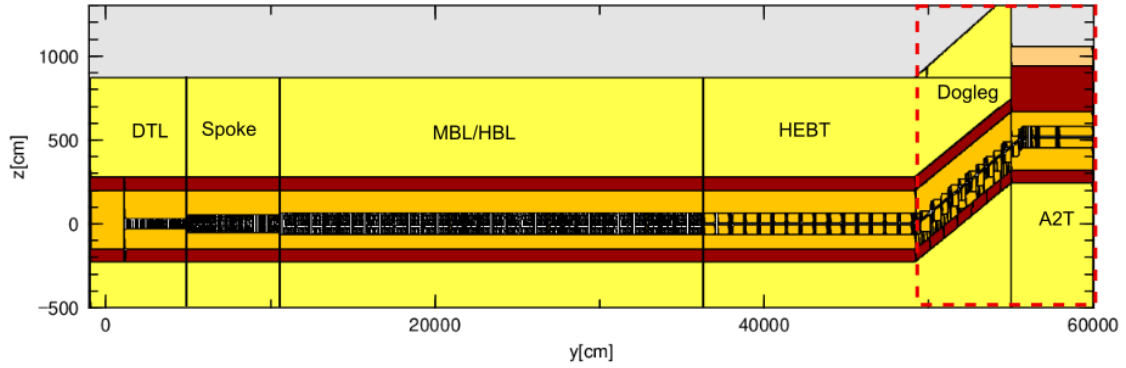
## 4 Results and Discussion

### 4.1 Model

A Comblayer model of the accelerator tunnel had already been constructed up to, and including the high-energy Beam Transport (HEBT) section. At the end of this section, the tunnel splits into two parts. One part leads to the target, and the other part leads to the beam dump. The Dogleg section is part of the former, leading the particles upwards at an angle from the tunnel to the target building. Between the Dogleg and target sits the Accelerator-To-Target (A2T). Both these sections were added to the model, which are shown in the region marked by red dashed lines in figure 2. The accelerator tunnel is 6 m wide and 3.5 m tall. Concrete shielding surrounds the tunnel in all directions, with the walls being 0.5 m wide, the roof being 0.8 m tall and the floor being 0.75 m deep. The rest of the model consists of the surrounding earth around the tunnel, which acts as additional shielding from radiation. The Dogleg consists of an around 60 m long tunnel, which connects to the HEBT. Unlike the rest of the tunnel, which runs parallel to the ground, the Dogleg is tilted at a 4.6-degree angle upwards. The A2T section is connected to the other end of the Dogleg, and stretches for 50 m.



(a)



(b)

Figure 2: Snapshots of the Comblayer accelerator model taken from different locations. 2a shows a top-down view of the tunnel up to and including part of the Dogleg. The A2T is not shown in this view due to its higher elevation. 2b shows the whole tunnel model as viewed from the side. The regions marked by red dashed lines are the parts that were constructed during the project. The orange section is the inside of the tunnel, and consists of air. The red regions are concrete shielding. The black sections in the middle of the tunnel represents the accelerator components. The beige section above the A2T section is a heavier concrete shielding compared to the rest. The yellow region is earth around the tunnel. The gray region is void around the model.

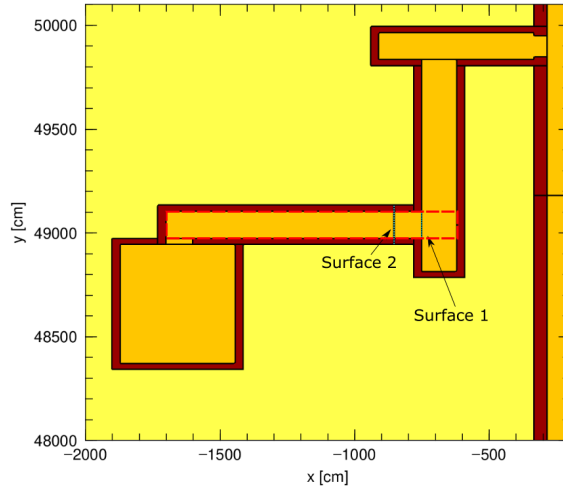


Figure 3: Top-down view of emergency exit 4 in the accelerator model. The two marked surfaces are the locations at which particle tallies have been placed during particle transport simulations. The shape of the T-track tally is marked by the dashed red lines.

The fourth emergency exit in the accelerator tunnel, which is connected to the Dogleg, was added to the model. This section would likely make for a good location to use KDE resampling in. It stretches some distance from the accelerator, so there should be low statistics in the far ends of it. It is also close enough to the tunnel that high-energy particles can penetrate the shielding and enter it. Some particles, which are not energetic enough to travel through the shielding, will enter the emergency exit via the doorway connecting it to the accelerator tunnel. This gives the energy spectrum in the emergency exit both a direct component and a streaming component, providing a wide range of energies that vary by several orders of magnitude.

A simplified model of the beamline was implemented in the Dogleg and A2T, containing the following components: The vacuum pipe in which the high-energy protons are contained and transported, the quadrupole magnets (see figure 4 for a cross-section) which focus protons along the beamline, and the dipole bending magnets which sit at each end of the Dogleg and tilt the trajectory of the protons when entering or leaving this section. These components have already been created in the existing parts of the accelerator, and could therefore be reused in these sections. Some modification was necessary when implementing them into the Dogleg, as the tilt had to be accounted for. The component planes parallel to the ground were tilted with the same angle as the Dogleg, while the rest of the planes were kept the same as before (see figure 2b).

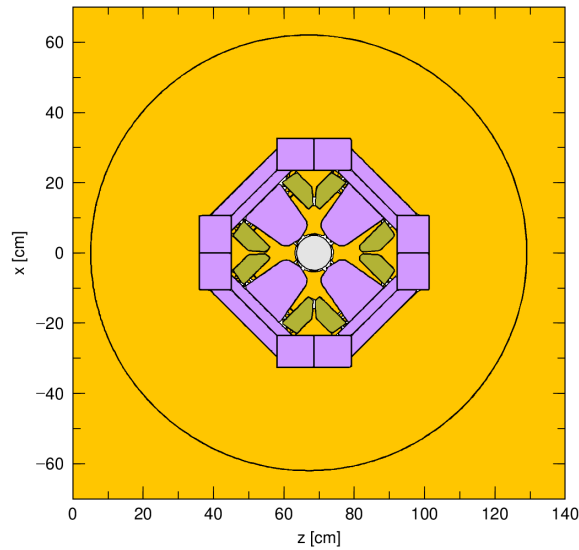


Figure 4: Cross-section of a quadrupole magnet used in the model of the beamline. The gray section is the vacuum pipe going through the magnet. The green sections represent the magnet coils, and the purple parts are the yokes holding the magnet together. The Orange region is air around the magnet. The black circle is used to mark where the magnet should be placed in the model, and does not represent anything physical.

In addition to the main model, a simple cylinder with two regions was created to test that the methodology worked. See appendix A for more details.

## 4.2 Source and Tallies

A linear source with a 1 W/m loss was placed in the beamline in the Dogleg section. This is a simplification of the beam loss caused by halo and other effects, but it gives a rough estimate of how much the beam power will decrease over distance [24]. A T-cross tally was placed at the surface connecting the second leg to the third leg (surface 1 in figure 3). A T-track tally, filling the whole third leg and reaching to the right wall in the second leg, was placed in the model in order to determine simulated dose rates over distance from the accelerator. Finally, a T-track was placed covering the whole emergency exit and part of the accelerator tunnel in order to create dose rate maps in the region. A simulation was run to have a reference of the dose rates look like without any resampling or modification to the model. Dose rate calculations



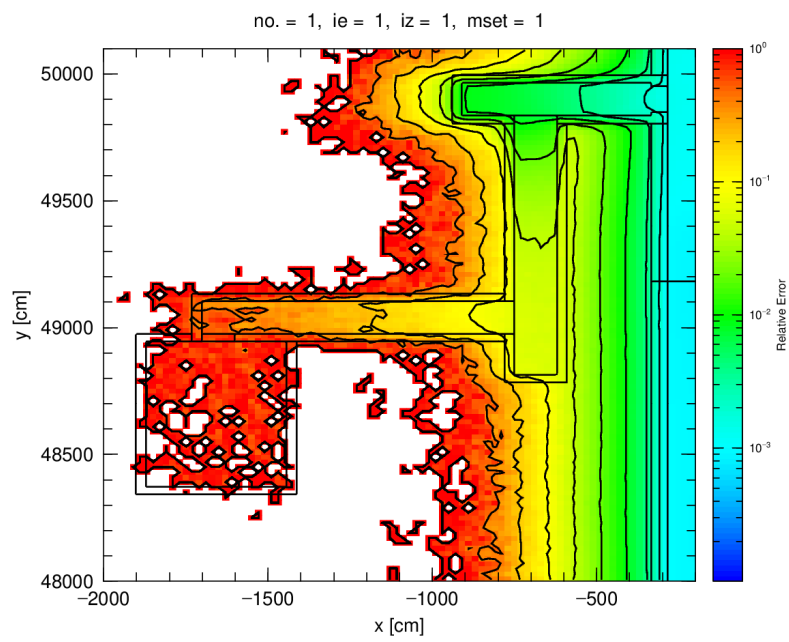
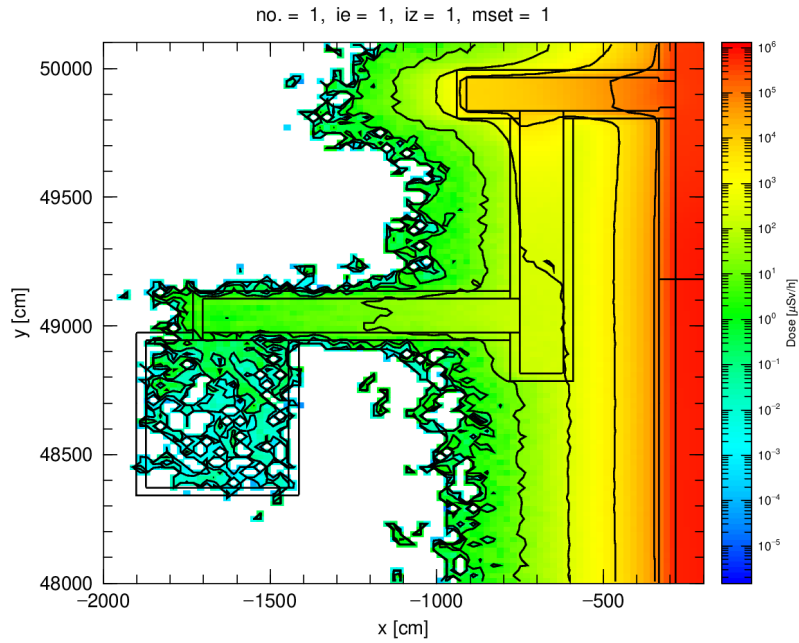


Figure 5: Dose rate maps of the preliminary shielding calculations done in emergency exit 4. 5a shows the simulated dose rates, and 5b shows the corresponding uncertainties. The yellow region is at around 10 % error, which is high enough that variance reduction is needed to get accurate results. This makes the emergency exit a good location to attempt the KDE method.

were then run for 70 hours with 319 threads in emergency exit 4, to investigate the uncertainty in the region. Figure 5b shows the uncertainty of the dose rates in the emergency exit. In the third leg of the exit, the error is around 10 % at the start, and rises up to almost 100 % uncertainty further into this section. This is high enough that variance reduction is needed to determine the dose rates in this region. Thus, this part of emergency exit 4 was selected as the location to run shielding calculations in.

### 4.3 KDSOURCE Resampling

A method to determine how well KDSOURCE captured the properties of the original source was needed in order to evaluate its performance. The method decided on was to make distribution histograms of energy, momenta in all three directions, as well

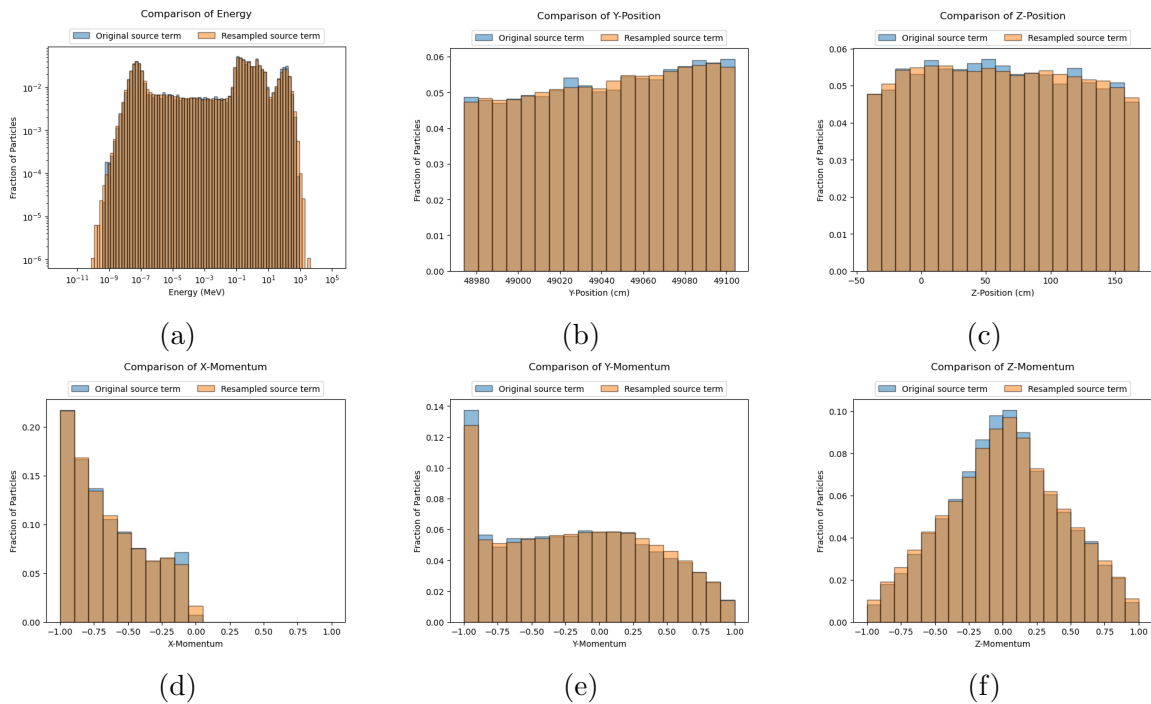


Figure 6: Distributions of energy, positions, and moment of the tallied particles compared to the resampled particles. KDSOURCE manages to capture many of the properties of the original source. However, the energy spectrum in 6a does stretch a little further than the tallied one. Note that the x-, y-, and z-momentum plots correspond to the x-, y-, and z-components of the particles' respective momentum vectors.

as the y and z positions. The x-position was excluded since the resampled source had to have the same dimensions as the T-cross tallies, which were all 2D surfaces parallel to the yz-plane. An example of this can be seen in figure 6, which shows the histograms of particles recorded at surface 1. The resampled energy spectrum in 6a is a bit broader than the tallied spectrum, but otherwise they are pretty similar.

A likely cause for the differences in the distributions is the bandwidth. As mentioned in section 2.5, it greatly affects how well a distribution is estimated. This is illustrated in appendix B where it was investigated how a poorly selected bandwidth can affect the estimation process. There are several ways the bandwidth selection process could be improved. One way would be to increase the relative importance of energy in the resampling process (section 3.3). For the simulations performed, the energy had an importance three times higher than the other variables. However, it could be that an even higher importance is needed. Splitting the particles into groups based on energy levels and resampling the groups individually might also be a way to improve the results.

## 4.4 KDE Source Dose Rates

### 4.4.1 Surface 1

One potential issue that can cause errors in dose calculations is particles being counted twice. For instance, if a particle is reflected during an interaction in the leg, it could then be reflected a second time outside the leg, causing it to go back inside. It will then be tracked a second time by the tally, which then increases the dose calculated by PHITS. This problem was circumvented by setting a region of the exit to have the material type “outer void”. Any particles crossing into a cell, defined as outer void, will be removed from the model. This is mostly used at the edge of a simulation area in order to remove particles that travel too far away. However, any cell can be defined as outer void, which was utilized in the exit. A region covering the whole third leg, but not the surrounding shielding, and limited by surface one was defined to be of outer void. The T-cross tally was kept at surface 1 so that crossing particles were tracked before being removed. The positions, momenta, and energy of the tracked particles were used to resample new particles using KDSource. They were placed in the same location as the tally and used as the source term in a new simulation.

Figure 7 shows the simulated dose rates from the resampled source compared to the full model doses over distance to the accelerator. There is a discrepancy in the dose rates determined here. Ideally, the ratio should be around one, but there are differences as high as 35 % at some of the points. A potential reason for this is particles

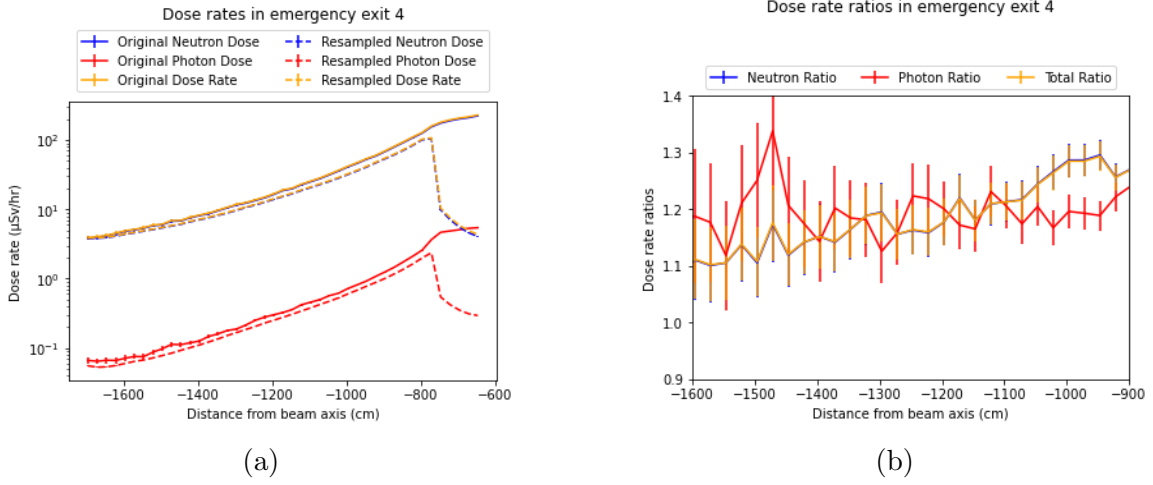


Figure 7: Dose rates simulated in emergency exit 4 from the 1 W/m source compared to ones from the source term resampled from particles at surface 1. 7a shows the dose rates and 7b shows the ratio between them. Note that the dose rates before -800 cm are not significant for the resampled source.

entering the emergency exit without passing the T-cross tally. Some particles from the beam source are energetic enough that they can penetrate through the earth and concrete shielding around the exit. If they travel from further down the beam source, they can enter the emergency exit without passing the T-cross tally, leading to a lower dose rate when the source is resampled.

#### 4.4.2 Surface 2

To account for any high-energy particles entering from further away, the outer void region was redefined to catch more of them. It was pushed one meter back into the leg, and extended to cover the concrete shielding around that region. The T-cross tally was also moved the same amount (see surface 2 in figure 3). It was also investigated if splitting the particles into groups based on their energies, and then resampling each group individually, before combining them into one source would improve the bandwidth selection. The correlation between particle energy and angles relative to the surface 2 normal was used to determine where the energy spectra should be split (geometry shown in figure 8). The angles are calculated using the momenta of each particle. Figure 9 shows the distributions of each property, as well as the correlations between them. One can observe three bands in the correlation plots that include energy. Two strong ones, one below  $10^{-6}$  MeV and one above 10 MeV, as well as a weaker band between 0.1 MeV and 10 MeV. The high energy particles correlate with

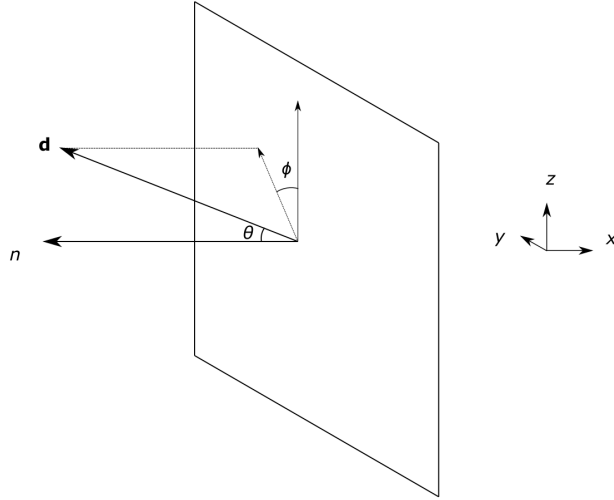


Figure 8: Angles  $\theta$  and  $\phi$  of a beam source particle traveling with momentum vector  $\mathbf{d}$  relative to the normal  $\mathbf{n}$  of surface 2.

$0^\circ < \theta < 45^\circ$ , giving them a more narrow travel direction around the surface normal. This is expected as they travel more directly towards the tally surface from the accelerator. The low energy particles have a comparatively more uniform correlation with  $\theta$ , since they are involved in scattering interactions as they stream around the chicane. There is also slight correlation for particle between 0.1 MeV and 10 MeV in this range for  $\theta$ . The correlation between energy and  $\phi$  is uniform for all energy levels. Based on this, it would seem that splitting the energy spectra at  $10^{-6}$ , 0.1 and 10 MeV would be a possible method for improving the resampling process.

Resampling with different amounts of particle groups in the energy produced the spectra seen in figure 10. To evaluate if grouping the energy improved the estimation of the distributions, the sum of the weighted square difference  $\sigma$  of the energy histograms were determined in the following way

$$\sigma = \sqrt{\sum_j w^2 (H_1 - H_2)_j^2} \quad (9)$$

where  $j$  denotes each bin in the histograms,  $H_1$ ,  $H_2$  are the contents in each bin, and  $w$  are weights, determined by interpolating the middle points of the bins to an ESS dose rate conversion table [25]. To simplify the test, only neutrons were used here.

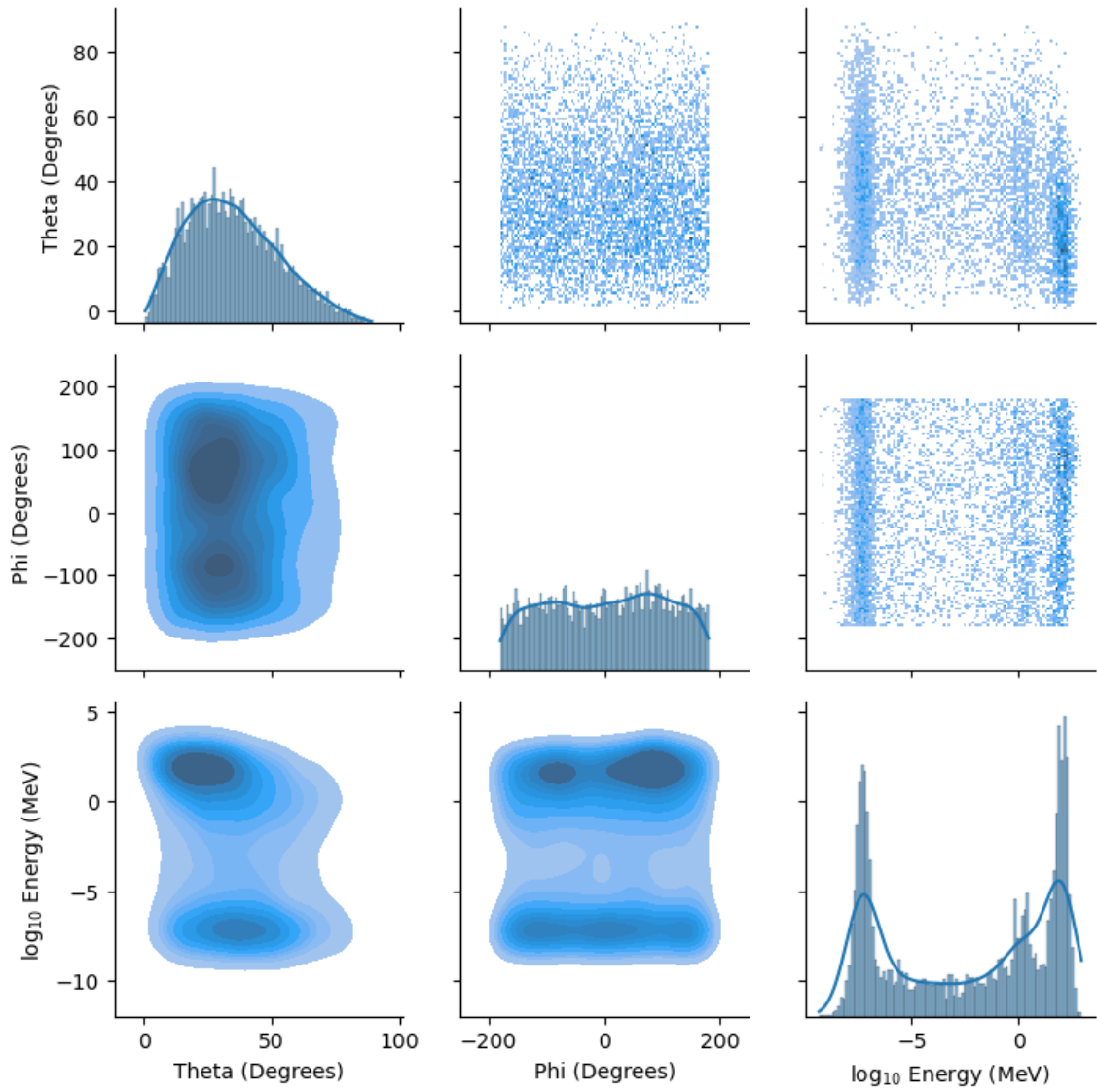


Figure 9: Different types of correlation maps between energy and inbound angle of particles recorded at surface 2. The diagonal shows the distributions of the individual parameters. Plots above the diagonal shows correlation determined using histograms. Plots below the diagonal shows correlation determined using KDE. Darker colors mean higher correlation.

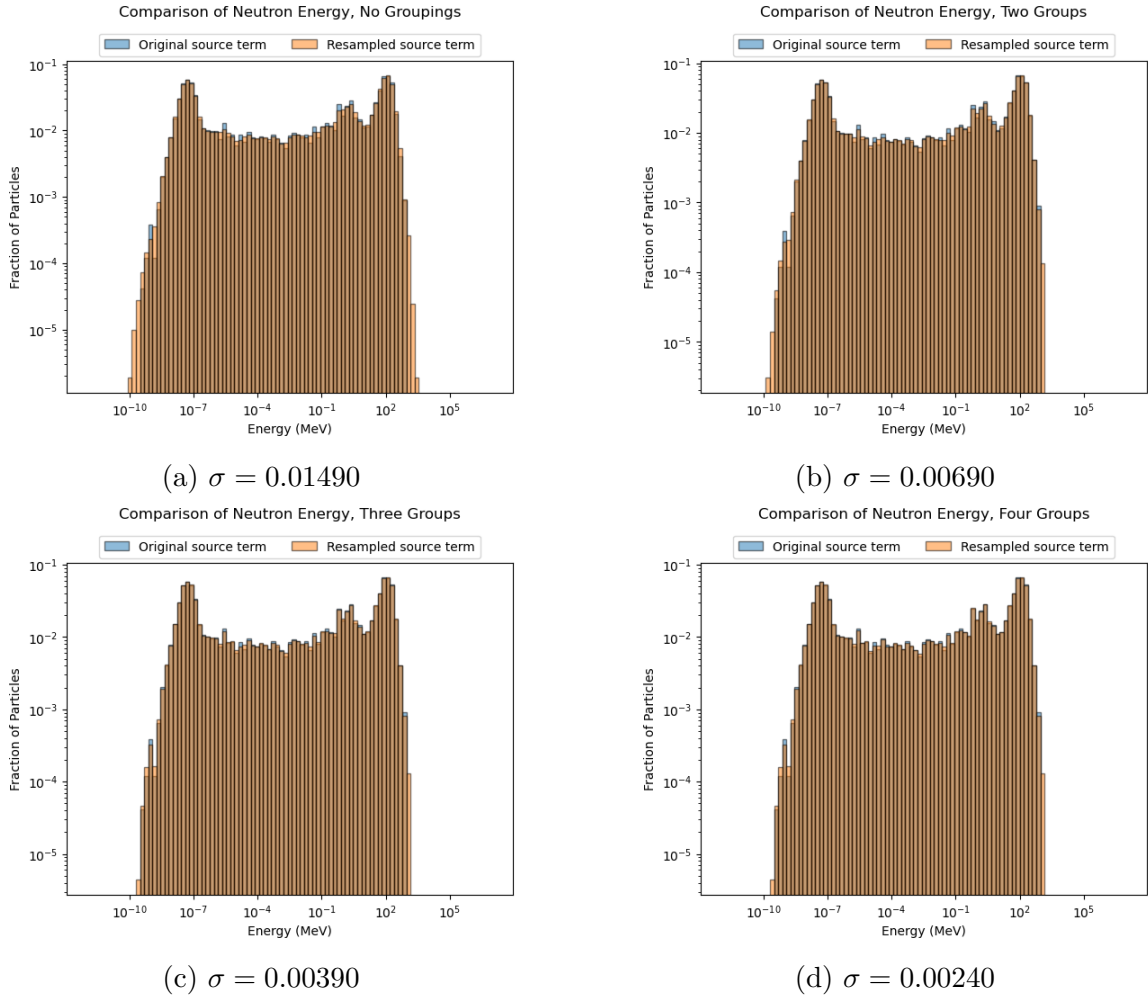


Figure 10: Comparison of energy spectra where the resampling has been performed on certain ranges of energy. 10a is the energy spectra acquired from the surface 2 data without any splits, 10b has one split at 10 MeV, 10c has a split at 10 MeV and one at  $10^{-6}$  MeV, 10d has the same splits as 10c plus an additional one at 0.1 MeV.

Figure 10 shows the energy spectra determined using different amounts of groupings in the energy. Each energy group decreased  $\sigma$  by different amounts. Splitting the spectrum into two groups decreased it by more than half. Specifically, it was 46.3 % of the weighted square for no groupings. Adding further groups lowered  $\sigma$  further to be 26.3 % and 16.4 % of its initial value. This means that adding energy groups will improve the estimate of the distributions. However, there are diminishing returns for each group added. At the same time, more groups increases computation time as KDSOURCE has to run more resampling processes. Thus, a balance has to be struck

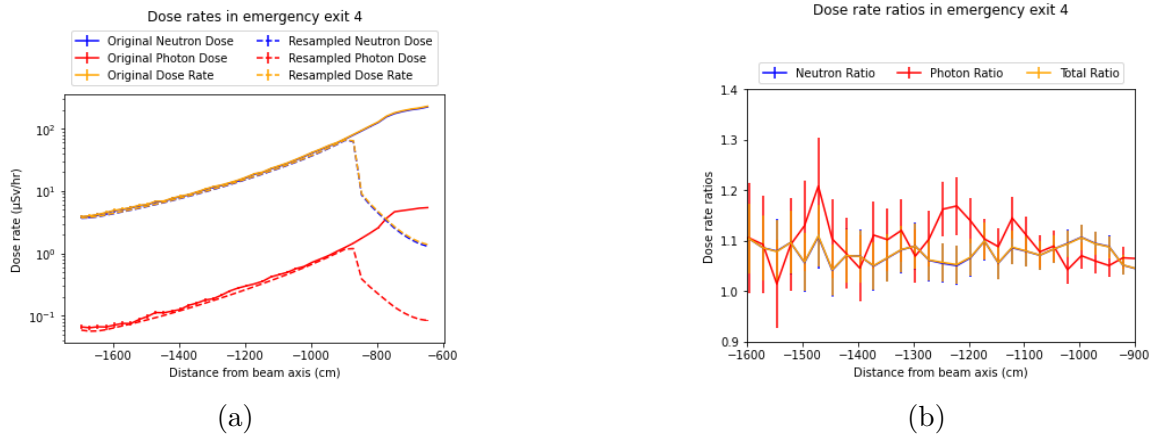


Figure 11: Dose rates from the resampled source created using particles simulated at surface 2 compared to the full model. 11a shows the dose rates and 11b shows the ratio between them. Compared to figure 7, the ratio has been lowered to be around 5-15 %, which is a significant improvement.

to make the software run in a reasonable time. Based on the results here, using three groups would be a good choice, as it gives a good compromise between speed and accuracy.

The particles resampled with three energy groupings were used as the source for dose rate calculations. The results can be seen in figure 11. The ratios of the different dose rates are now closer to one, spanning 5-15 % for the most parts. Compare this to the previous test (figure 7), where the ratios were around 15-35 %. This shows a significant improvement, and implies that grouping energy spectra before resampling will give a better estimate of the particles' distribution.

Finally, the doses in the whole emergency exit were plotted in a 2D map together with the corresponding uncertainties in figure 12 using the source at surface 2 resampled with three energy groups. Note that systematic errors of the intermediate source are not propagated into these results. In the regions where the error was above 10 % without any resampling, it is now 5 % and below, depending on proximity to the resampled source. This shows that KDE resampling can be used for variance reduction in shielding calculations.



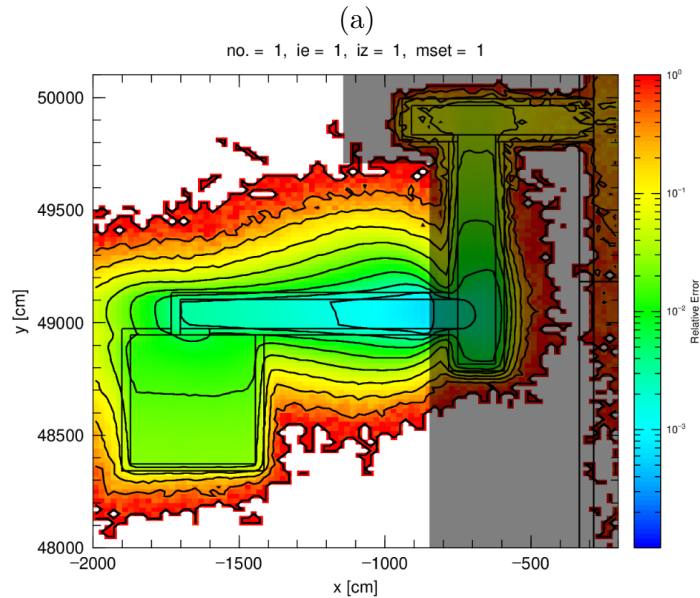
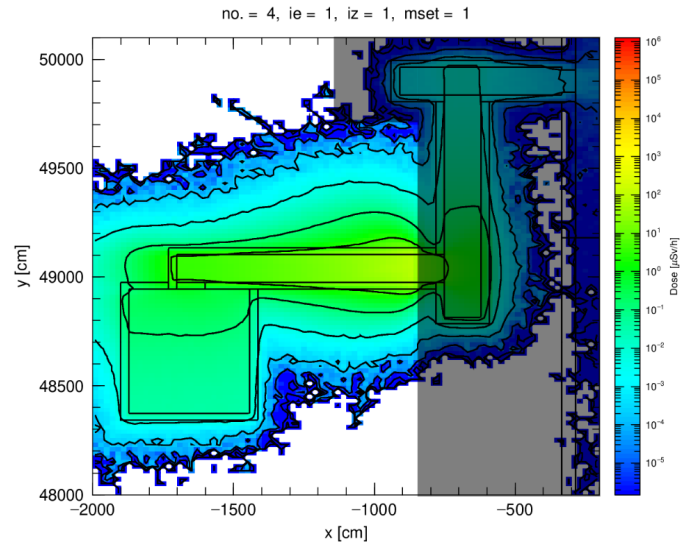


Figure 12: Dose rate map from the calculations where particles were grouped based on energy level before being resampled. 5a shows the simulated dose rates, and 5b shows the corresponding uncertainties. The darkened region is not relevant, due to the position of the resampled source. Compared to figure 5, the error has been reduced to be below 5 % in the regions where the error was 10 % or higher. This shows that KDE resampling is viable as a variance reduction method.

## 5 Summary

In this project, Kernel Density Estimation techniques has been used to reduce variance in shielding calculations around the ESS linear accelerator. Part of a model of the accelerator was constructed, and particle transport simulations were run in said model. Simulated particles crossing into an emergency exit in the accelerator tunnel were tracked, and their properties were used to resample new particles. New calculations were performed, using the resampled particles as sources, and the simulated dose rates in the region were determined. Simulations with added regions that removed entering particles in order to more accurately resample particles were also performed.

The results show that Kernel Density Estimation is a viable variance reduction method. It was possible to resample particles that kept many of the properties of the 1 W/m source. Calculating dose rates using this method gave smaller uncertainties in regions where the beamline source had high error. It was also found that splitting the particles into groups and then resampling said groups individually made the estimate of the distribution better.

## References

- [1] European Spallation Source. <https://europeanspallationsource.se/about>. Accessed: 2024-06-17.
- [2] N.S Schmidth, et al. KDSOURCE, a tool for the generation of Monte Carlo particle sources using kernel density estimation. *Annals of Nuclear Energy*, 177, 2022.
- [3] D. Richter and T. Springer. A twenty years forward look at neutron scattering facilities in the oecd countries and russia. Technical report, European Science Foundatio and Organisation for Economic Co-operation and Development Megascience Forum, 1998.
- [4] R. Garoby, et al. The European Spallation Source Design. *Physica Scripta*, 2018.
- [5] M. Eshraqi, et al. The ess linac. In *5th International Particle Accelerator Conference*, 2014.
- [6] Shielding aspects of Accelerators, Targets and Irradiation Facility, May 2024. INFN Laboratori Nazionali di Frascati.
- [7] D. P. Kroese, et al. Why the Monte Carlo method is so important today. *WIREs Computational Statistics*, 6:386 – 392, 2014.
- [8] J. Stenander. Geant4 simulations of neutron flux using automated weight windows with applications to ess, 2015.
- [9] X-5 Monte Carlo Team. *MCNP — A General Monte Carlo N-Particle Transport Code, Version 5*. Los Alamos National Laboratory, 2003.
- [10] A. Gramacki. *Nonparametric Kernel Density Estimation and Its Computational Aspects*, volume 37. Springer International Publishing, 2018.
- [11] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall/CRC, 1 edition, 1994.
- [12] N. Agarwal and N. R. Aluru. A data-driven stochastic collocation approach for uncertainty quantification in mems. *International Journal for Numerical Methods in Engineering*, 83(5), 2010.
- [13] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1998.
- [14] G. R. Terrell and D W. Scott . Variable Kernel Density Estimation. *The Annals of Statistics*, 20:1236–1265, 1992.

- [15] P. Refaeilzadeh et al. On comparison of feature selection algorithms. *Proceedings of AAAI workshop on evaluation methods for machine learning II*, 3(4), 2007.
- [16] T. Hastie, et al. *The Elements of Statistical Learning*. Springer, second edition, 2009.
- [17] D. Julio, et al. Radiation transport calculations for the European Spallation Source accelerator environmen. 14th International Conference on Radiation Shielding and 21st Topical Meeting of the Radiation Protection and Shielding Division (ICRS 14/RPSD 2022) meeting, 2022.
- [18] J. K. Shultis and R. E. Faw. *An MCNP Primer*. Dept. of Mechanical and Nuclear Engineering, Kansas State University, 2011.
- [19] S. Ansell. CombLayer : A fast parametric MCNP(X) model constructor. In *Proceedings of the 21st Meeting of the International Collaboration on Advanced Neutron Sources (ICANS-XXI)*, 2014.
- [20] T. Sato, et al. Features of Particle and Heavy Ion Transport code System (PHITS) version . *Journal of Nuclear Science and Technology*, 55(6):684–690, 2018.
- [21] G. Battistoni, et al. The FLUKA Code: An Accurate Simulation Tool for Particle Therapy. *Frontiers in Oncology*, 6, 2016.
- [22] T. Kittelmann, et al. Monte Carlo Particle Lists: MCPL. *Computer Physics Communications*, 218:17–42, 2017.
- [23] PHITS Office. *PHITS User’s Manual*. Version 3.33, 2023.
- [24] N. Mokhov and W. E. Chou. In *Beam Halo and Scraping*. Proc. 7th ICFA workshop on high intensity high brightness hadron beam, 1999.
- [25] ESS Procedure for designing shielding for safety, 2017. ESS-0019931.

# A Test Problem

## A.1 Test Problem

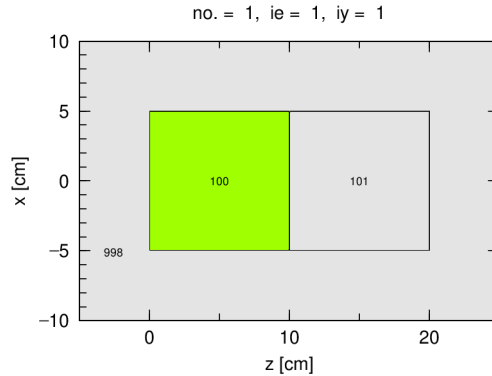
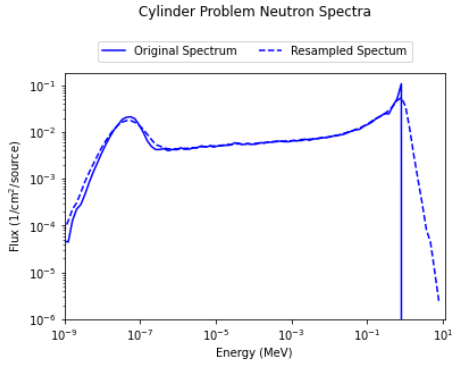


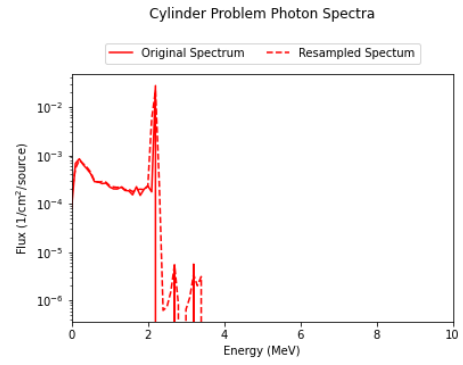
Figure A.1: Cross-sectional geometry of the cylinder used to test the methodology. The green section is water and the gray sections are vacuum. The point source is placed in the middle of the water section.

To investigate if the methodology worked as intended, a simple test problem was used. A cylinder split into two sections was modeled (see figure), with different materials in each. The left section had water and the right one had void. An isotropic neutron point source with energy 1 MeV/u was placed in the middle of the left section of the cylinder. A T-track tally was placed in the cylinder at the plane separating the two sections. The tally output was then used as a source without resampling to see if any normalization was needed. The amplitude of the energy spectra differed by one order of magnitude, so a method for selecting a normalization factor was needed. In order to determine it, a T-cross tally was placed at a new surface 0.1 cm into the vacuum section. This was done so that the particle currents through this region could be recorded. The ratio between the particle currents of the point source and the resampled source were calculated, and then used as normalization factors. Doing this gave the factor as 0.14260. New spectra were then generated after resampling the source term.

Figure A.2 shows a comparison between the spectra of the original source term, and the resampled source term for neutrons and protons. They follow each other closely for the most part. It is only towards the higher energies where a large difference is noticeable. Namely, that the resampled spectra extend a bit further than the original one. This could be caused by the discrete source used in the problem. The current version of KDSOURCE does not support removing particles above a certain threshold,



(a)



(b)

Figure A.2: The comparisons of the energy spectra of the different particle types in the cylinder problem. A.2a shows the neutron spectra, while A.2b shows the photon spectra. The spectra of both particle types are similar for a majority of the spectra. The main difference is that the resampled spectra extend a bit further than the original ones

and therefore cannot account for the particles having a maximum energy. Redoing the test using a continuous source instead would make it easier for KDSource to estimate the distribution.

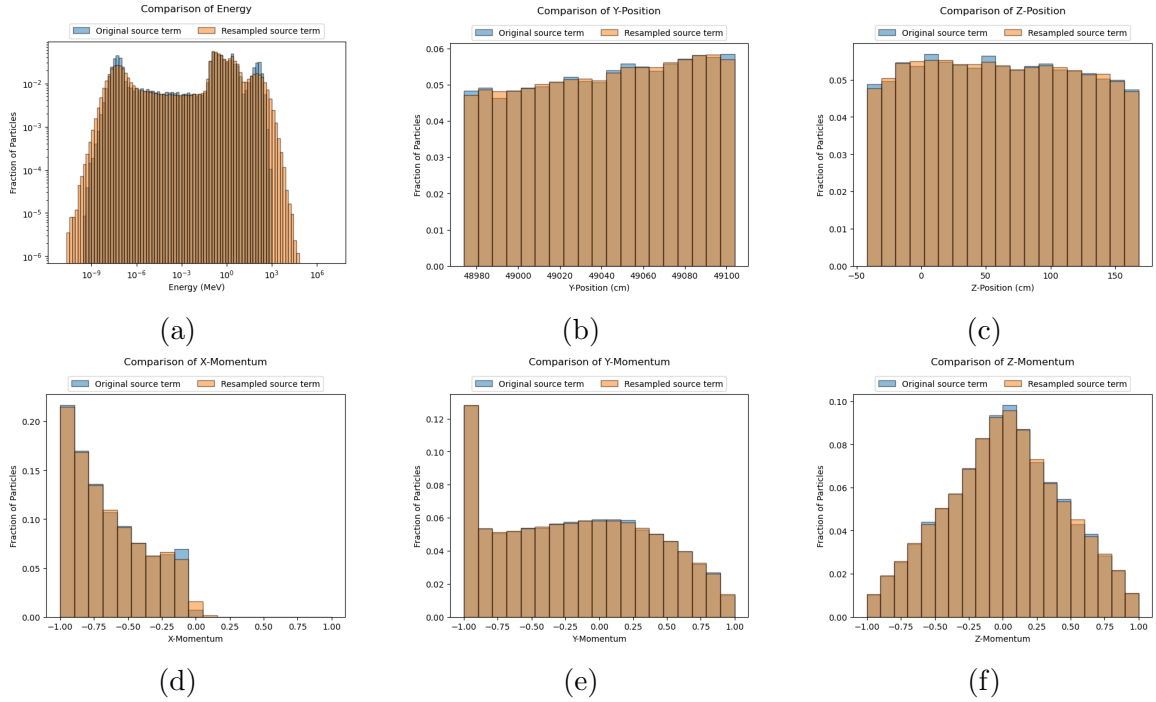


Figure B.3: Distributions of energy, positions, and momenta of the tallied particles compared to the resampled particles for a poorly selected bandwidth.

## B Bandwidth Test

A test where an intentionally poor bandwidth was selected was performed. This was done in order to check how the distribution estimate would be affected. The energy spectrum was set to have the same importance as all the other properties. Figure B.3 shows the different properties simulated with the resampled source and the beam source. The positions and momenta of the resampled source follows that of the tallied particles pretty closely. However, the energy spectrum extends further than the tallied one by several orders of magnitude, and it does not follow the peaks, having differences as large as 25 %. The wider spectrum can have a large impact on the dose rate calculations, as dose is a measure of absorbed energy from ionizing radiation, therefore, increased energy levels leads to a higher dose. That is why it is important to select parameters carefully before resampling with KDE.