Novel Application of Density Estimation Techniques in Muon Ionization Cooling Experiment

Tanaz Angelina Mohayai,
Illinois Institute of Technology, Chicago, IL, USA,
Fermilab, Batavia, IL, USA

Abstract

Cooled muon beams are essential to the production of high flux, background-free neutrino beams at the Neutrino Factory (NF) and high-luminosity collisions at the Muon Collider (MC). When pions decay into muons, they form beams with large phase-space volumes. To optimize muon yield and fit the beam into cost-effective accelerators, the beam phase-space volume needs to be reduced. Ionization cooling is the only technique that can reduce the beam phase-space volume within the short muon lifetime, and the international Muon Ionization Cooling Experiment (MICE) will be the first experiment to demonstrate this. A figure of merit for beam cooling is the transverse root-mean-square (RMS) emittance reduction. However, RMS emittance can be sensitive to non-linear effects in beam optics. We study an alternative measure of cooling where a measurement of phase-space density and volume is made through the novel application of the Density Estimation (DE) techniques to MICE, in specific the Kernel Density Estimation (KDE) method.

1 Introduction to Univariate Density Estimation Techniques

The research question presented in this note is whether the MICE muon beam can be cooled. To answer this question, development of an unbiased measurement technique is necessary. The RMS emittance beam cooling is sensitive to non-linear effects in beam optics, therefore in this report, an investigation is carried out to study the changes and evolution of the phase-space density and volume in the MICE cooling channel.

Phase-space density and volume are important concepts in describing the state of a system of particles. The phase-space density can be described in terms of the probability density function (PDF) which is the probability distribution of a random variable. The probability that a random variable falls within a particular interval is found by calculating the area under the distribution

curve within this particular interval. When the beam optics is non-linear and the beam distribution is long-tailed and distorted, defining a parametric model for the distribution becomes challenging. To overcome this challenge, the underlying PDF can be found using density estimation (DE) techniques. There are two DE methods. One is called parametric, in which an assumption about the underlying PDF is made and the distribution parameters are estimated. The other is the non-parametric DE method where fewer assumptions are made about the underlying distribution [1]. In general, the aim is to use a DE technique where the individual data points are allowed to speak for themselves and the difference between the estimated density and the true density is the smallest (this difference is typically known as the error of the estimated density). Non-parametric DE techniques typically rely on some smoothing technique to estimate the density (i.e, kernels, series, and splines) and the level of smoothing is generally tuned by a parameter. In this report, the kernel smoothing technique (KDE) is used to estimate the density of muons in the position-momentum phase space. Please note that throughout this report, notation X_i represents the i^{th} data point, h is the bin width or the bandwidth for the density estimator under study, and $\hat{f}(x)$ is the estimated PDF or density evaluated at some arbitrary point x. Please also note that in this report, I tend to use PDF and density interchangeably. Below, I describe some of the existing DE methods.

One of the most widely-used and oldest DE techniques is the histogram. Suppose we construct an m number of bins and place the first bin at some point x_0 on the coordinate axis. x_0 can be positive or negative and is usually the minimum data point value. Then, we can construct bins of widths h and to estimate the density, we count the number of data points which reside inside an interval of $x_0 + mh$ and $x_0 + (m+1)h$,

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} X_i \text{ where } X_i \in [x_0 + mh, x_0 + (m+1)h], \tag{1.1}$$

where n is the sample size, x_0 is some arbitrary point on the coordinate axes, h is the bin width, X_i are the data points that fall within the bins, and m is an integer representing the bin number. This approach to constructing the bins in histogram estimator method ensures equal intervals for each bin, However, in this approach, the PDF can vary substantially with changes in the bin width, h and the choice of x_0. These variations can cause the same data to be interpreted differently. One may suggest that by choosing a very narrow bin, the effect of bin placement (choice of x_0) can be minimized, however in such situtations, the level of noise in the density would prevent any useful data interpretation. In addition to these constraints, histograms come with several other challenges especially when the data sets are multi-dimensional [1]. In MICE, the typical space under study is the four-dimensional position-momentum phase space represented by the coordinates (x, p_x, y, p_y). These coordinates are coupled in the presence of the solenoidal fields of the MICE Spectrometer Solenoids (SS) and Absorber Focus Coils (AFC) requiring that the density estimation techniques be applied in four dimensions. In addition, when the density estimation technique is used as a mediator to another analysis technique, the histogram may not be the most appropriate density estimation technique [1] - in this report, the estimated densities are used in a subsequent Monte Carlo technique for measuring the phase-space volume. There is an overview on an alternative DE technique in the following paragraphs. This alternative technique is an improvement to the histogram estimator and forms the basis of the kernel density estimation (KDE) technique.

From the general definition of PDF, the density $f(x)$ of a random variable x is defined as,

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < x < x + h), \tag{1.2}$$

where $P(x-h<x<x+h)$ represents the probability that x lies in an interval of $2h$. To estimate the density, we can estimate that $P(x-h<x<x+h)$ is the probability that certain number of data points out of the entire sample size n end up inside an interval $2h$ [1]. Following is the mathematical notation of such estimated density,

$$\hat{f}(x) = \frac{1}{2nh}\sum_{i=1}^{n} x-h<X_i<x+h. \tag{1.3}$$

Eq. 1.3 can be written in terms of a weight function. A weight function would assign a weight of 1 to a data point within the interval of $2h$ and a weight of 0 to a data point outside of $2h$. Let us define this weight function in the form of a step function,

$$w(x) = \begin{cases} \frac{1}{2} \text{ if } |x|<1 \\ 0 \text{ otherwise} \end{cases}. \tag{1.4}$$

Eq. 1.3 re-written in terms of this weight function expression (Eq. 1.4) is:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}w\left(\frac{x-X_i}{h}\right). \tag{1.5}$$

I shall refer to this new density estimator expression (Eq. 1.5) as the weight function estimator. It is also referred to as the naive estimator in [1]. Please note that the factor of $\frac{1}{2}$ is no longer displayed in Eq. 1.5; this factor is now included in the definition of $w(x)$. We can re-write the weight function expression from Eq. 1.4 with the function variable $\frac{x-X_i}{h}$ (X_i is the i_{th} data point),

$$w\left(\frac{x-X_i}{h}\right) = \begin{cases} \frac{1}{2} \text{ if } |x-X_i|<h \\ 0 \text{ otherwise} \end{cases}. \tag{1.6}$$

According to the step function expression in Eq. 1.6, a weight of $\frac{1}{2}$ is assigned to the individual X_i data points and these data points contribute to the estimated density at the arbitrary point x as long as their differences with x is less than h. In other words, the weight function estimator defined in Eq. 1.5 estimates the density by placing a box of width $2h$ and height $\frac{1}{2hn}$ on an arbitrary point x each time a data point X_i satisfies the first condition of Eq. 1.6.

It is worth comparing the weight function estimator with the histogram estimator and understand in which ways the weight function estimator is an improvement to the histogram estimator. To estimate the density using histogram estimator, we can construct histogram bins of widths $2h$ similar to the weight function estimator. Eq. 1.6 would then ensure that the arbitrary point x is always located at the center of the bin. This is in contrast with the histogram estimator defined in Eq. 1.1, where the placement of the bins along the coordinate axes would vary with the choice of x_0. An example can help clarify this further. The aim is to show that Eq. 1.6 forces a bin of width $2h$ to be centered at an arbitrary point x. Suppose h is 2 and the arbitrary point x is 4, and there are 4 data points with values $X_1 = 5.5$, $X_2 = 6.5$, $X_3 = 3.5$, and $X_4 = 1.5$. The first and the third data points X_1 and X_3 satisfy the first condition of Eq. 1.6 since $|4-5.5|<2$ and $|4-3.5|<2$. Both X_1 and X_3 are placed within $2h$ of x with X_1 placed to the right of x and X_3 placed to its left. The second and fourth data points, X_2 and X_4 satisfy the second condition of Eq. 1.6 since $|4-6.5|>2$ and $|4-1.5|>2$. So as we move towards the right or the left of x along the coordinate axes by more than $2h$, the data points no longer contribute to the estimated density at x. This example illustrates how Eq. 1.6 enforces the arbitrary point x (the point at which the estimated density is being evaluated) to be centered within the bin width.

The weight function density estimation approach described in Eq. 1.5 has a disadvantage from the point of view of data representation. It has a step-wise nature and is discontinous at

points $X_i \pm h$. However, one can overcome such discontinous nature by replacing the weight function by a smooth function $k(x)$ which satisfies the following condition.

$$\int_{-\infty}^{\infty} k(x)dx = 1. \tag{1.7}$$

This new weight function $k(x)$ is called a kernel function, and in one dimension, the kernel density estimation (KDE) evaluated at some arbitrary point x can be written in terms of this kernel function,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{x - X_i}{h}\right), \tag{1.8}$$

where h, as previously defined, is the width of the kernel function, commonly referred to as the bandwidth parameter. In the terminology of Gaussian statistics, h is the RMS width of the kernel function typically multiplied by some data-dependent factor. The bandwidth parameter indicates the level of smoothing of the estimated density and it is further explained in the paragraphs which follow.

It is worth comparing the weight function estimator in Eq. 1.5 with KDE. Eq. 1.5 is a special case of Eq. 1.9. To estimate the density using Eq. 1.5, a bin of width $2h$ is centered at an arbitrary point x and the number of data points X_i within the bin are counted. Each data point that ends up inside the bin contribtes a weight of $\frac{1}{2}$ and the individual contributions are then summed to result in an estimated density at point x. Similarly, the kernel function in Eq. 1.9 acts as a weight function and although, this time we may not be directly counting the number of data points that end up inside a kernel function, still the data points which are closer to the arbitrary point x contribute more to the estimated density at x than those data points which are further away. In other words, in a region of space where there are more closely-spaced and a larger number of data points, the estimated density at x would be larger than a region of space where there are more scattered and a fewer number of data points. This is illustrated in Fig. 1.1, where the individual kernel functions are centered to each point and the contribution from each data point is then summed to estimate the underlying PDF everywhere.

Fig. 1.1 uses Gaussian kernel functions of widths h and centers them at some arbitrary location x along the one-dimensional coordinate axis. The arbitrary point which is sometimes referred to as the reference point can be either a data point or a grid point − an example one-dimensional calculation is carried out in the appendix (section 6.1). For the analysis in this report, the data points are used as the reference points. The width of the kernel function is determined from the bandwidth parameter h that consists of a bandwidth factor obtained from the existing bandwidth selection rules (more explicitly explained below) and a data-dependent scale parameter which in one dimension is typically the RMS width or the standard deviation. In general, a kernel function should satisfy a certain set of conditions for the estimated density to satisfy the requirements of a PDF [3]:

1. Non-negative: $k(x) \geqslant 0$.
2. Integrates to 1: $\int_{-\infty}^{\infty} k(x)dx = 1$.
3. Symmetric about its center: $\int_{-\infty}^{\infty} xk(x)dx = 0$.
4. Finite second moment condition: $k_2(k) = \int_{-\infty}^{\infty} x^2 k(x)dx > 0$.

Here $k_j(k) = \int x^j k(x)dx$ defines the moment integral of a kernel function. The first non-zero moment indicates the order of the kernel. For instance, if the first non-zero moment of the

kernel function is the second then the kernel is of second order. In this report, the second order ($j = 2$) Gaussian kernel functions are used and in one dimension, it has the following form,

$$k\left(\frac{x - X_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right] \tag{1.9}$$

As explained earlier, imposing conditions 1 through 4 on the kernel function $k(x)$ are necessary to ensure that the resulting estimated density $\hat{f}(x)$ would satisfy the requirements of a PDF. For instance, the second condition (Eq. 1.7) is necessary to ensure that the estimated density would integrate to unity for any arbitrary window width or bandwidth h. The proof of this is below:

$$\begin{gathered} \int_{-\infty}^{\infty} \hat{f}(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - X_i}{h}\right) dx = \\ \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h} k\left(\frac{x - X_i}{h}\right) dx = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} k(u)\,du = \frac{1}{n} \sum_{i=1}^{n} 1 = 1, \end{gathered} \tag{1.10}$$

where we have made the substitution $u = \frac{x - X_i}{h}$.

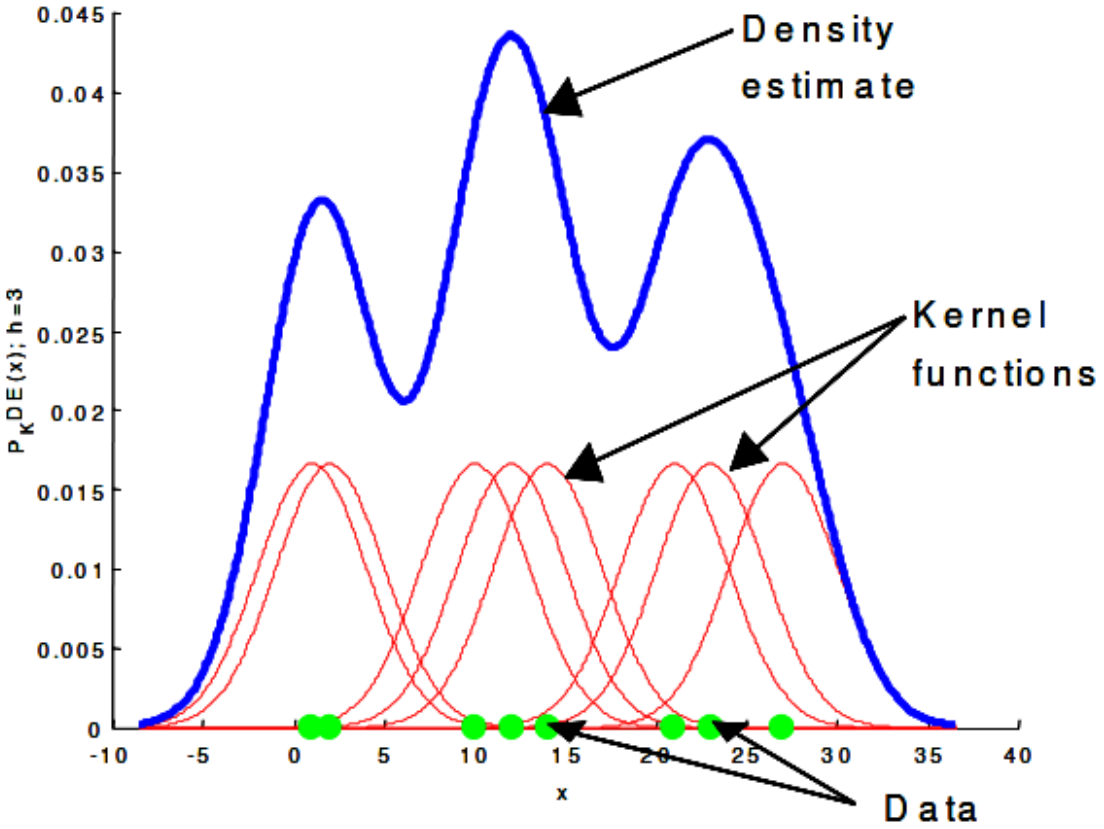

Figure 1.1: An illustration of the KDE technique in one dimensions with a bandwidth parameter of 3 and Gaussian kernel functions(from [2]).

When choosig the DE technique, one needs to ensure that the error associated with the estimated density is small or the discrepancy between the estimated PDF and the true PDF is minimum. In general, in the context of density estimation, error is the difference between the

estimated density and the true density. The typical measure of error in DE analysis is the mean square error (MSE) which can be written in terms of the bias and the variance associated with the estimated density. Bias is the systematic error associated with the estimated density and it is the measure of accuracy of the estimated density. Variance is the random error and is the measure of precision of the estimated density. If we have a large bias in the estimated density, it means that the estimated density is overly smooth and the measurement may not be accurae. If the estimated density has a large variance, it means that the estimated density is overly noisy or spiky and the estimated density is not precise. The mathematical forms of bias and variance are derived in the following paragraphs and the methamtical form of MSE is below [3]:

$$\mathrm{MSE}(\hat{f}(x)) = \mathrm{E}\left(\hat{f}(x) - f(x)\right)^2. \tag{1.11}$$

As mentioned above, MSE is a measure of how close the estimated PDF is to the true PDF. According to Eq. 1.11, the mean square error is obtained when the mean of the error sqaured (square of the difference between true density and estimated density) is obtained through the expectation value operation (E term in Eq. 1.11). MSE can also be re-written in terms of the variance and the bias of the estimator [5],

$$\mathrm{MSE}(\hat{f}(x)) = \left(\mathrm{bias}(\hat{f}(x))\right)^2 + \mathrm{variance}(\hat{f}(x)), \tag{1.12}$$

where the bias and variance of the estimated PDF are defined as,

$$\mathrm{bias}(\hat{f}(x)) = \left[E\hat{f}(x) - f(x)\right], \tag{1.13}$$

$$\mathrm{variance}(\hat{f}(x)) = \left[E\left(\hat{f}^2(x)\right) - \left(E\hat{f}(x)\right)^2\right]. \tag{1.14}$$

where E is the expectation value operation.

One can observe from Eq. 1.12 that MSE only measures the error of the estimated PDF at one single point x, while what is needed is a measure of error for all values of x. For a more global measure of error, the mean integrated square error (MISE) is used. MISE is written in terms of the integrated bias and variance terms [1],

$$\begin{aligned}\mathrm{MISE}(\hat{f}(x)) &= \int_{-\infty}^{\infty} \left[\mathrm{bias}(\hat{f}(x))\right]^2 dx + \int_{-\infty}^{\infty} \mathrm{variance}(\hat{f}(x))dx \\ &= \int_{-\infty}^{\infty} \left[E\hat{f}(x) - f(x)\right]^2 dx + \int_{-\infty}^{\infty} \left[E\left(\hat{f}^2(x)\right) - \left(E\hat{f}(x)\right)^2\right] dx.\end{aligned} \tag{1.15}$$

where $E\hat{f}(x)$ is the expection of the estimated PDF, sometimes referred to as the convolution of the estimated PDF with the true PDF [1],

$$E\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} E\frac{1}{h}k\left(\frac{x - X_i}{h}\right) = \int_{-\infty}^{\infty} \frac{1}{h}k\left(\frac{x-z}{h}\right) f(z)dz. \tag{1.16}$$

We can use Eq. 1.16 to expand the variance term (Eq. 1.14) [1],

$$\mathrm{variance}(\hat{f}(x)) = \frac{1}{n}\int_{-\infty}^{\infty} \frac{1}{h^2}k\left(\frac{x-z}{h}\right)^2 f(z)dz - \frac{1}{n}\left(\frac{1}{h}\int_{-\infty}^{\infty} k\left(\frac{x-z}{h}\right) f(z)dz\right)^2. \tag{1.17}$$

After substituting equations 1.16 and 1.17 into 1.15, we obtain [1],

$$\begin{aligned}\mathrm{MISE}(\hat{f}(x)) = &\left(\int_{-\infty}^{\infty} \frac{1}{h}k\left(\frac{x-z}{h}\right) f(z)dz - f(x)\right)^2 + \\ &\frac{1}{n}\int_{-\infty}^{\infty} \frac{1}{h^2}k\left(\frac{x-z}{h}\right)^2 f(z)dz - \frac{1}{n}\left(\frac{1}{h}\int_{-\infty}^{\infty} k\left(\frac{x-z}{h}\right) f(z)dz\right)^2.\end{aligned} \tag{1.18}$$

To solve for an analyzable form of MISE in Eq. 1.18, we have to make asymptotic approximations (the limit in which estimated density approaches true density) [1]. In addition, we take the kernel function requirements stated on page 13 into account. The assymptotic form of MISE is commonly referred to as the assymptotic mean integrated square error (AMISE).

Let us first take a look at the bias term (Eq. 1.13) in the assymptotic approximations,

$$\text{bias}(\hat{f}(x)) = \int_{-\infty}^{\infty} \frac{1}{h} k\Big(\frac{x-z}{h}\Big) f(z)dz - f(x). \tag{1.19}$$

We can make change of variable $z = x + hu$ and re-write Eq. 1.19,

$$\text{bias}(\hat{f}(x)) = \int_{-\infty}^{\infty} k(u) f(x+hu)du - f(x). \tag{1.20}$$

We can Taylor expand $f(x+hu)$ [1],

$$f(x+hu) = f(x) + f'(x)hu + \frac{1}{2!}f''(x)h^2u^2 + \frac{1}{3!}f^{(3)}(x)h^3u^3 + ... \tag{1.21}$$

Let us substitute the Taylor expansion into the first term of Eq. 1.20 (in the right hand side of the equation) and make use of the kernel function requirements on page 13 to obtain [1],

$$\begin{aligned} \int_{-\infty}^{\infty} k(u) f(x+uh)du = f(x) + f'(x)hk_1(k) + \frac{1}{2}f''(x)h^2k_2(k) \\ + \frac{1}{3!}f^{(3)}(x)h^3k_3(k) + ...) = f(x) + \frac{1}{2!}f''(x)h^2k_2(k) + \text{higher order terms of } h. \end{aligned} \tag{1.22}$$

The bias is then obtained by subtracting Eq. 1.22 from the true PDF,

$$\text{bias}\big(\hat{f}(x)\big) = \frac{1}{2}f''(x)h^2k_2(k) + \text{higher order terms of } h. \tag{1.23}$$

For the second order kernel, the bias has a dependence on the square of the bandwidth parameter, h^2, hence the larger the bandwidth, the greater the bias of the estimated PDF. In other words, a large bandwidth parameter leads to a smooth estimated PDF. We can re-write Eq. 1.23 in the context of the first term in Eq. 1.18 (on the right hand side of the equation) [1],

$$\left(\int_{-\infty}^{\infty} \frac{1}{h} k\Big(\frac{x-z}{h}\Big) f(z)dz - f(x)\right)^2 \approx \frac{1}{4}h^4k_2^2(k)\int_{-\infty}^{\infty} f''(x)^2dx. \tag{1.24}$$

Let us now take a look at the variance term,

$$\begin{aligned} \text{variance}(\hat{f}(x)) &= \frac{1}{n}\int_{-\infty}^{\infty} \frac{1}{h^2} k\Big(\frac{x-z}{h}\Big)^2 f(z)dz - \frac{1}{n}\big(f(x) + \text{bias}(\hat{f}(x))\big)^2 \\ &\approx \frac{1}{n}\int_{-\infty}^{\infty} \frac{1}{h} f(x+hu)k(u)^2du - \frac{1}{n}(f(x) + \text{O}(h^2))^2. \end{aligned} \tag{1.25}$$

where we have written the second integral of Eq. 1.25 in terms of the bias term. We can substitute the Taylor series expansion from Eq. 1.21 into Eq. 1.25 and assume that h approaches zero as n approaches infinity − assymptotic approximation where the limit of the estimated PDF approaches the true PDF − to re-write Eq. 1.25 [1],

$$\text{variance}(\hat{f}(x)) = \frac{1}{n}\frac{1}{h}f(x)\int_{-\infty}^{\infty} k(u)^2du + \text{O}\left(\frac{1}{n}\right) \approx \frac{1}{n}\frac{1}{h}f(x)\int_{-\infty}^{\infty} k(u)^2du. \tag{1.26}$$

Note that $f(x)$ vanishes in the above as its integral is equal to unity. We can write the variance term in the context of the last two terms of Eq. 1.18 (the right hand side of the equation),

$$\int_{-\infty}^{\infty} \text{variance}(\hat{f}(x))dx = \frac{1}{n}\frac{1}{h}\int_{-\infty}^{\infty} k(u)^2 du. \tag{1.27}$$

The finalized form of the MSE is below [1],

$$\begin{aligned} \text{MSE}(\hat{f}(x)) &= \text{E}\left(\hat{f}(x) - f(x)\right)^2 = \text{bias}(\hat{f}(x))^2 + \text{variance}(\hat{f}(x)) \\ &\simeq \left(\frac{1}{2}f''(x)h^2u^2k_2(k)\right)^2 + \frac{1}{n}\frac{1}{h}f(x)\int_{-\infty}^{\infty} k(u)^2 du = \text{AMSE}(\hat{f}(x)), \end{aligned} \tag{1.28}$$

where AMISE (assymptotic mean integrated square error) is,

$$\begin{aligned} \text{AMISE}(\hat{f}(x)) = \frac{1}{4}h^4k_2^2(k)\int_{-\infty}^{\infty} f''(x)^2 dx \\ \frac{1}{n}\frac{1}{h}\int_{-\infty}^{\infty} k(u)^2 du. \end{aligned} \tag{1.29}$$

We can see that Equations 1.28 and 1.29 demonstrate a bias-variance trade-off. If an attempt

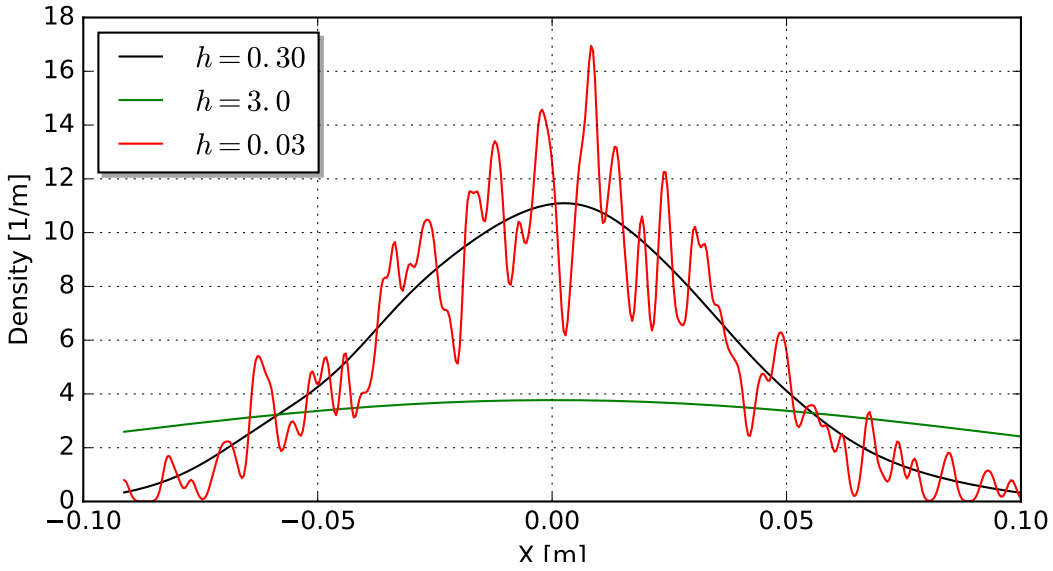

Figure 1.2: Effects of changing the bandwidth parameter on the estimated PDF [2].

is made to reduce the bias by choosing a small value of the bandwidth parameter, h, then the variance becomes large. In general, changing the bandwidth parameter can affect the estimated PDF; a large bandwidth parameter results in an over-smoothed density, while a small one results in an under-smoothed density. This is illustrated in Figure 1.2 above. The KDE technique is applied to the x coordinates of a subsample of 500 muons at the entrance of the MICE upstream tracker (the distribution is described in the following section). The x distribution is approximately Gaussian and the optimal bandwidth parameter, $h = 0.3$, approximately reveals a Gaussian distribution. When the bandwidth parameter is 3.0, the estimated PDF is over-smoothed (green curve in Fig. 1.2). A smaller factor of 0.03 in the bandwidth parameter leads to a noisier estimated density (red curve in Fig. 1.2). Please note that the large differences

between the selected bandwidth factors and the relatively small sample size were intentional; the aim was to show an exagerated change in density with a change in the bandwidth parameter.

The bandwidth parameter is optimized when AMISE is the smallest. This optimal bandwidth parameter for a minimized discrepancy is often referred to as the asymptotically optimal bandwidth and is found through minimizing the AMISE [4]. For a second order kernel function in one dimension, the optimal bandwidth is [1]:

$$\frac{d}{dh}\mathrm{AMISE}\left(\hat{f}(x)\right) \approx \frac{d}{dh}\left(\frac{1}{4}h^4 k_2^2(k)\int_{-\infty}^{\infty} f''(x)^2 dx + \frac{1}{n}\frac{1}{h}\int_{-\infty}^{\infty} k(u)^2 du\right) = 0, \tag{1.30}$$

$$h_{\text{optimal}} = k_2(k)^{-\frac{2}{5}}\left(\int_{-\infty}^{\infty} k(u)^2 du\right)^{\frac{1}{5}}\left(\int_{-\infty}^{\infty} f''(x)^2 dx\right)^{\frac{-1}{5}} n^{-\frac{1}{5}}. \tag{1.31}$$

Equation 1.31 above illustrates that the ideal bandwidth parameter approaches zero with increasing sample size. The $\left(\int_{-\infty}^{\infty} f''(x)^2 dx\right)$ term represents the curvature or the rate of fluctuations of the true density, and one can observe that smaller values of h are more optimal for a highly curved or a rapidly fluctuating density curve.

1.1 Bandwidth Parameter in One Dimension

The unkown curvature term $\left(\int_{-\infty}^{\infty} f''(x)^2 dx\right)$ from Eq. 1.31 above can be solved for known distribution families. In particular, the standard Gaussian distribution (Gaussian distribution where mean of the distribution is zero) can be used. The method of using the standard Gaussian distribution is known as the normal reference rule, since there is reference to the normal or Gaussian distribution. This method is also sometimes referred to as the rule-of-thumb or the plug-in method [1]. With the true density, $f(x)$ written in terms of $\phi(x)$ which is the standard Gaussian density with a variance of σ^2 and a mean of zero, one can first solve for $\left(\int_{-\infty}^{\infty} f''(x)^2 dx\right)$ [1]:

$$\int_{-\infty}^{\infty} f''(x)dx = \sigma^{-5}\int_{-\infty}^{\infty} \phi''(x)^2 dx = \frac{3}{8}\pi^{-1/2}\sigma^{-5} \approx 0.212\sigma^{-5}. \tag{1.32}$$

To obtain the bandwidth parameter, one should substitute Eq. 1.32 and the Gaussian kernel function into Eq. 1.9. The optimal bandwidth parameter in one dimension, would then be [1]:

$$h_{\text{optimal}} = (4\pi)^{-1/10}\left(\frac{3}{8}\pi^{-1/2}\right)^{-1/5}\sigma n^{-1/5} = \left(\frac{4}{3}\right)^{\frac{1}{5}}\sigma n^{-1/5} = 1.06\sigma n^{-1/5}. \tag{1.33}$$

2 Multivariate Kernel Density Estimation

In MICE, the six-dimensional position momentum coordinates are used to represent individual muons in phase space. Muon position and momentum values $\vec{X}_i = (x_i, p_{x_i}, y_i, p_{y_i})$ are reconstructed one by one in the tracking detectors [6], where i runs from 1 to n and n represents the total number of muons in the beam sample under study. The KDE technique can then be applied in four dimensions to the muons in the MICE beam. Similarly to the previous section, to find h_{optimal} in multiple dimensions, the AMISE, which is defined in terms of the bias and variance, needs to be minimized. We can find new d-dimensional bias and variance terms, through use

of the multi-dimensional Taylor series expansion and a similar set of simplifications as those in section 2 [1]:

$$\text{bias}(\hat{f}(\vec{x})) \approx \frac{1}{2}h^2\vec{k}_2(k)\bigtriangledown^2 f(\vec{x}), \tag{2.1}$$

$$\text{variance}(\hat{f}(\vec{x})) \approx n^{-1}h^{-d}f(\vec{x})\int_{-\infty}^{\infty}k(u)^2du, \tag{2.2}$$

where $k(u)$ is the kernel function in terms of the substitution variable $u = \frac{x-z}{h}$ (explicitly explained in the previous section). The new terms (equations 2.1 and 2.2) are now mlti-dimensional.

To obtain AMISE, we combine the integrations of equations 2.1 and 2.2:

$$\text{AMISE}(\hat{f}(\vec{x})) \approx \frac{1}{4}h^4\vec{k}_2(k)\int_{-\infty}^{\infty}\left(\bigtriangledown^2 f(\vec{x})\right)^2dx + n^{-1}h^{-d}\int_{-\infty}^{\infty}k(u)^2du. \tag{2.3}$$

Minimizing Eq. 2.3 leads to the optimal bandwidth parameter in multiple dimensions [1],

$$h_{\text{optimal}} = \left(d\vec{R}(k)\vec{k}_2(k)^{-2}\right)^{\frac{1}{d+4}}\left(\int_{-\infty}^{\infty}\left(\bigtriangledown^2 f(\vec{x})\right)^2dx\right)^{\frac{-1}{d+4}}n^{\frac{-1}{d+4}}. \tag{2.4}$$

$$h_{\text{optimal}} = \left(\frac{4}{d+4}\right)^{-\frac{1}{d+4}}\Sigma n^{-\frac{1}{d+4}} = h_{\text{factor}}\Sigma. \tag{2.5}$$

The bandwidth parameter is Eq. 2.5 is the used for analysis with MICE muon beam, and dimension variable d is 4. The finalized multi-dimensional (multivariate) KDE term is then:

$$\hat{f}(\vec{x}) = \frac{|\Sigma|^{-1/2}}{nh_{\text{factor}}^d\sqrt{(2\pi)^d}}\sum_{i=1}^{n}\exp\left[-\frac{(\vec{x}-\vec{X}_i)^T\Sigma^{-1}(\vec{x}-\vec{X}_i)}{2h_{\text{factor}}^2}\right], \tag{2.6}$$

where Σ is the covariance matrix of the data set. The elements of this covariance matrix represent the covariances of each of the $\vec{X}_i$ coordinates. The estimated density, $\hat{f}$, at an arbitrary point $\vec{x} = (x, p_x, y, p_y)$ in phase space is then determined by summing the contributions of the transverse coordinates $\vec{X}_i$ of all muons.

3 Novel Application of KDE in MICE

A figure of merit for muon beam cooling in MICE is the measurement of the normalized root-mean-square (RMS) emittance reduction. In general, beam cooling through ionization energy loss reduces the transverse emittance with simultaneous increase in longitudinal emittance. This is because of the statistical nature of the ionization energy loss and the deviations in energy losses (straggling) around the mean value [1]. There are, however, proposed methods with which one can have full six−dimensional beam cooling where both the transverse and longitudinal emittances are reduced (briefly explained in the "Conclusion and Next Steps" section of this report).

As shown in Fig. 3.4, MICE has two tracking detectors, upstream and downstream of the absorber (upstream and downstream trackers) [6]. Each tracker is composed of five scintillating-fiber planar stations with three doublet fiber layers immersed in the solenoidal fields of the Spectrometer Solenoids, SS. The upstream and downstream SS consist of five superconducting coils with two responsible for matching the MICE muon beam into and out of the absorber (Match 1 and Match 2 coils), and three responsible for keeping the fields constant inside the tracking volumes (End 1, Center, and End 2 coils) [6]. For measurement of beam cooling in

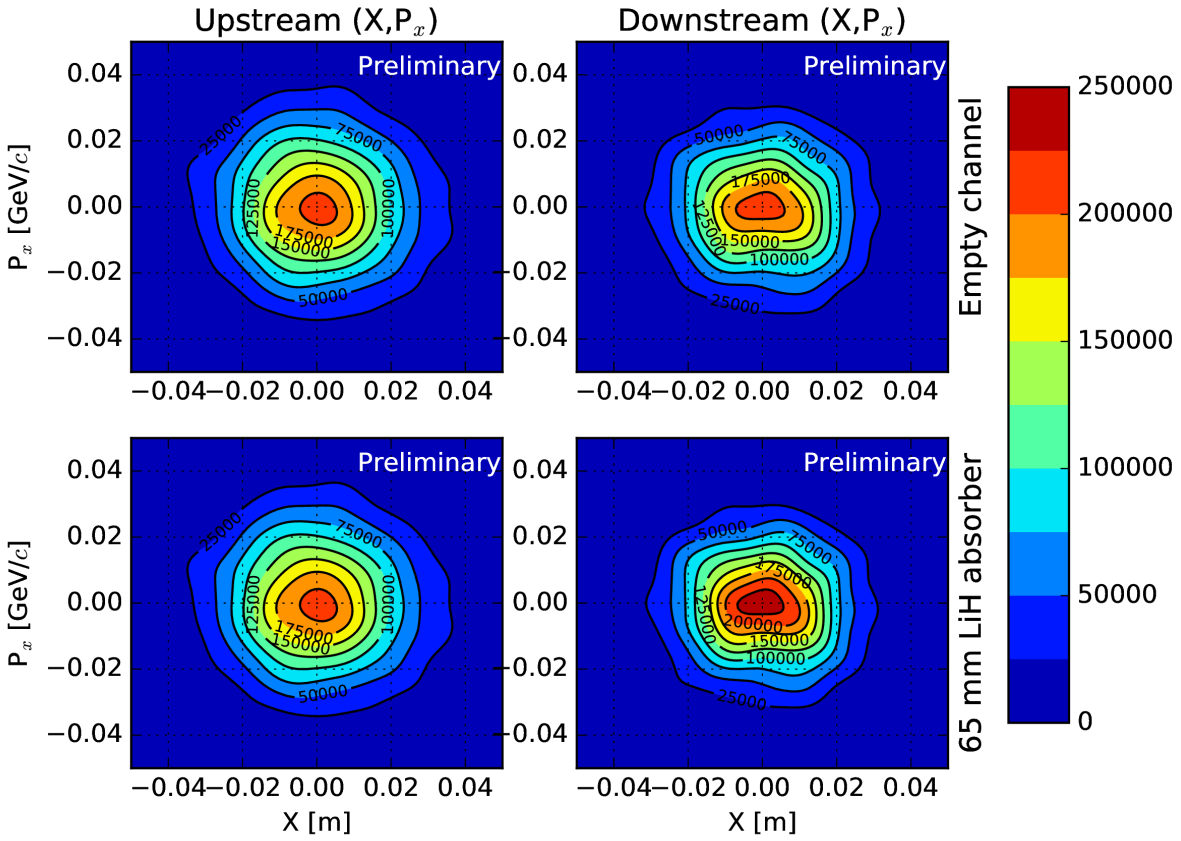

Figure 3.1: Phase-space contours located upstream and downstream of the absorber [7].

MICE, the input and output beam distributions at the tracker stations immediately upstream and downstream of the absorber (tracker reference planes) are compared.

To generate the initial muon distribution at the upstream tracker, the Monte Carlo simulation routines in MICE Analysis User Software (MAUS) [10] and Xboa [11] are used. The initial beam can be chosen to be Gaussian in the transverse direction and matched to the field of the upstream SS. I have done two different simulations, one with an idealized MICE lattice and the other with a realistic one. In the idealized lattice, the SS modules have 4 T fields and the input beam emittance and momentum are set to 4.2 π mm·rad and 140 MeV/c, respectively [7]. This initial muon distribution is in the form of a beam file, which is then fed into a G4beamline MICE Step IV lattice [12] for simulation of 10,000 muons from upstream to downstream trackers. In G4beamline, the currents in SS modules and AFC modules (Table 1 in [7]) are configured to flip sign at the center of the LiH absorber to prevent the build-up of net angular momentum—"flip mode" [7]. In the analysis performed in this report, to ensure an accurate measurement of phase-space density, I select only muon tracks which make it to the downstream tracker, and for the idealized MICE geometry under study, this selection results in 100% of muons being transmitted [7].

The contour plots of upstream and downstream (x, p_x) distributions for the idealized cooling channel are shown in Fig. 3.1. The `gaussian kde` module in `Scipy` version 0.13.3 [8] from Python 2.7 [13] and Scott's Normal reference rule or rule-of-thumb bandwidth parameter [3] were used to estimate the PDF of the muons everywhere in the four-dimensional phase space. As shown in Fig. 3.1, the cooling performance of the lattice is qualitatively demonstrated as an

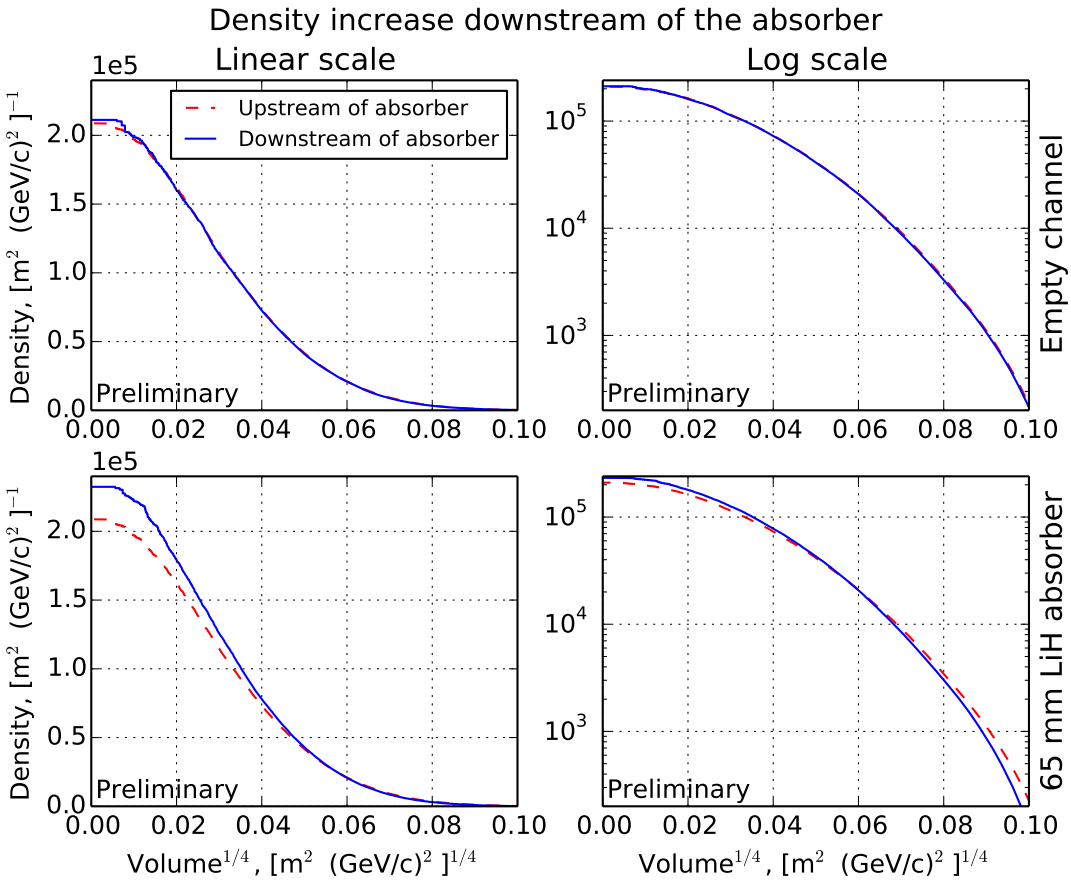

Figure 3.2: Density versus beam radius [7].

increase in the estimated density from the upstream tracker reference plane to the downstream. This increase in the estimated density is better observed for muons located in the core of the beam since muons which were once outliers in the upstream distribution move towards the beam center region after being cooled in the absorber. The contour lines in Figure 3.1 represent constant density slices of the data set, and particles residing within each contour line have estimated densities greater than or equal to the contour line's constant estimated density [7].

Figure 3.2 illustrates the change in the density values from the beam center to the periphery of the beam. The upstream and downstream curves refer to the tracker reference planes immediately upstream and downstream of the absorber. The center of the beam is at volume= 0 and the beam periphery is reached as the value of volume increases [7]. In a channel with a 65 mm LiH absorber, the density at the core of the beam increases. In an empty channel, no change in density is observed as the beam passes through the absorber. Volume$^{1/4}$, in the horizontal axes of the plots in Fig. 3.2 approximately represents the mean radius of the multi-dimensional ellipsoid, with zero representing the beam center. The density value at each radius represents the densities corresponding to the phase-space contours, and contours closer to the center have a higher density of muons [7].

As mentioned, the analysis in the above paragraphs was done with an idealized MICE cooling channel for understanding the KDE technique in MICE Step IV. Since the failure of the Match 1 coil of the downstream SS during 2015 magnet commissioning, the collaboration has come up with new running configurations which are different from the previous ideal run conditions and below I study the realistic MICE Step IV lattice with the KDE technique using these new running configurations.

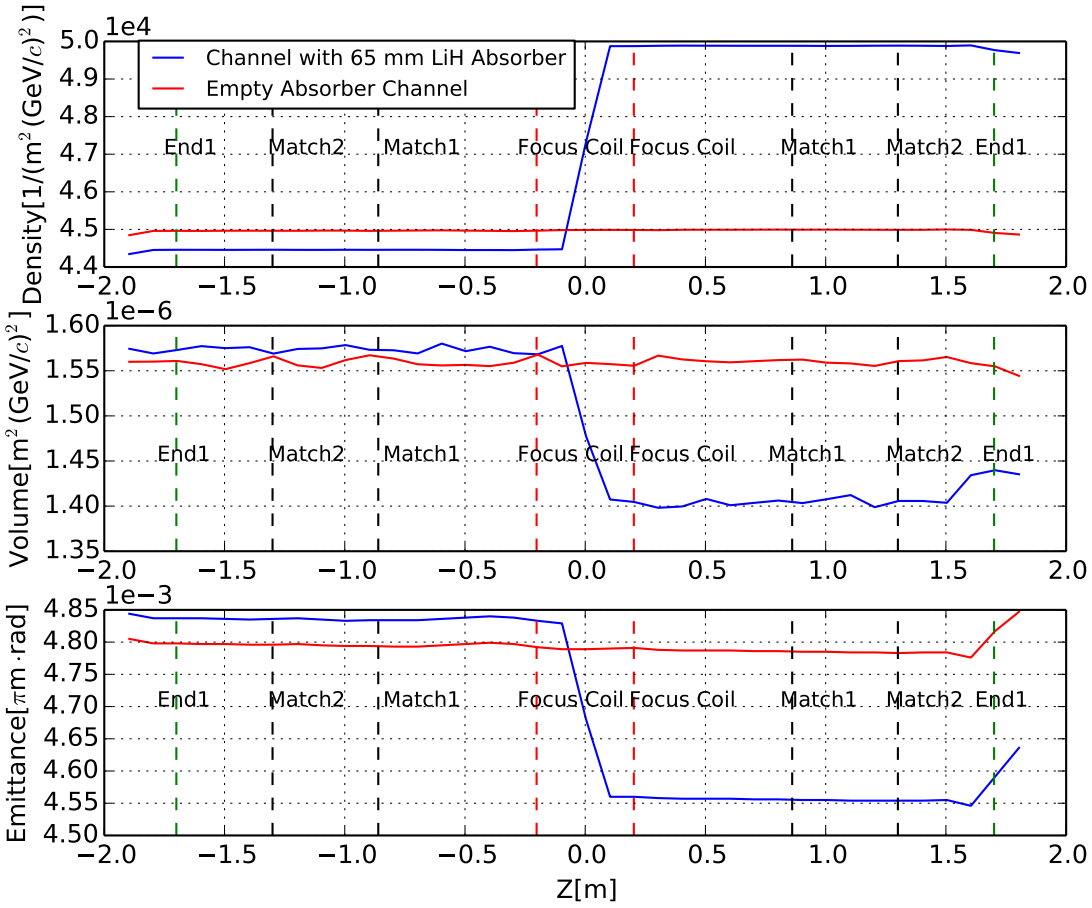

Figure 3.3: Density, volume and RMS emittance evolution in the MICE cooling channel [14].

Similar to the above, the initial beam was chosen to be Gaussian in the transverse direction but this time it was matched to the 3 T fields of the upstream SS, rather than 4 T fields. The input beam emittance and momentum were 6.0 π mm·rad and 200 MeV/c and 100,000 muons are tracked from upstream to downstream trackers in the MICE Step IV lattice using G4beamline. The currents in the SS and AFC modules are the default coil parameters from the MAUS geometry files but scaled by 3/4 to produce the 3 T fields in the modules [14]. Unlike the idealized lattice, a realistic MICE Step IV lattice is studied, where the fields in the downstream Match coils, Match 1 and Match 2 are set to zero. In addition, we only select muon tracks which make it to the downstream tracker and this selection results in a transmission efficiency of 85% in the cooling channel and 84% in the channel without an absorber [14]. The cooling channel without an absorber demonstrates larger muon loss due to its larger multiple Coulomb scattering effects compared with the cooling channel which has a 65 mm LiH absorber. The different scattering effects in the two cooling channels lead to their different transmission efficiencies and their different muon beam input volume, density, and emittance values.

Figure 3.3 shows the evolution plots of kernel-based density and volume, as well as the RMS emittance for channels with and without LiH absorbers. The information on each muon was obtained from G4beamline NTuple hits, recorded at every 0.1 m intervals in the MICE Step IV cooling lattice. To obtain the density evolution plot, Python's `Scipy` package with Scott's rule of thumb bandwidth parameter was used. To compute the phase-space volume, a separate Monte Carlo (MC) method for volume calculation was implemented, where large number of random points were thrown uniformly at a multi-dimensional rectangle in phase space, bounded by the minimum and maximum values of $(x_i, p_{x_i}, y_i, p_{y_i})$. Then, the volume

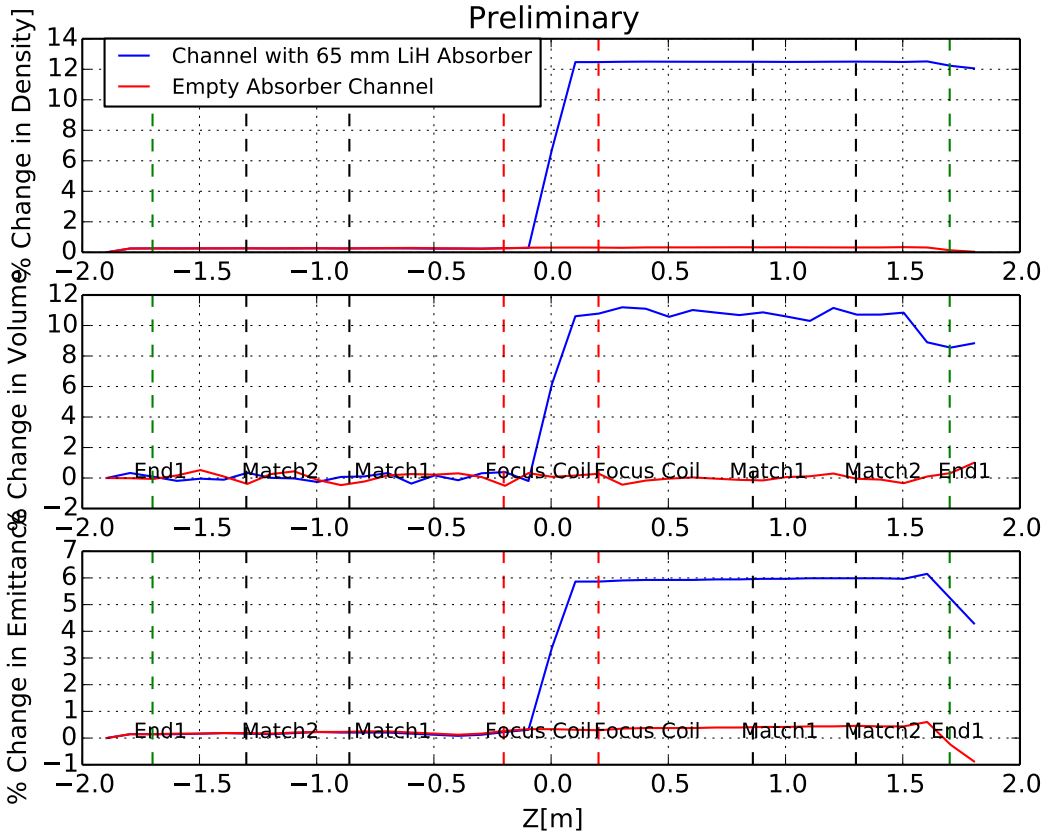

Figure 3.4: Evolution curves representing the percent change values of phase-space density, volume and RMS emittance.

of the constant-density contour was obtained by multiplying the volume of the rectangle by the ratio of the number of MC points within the contour to the total number of MC points [14]. The volume and density curves in Fig. 3.3 correspond to the volume of the contour containing 9% of the total muons, representing the contour closest to the 1σ or the RMS of the beam. The Ecalc9 analysis routine in the ICOOL simulation package [15] was used to compute normalized transverse emittance shown in Fig. 3.3. The emittance and volume curves show decreases of 4% and 9%, respectively. The density increases as a result of cooling by 12% [14]. As expected by the Liouville theorem, the curves corresponding to an empty absorber lattice show conservations of these three quantities [14]. The noise in the volume curve is statistical and is due to the randomness of the MC process used in the volume calculation [14]. The spikes in the evolution curves at the downstream Match 2 coil are currently under investigation [14]. Figure 3.4 illustrates the percent change values of phase-space density, volume and emittance. The percent changes, in the context of this analysis are percent differences between the density, volume and emittance values at every region in the MICE cooling channel and the density, volume and emittance values at the very first region at $Z \approx -1.8$m.

Figure 3.5 is similar to Fig. 3.2 and represents the densities at the locations of the upstream and downstream tracker planes closest to the absorber. In a channel with 65 mm LiH absorber, the density at smaller volume values increases as the beam passes through the absorber while in an empty channel, no change in density is observed. Volume$^{1/4}$ again approximately represents the radius of the multi-dimensional particle distribution [1].

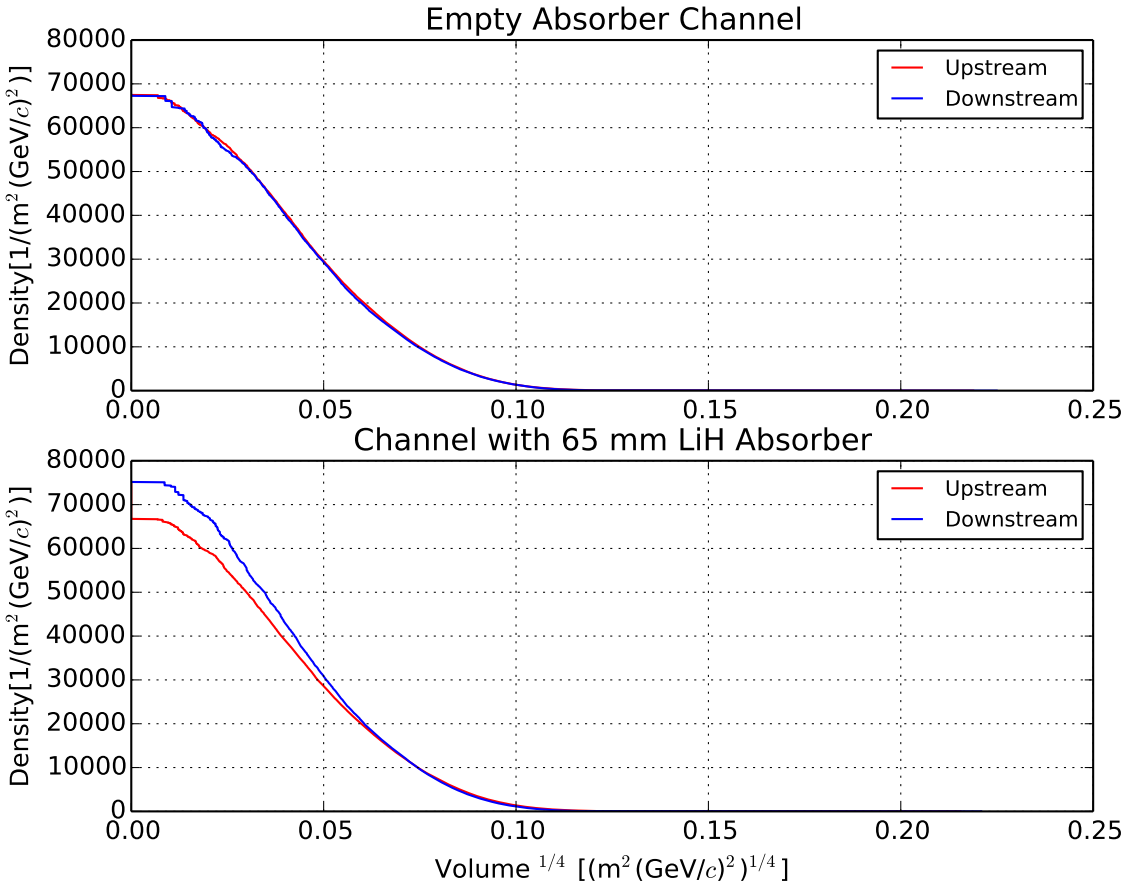

Figure 3.5: Density versus beam radius in four dimensions.

4 Conclusion and Future Prospects

The ideal MICE Step IV lattice, where the problematic Match 1 coil in the downstream spectrometer is assumed operable demonstrates cooling through an increase in phase-space density as estimated by KDE algorithm (Figures 3.1 and 3.2).

In addition, the realistic MICE Step IV lattice with zero fields in the Match 2 and the inoperable Match 1 coils in the downstream Spectrometer Solenoids show cooling through phase-space density increase and phase-space volume decrease using the KDE technique. In the study of this realistic lattice, we observe density, emittance and volume spikes in the evolution curves of Fig. 3.3 at the location of the disabled Match 2 coil in the downstream Spectrometer Solenoid. As the beam drifts in the solenoidal field, the transverse phase-space becomes long-tailed, and although KDE, as a measure of cooling performs better than RMS emittance, it is still sensitive to tails when it is used in its standard form and the same bandwidth parameter is applied to every phase-space contour [1]. Methods exist for a more systematic adaptation of the level of smoothing to the localization of data. Examples of these methods are the nearest neighbor, the variable KDE, and the adaptive KDE. I am currently investigating these techniques and the next step would be to apply them to data and simulation.

KDE algorithm relies on approximations that are dependent on the number of muons and the choice of bandwidth parameter. Studies are under way to assess systematic and random errors and there exist analytical relations between the bandwidth parameter and the volume of the phase space as well as the bandwidth parameter and the density. I am planning to understand, develop, and test these analytical relations and along the process, develop data-dependent

bandwidth selection techniques. I also plan to extend KDE to six dimensions and investigate the count of the number of muons inside a fixed contour using KDE, as another figure of merit for cooling.

KDE technique is applied to the recent MICE data which was taken with a LiH absorber for multiple scattering studies and later, it will be applied to MICE cooling-measurement data.

REFERENCES

[1] B. W. Silverman, "Density Estimation for Statistics and Data Analysis", Monographs on Statistics and Applied Probability, Chapman and Hall(1986).

[2] R. Gutierrez Osuna, "Kernel density estimation", CSCE 666 Pattern Analysis, Texas AM University (2013).
http://www.stat.washington.edu/courses/stat539/spring13/Handouts/tamu-csce-666-pr_l7-annotated.pdf

[3] D. W. Scott, "Multivariate Density Estimation: Theory, Practice, and Visualization", John Wiley and Sons (1992).

[4] B. E. Hansen, "Lecture Notes on Nonparametrics", University of Wisconsin (2009).

[5] D. Wackerly, W. Mendenhall, "Mathematical Statistics with Applications", Thomson Higher Education (2008).

[6] M. Ellis, *et. al.,* "The design, construction and performance of the MICE scintillating fibre trackers", NIM (2011).

[7] T. A. Mohayai, *et al.,* "Simulated Measurements of Cooling in Muon Ionization Cooling Experiment", Proc. IPAC'16, IPAC-2016-TUPMY011 (2016).

[8] G. Varoquaux *et. al.,* "Scipy Lecture Notes" (2015)
https://www.github.com/scipy/scipy/blob/master/scipy/stats/kde.py

[9] S. J. Sheather, "Density Estimation", Statistical Science (2004).

[10] C. D. Tunnell, *et al.,* "MAUS: MICE Analysis User Software", Proc. IPAC'11, IPAC−2011−MOPZ013 (2011).

[11] C. Rogers, "XBOA: A multiparticle tracking postprocessor library for accelerator physicsists"
http://micewww.pp.rl.ac.uk/projects/x-boa/wiki

[12] T. Roberts, "G4beamline User's Guide", Muons Inc (2013)
http://www.muonsinc.com

[13] Python Software Foundation. Python Language Reference, version 2.7.
http://www.python.org

[14] T. A. Mohayai, *et al.,* "Simulated Measurements of Beam Cooling in Muon Ionization Cooling Experiment", Proc. NA-PAC'16, NA-PAC-2016-WEPOA036 (2016).

[15] R. C. Fernow, "Physics analysis performed by ECALC9", MUC-NOTE (2003). J. Scott Berg
http://www.cap.bnl.gov/ICOOL/

5 APPENDIX

5.1 A WALK-THROUGH

Let us now walk through the computation of kernel function for a subset of three data points, with x representing the positions of muons in phase space. The data set under study consists of a subsample of muons at the entrance of the upstream tracker in MICE Step IV geometry. I estimate the density of the x position coordinates using the Gaussian kernel function,

$$K\left(\frac{\vec{x}-\vec{x}_i}{h}\right) = \frac{1}{(2\sigma\pi)^{\frac{1}{2}}}\sum_{i=1}^{n}\exp\left(-\frac{|\vec{x}-\vec{x}_i|^2}{2\sigma^2h^2}\right). \tag{5.1}$$

where n for the sample space under study is three. The optimal bandwidth factor is selected based on Eq. 1.31 and it is about 0.80. Let us take the first arbitrary point x to be the x position of the one of the muons in the data set, $x = -0.06$ m. The standard deviation of the three muons is 0.039. With information on the bandwidth factor and the standard deviation σ, the kernel function for the muon at x position of -0.06 m can be computed by including the contribution from each of the other two muons in the sample space under study. The x position of the three muons in the sample space under study is written in the form of an array, $[-0.060, -0.064, 0.020]$ meters, and their corresponding estimated kernel functions are,

$$\begin{aligned} K\left(\frac{\vec{x}-\vec{x}_i}{h}\right) &= \frac{1}{(2\sigma^2\pi)^{\frac{1}{2}}}\sum_{i=1}^{n}\exp\left(-\frac{|\vec{x}-\vec{x}_i|^2}{2\sigma^2h^2}\right) \\ &= \frac{1}{\left(2(0.039)^2\pi\right)^{1/2}} \end{aligned} \tag{5.2}$$

$$\left(\exp\left(\frac{-(-0.06+0.06)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(-0.06+0.064)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(-0.06-0.020)^2}{2(0.039)^2(0.80)^2}\right)\right) = 20.6,$$

$$\begin{aligned} K\left(\frac{\vec{x}-\vec{x}_i}{h}\right) &= \frac{1}{(2\sigma^2\pi)^{\frac{1}{2}}}\sum_{i=1}^{n}\exp\left(-\frac{|\vec{x}-\vec{x}_i|^2}{2\sigma^2h^2}\right) \\ &= \frac{1}{\left(2(0.039)^2\pi\right)^{1/2}} \end{aligned}$$

$$\left(\exp\left(\frac{-(-0.064+0.06)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(-0.064+0.064)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(-0.064-0.020)^2}{2(0.039)^2(0.80)^2}\right)\right) = 19.2, \tag{5.3}$$

$$\begin{aligned} K\left(\frac{\vec{x}-\vec{x}_i}{h}\right) &= \frac{1}{(2\sigma^2\pi)^{\frac{1}{2}}}\sum_{i=1}^{n}\exp\left(-\frac{|\vec{x}-\vec{x}_i|^2}{2\sigma^2h^2}\right) \\ &= \frac{1}{\left(2(0.039)^2\pi\right)^{1/2}} \end{aligned} \tag{5.4}$$

$$\left(\exp\left(\frac{-(0.020+0.06)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(0.020+0.064)^2}{2(0.039)^2(0.80)^2}\right) + \exp\left(\frac{-(0.020-0.020)^2}{2(0.039)^2(0.80)^2}\right)\right) = 9.9.$$