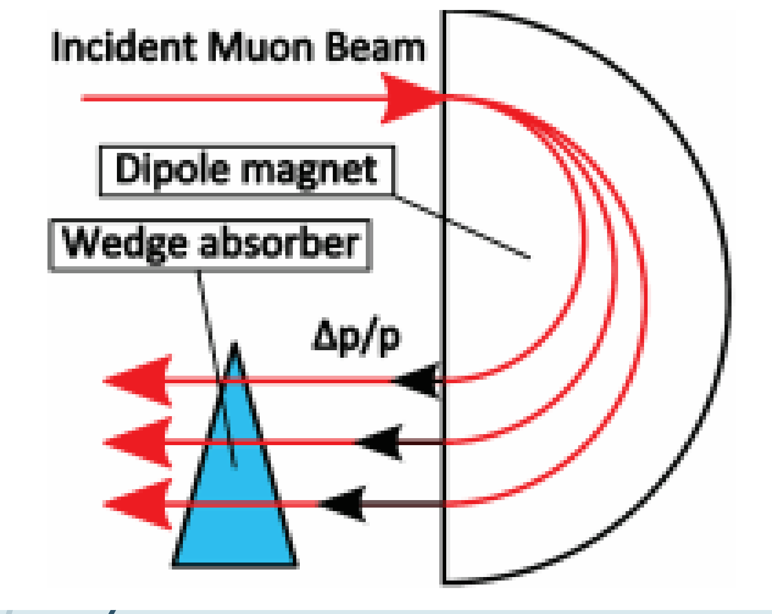
Emittance Exchange in MICE

Craig Brown

Brunel University

12 September 2019

Aims



- Demonstrate Emittance Exchange and Reverse Emittance Exchange in the Wedge using MICE data
- Emittance Exchange can be demonstrated by looking at the change in phase space density of the particle selection before and after having passed through a Wedge absorber
- Emittance Exchange is shown by a decreased transverse phase space density (x, px, y, py) and increased longitudinal phase space density (z, pz), (and vice versa for Reverse Emittance Exchange)
- Can use a number of techniques to calculate phase space density: KDE, KNN, Voronoi Tessellations, etc.
- MICE beam only has a small natural dispersion → Use beam reweighing techniques to select beams with desired dispersion

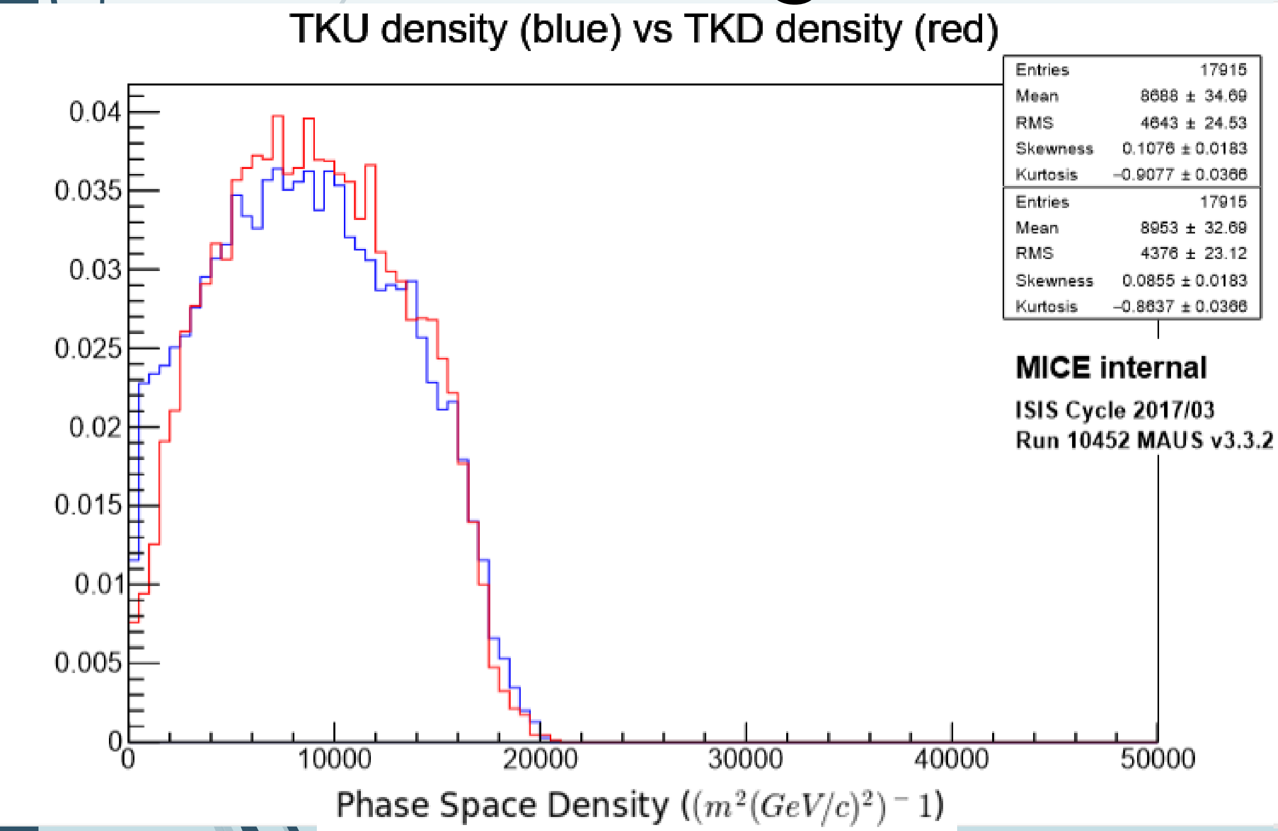
Previously: Particle Selection – 4D transverse

- Will look at a number of selections for when the wedge is present/absent and see the advantages/disadvantages of selection cuts
- All will include:
  - TOF01 cut
  - Radius cut < 150 mm
  - Momentum cut 130 -150 MeV/c
  - Single track in the Upstream Tracker and a single track in the Downstream Tracker
- Will compare this cut with the selection for when there is an Upstream Track but no Downstream Track, to look at selection bias.

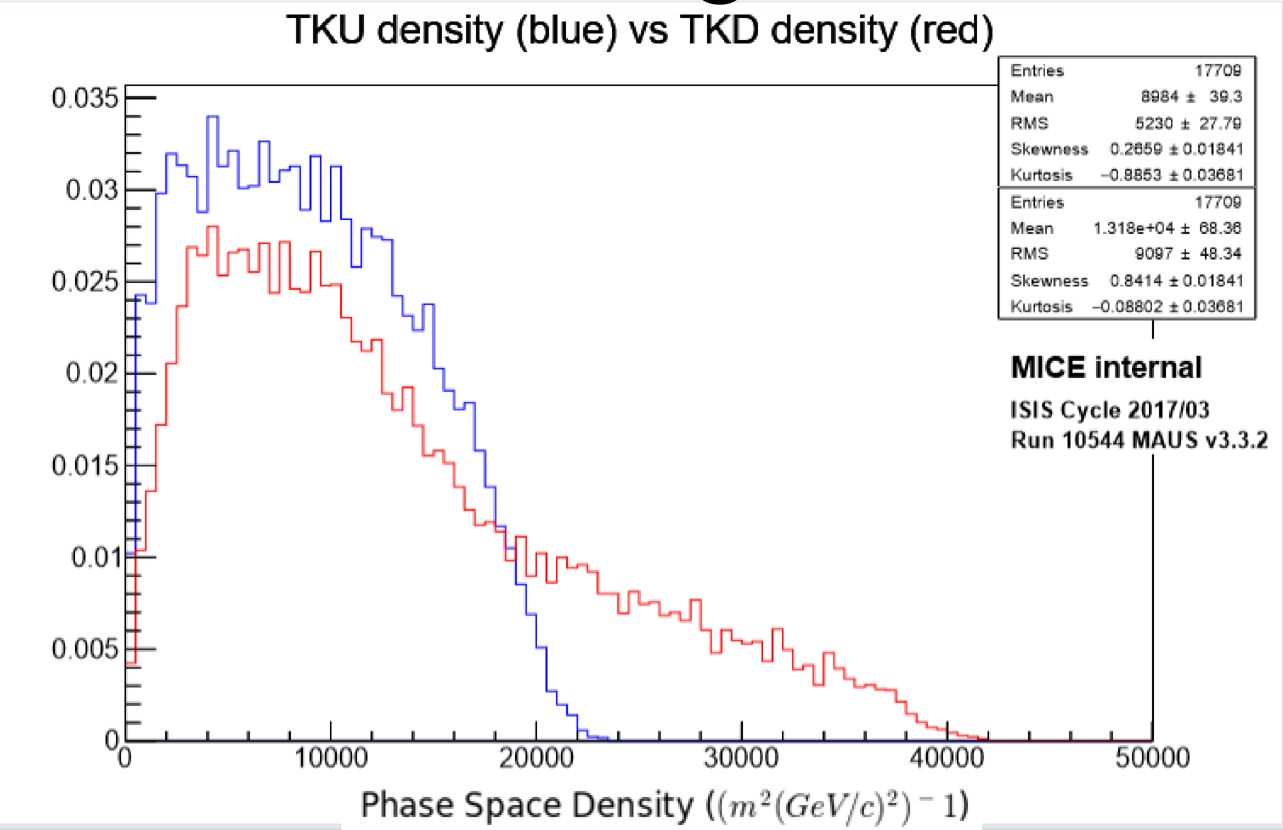
10-140 4D Transverse phase space density

- Single track that has gone both through Upstream and Downstream Tracker

No Wedge



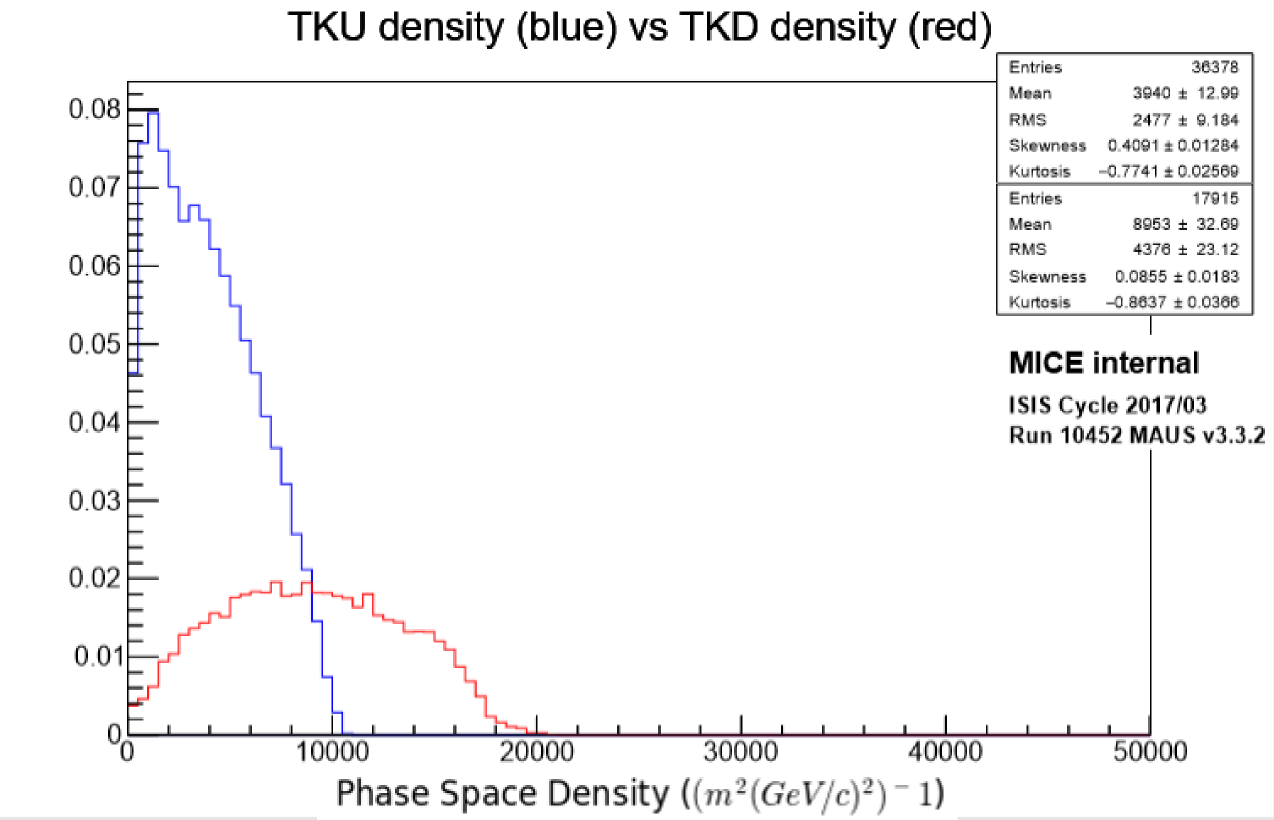
Wedge



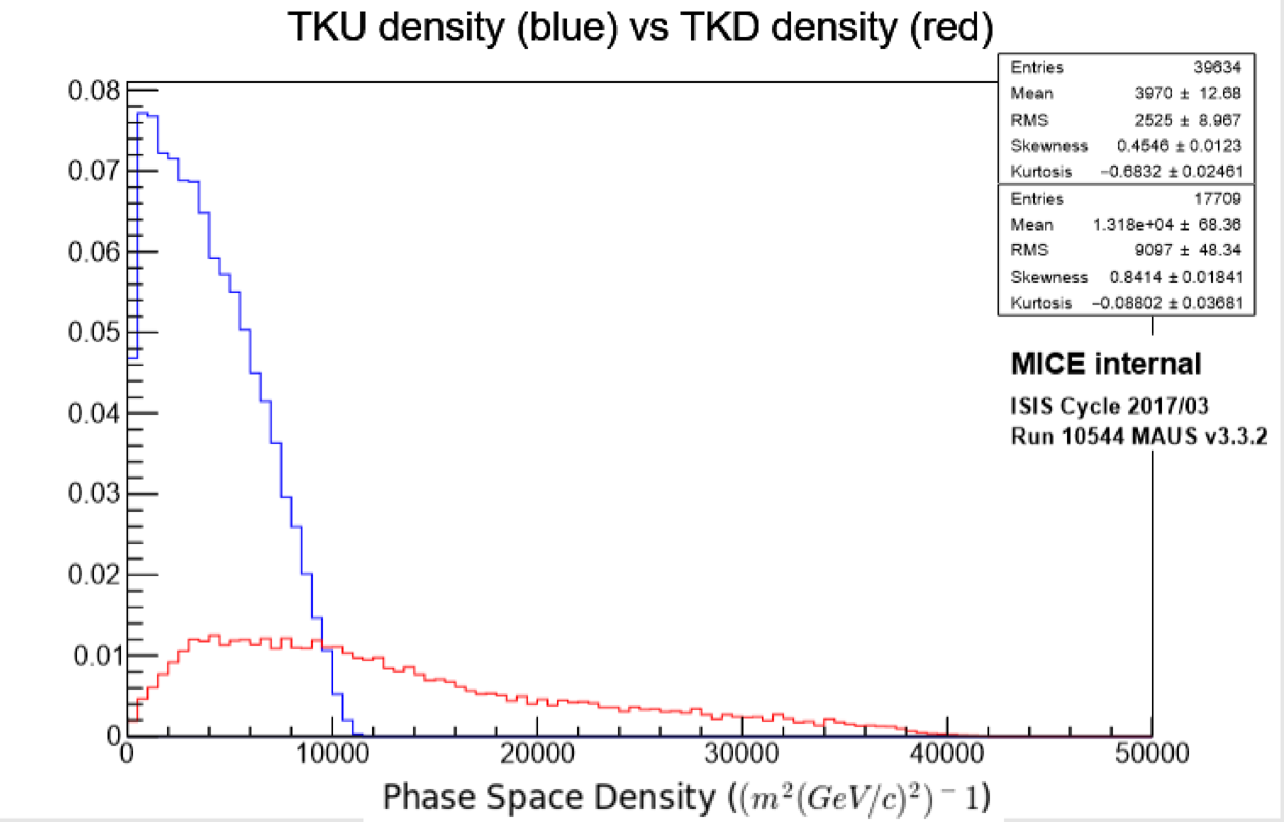
10-140 4D Transverse phase space density

- Single track that has gone both through Upstream and Downstream Tracker
- And single track that has only gone through Upstream Tracker

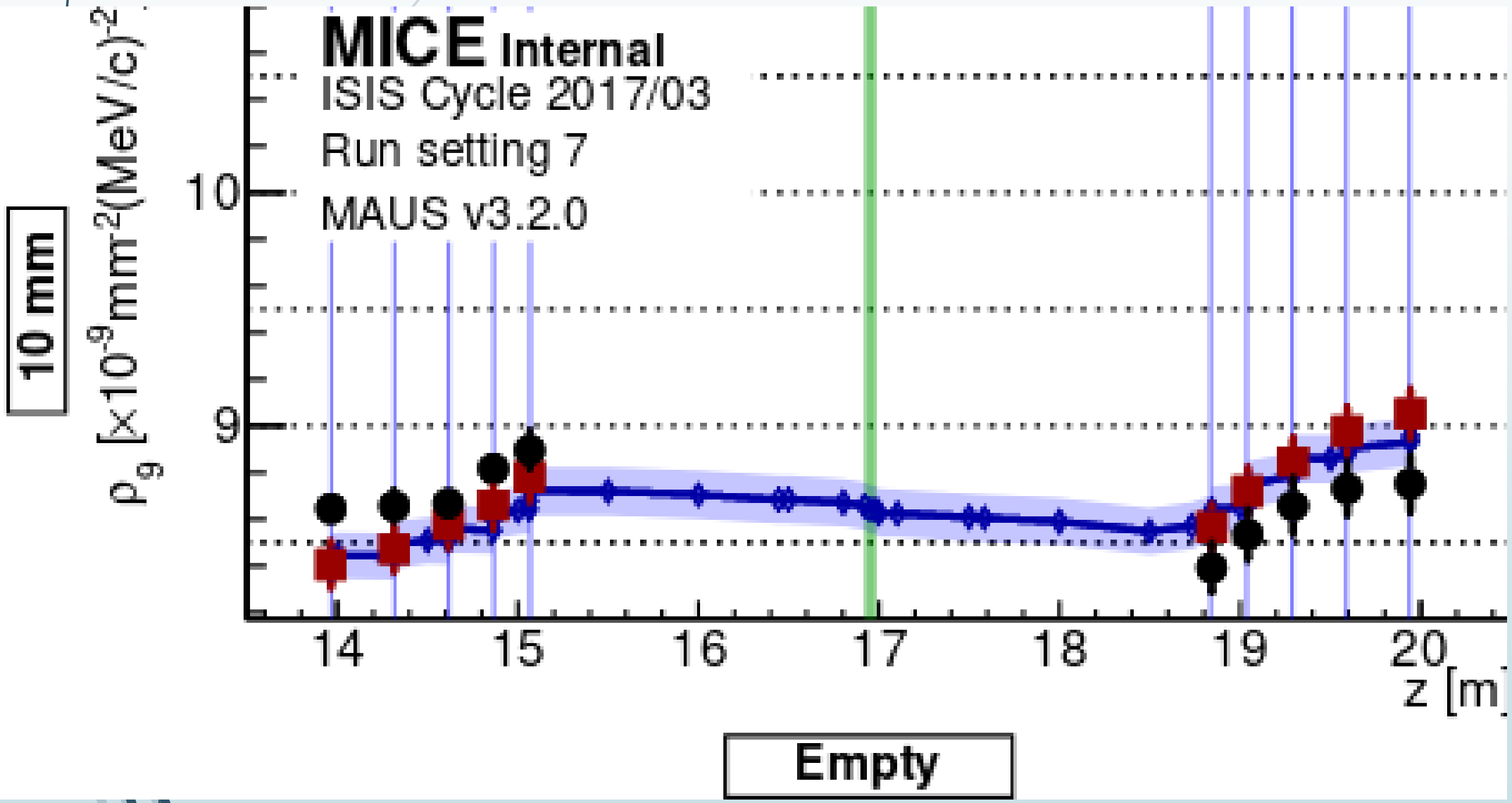
No Wedge



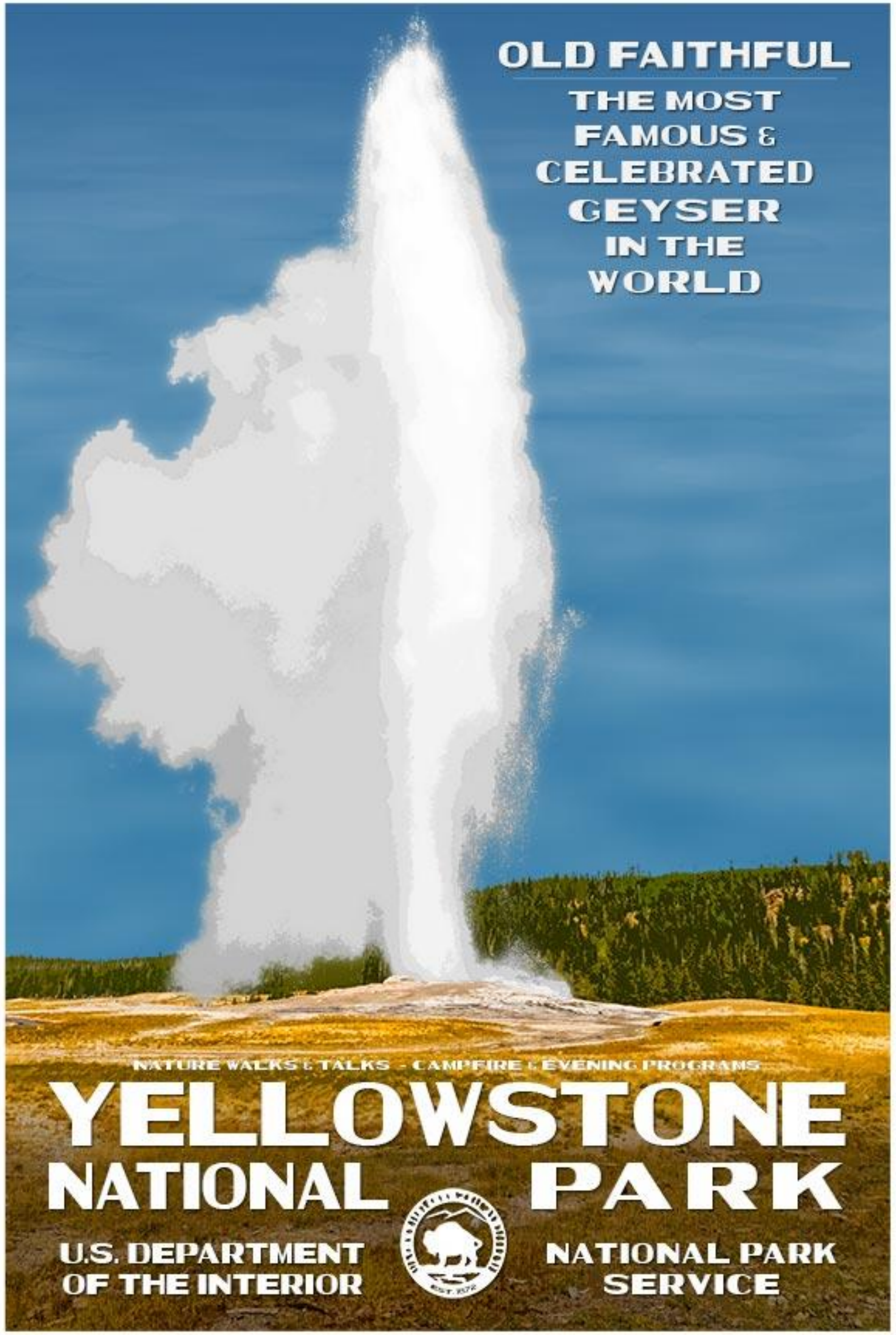
Wedge



Is KDE a poor Estimator?



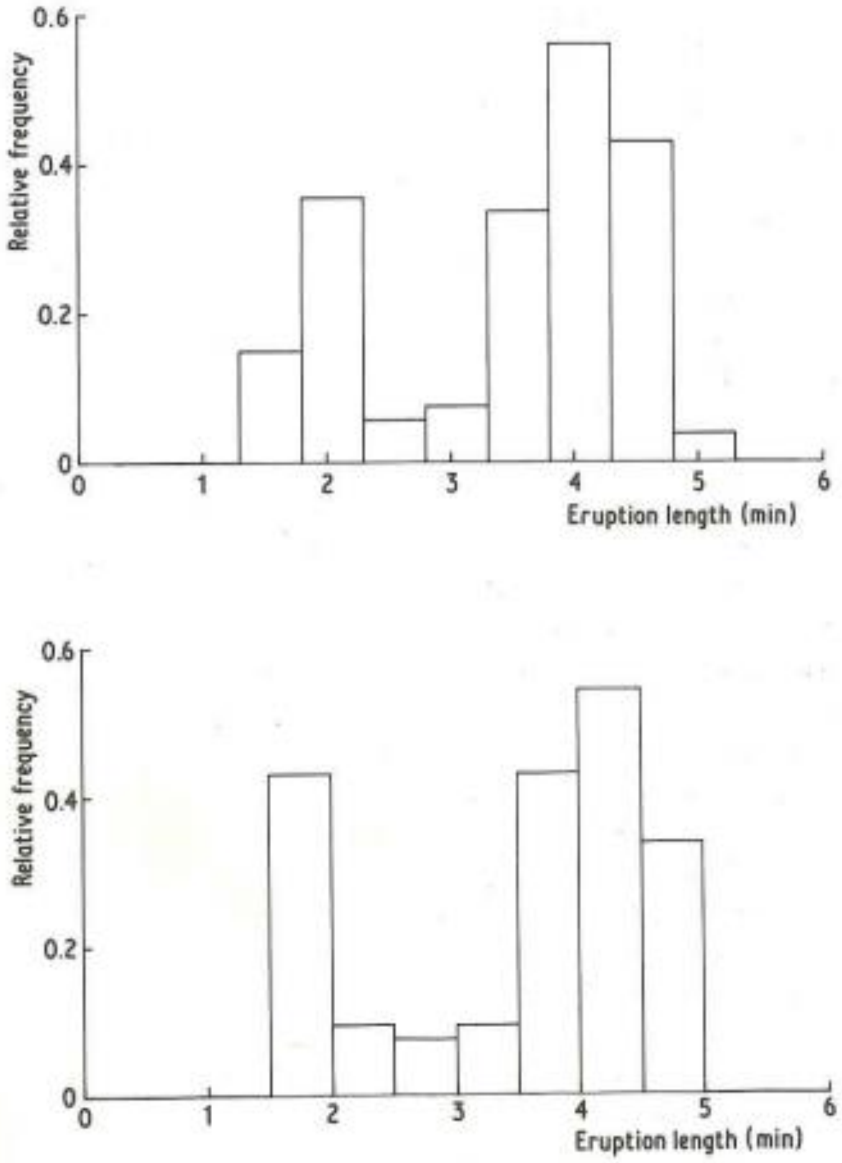
- Produces very different results depending on input selection
- It does show agreement with Francois' KNN estimate for the 9th percentile of the no absorber upstream sample where the data is comparable (bar for any small differences in magnetic fields)
- Why the different results?



Histograms, KDE and KNN Old Faithful Geyser Eruptions

- Highly predictable geothermal feature
- Spews boiling hot water 100 – 180 feet into the air
- Erupts 20 times a day. Eruptions can be predicted to within a 90% accuracy in a 10 minute interval
- Eruptions typically last 1.5 to 5 minutes
- Shows distinct bimodal feature
- The following will look at a sample of the data which should follow the parent distribution i.e. all eruptions in time
- This will be basis to determine if a density estimate follows the true underlying density

Histograms, KDE and KNN – Basics (from Silverman)



- Probability density function gives the probability a quantity is found in the interval:

$$P(a < X < b) = \int_a^b f(x)dx \qquad for\ all\ a < b$$

- The m^{th} histogram interval for origin x_0 and bin width h is given by:

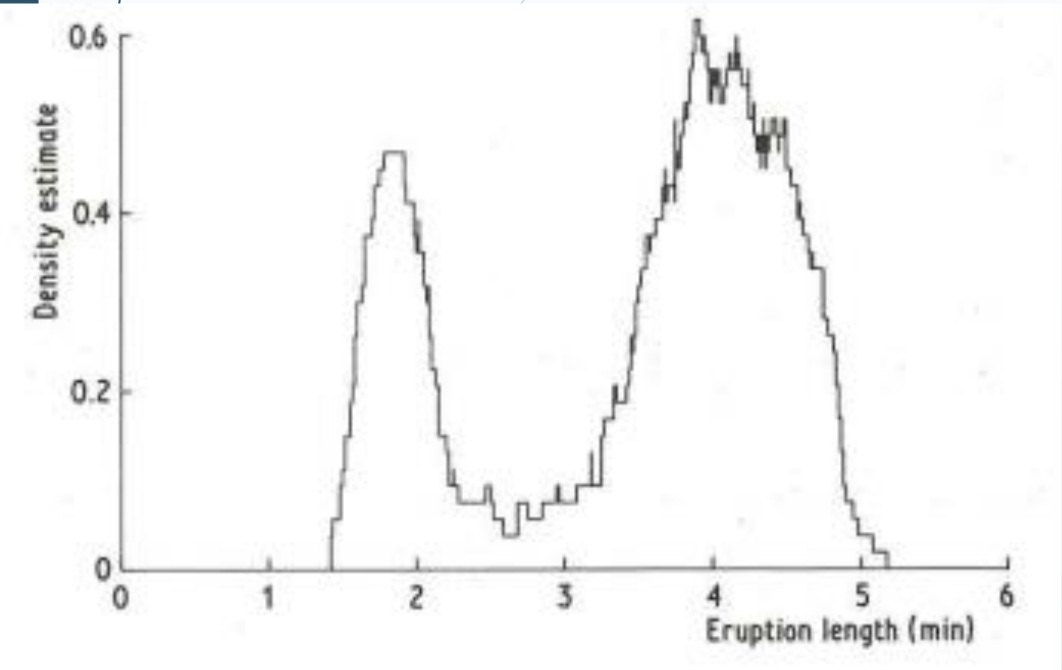
$$[x_0 + mh, x_0(m+1)h)$$

- The histogram is then defined by:

$$\hat{f}(x) = \frac{1}{nh}(no. of\ X_i\ in\ the\ same\ bin\ as\ x)$$

- Choice of origin and bin width can give “apparent structure effects” that are due to random error
- Discontinuity of histograms can cause difficulty if derivatives of the estimate are required

Naive Estimator



- For a random variable X with density f, then:

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

- Then

$$\hat{f}(x) = \frac{1}{2hn} [no. of\ X_i, \ldots, X_n\ falling\ in\ (x - h, x + h)]$$

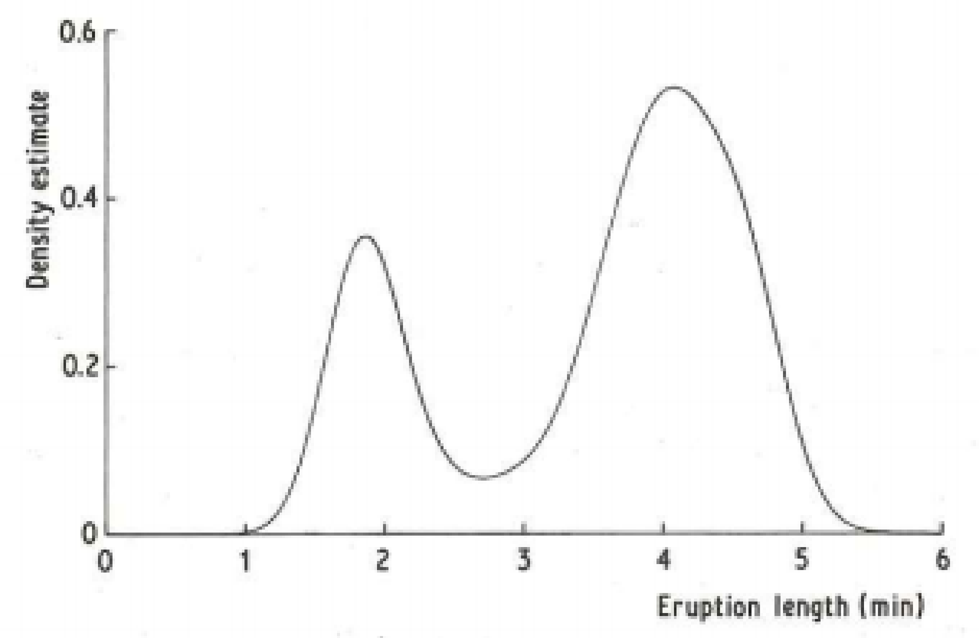
- Define weight function w by

$$w(x) = \begin{cases} 1/2 & if\ |x| < 1 \\ 0 & otherwise \end{cases}$$

- Then

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

Kernel Estimator



- The kernel estimator is obtained by replacing the weight function of the naive estimator by a kernel function K satisfying:

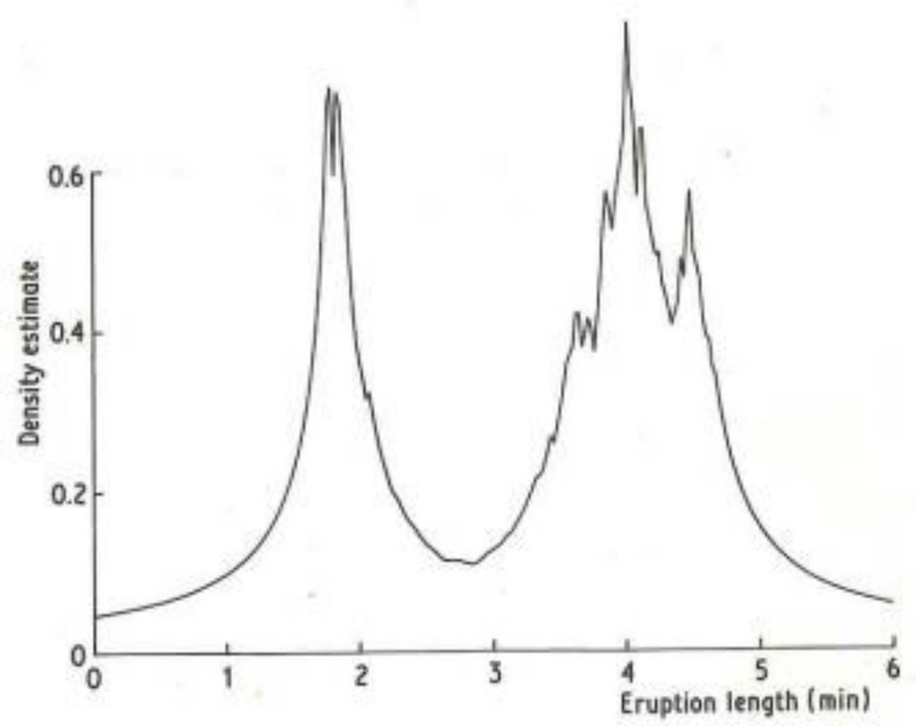
$$\int_{-\infty}^{\infty} K(x)dx = 1$$

- The kernel estimator of bandwidth h is then defined by

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

- Varying the bandwidth h determines the level of smoothing, as h tends to zero, the smoothing becomes a sum of Dirac delta spikes, but if h becomes large, all detail is obscured.
- If K is non-negative everywhere, then $\hat{f}$ itself will be a probability density. The probability density function of the sample has been convolved with the kernel
- This can lead to non-negative tails to naturally positive data, especially when the data distribution is long-tailed. Parameter choice can be used to minimize the undesired effects

K-nearest neighbour



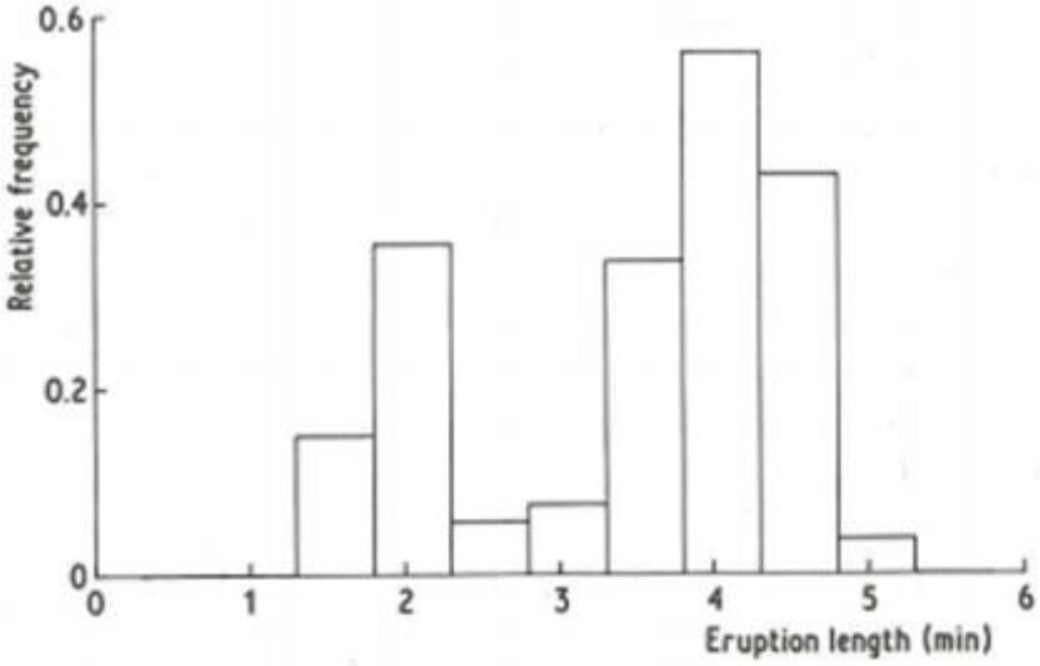
- Define the distance $d(x,y)$ between two points on the line to be $|x-y|$ and for each t define $d_1(t) \le d_2(t) \le \cdots \le d_n(t)$ to be the distances arranged in ascending order.
- The k^{th} nearest neighbour density estimate is then given by
$$\hat{f}(t) = \frac{k-1}{2nd_k(t)}$$
- i.e. $(k-1)$ observations fall in the interval $[t-d_k(t), t+d_k(t)]$
- The nearest neighbour estimate is inversely proportional to the size of box needed to contain it -> undersmoothing in tails should be reduced
- $\hat{f}(t)$ is positive and continuous everywhere, but its derivative will be discontinuous at all the same points as d_k
- The nearest neighbour estimate will not be a probability density (but only an approximation) as it does not integrate to unity
- For t less than the smallest data point, $d_k(t) = X_k - t$ and for $t > X_n$: $d_k(t) = t - X_{(n-k+1)}$, thus $\int_{-\infty}^{\infty} \hat{f}(t)dt$ is infinite and the tails of $\hat{f}$ die away slowly

KNN relation to KDE

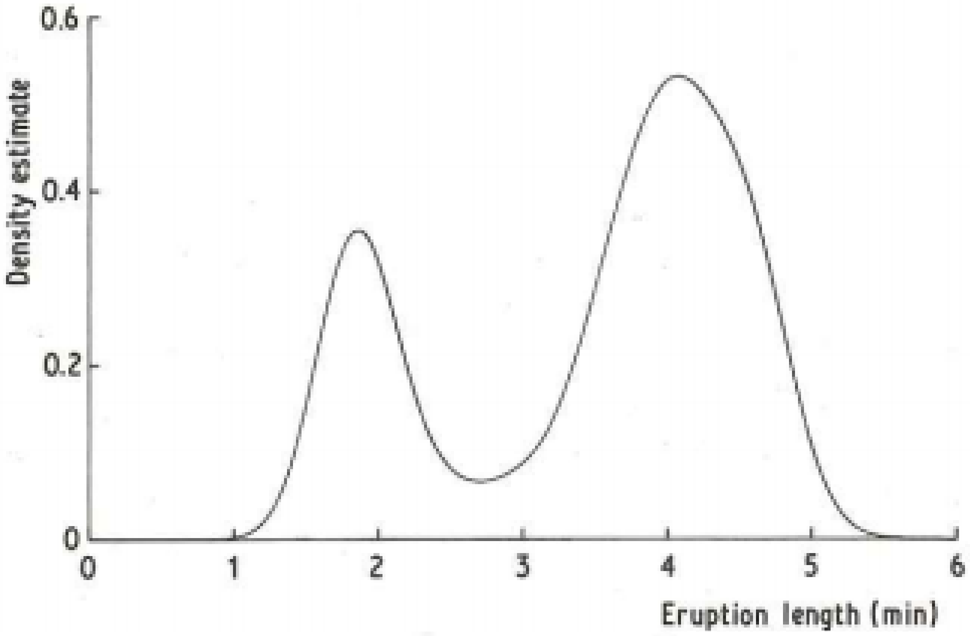
- Let $K(x)$ be a kernel function integrating to one
- The k^{th} nearest neighbour estimate is given by

$$\hat{f}(t) = \frac{1}{nd_k(t)} \sum_{i=1}^{n} K\left(\frac{t - X_i}{d_k(t)}\right)$$

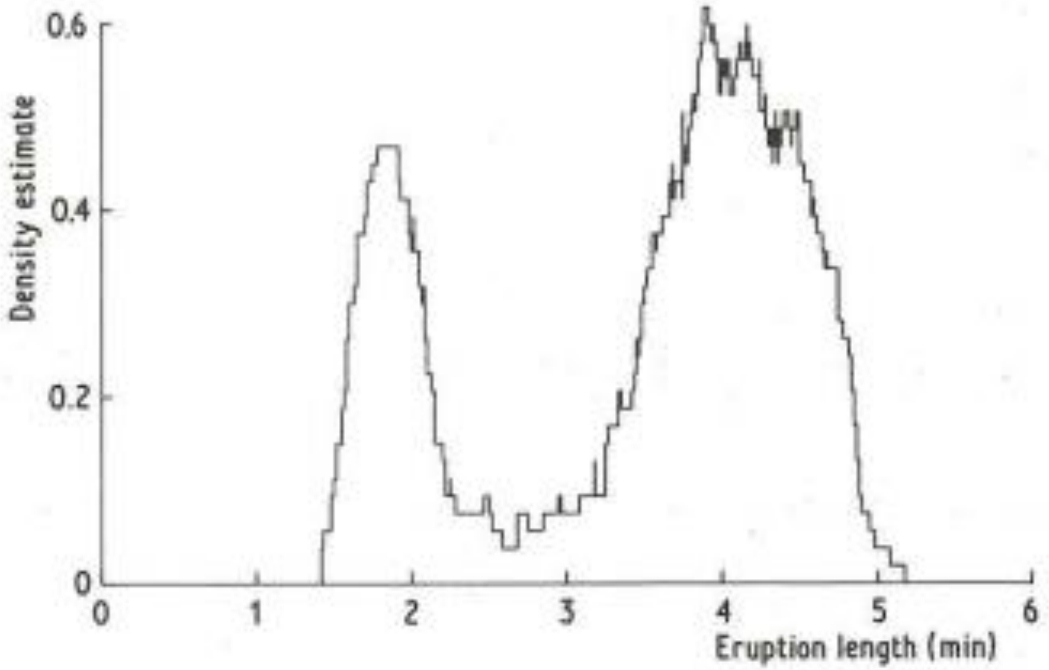
- $\hat{f}(t)$ is the kernel estimate evaluated at t with window width $d_k(t)$ where the choice of k governs the smoothing.



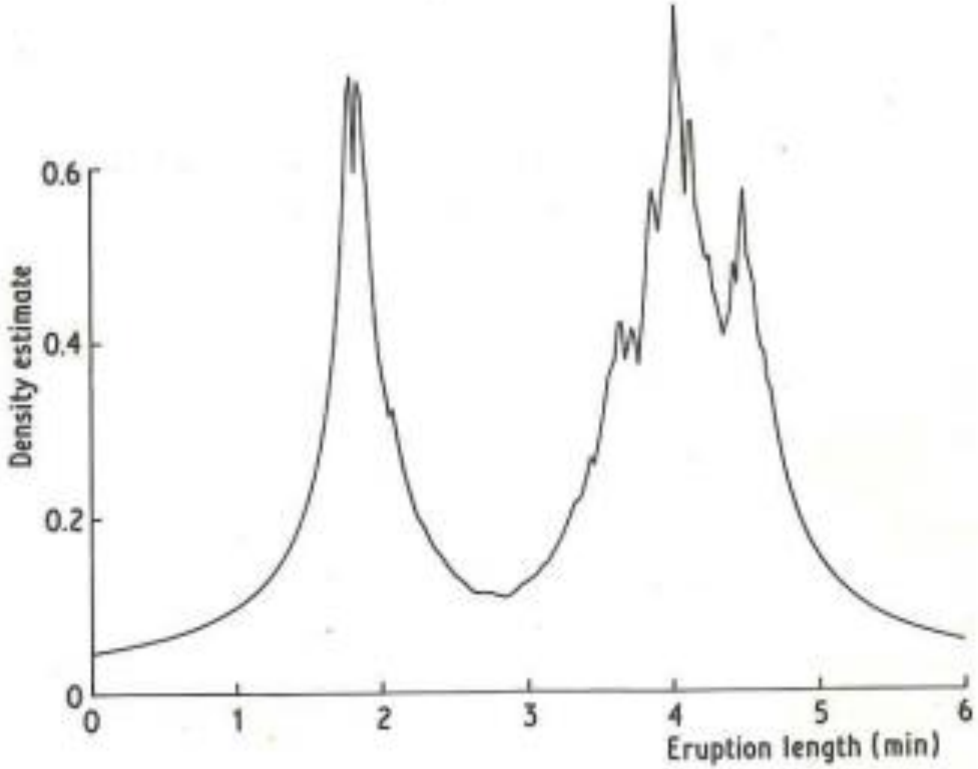
0.6
0.4
0.2
0
Relative frequency
0
1
2
3
4
5
6
Eruption length (min)



0.6
0.4
0.2
0
Density estimate
0
1
2
3
4
5
6
Eruption length (min)



0.6
0.4
0.2
0
Density estimate
0
1
2
3
4
5
6
Eruption length (min)



0.6
0.4
0.2
0
Density estimate
0
1
2
3
4
5
6
Eruption length (min)

1-D estimate to n-dimensional estimate

- KDE in 1-D

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

Becomes in n-D:

$$\hat{f}(\vec{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left\{\frac{1}{h}(\vec{x} - \vec{X_i})\right\}$$

where $\int_{R^d} K(\vec{x})dx = 1$ for a d-dimensional space and h^d is the smoothing parameter for each particular dimension. h^d can also be given by a smoothing matrix e.g. the covariance matrix if it is representative of the underlying distribution.

- The choice of kernel only has a minor effect (slightly different efficiencies), and thus a gaussian kernel (most common) will be used to retain the differentiability of $\hat{f}(\vec{x})$. The gaussian kernel is given by:

$$K(\vec{x}) = (2\pi)^{-d/2}\exp(-\frac{1}{2}\vec{x}^T\vec{x})$$

1-D estimate to n-dimensional estimate

- KNN in 1-D

$$\hat{f}(t) = \frac{k-1}{2nd_k(t)}$$

Becomes in n-D (from Francois):

$$\vec{f}(x) = \frac{k}{n\kappa_d R_k^d} = \frac{k\Gamma(\frac{d}{2}+1)}{n\pi^{\frac{d}{2}} R_k^d}$$

where $d_k(t)$ is now the Euclidean distance $R_i = \left|\left|\vec{x} - \vec{x_i}\right|\right| = \sqrt{(\vec{x} - \vec{x_i})^T(\vec{x} - \vec{x_i})}$, κ_d is the volume of a unit d-ball (in 1-D it is equal to two), $\Gamma(\frac{d}{2}+1)$ is Euler's gamma function, while k and (k-1) differ due to counting conventions of whether the test point is included.

Choice of smoothing parameter

- The Mean Integrated Squared Error (MISE) can describe the accuracy of $\vec{f}(t)$ as an estimator of f.
- $MISE(\vec{f}) = E\int\{\vec{f}(x) - f(x)\}^2 dx = \int\{E\vec{f}(x) - f(x)\}^2 dx + \int Var(\vec{f}(x))dx$
- $MISE(\vec{f}) = \int\left[Bias\left(\vec{f}(x)\right)^2 + Var(\vec{f}(x))\right]dx$
- As $\vec{f}(x) = \sum_{i=1}^{n}\vec{f}(x) \sim \frac{k}{r^d}$, the optimal choice of k is determined by a trade-off between the variance and the squared bias.

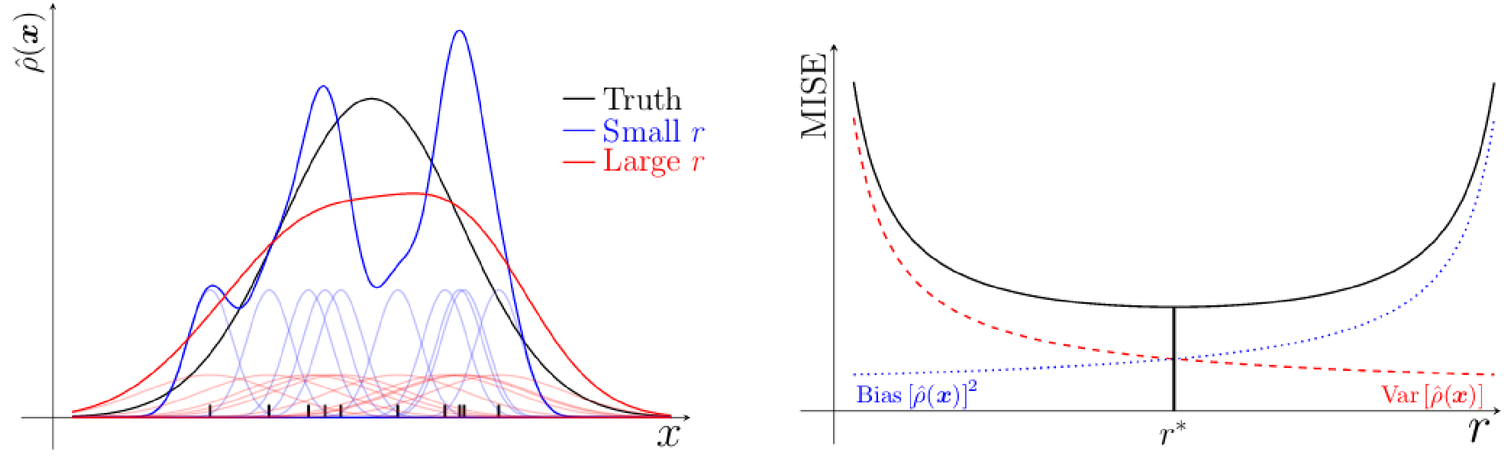


Figure 7.1: (Left) Illustration of the effect of the smoothing radius, r, on the behaviour of nonparametric density estimators. (Right) Schematic of the evolution of the bias, variance and MISE as a function of the smoothing radius, r.

For small r the estimate follows the data closely as its not biased but has a very large variance due to fluctuations.

For large r the estimate varies little as it is less sensitive to fluctuations, but becomes more biased.

Bias and Variance

- For KNN:

$$Bias[\vec{f}(x)] \cong \frac{\mu_2(w)\nabla^2 f(x)}{2(\kappa_d f(x))^{\frac{2}{d}}} \left(\frac{k}{n}\right)^2$$

$$Var[\vec{f}(x)] \cong \frac{f^2(x)}{k}$$

With $\mu_2(w)$ the second moment of the uniform kernel and $\nabla^2 f(x)$ the Laplacian of the density field. The MISE is of order:

$$MISE(k) = \mathcal{O}\left(\left(\frac{k}{n}\right)^{\frac{4}{d}} + \frac{1}{k}\right)$$

Which admits a minimum for a parameter k of order:

$$k \sim n^{-4/(4+d)}$$

The optimal rate of convergence for a KNN estimator is then:

$$MISE(k) = \mathcal{O}\left(n^{-4/(4+d)}\right)$$

Bias and Variance

- For KDE (with second-order kernels):

$$Bias_h(\vec{x}) \approx \frac{1}{2} h^2 \nabla^2 f(\vec{x}) \int t_1^2 K(\vec{t}) d\vec{t}$$

$$Var[\vec{f}(x)] \approx n^{-1} h^{-d} f(\vec{x}) \int K(\vec{t})^2 d\vec{t}$$

The MISE is then approximated by

$$\frac{1}{4} h^4 \left\{ \int t_1^2 K(\vec{t}) d\vec{t} \right\}^2 \int \{\nabla^2 f(\vec{x})\}^2 d\vec{x} + n^{-1} h^{-d} \int K(\vec{t})^2 d\vec{t}$$

The optimal window width to minimize MISE is given by

$$h_{opt}^{d+4} = d \int K(\vec{t})^2 d\vec{t} \left\{ \int t_1^2 K(\vec{t}) d\vec{t} \right\}^{-2} \left\{ \int \{\nabla^2 f(\vec{x})\}^2 d\vec{x} \right\}^{-1} n^{-1}$$

Therefore

$$MISE(h) = \mathcal{O}\left(n^{-4/(4+d)}\right)$$

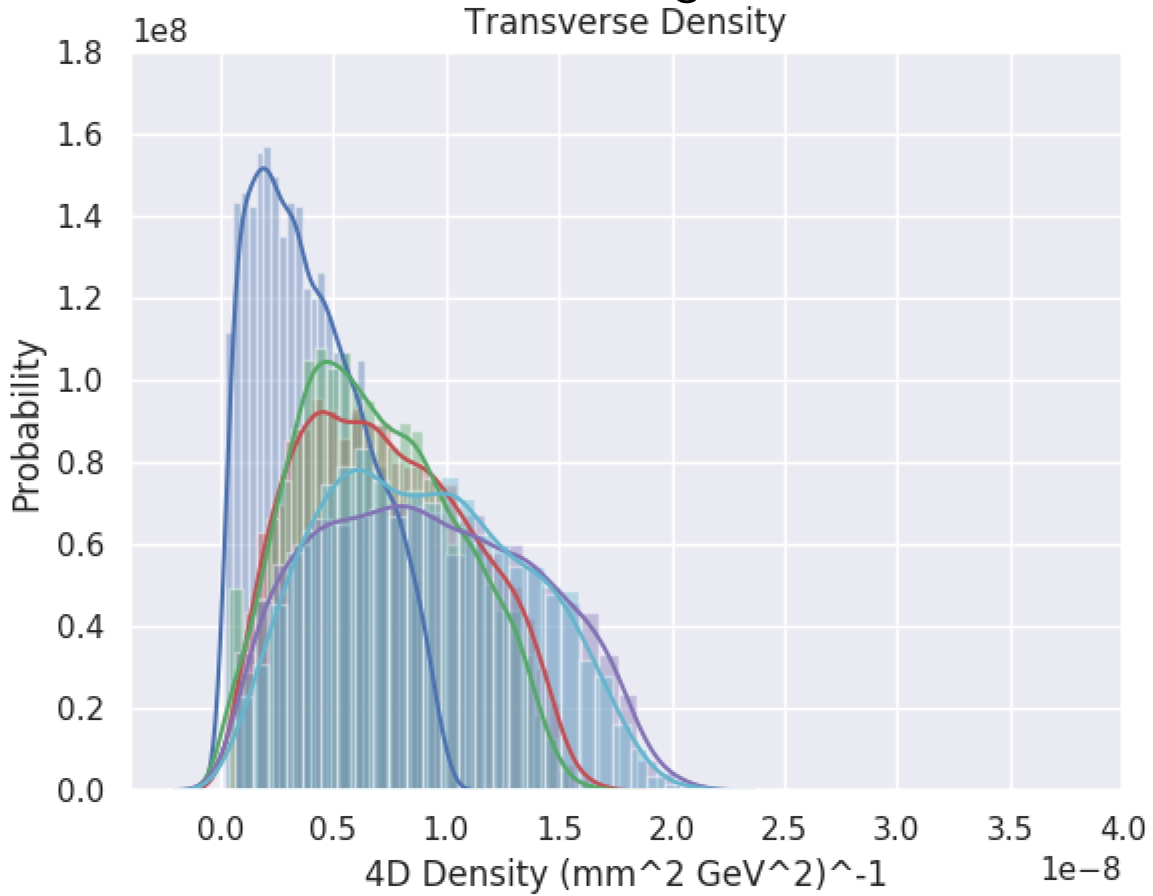
This is same as for KNN, that is the rate of convergence to the density estimate is the same for KNN and KDE. The rate of convergence for the histogram is given by

$$MISE(\Delta) = \mathcal{O}\left(n^{-2/(2+d)}\right)$$

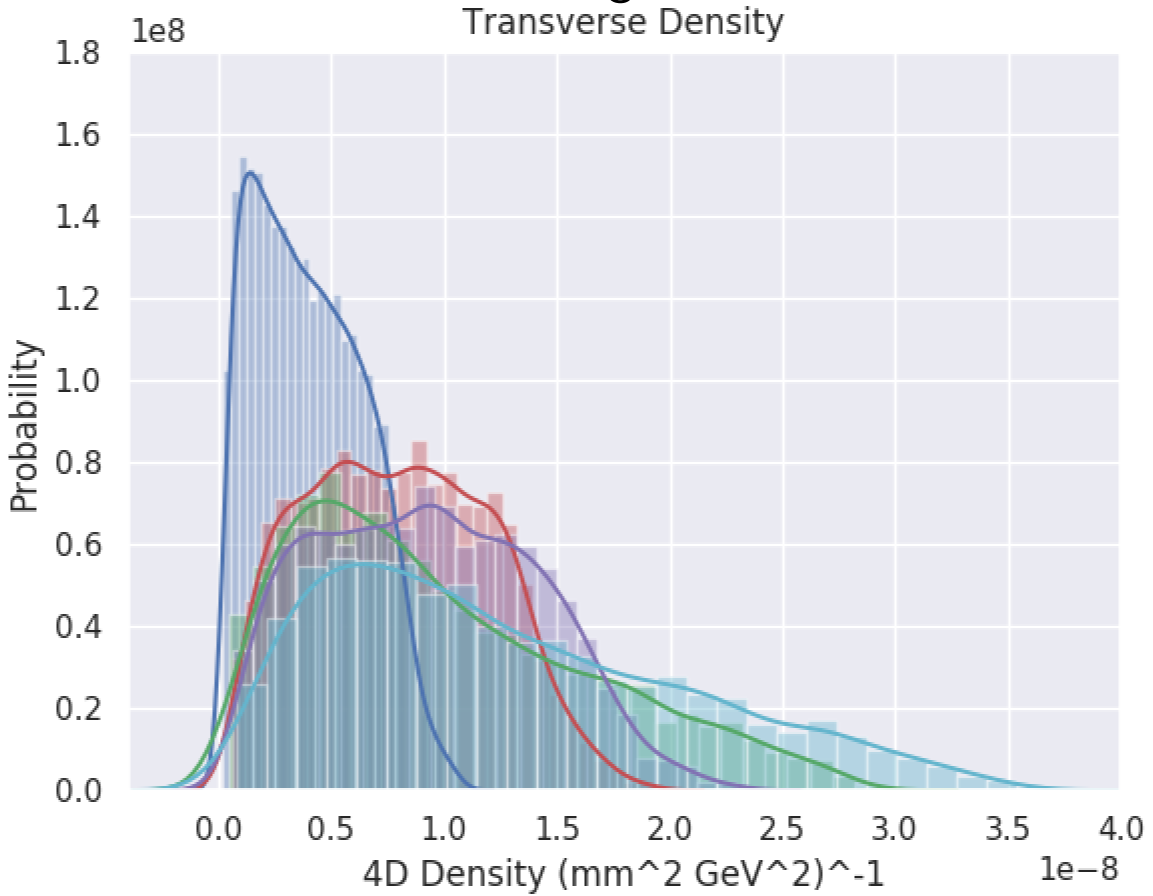
KDE Transverse Phase Space Density

Blue – Full Upstream sample
Orange – Upstream sample that made it Downstream
Green – Full Downstream sample
Red – Upstream sample that made it to TOF2
Magenta – Downstream sample that made it to TOF2

No Wedge



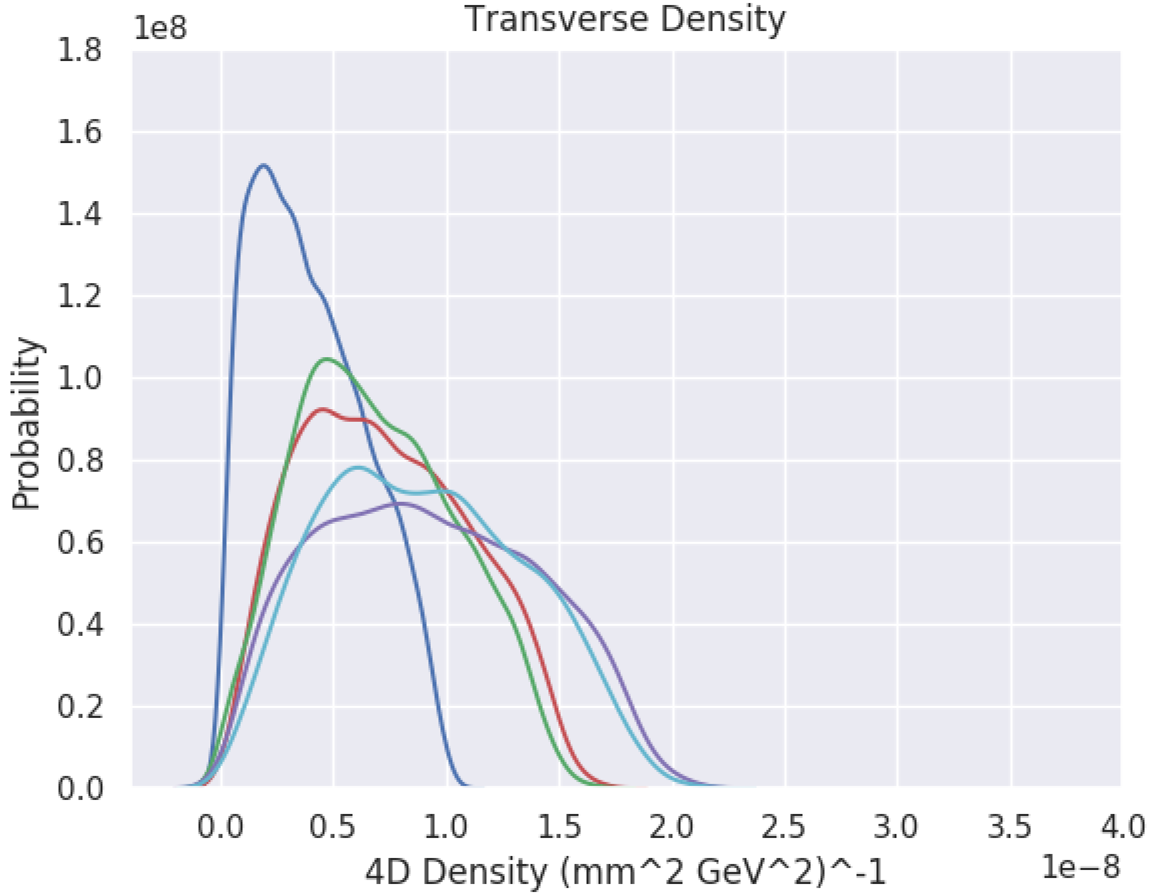
Wedge



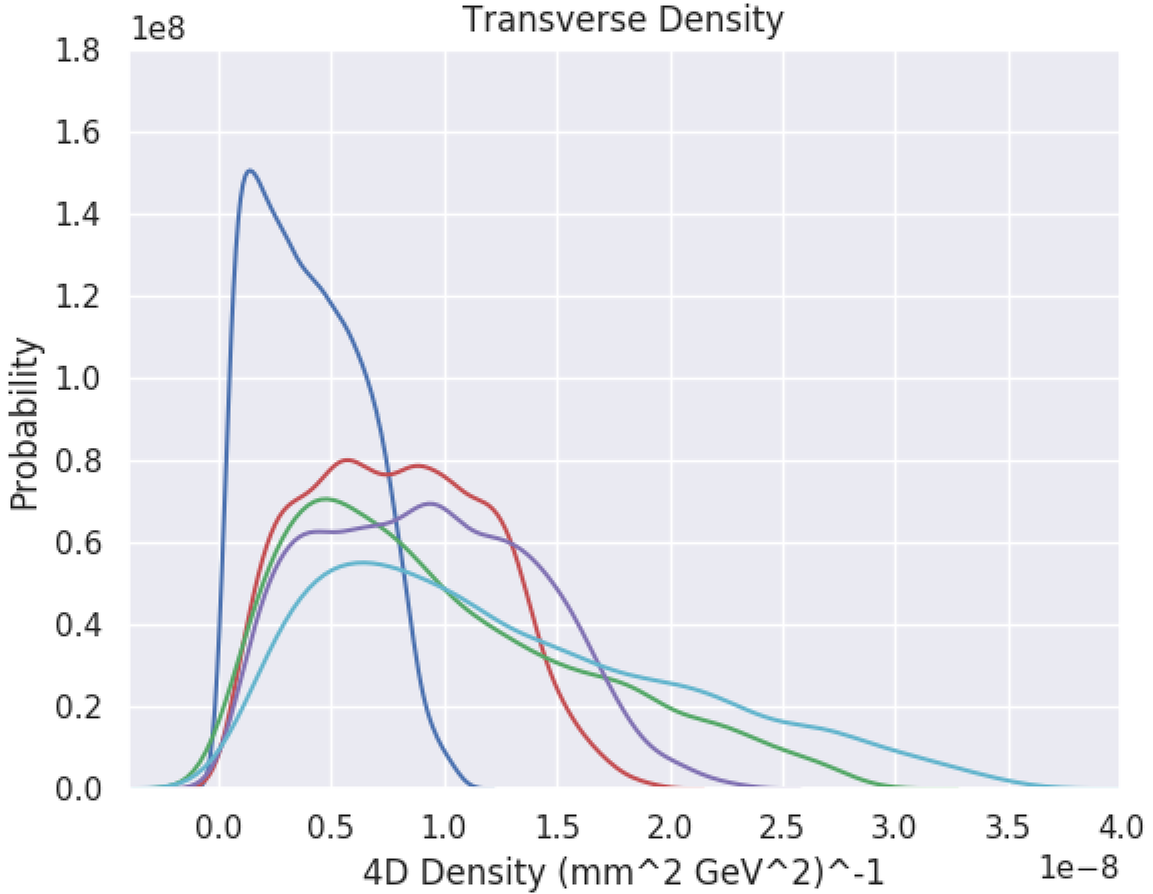
KDE Transverse Phase Space Density

Blue – Full Upstream sample
Orange – Upstream sample that made it Downstream
Green – Full Downstream sample
Red – Upstream sample that made it to TOF2
Magenta – Downstream sample that made it to TOF2

No Wedge



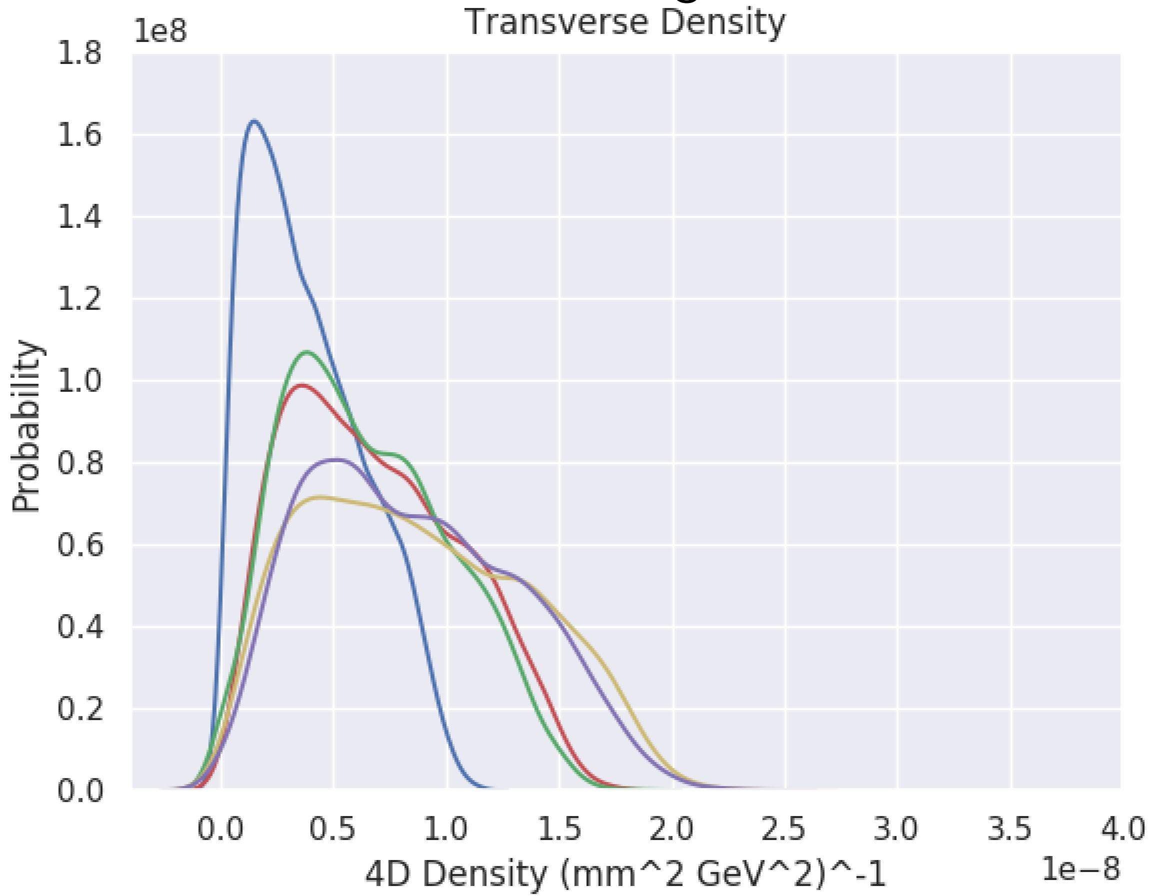
Wedge



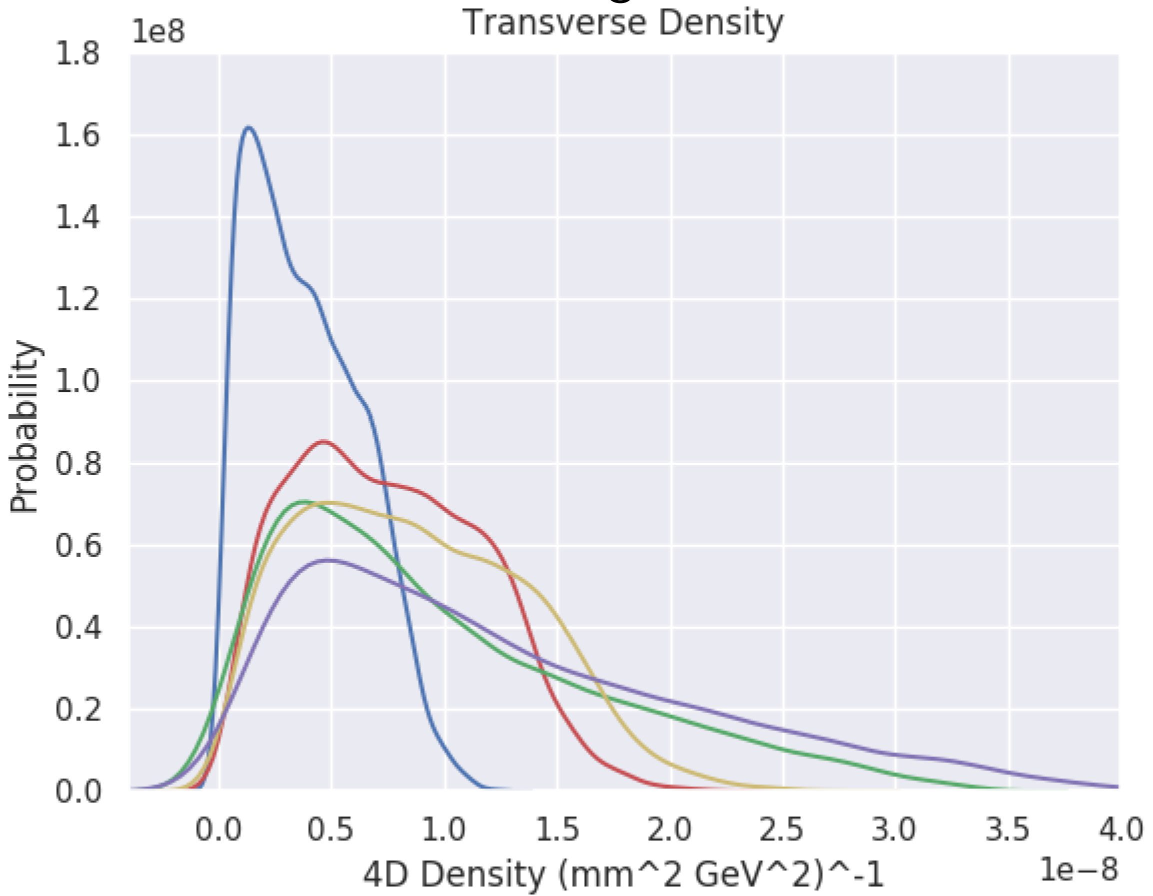
KNN Transverse Phase Space Density

Blue – Full Upstream sample
Orange – Upstream sample that made it Downstream
Green – Full Downstream sample
Red – Upstream sample that made it to TOF2
Magenta – Downstream sample that made it to TOF2

No Wedge



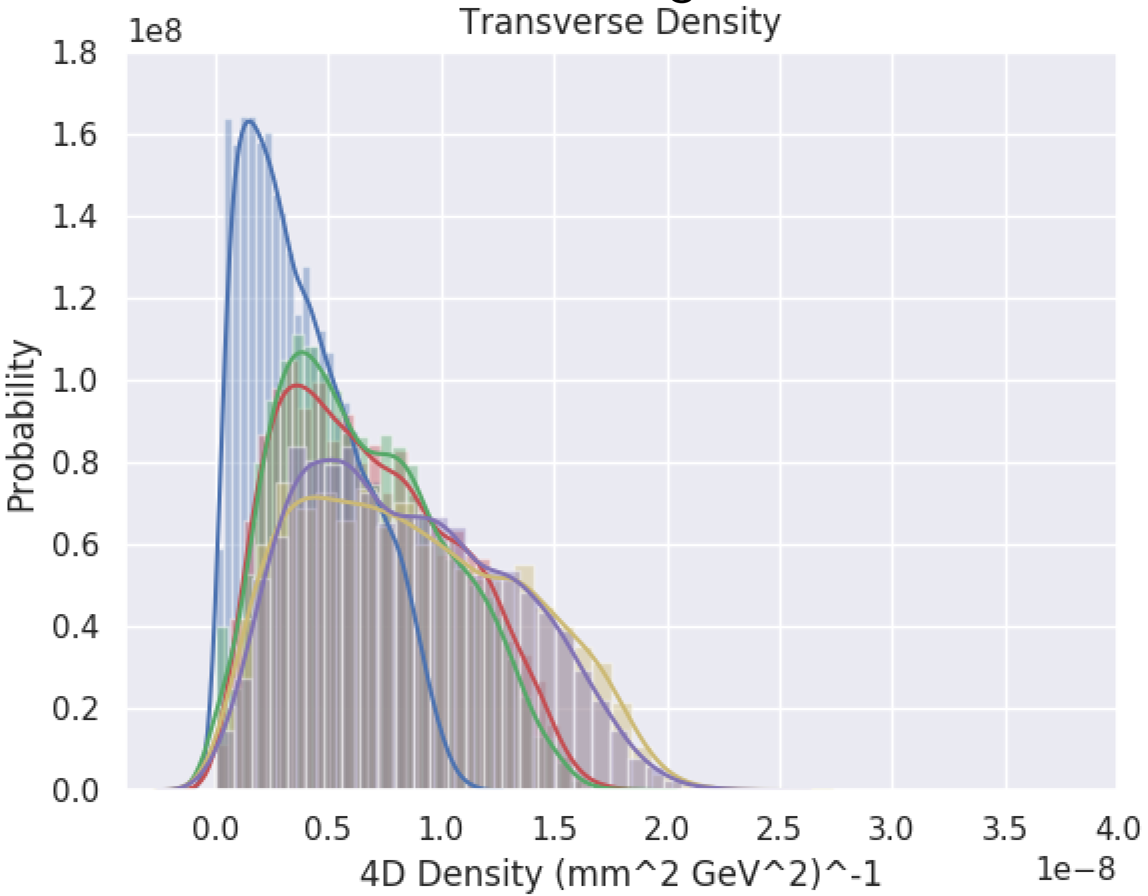
Wedge



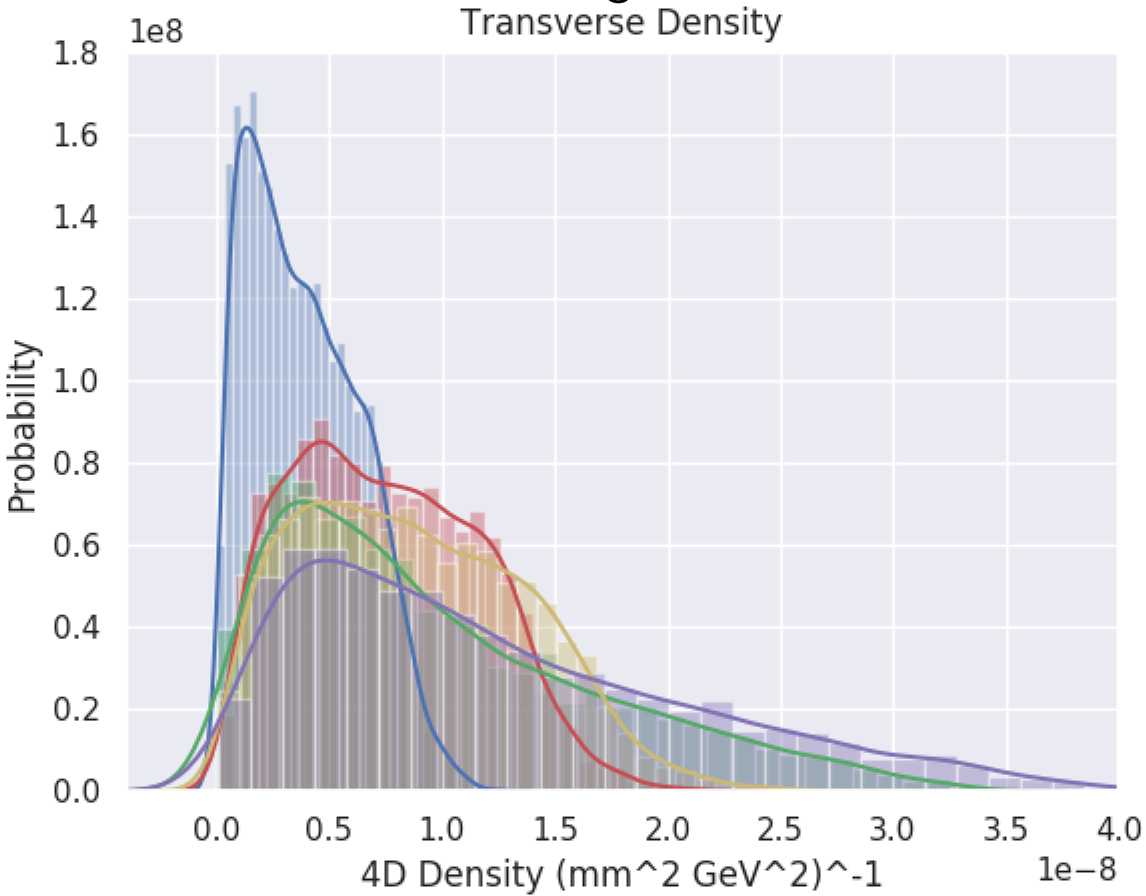
KNN Transverse Phase Space Density

Blue – Full Upstream sample
Orange – Upstream sample that made it Downstream
Green – Full Downstream sample
Red – Upstream sample that made it to TOF2
Magenta – Downstream sample that made it to TOF2

No Wedge



Wedge

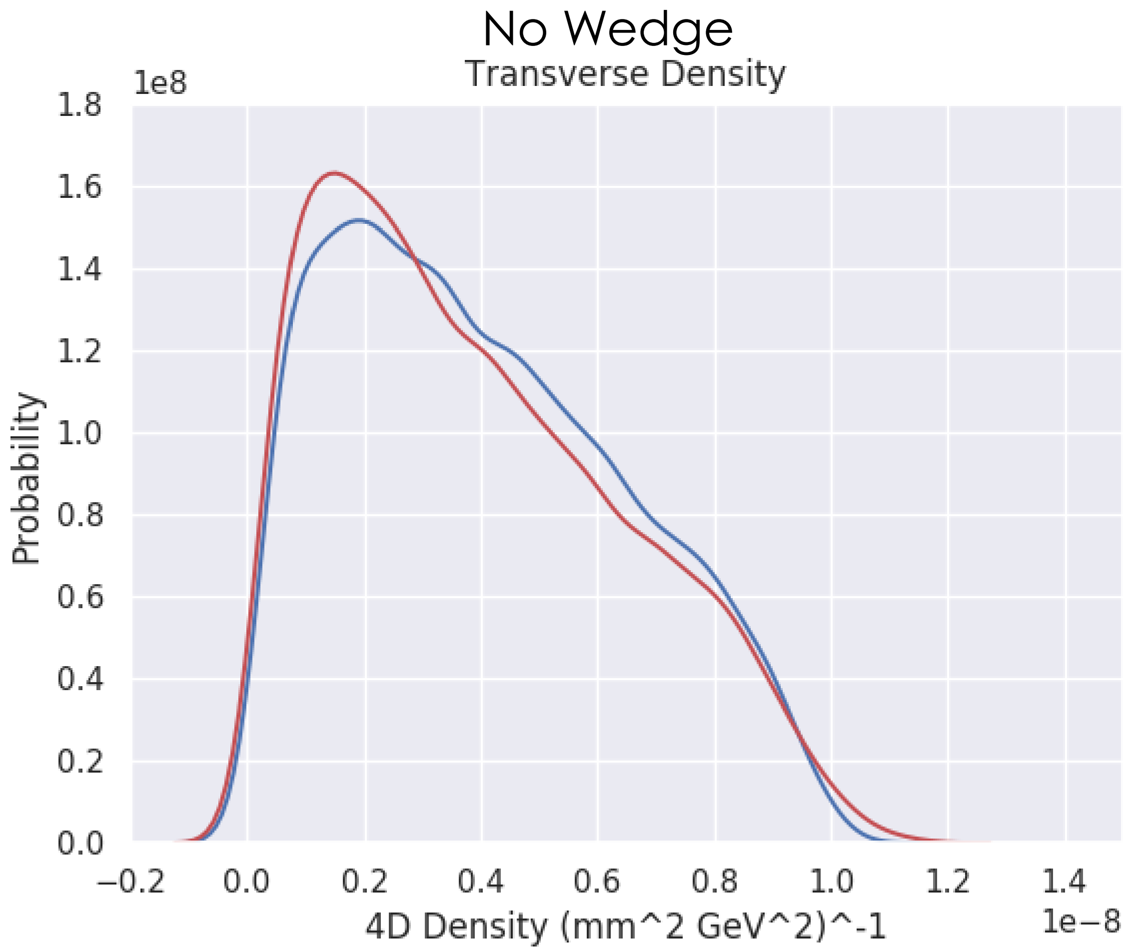


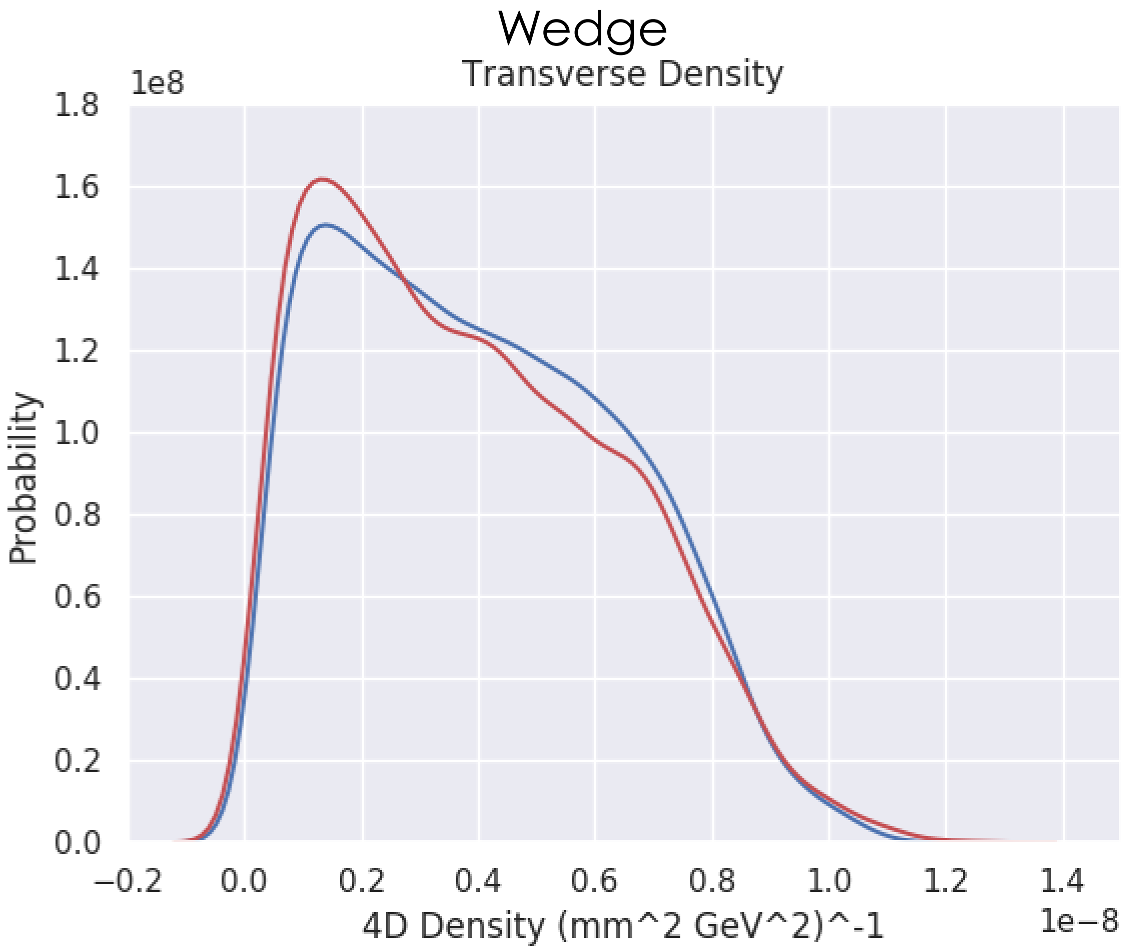
KDE vs KNN

Full Upstream Sample

Blue – KDE   Red – KNN

Slight differences due to KDE convolving the density with the kernel, while for KNN it has been smoothed to ensure area of graph is one





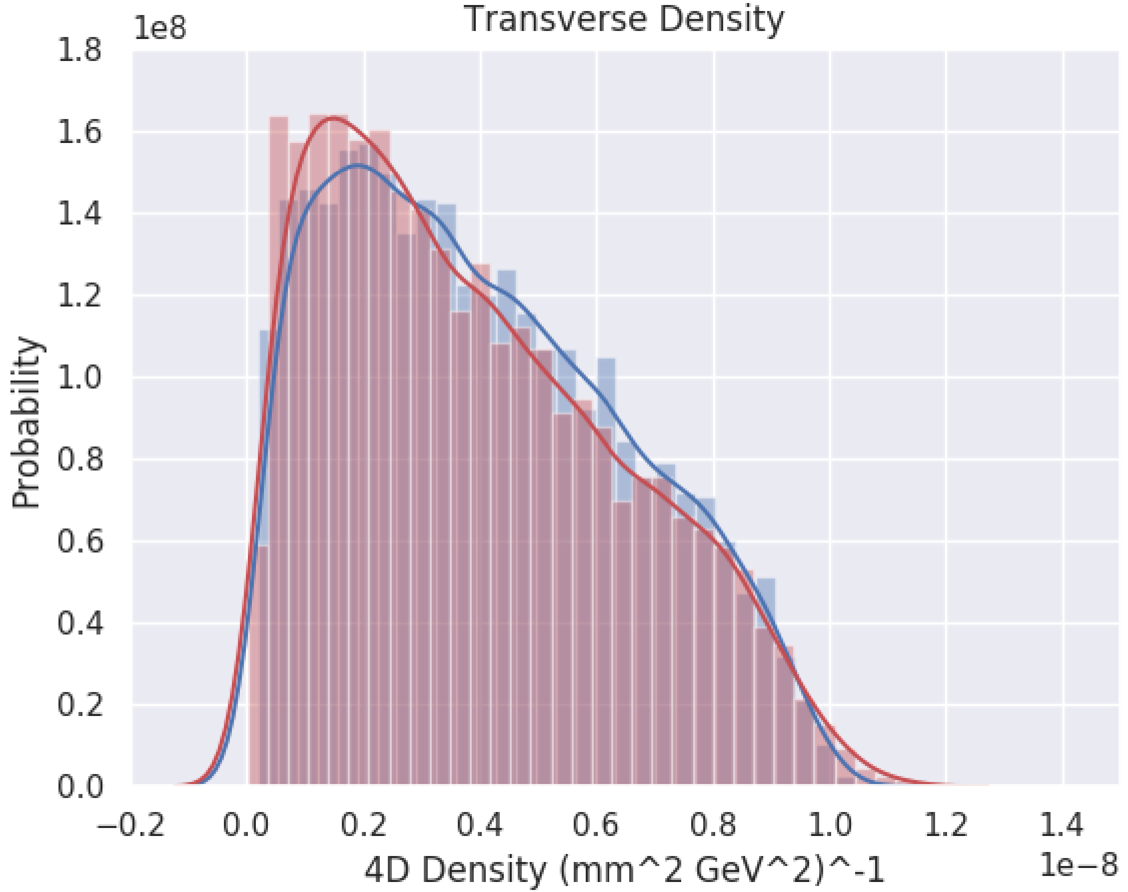
KDE vs KNN

Full Upstream Sample

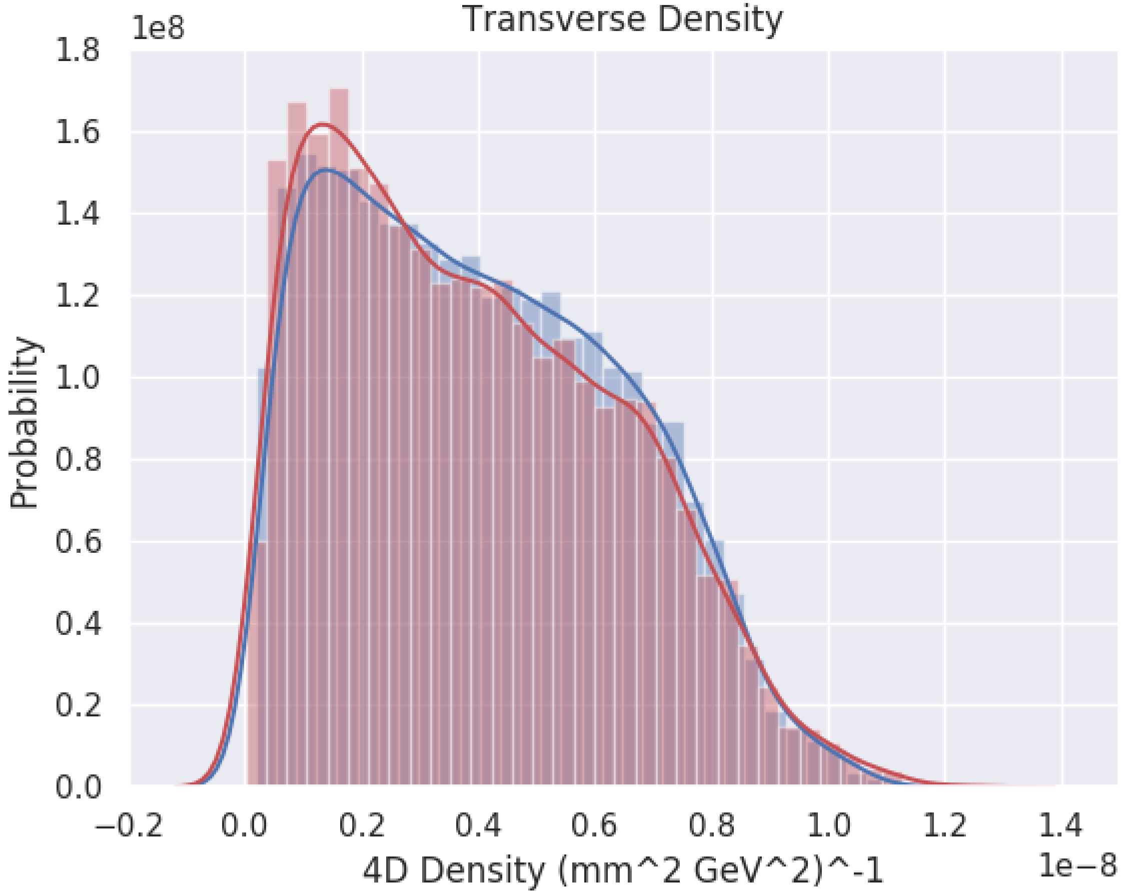
Blue – KDE   Red – KNN

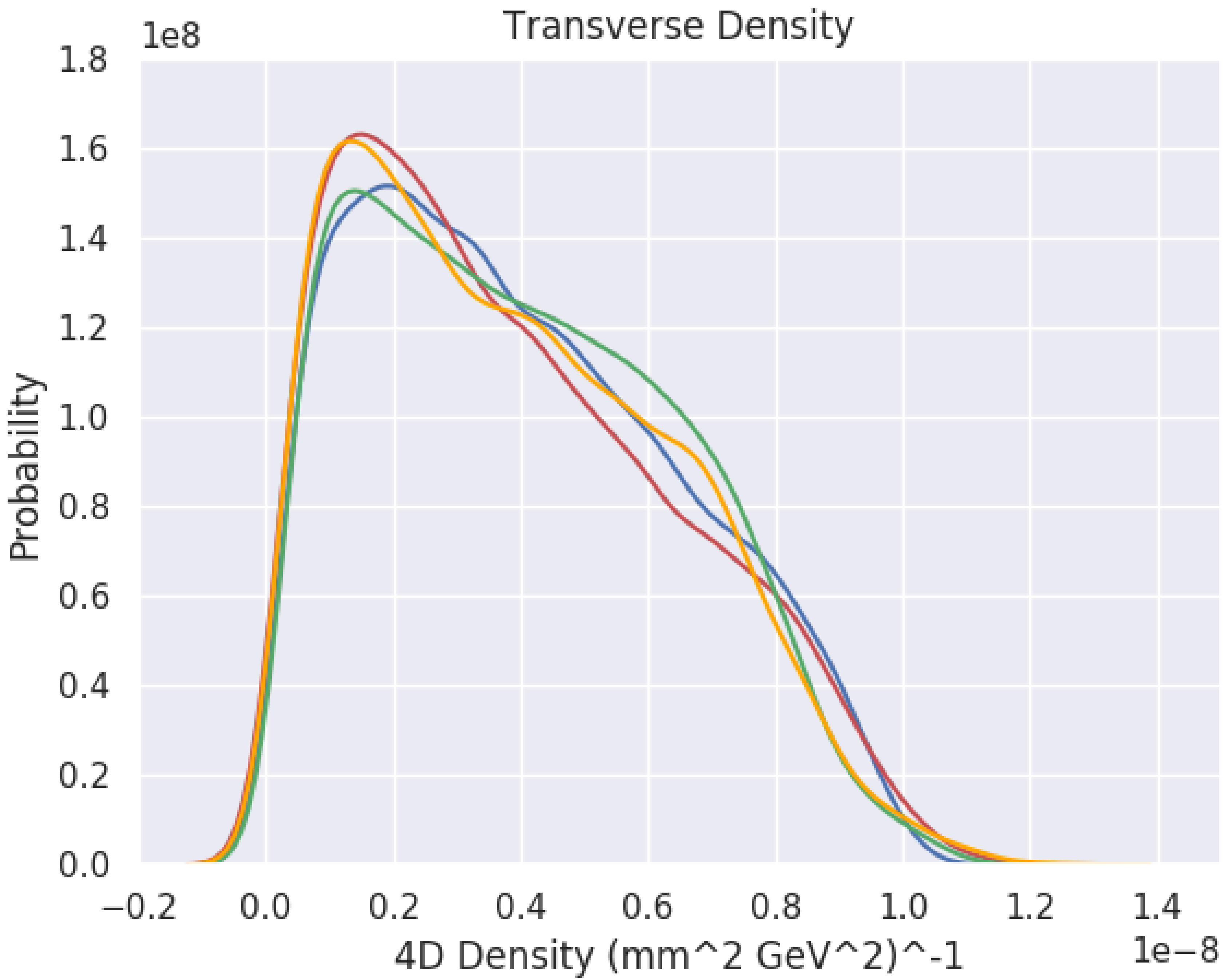
Slight differences due to KDE convolving the density with the kernel, while for KNN it has been smoothed to ensure area of graph is one

No Wedge



Wedge





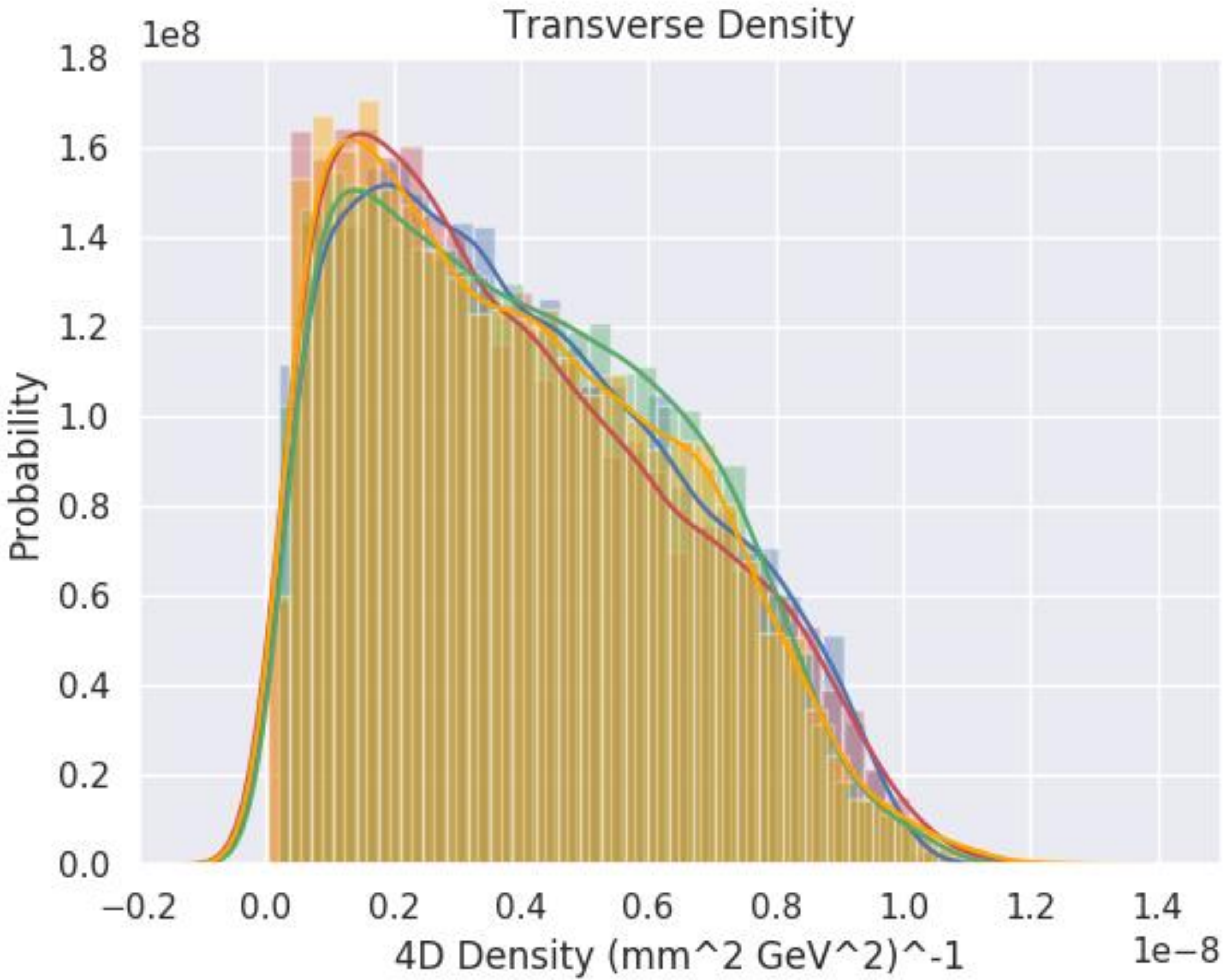
Full Upstream Sample

Blue – KDE – No Wedge
Red – KNN – No Wedge
Green – KDE – Wedge
Yellow – KNN – Wedge

Should be identical bar for any differences in smoothing due to KDE and KNN.

Wedge and No Wedge should have same input beam

Increased sample size may eliminate bumps in mid-density region



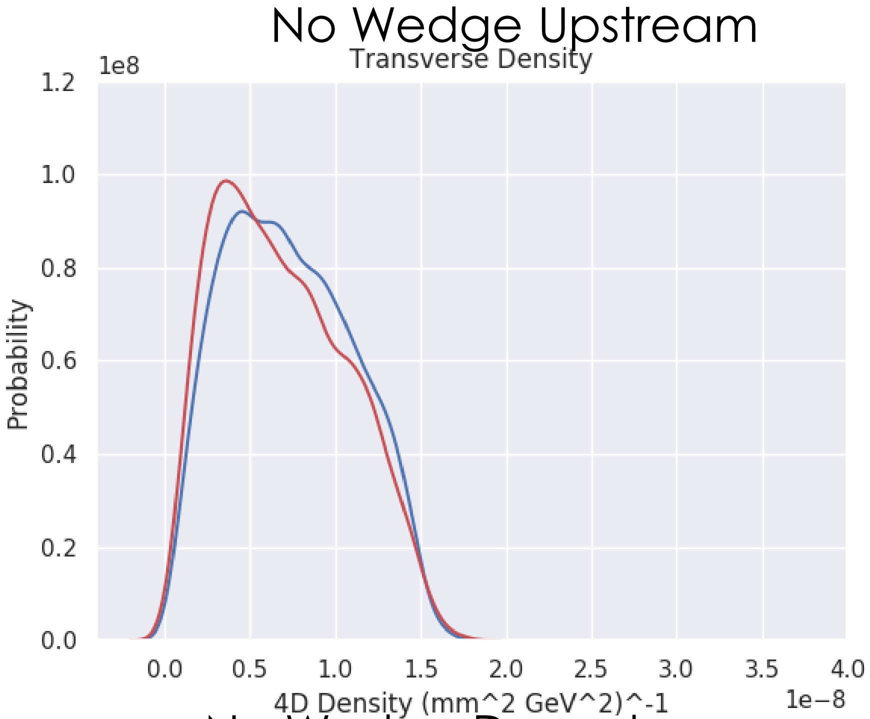
Full Upstream Sample

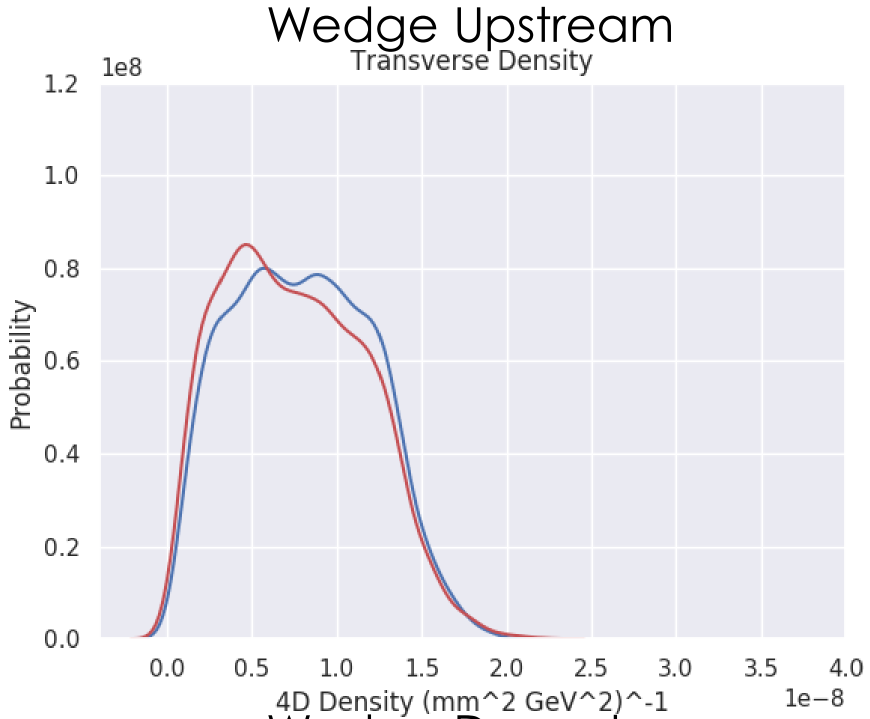
Blue – KDE – No Wedge
Red – KNN – No Wedge
Green – KDE – Wedge
Yellow – KNN – Wedge

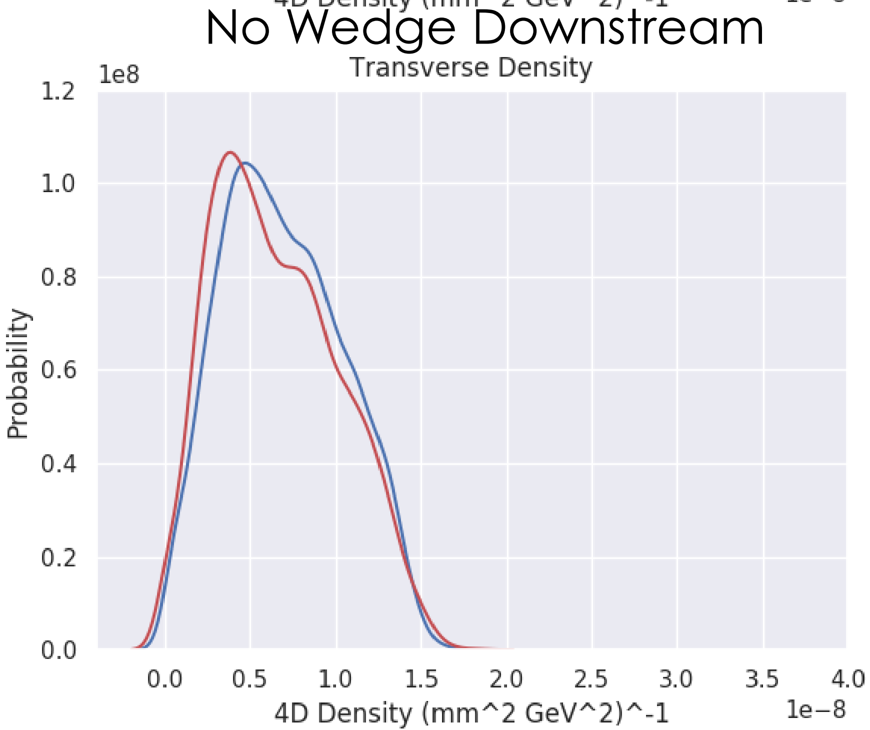
Should be identical bar for any differences in smoothing due to KDE and KNN.

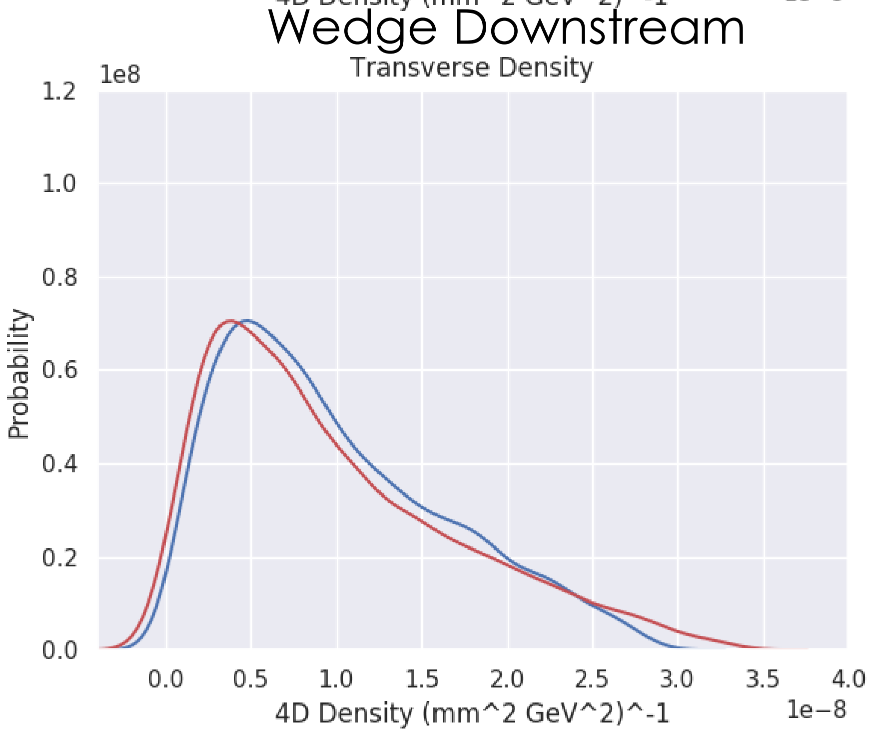
Wedge and No Wedge should have same input beam

Increased sample size may eliminate bumps in mid-density region









Top left and Right: Upstream sample which made it Downstream

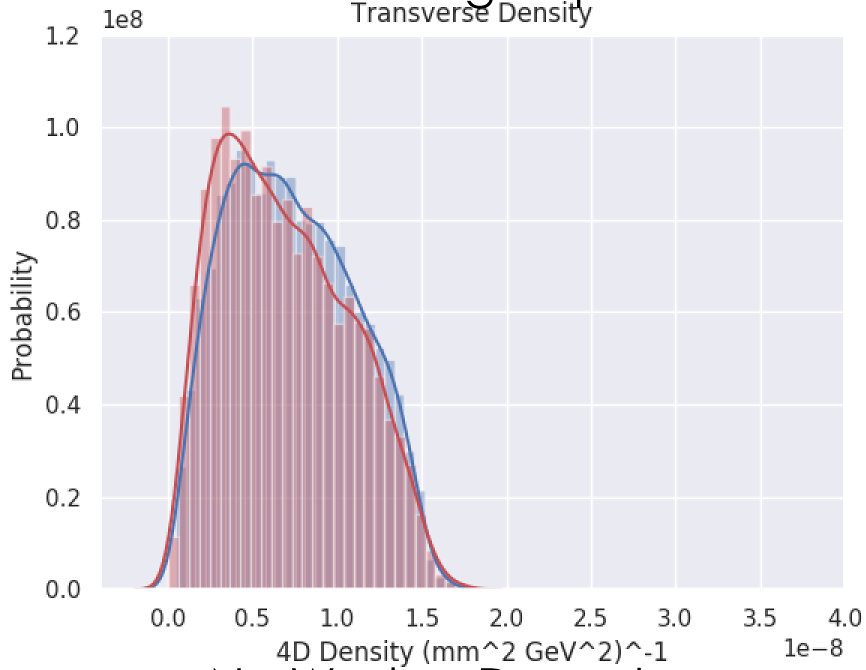
Bottom left and Right: Downstream sample

Blue – KDE
Red – KNN

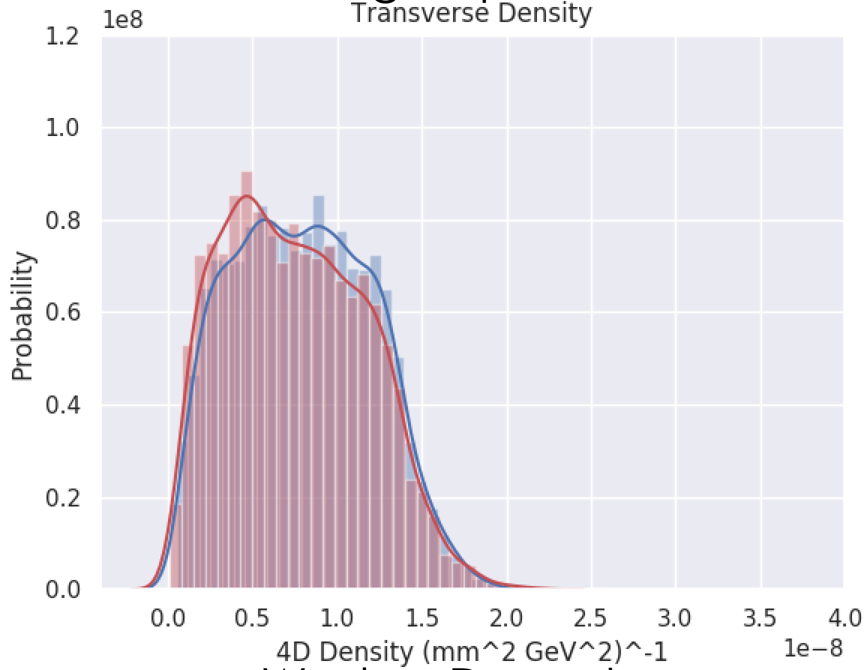
The No Wedge and Wedge Upstream samples are no longer comparable as it has been biased by the Downstream selection

The Upstream to Downstream samples do however show the change in phase space density for that selection

No Wedge Upstream



Wedge Upstream



Top left and Right:
Upstream sample which made it Downstream

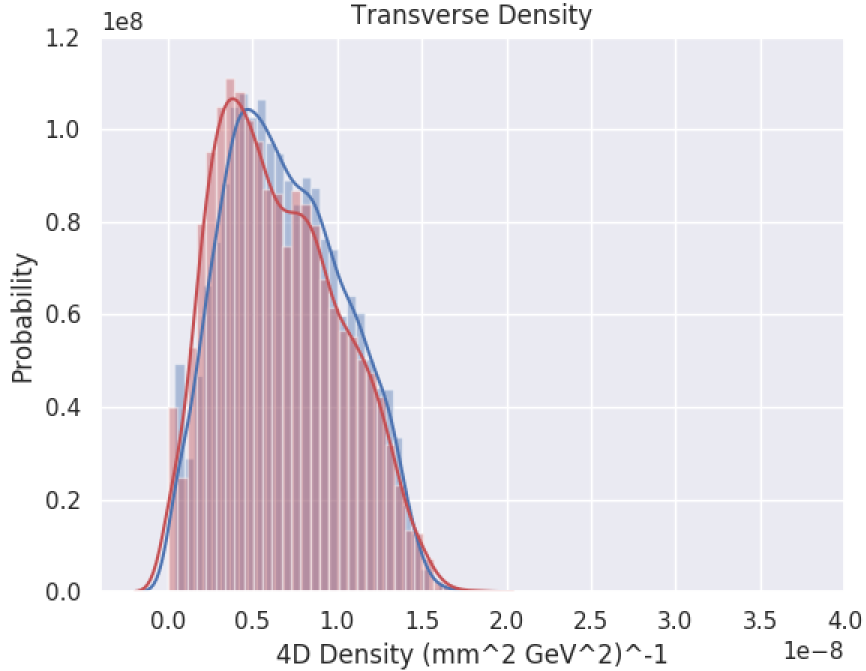
Bottom left and Right:
Downstream sample

Blue – KDE
Red – KNN

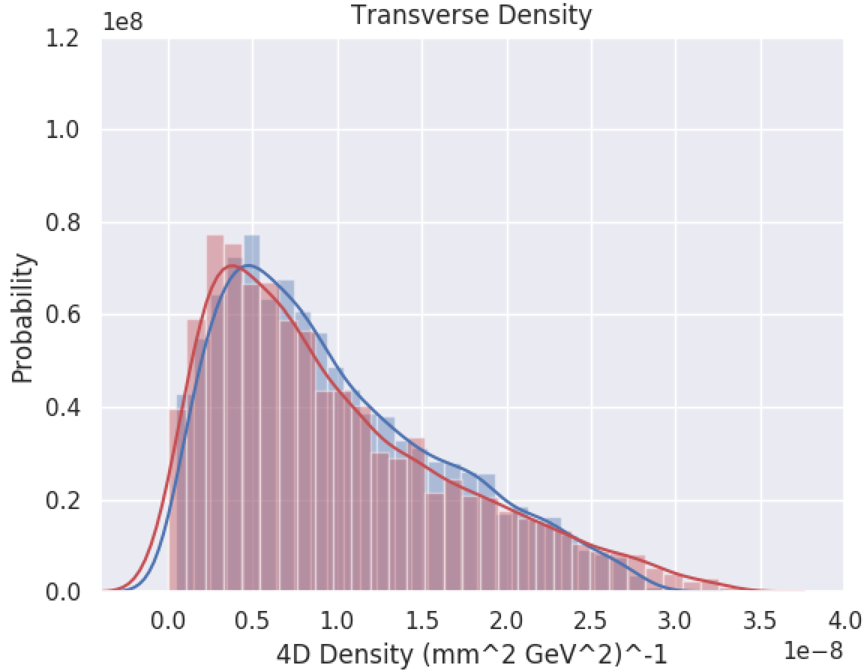
The No Wedge and Wedge Upstream samples are no longer comparable as it has been biased by the Downstream selection

The Upstream to Downstream samples do however show the change in phase space density for that selection

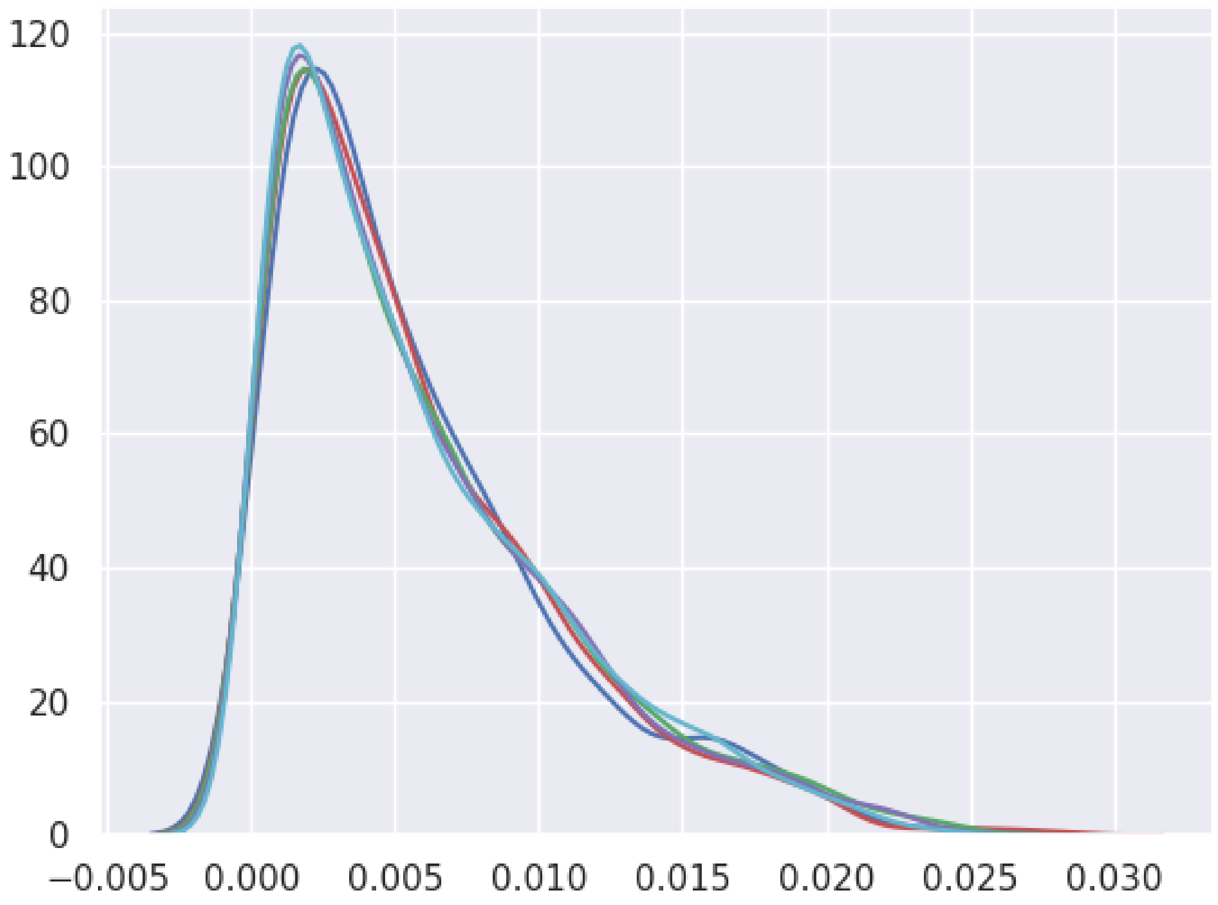
No Wedge Downstream



Wedge Downstream



Change in Sample Size – Toy Scenario



- See effect of change in sample size, as sample size increases, should approach underlying density of sample
- Random 4-D distribution with mean = 0, Standard Deviation = diag(1,1,1,1)

Blue: n =1000, k = 31

Red: n = 2000, k = 44

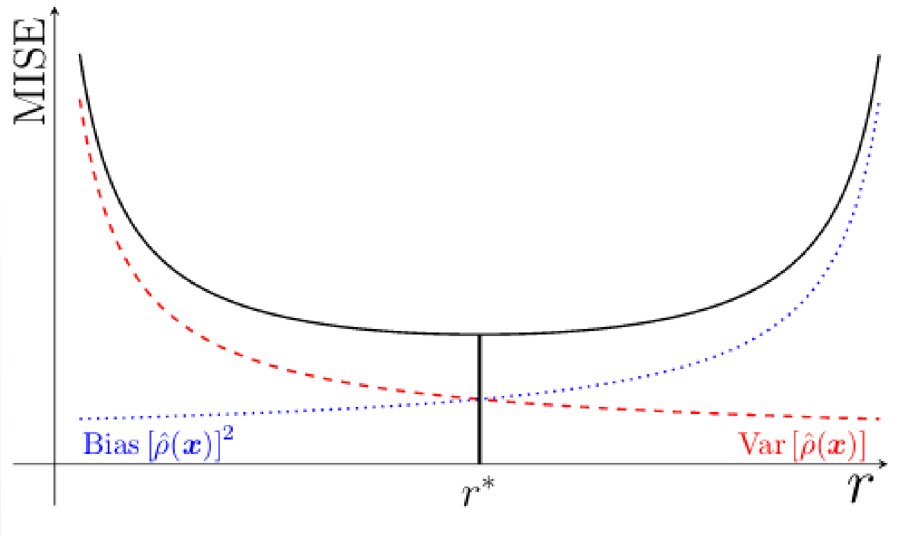
Green: n = 3000, k = 54

Magenta: n = 4000, k = 63

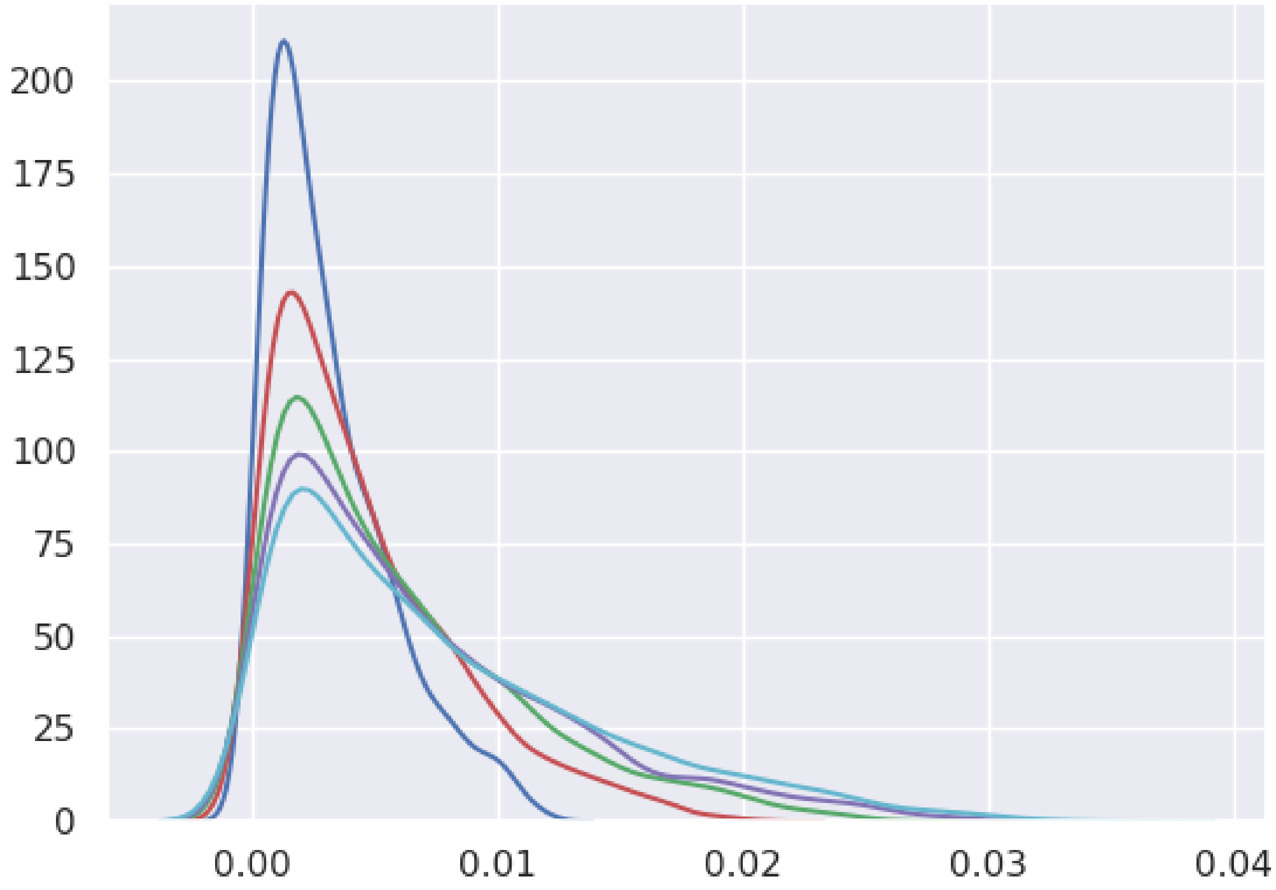
Cyan: n = 5000, k = 70

- Underlying sample density is approached as sample size increases, optimal k adjusts to reflect increase in sample size

Change in sample size, same k – Toy Scenario



- Changing sample size but keeping k constant increases MISE, as a suboptimal k is chosen
- D-dimensional radius for a test point increases/decreases as the test point needs to find more/less neighbours. This can give an apparent decrease/increase in the phase space density. As the sample size is increased, the phase space density becomes less susceptible to small changes in optimal k



Blue: n =1000, k = 54

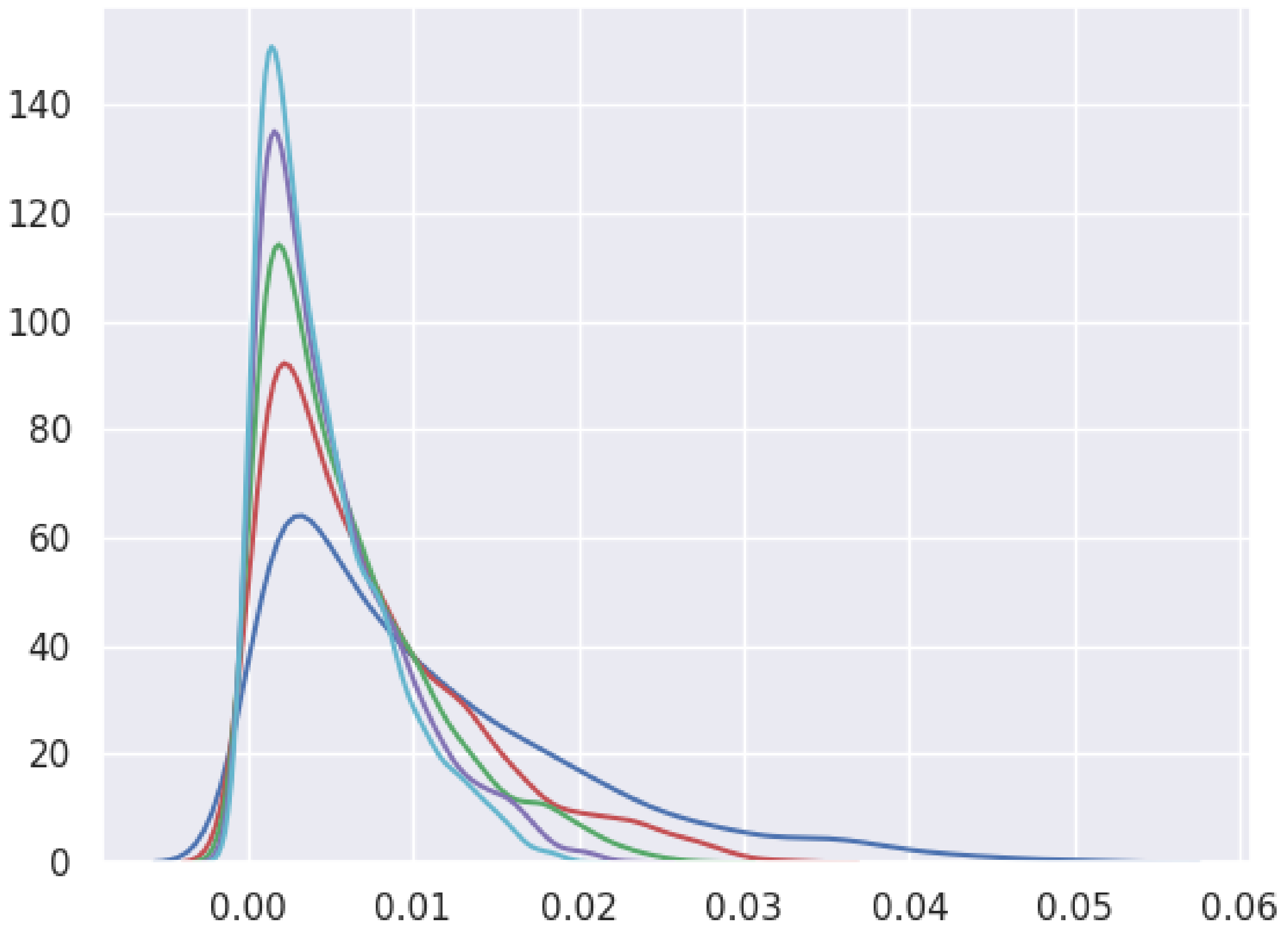
Red: n = 2000, k = 54

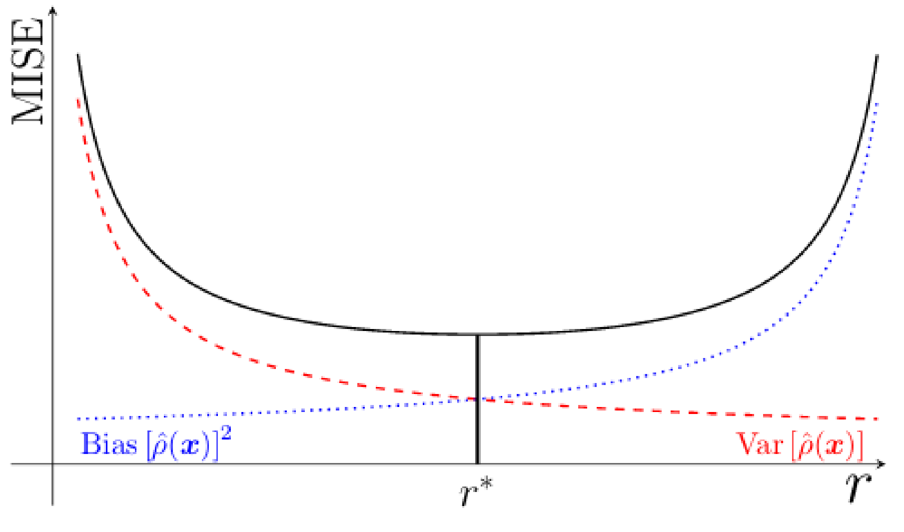
Green: n = 3000, k = 54

Magenta: n = 4000, k = 54

Cyan: n = 5000, k = 54

Change in k, same sample size – Toy scenario





- Choosing a suboptimal k leads to an increase in MISE
- When comparing data samples, one needs to use the same conditions for the sample i.e. use the same k to n relation e.g. $k{\sim}n^{-4/(4+d)}$
- A MISE that may not have been minimized may be desirable in areas that have been over or under smoothed

Blue: n =3000, k = 31

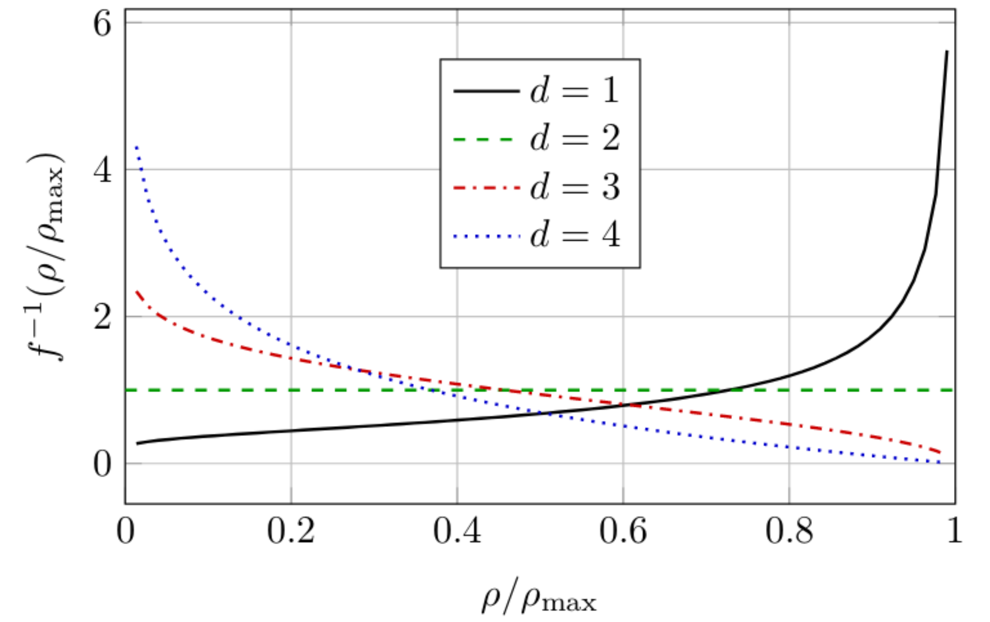
Red: n = 3000, k = 44

Green: n = 3000, k = 54

Magenta: n = 3000, k = 63

Cyan: n = 3000, k = 70

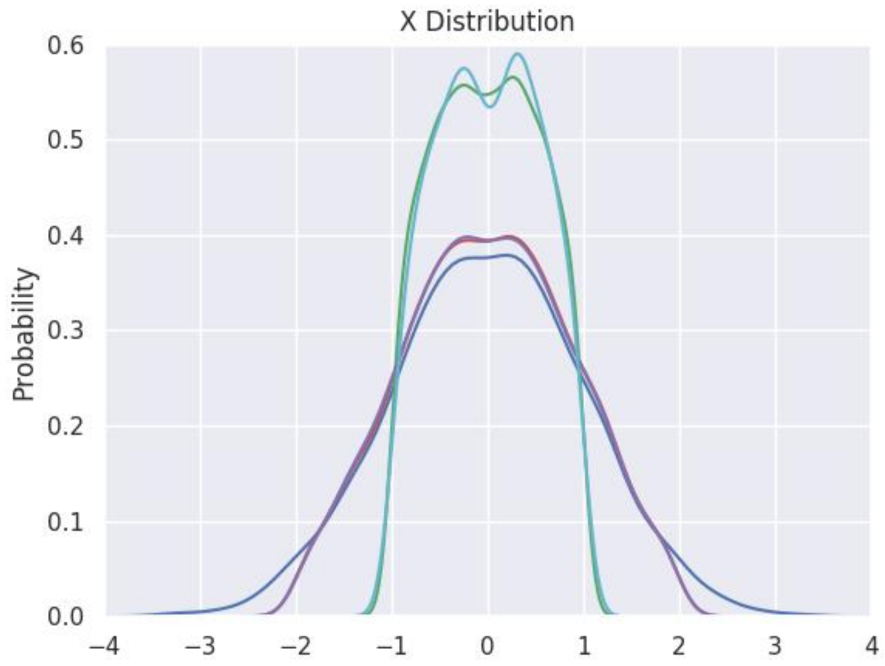
Missing Data - Toy Scenario
Scraping and Transmission Losses

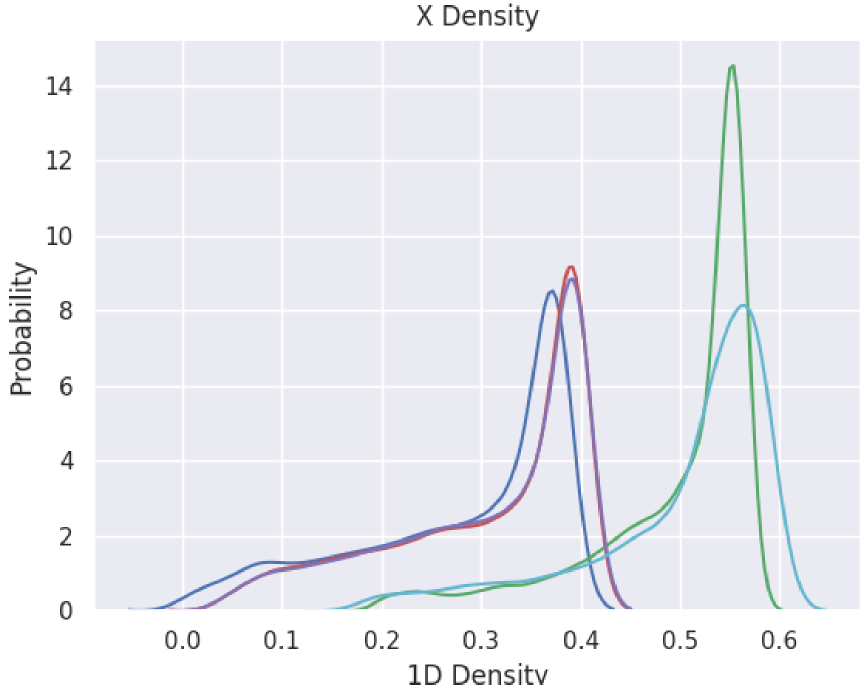


- Left – Expected Density for a Gaussian sample in each dimension normalized to the maximum density. As the dimension increases, particles more likely be found at a low phase space density

Toy example (next slides):

- 4D Gaussian sample – Mean = 0, Standard Deviation = diag(1,1,1,1)
- Full sample – No cuts – Blue
- Cut at +/- 2 sigma in one dimension called 'X' – red
- Cut at +/- 1 sigma in one dimension called 'X' – green
- Cut at +/- 2 sigma in each dimension – magenta
- Cut at +/- 1 sigma in each dimension – cyan





Full sample – No cuts – Blue

Cut at +/- 2 sigma in one dimension called 'X' – red

Cut at +/- 1 sigma in one dimension called 'X' – green

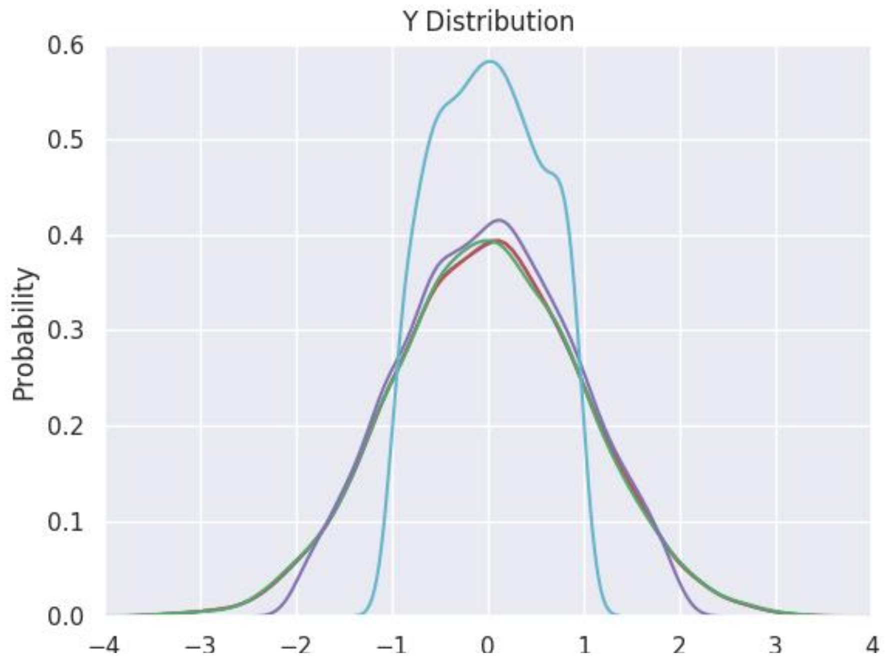
Cut at +/- 2 sigma in each dimension – magenta

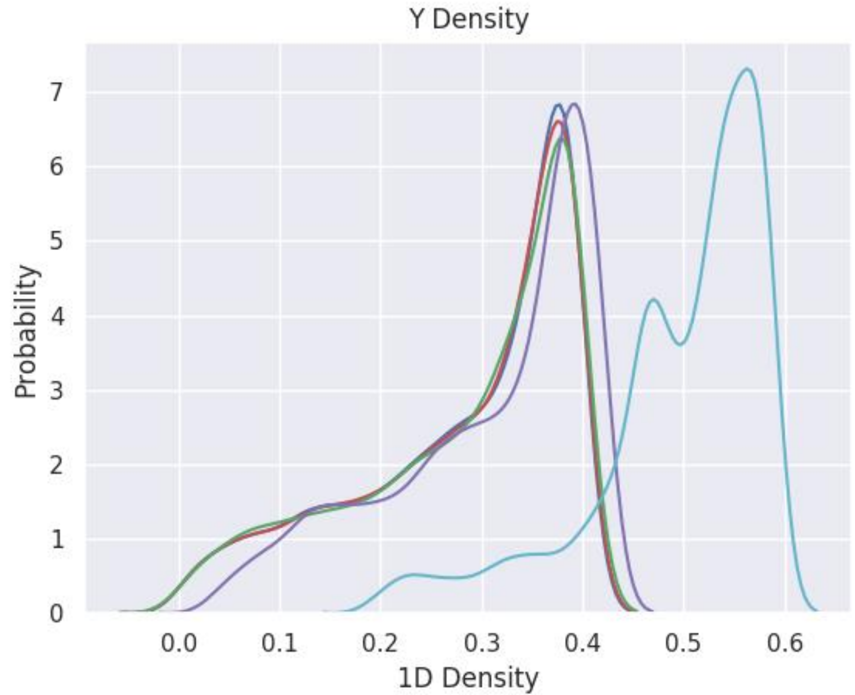
Cut at +/- 1 sigma in each dimension – cyan

2 sigma cut causes ~5% cut in 1D and ~17% in 4D which alters the density and distribution only slightly

1 sigma cut causes far greater change (~32% cut in 1D, ~79% cut in 4D)

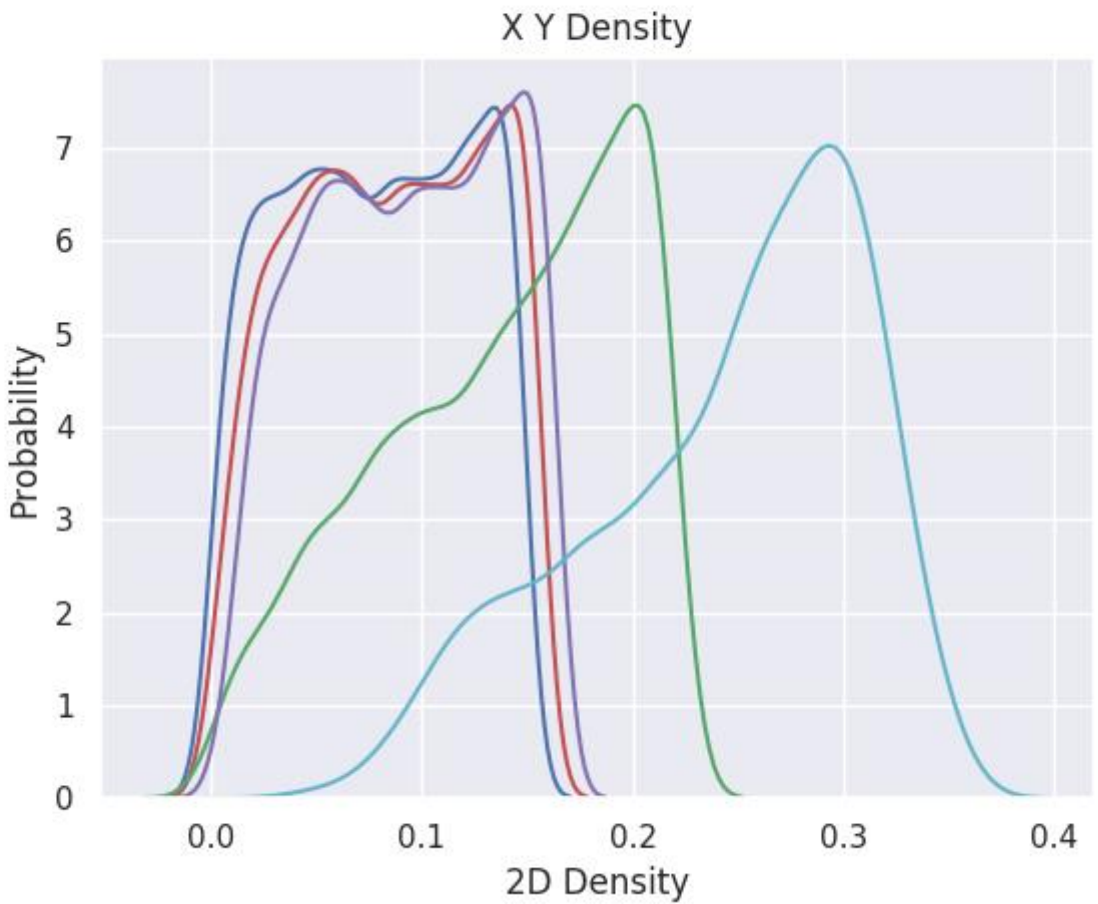
The k value is related to n, if the distribution is denser, than the calculated density will also be denser

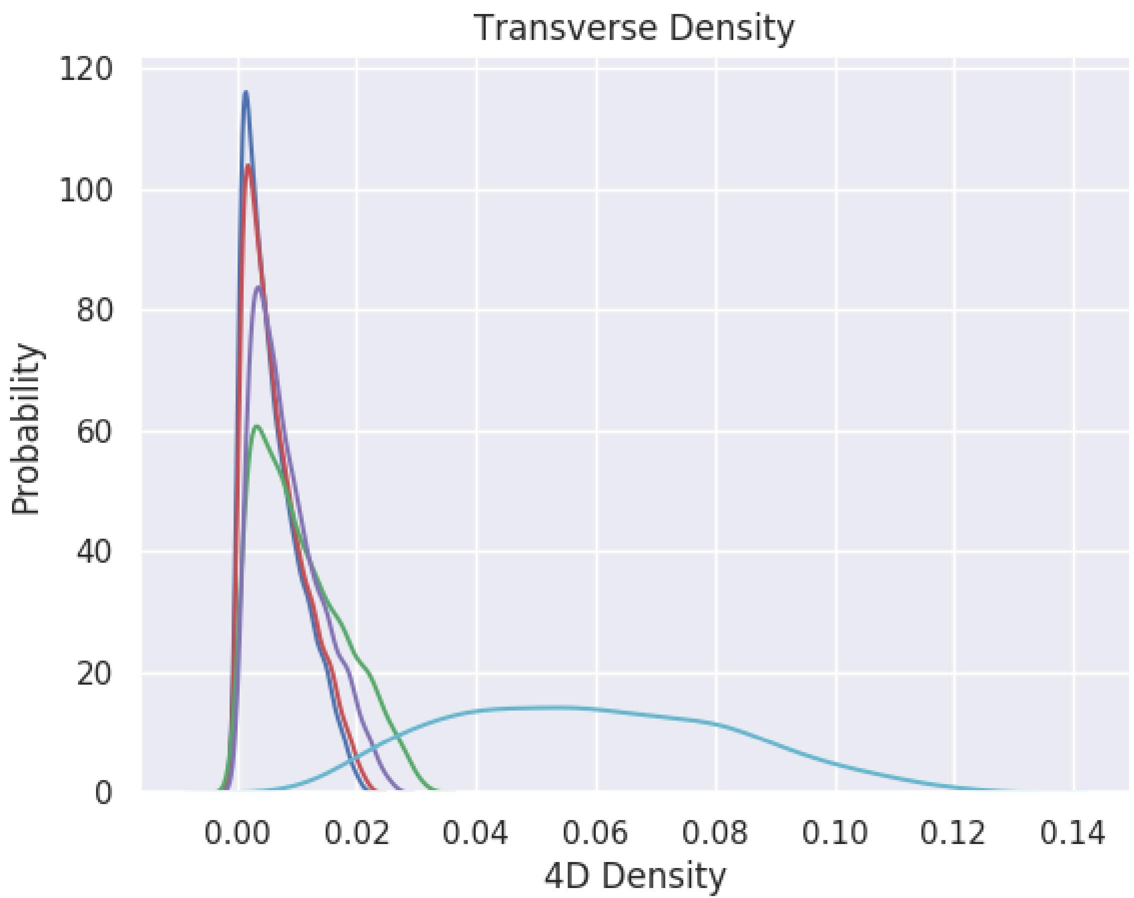




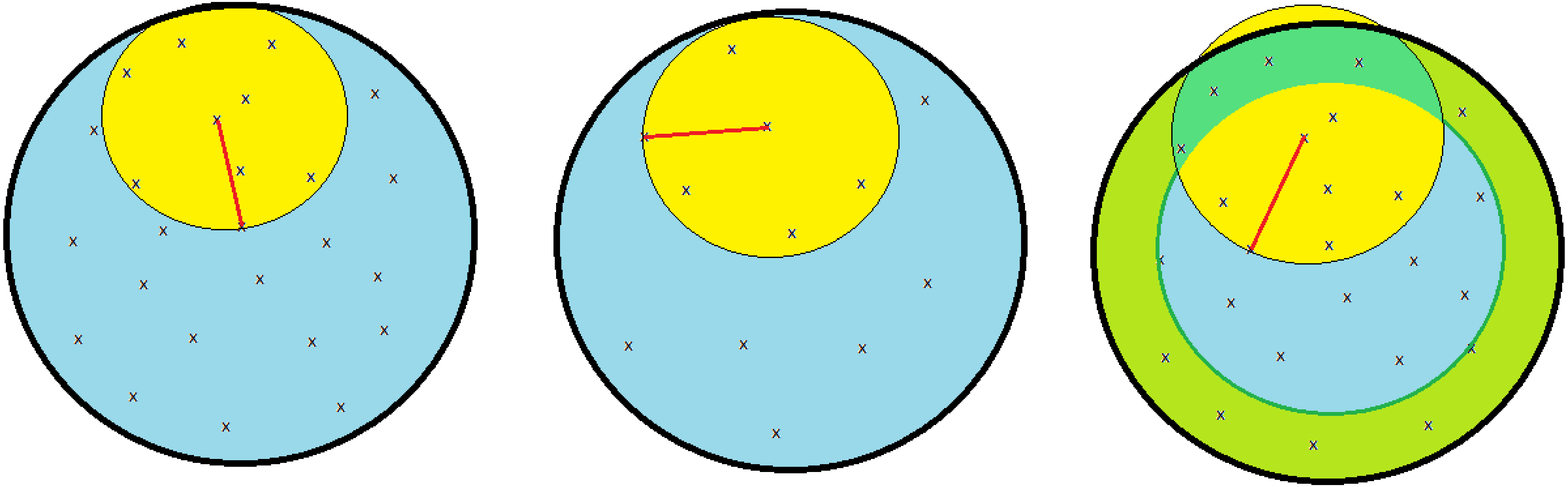
Missing Data Toy Scenario

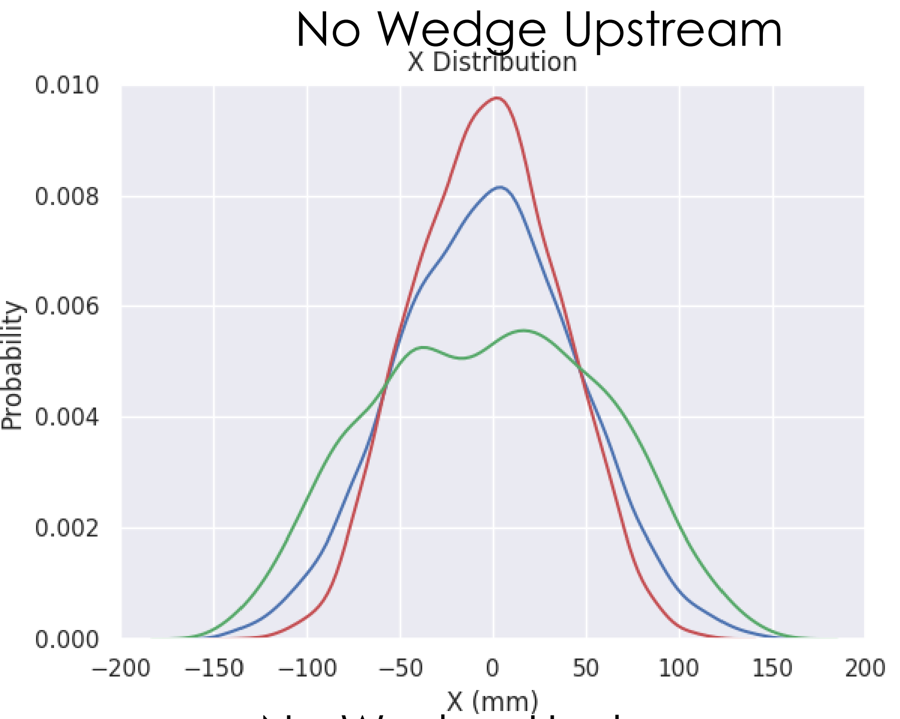
Full sample – No cuts – Blue
Cut at +/- 2 sigma in one dimension called 'X' – red
Cut at +/- 1 sigma in one dimension called 'X' – green
Cut at +/- 2 sigma in each dimension – magenta
Cut at +/- 1 sigma in each dimension – cyan

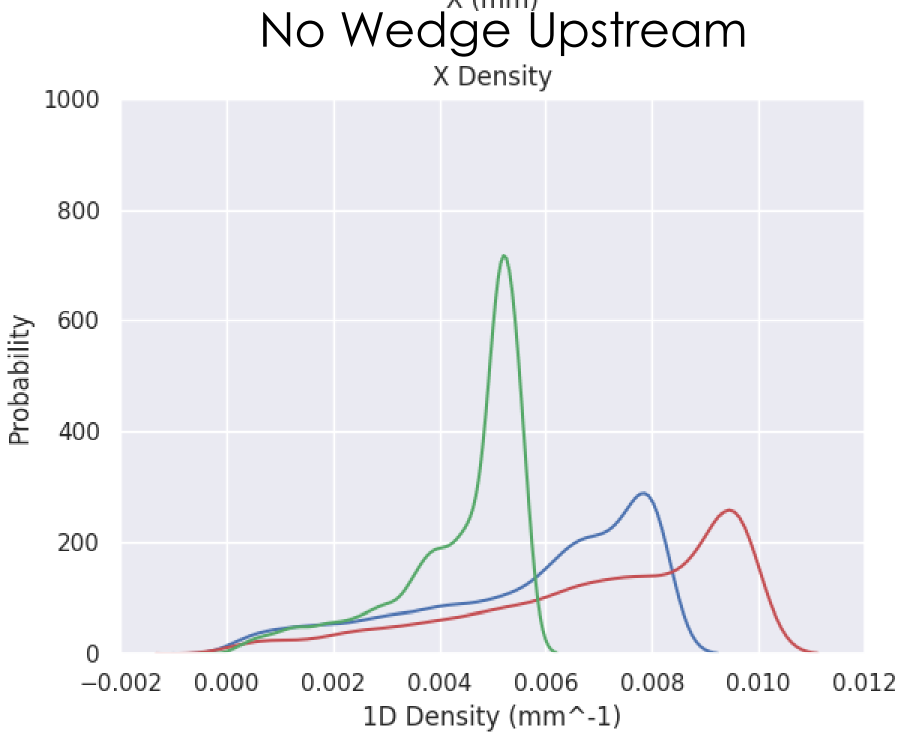




- Left - original sample – red line shows k-nearest neighbour for point at centre of yellow circle
- Middle - subsample from original – red line distance only has slight change, k adjusts according to n. As n becomes small the error increases
- Right – Aperture cut by the green sub-circle – points at large radius are removed. While n has reduced, the k is now ideal for the subsample distribution.
- For points with a bounding circle affected by the aperture cut, the k-nearest neighbour may be further away, while for points at the centre of the sample the nearest neighbour is closer as the k is reduced, but no close points are removed







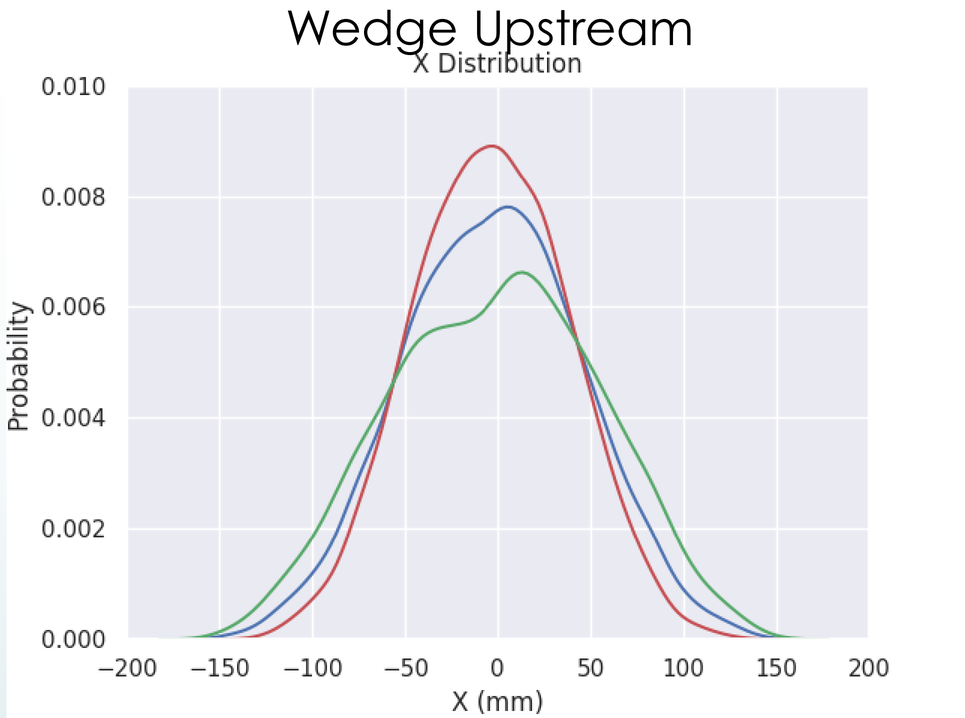
No Wedge (left) and Wedge (right)

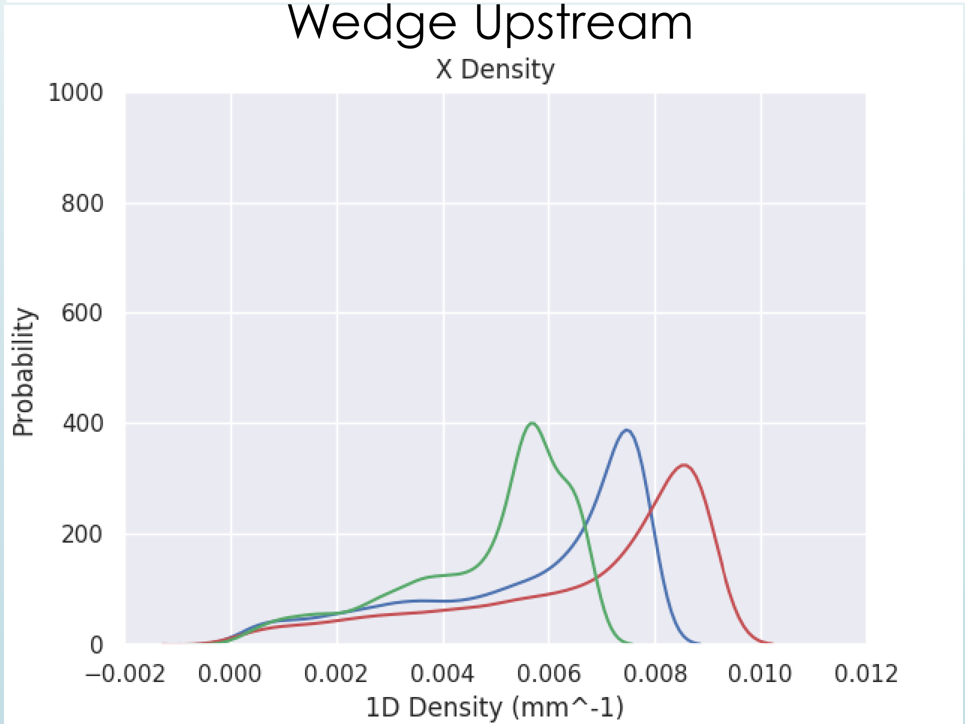
X Distribution (Top) and Density (Bottom)

Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

Small preference for larger magnitude x not to make it downstream

Wedge case shows slight directional bias as well. The Wedge does not transmit up to 15% of particles that would have made it downstream otherwise.

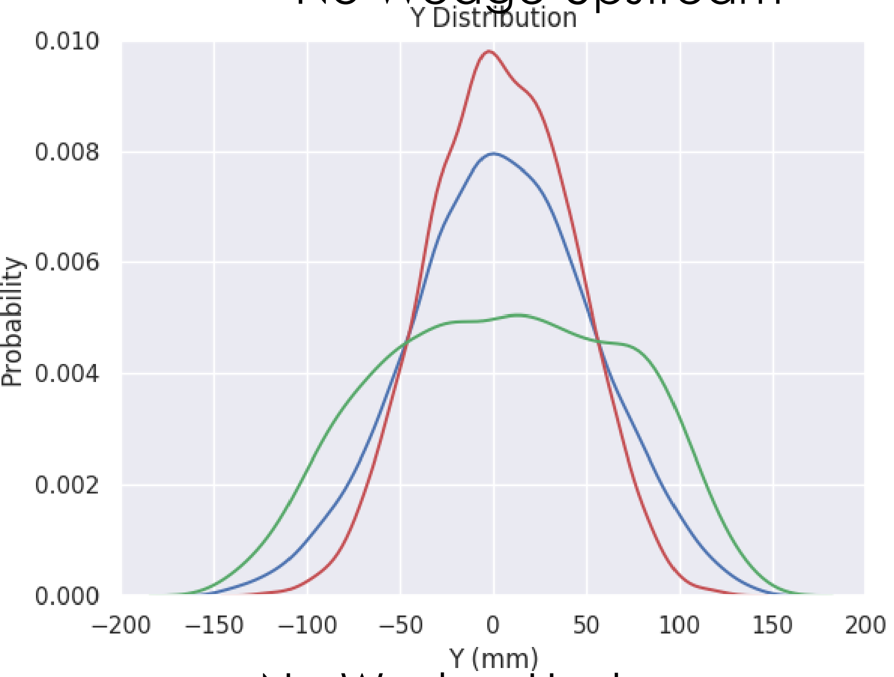




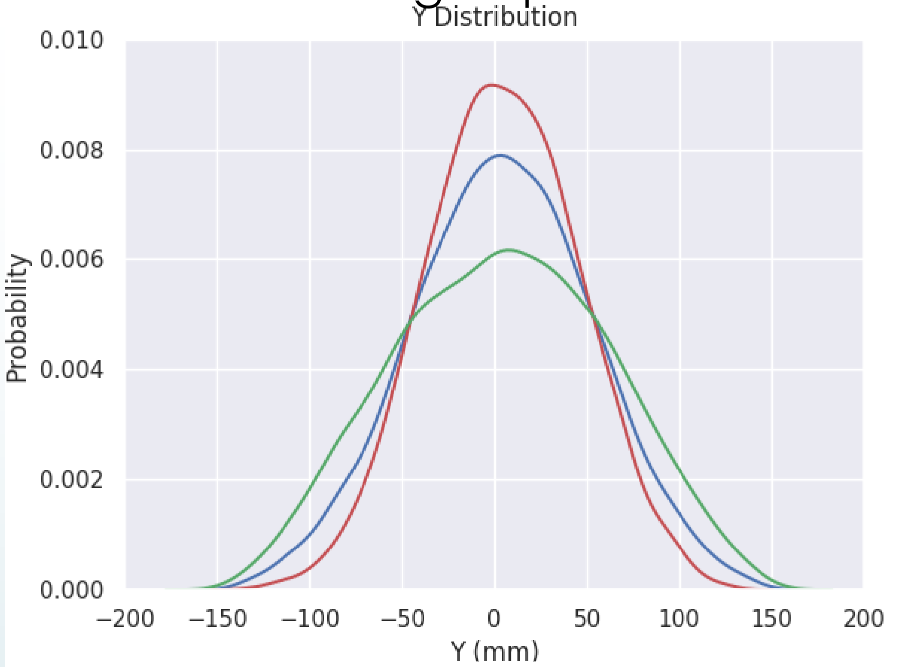
No Wedge (left) and Wedge (right)

Y Distribution (Top) and Density (Bottom)

No Wedge Upstream

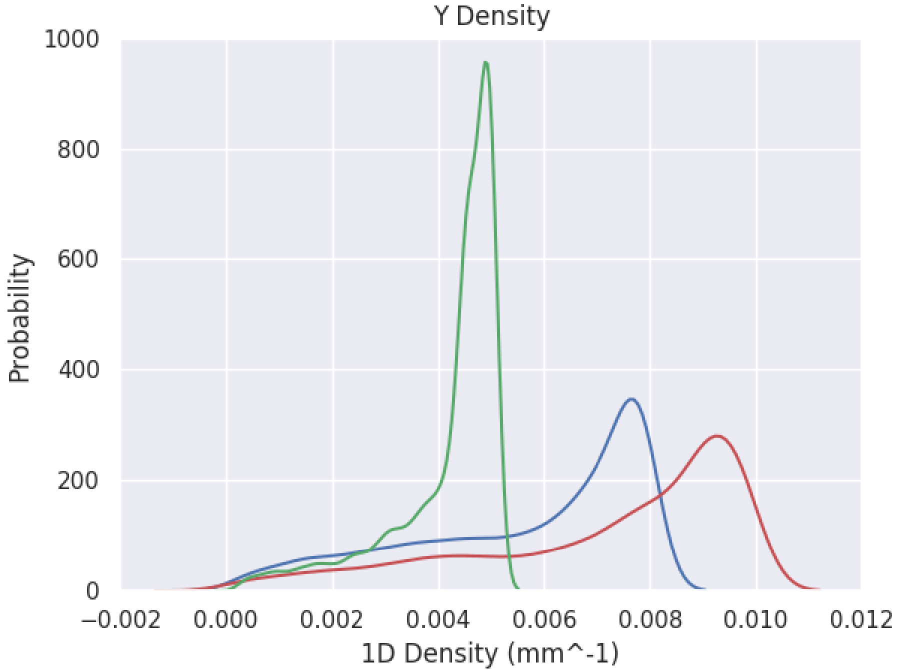


Wedge Upstream

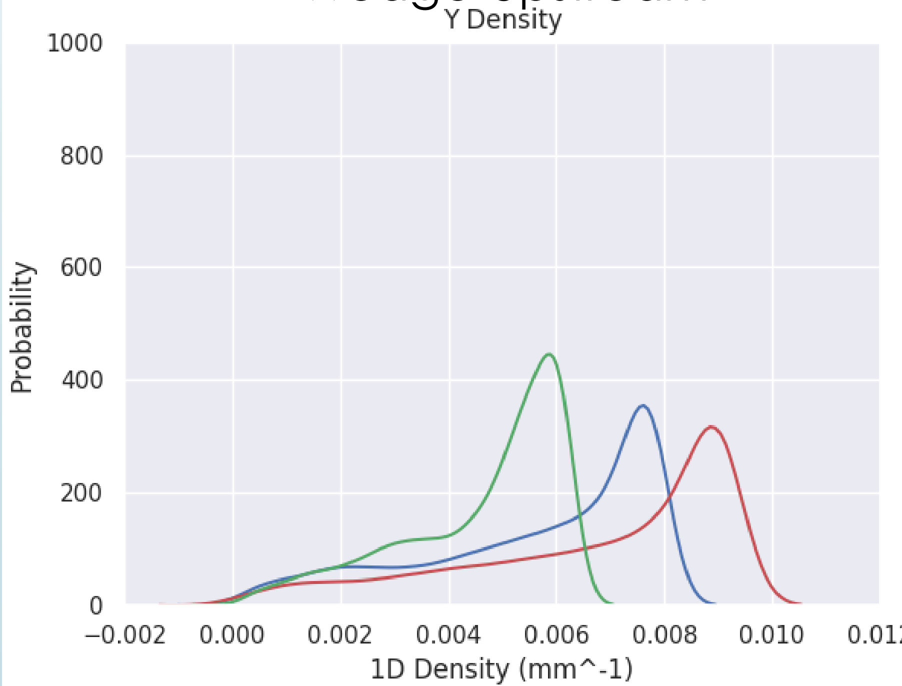


Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

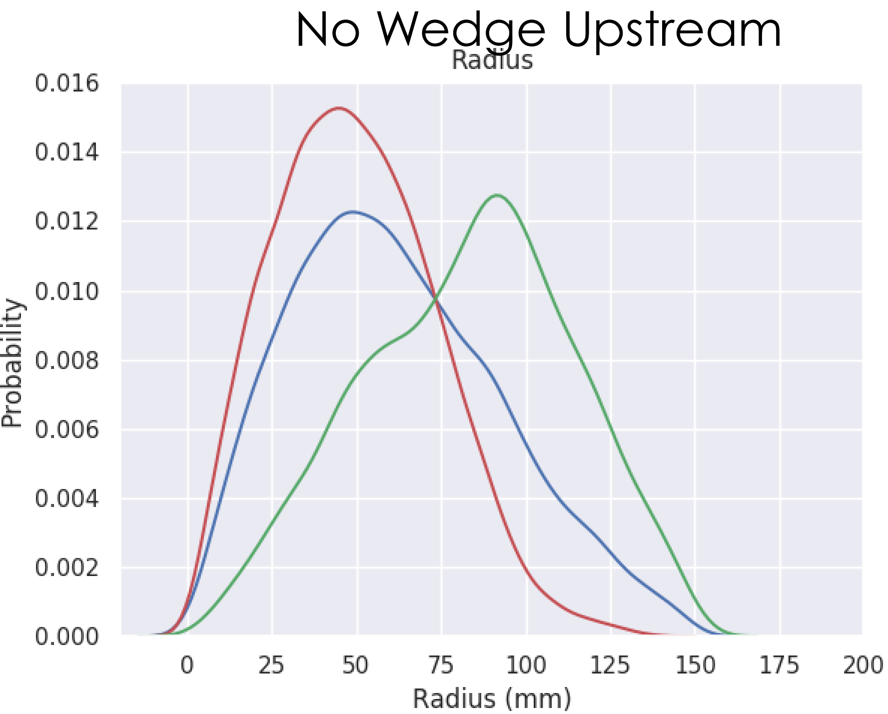
No Wedge Upstream

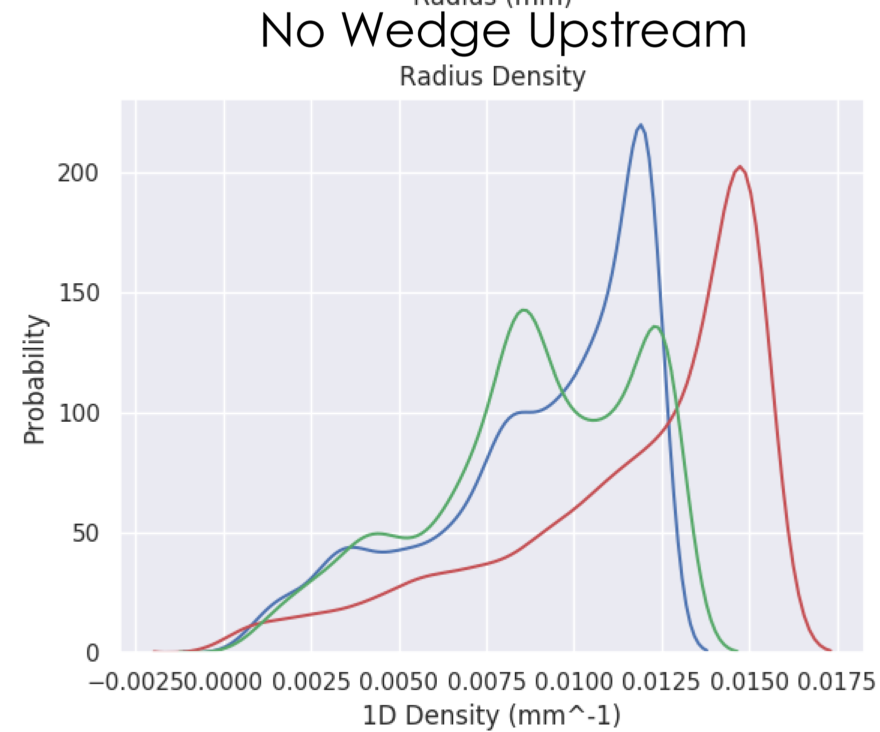


Wedge Upstream



The Wedge counteracts some of the aperture cut effects, so that both low and high density particles do not make it downstream. This results in more similar distributions, however it is direction dependent.





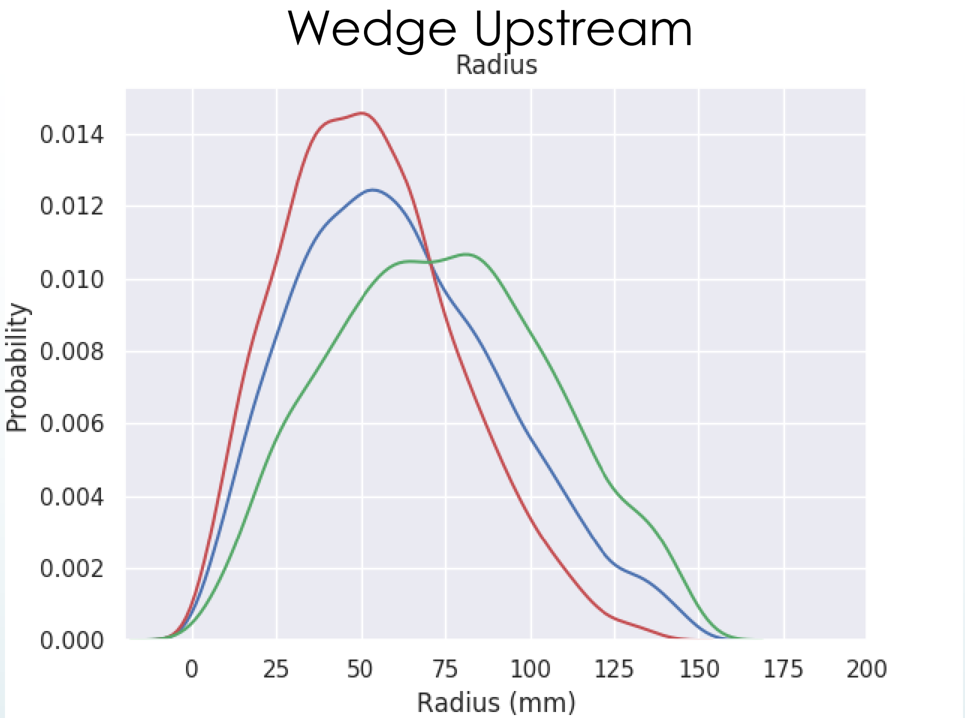
No Wedge (left) and Wedge (right)

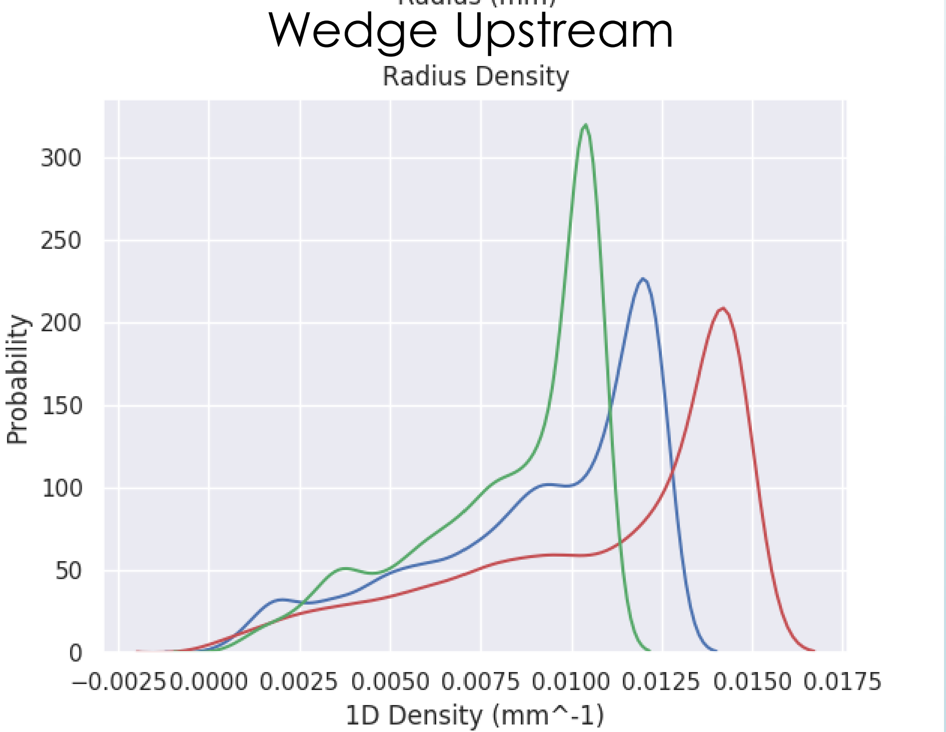
Radius Distribution (Top) and Density (Bottom)

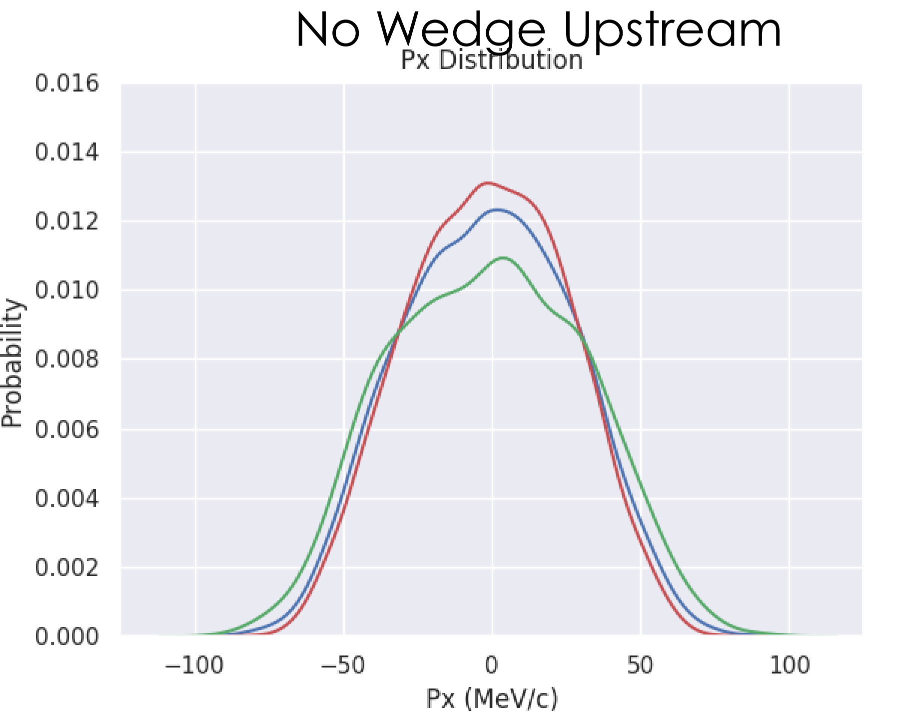
Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

Not only high radius particles are eliminated. It is more likely for low to mid radius particles to be eliminated as there are simply more of them.

The double peak is due to the triangular shape of the distribution.



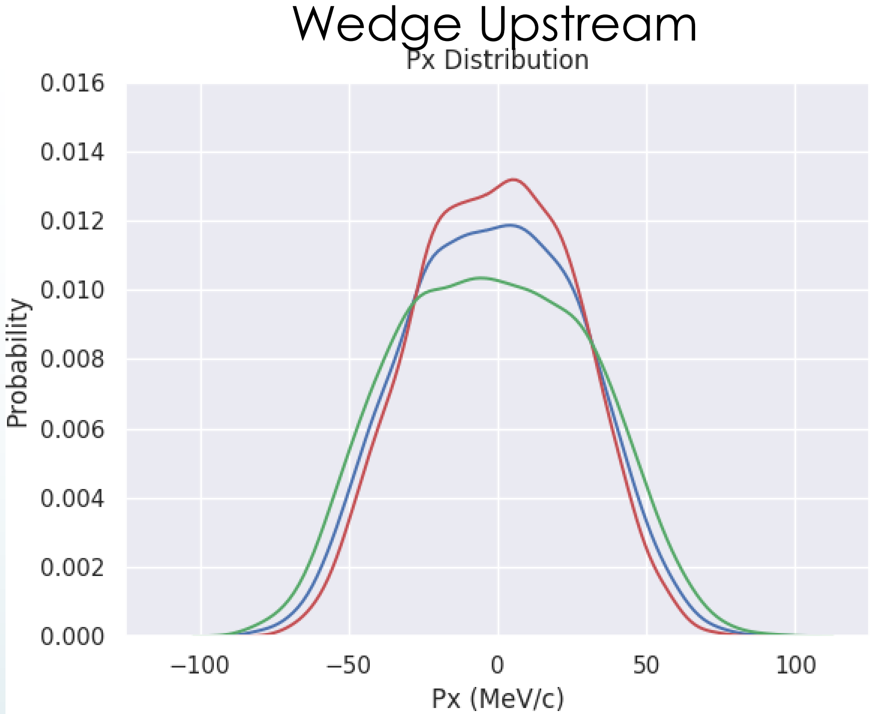


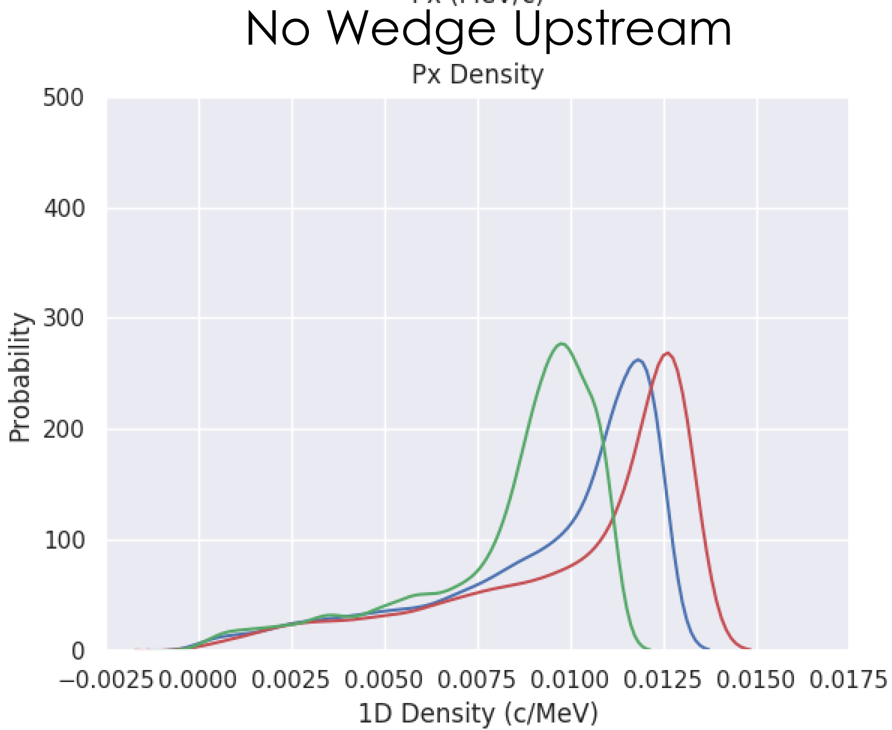


No Wedge (left) and Wedge (right)

Px Distribution (Top) and Density (Bottom)

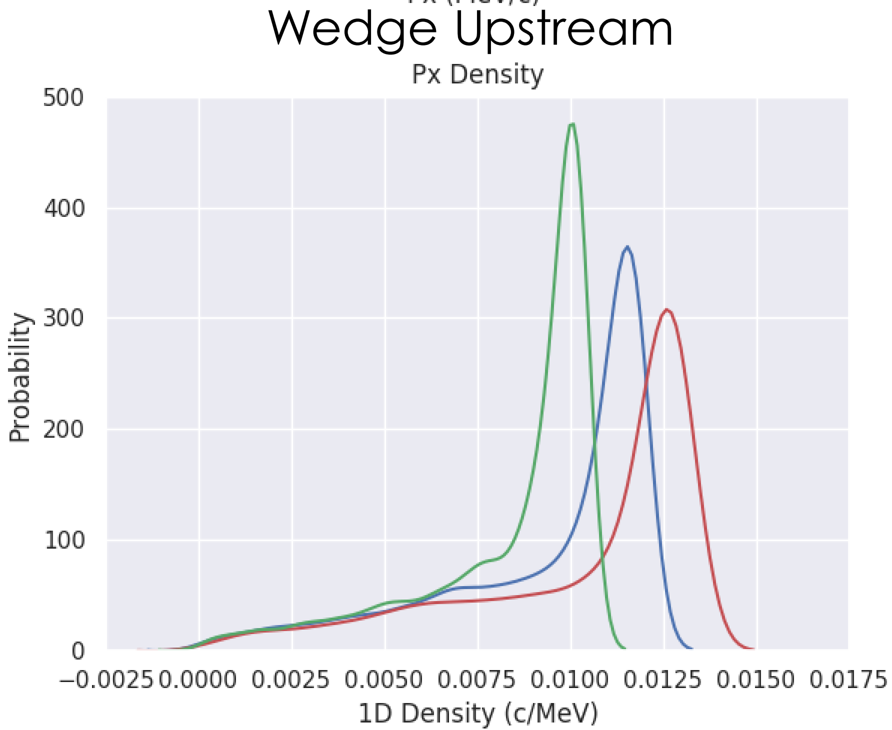
Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

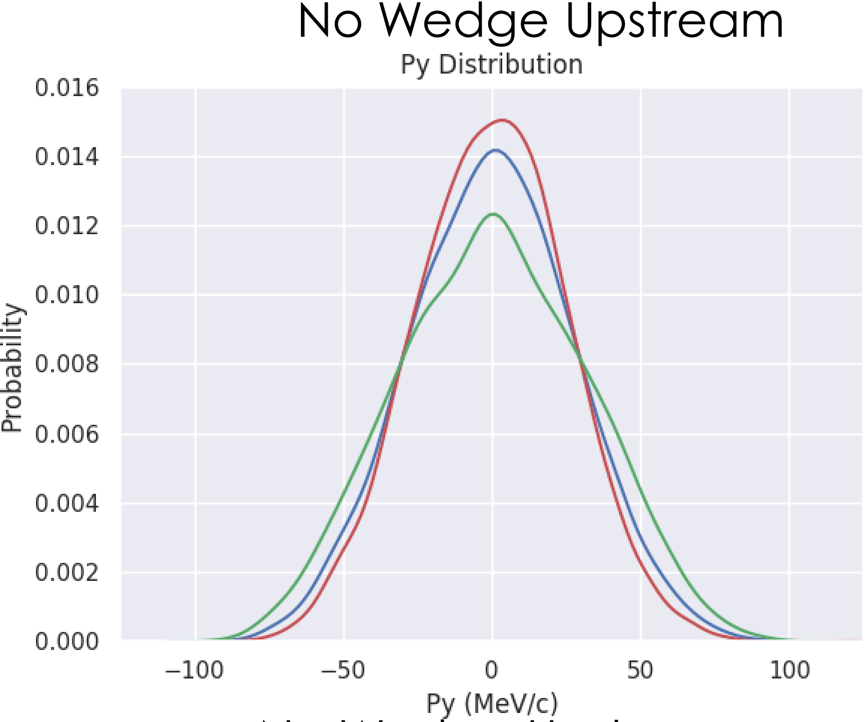




The Px and Py data are less affected by the aperture cut than the radius.

Px of higher density are more likely to be affected by the wedge than in the no wedge case



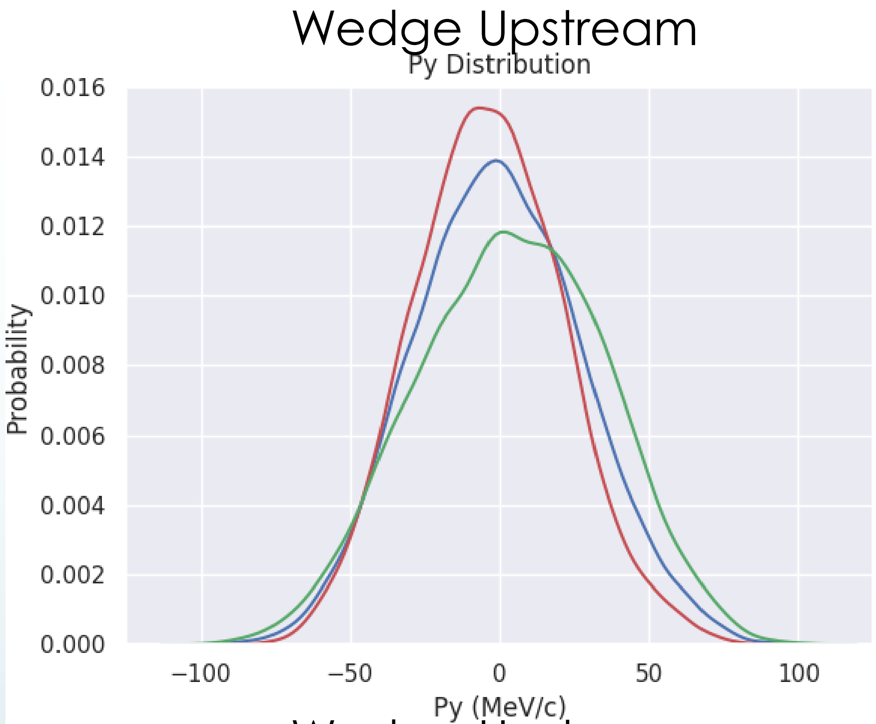


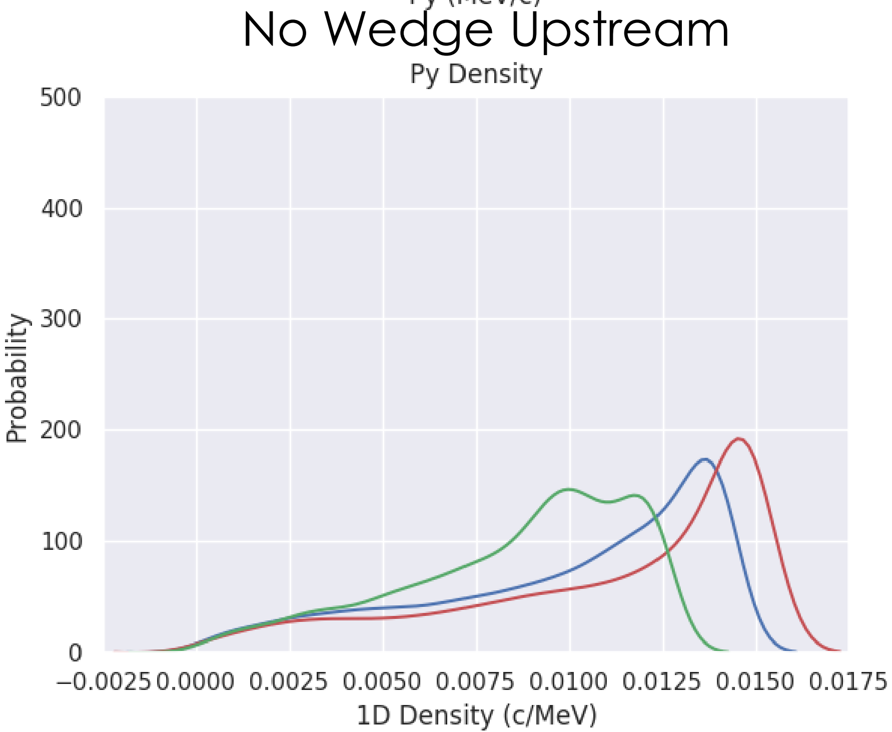
No Wedge (left) and Wedge (right)

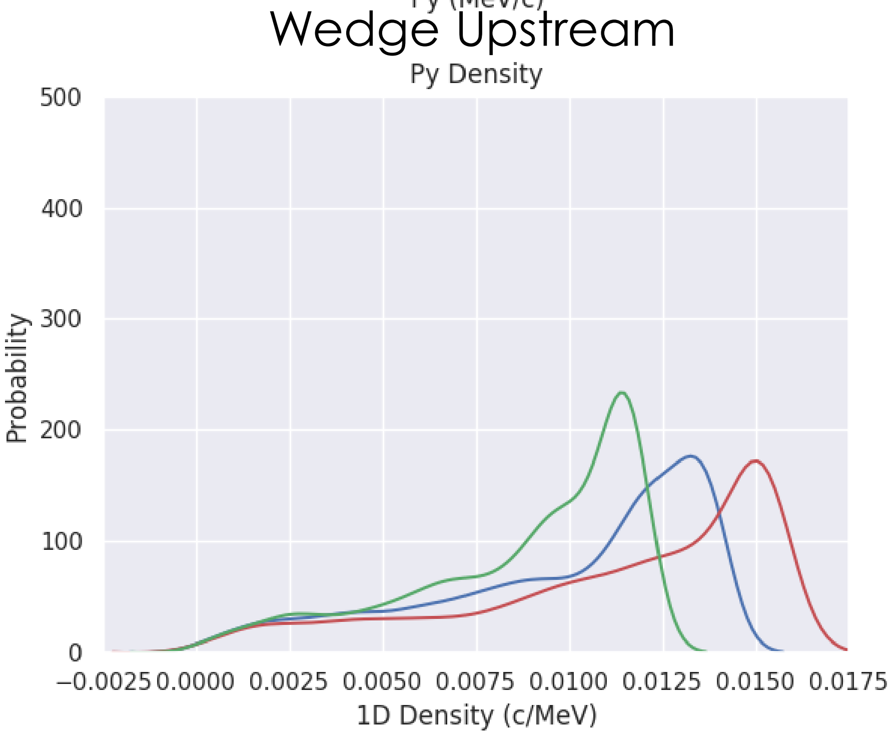
Py Distribution (Top) and Density (Bottom)

Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

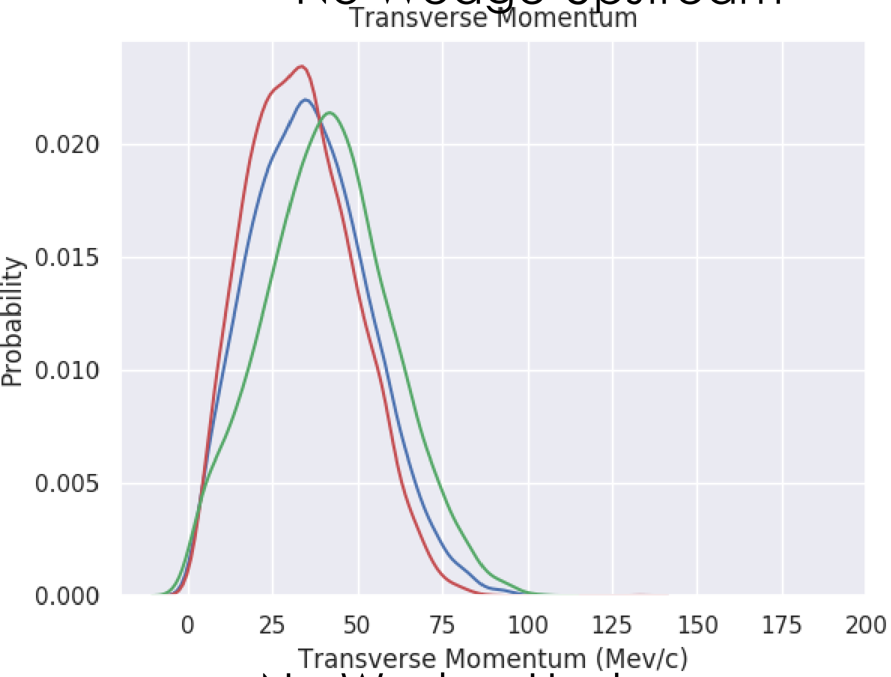
The Py distribution shows a directional preference for particles that don't make it downstream. This is due to the x-py correlation



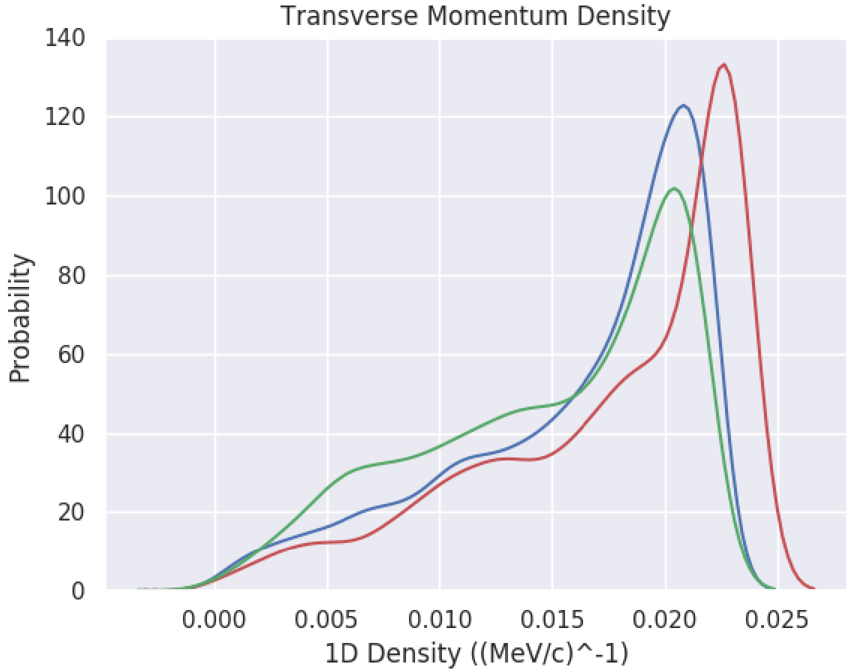




No Wedge Upstream



No Wedge Upstream



No Wedge (left) and Wedge (right)

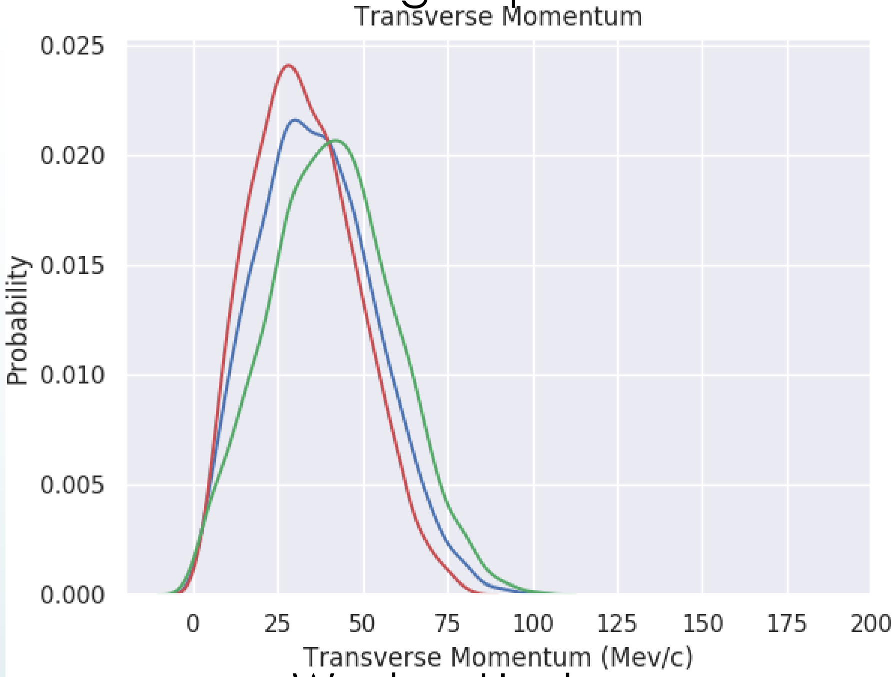
Pt Distribution (Top) and Density (Bottom)

Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

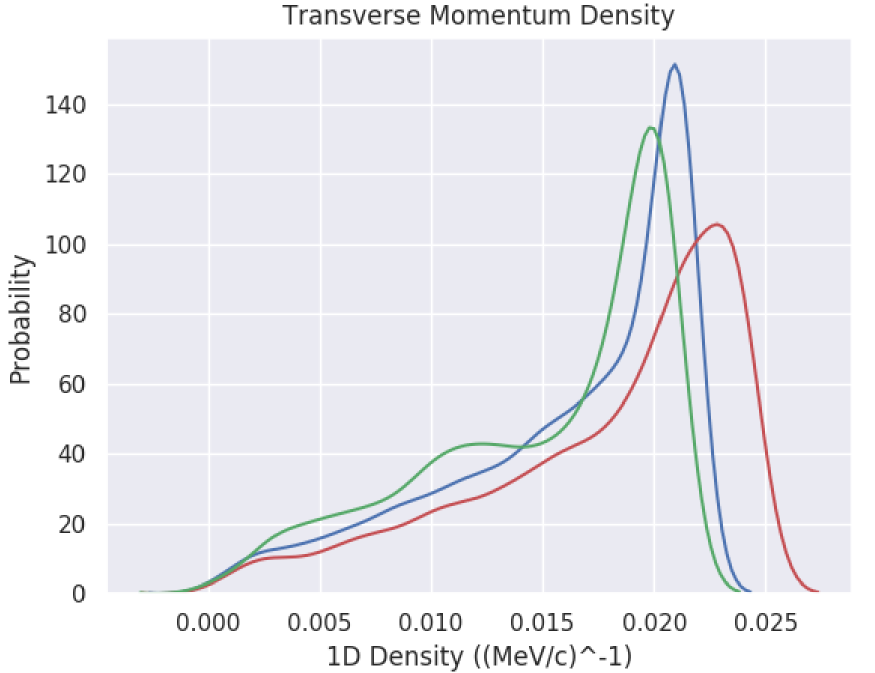
Higher Transverse momenta are less likely to make it downstream, but do not show the same distribution shape as for radius

This results in the upstream and downstream samples being affected more in two of the four dimensions.

Wedge Upstream



Wedge Upstream

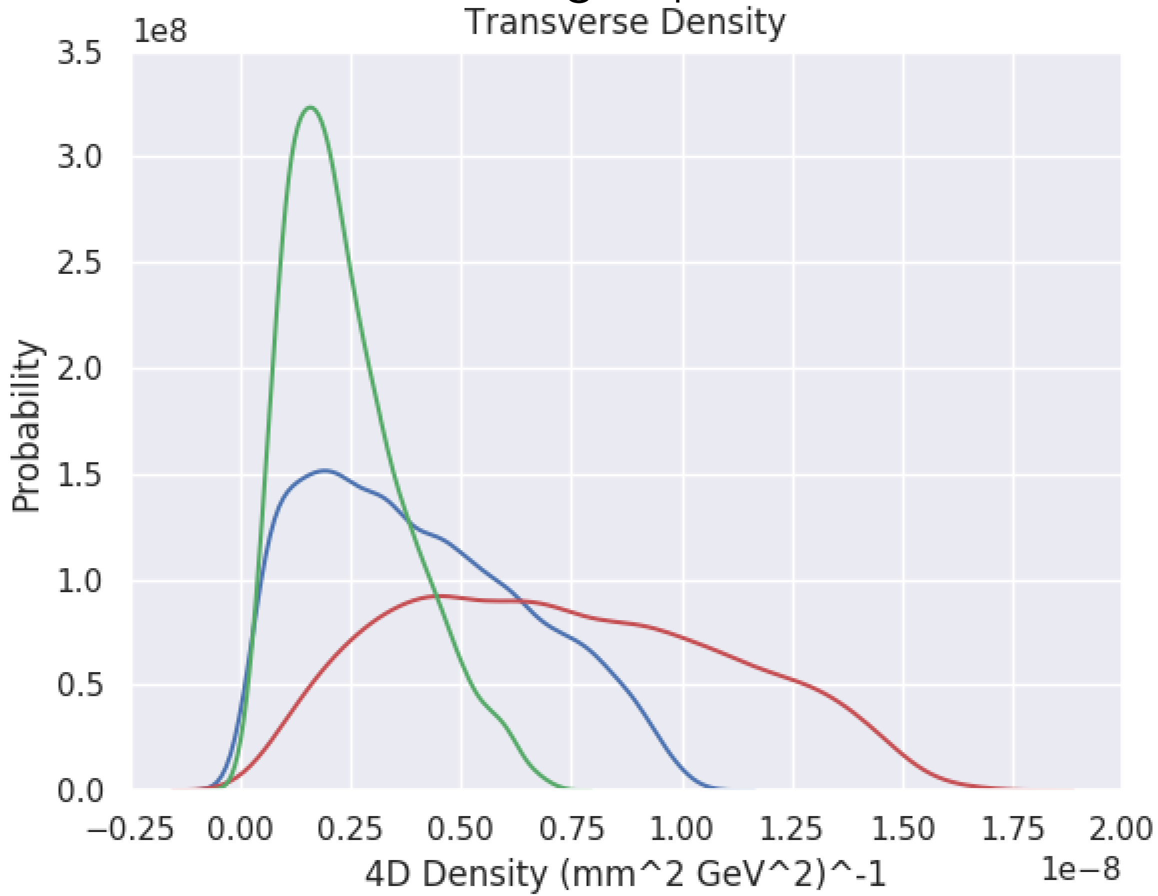


4D Transverse density

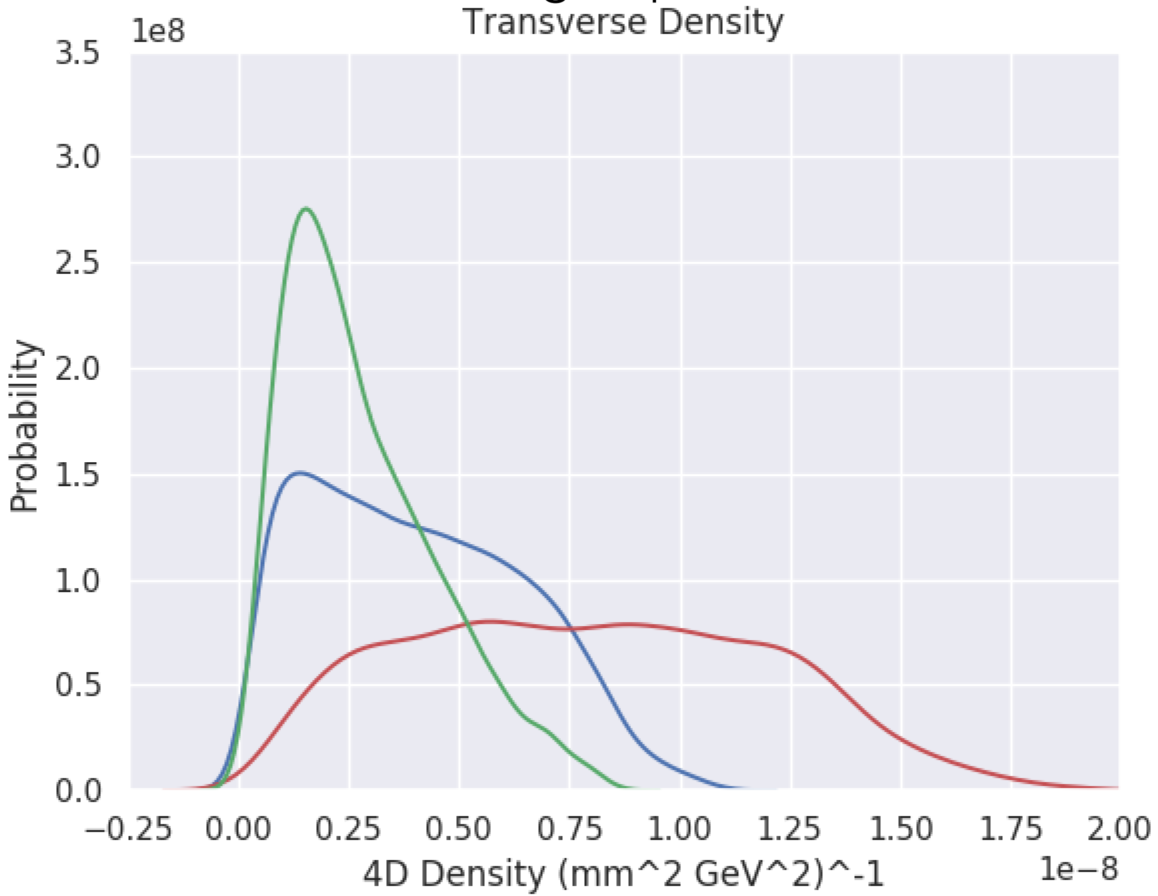
Blue – Full Upstream Sample
Red – Upstream Sample which makes it Downstream
Green – Upstream Sample which does not make it Downstream

Blue distributions are fairly similar, however the green distribution has become broader as some lower radius particles have been eliminated by the wedge

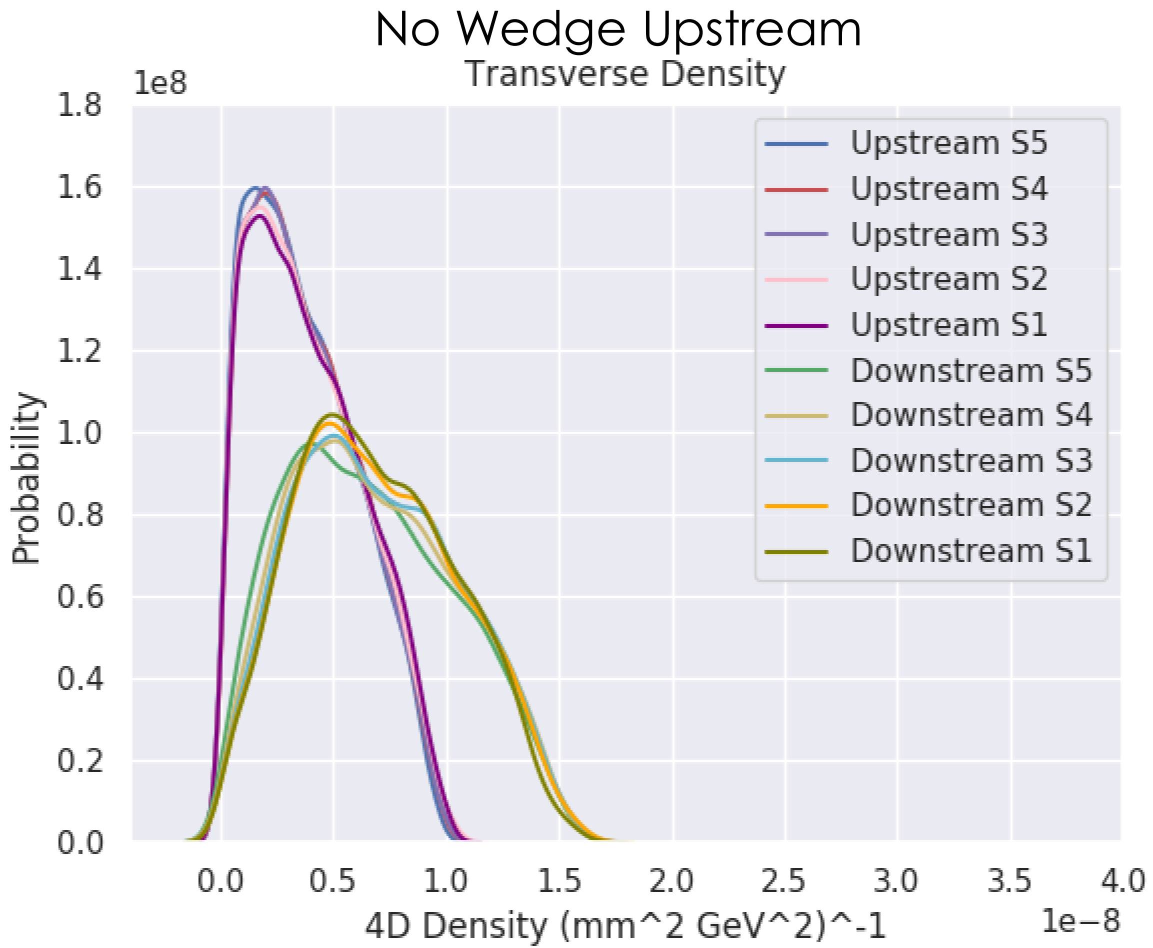
No Wedge Upstream

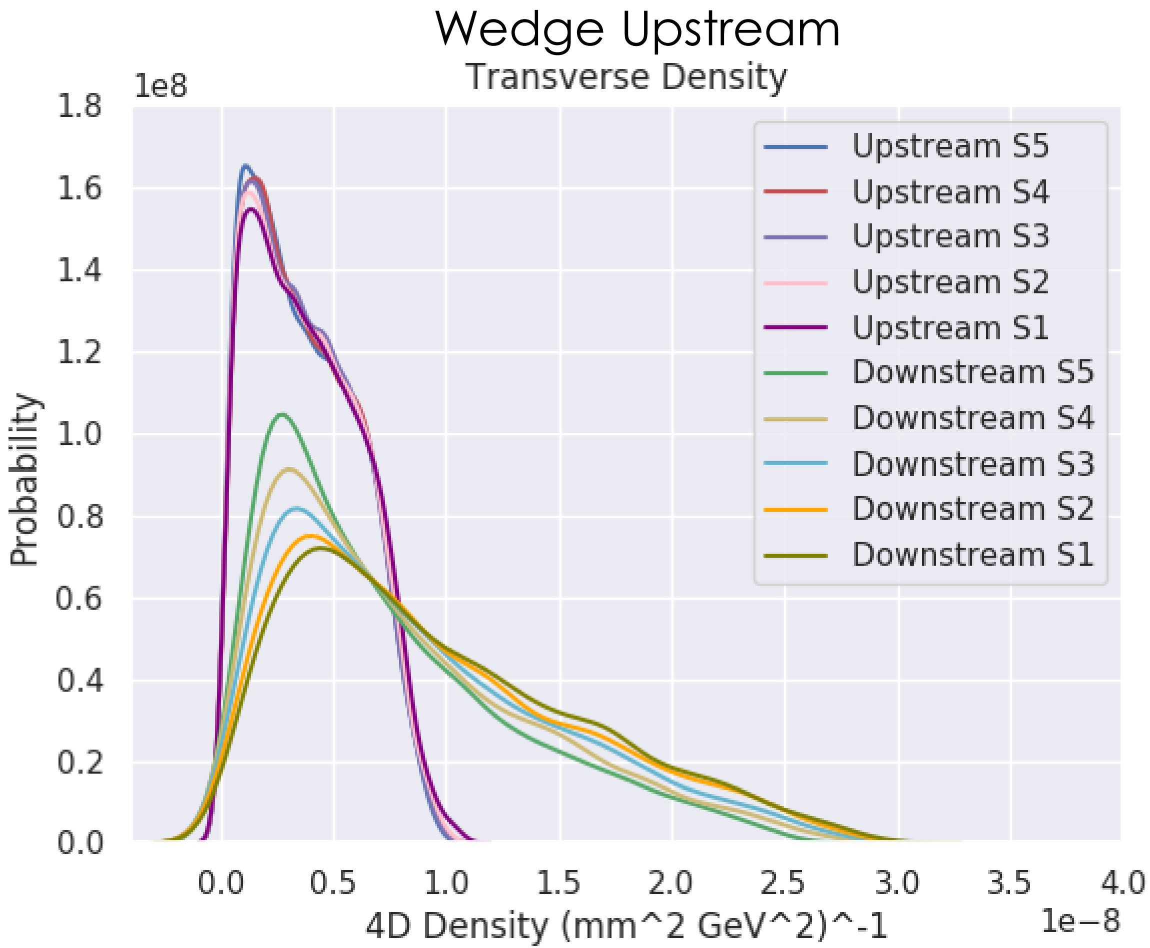


Wedge Upstream

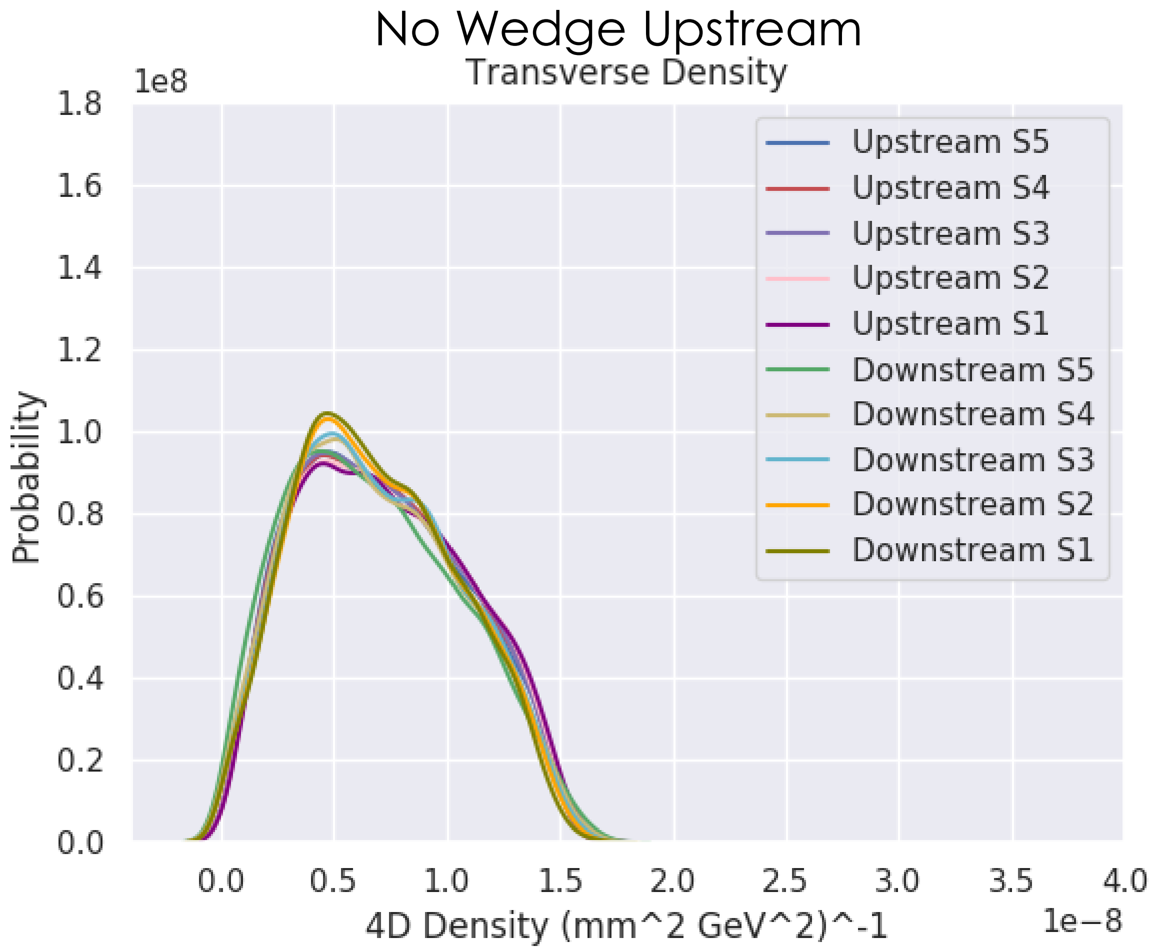


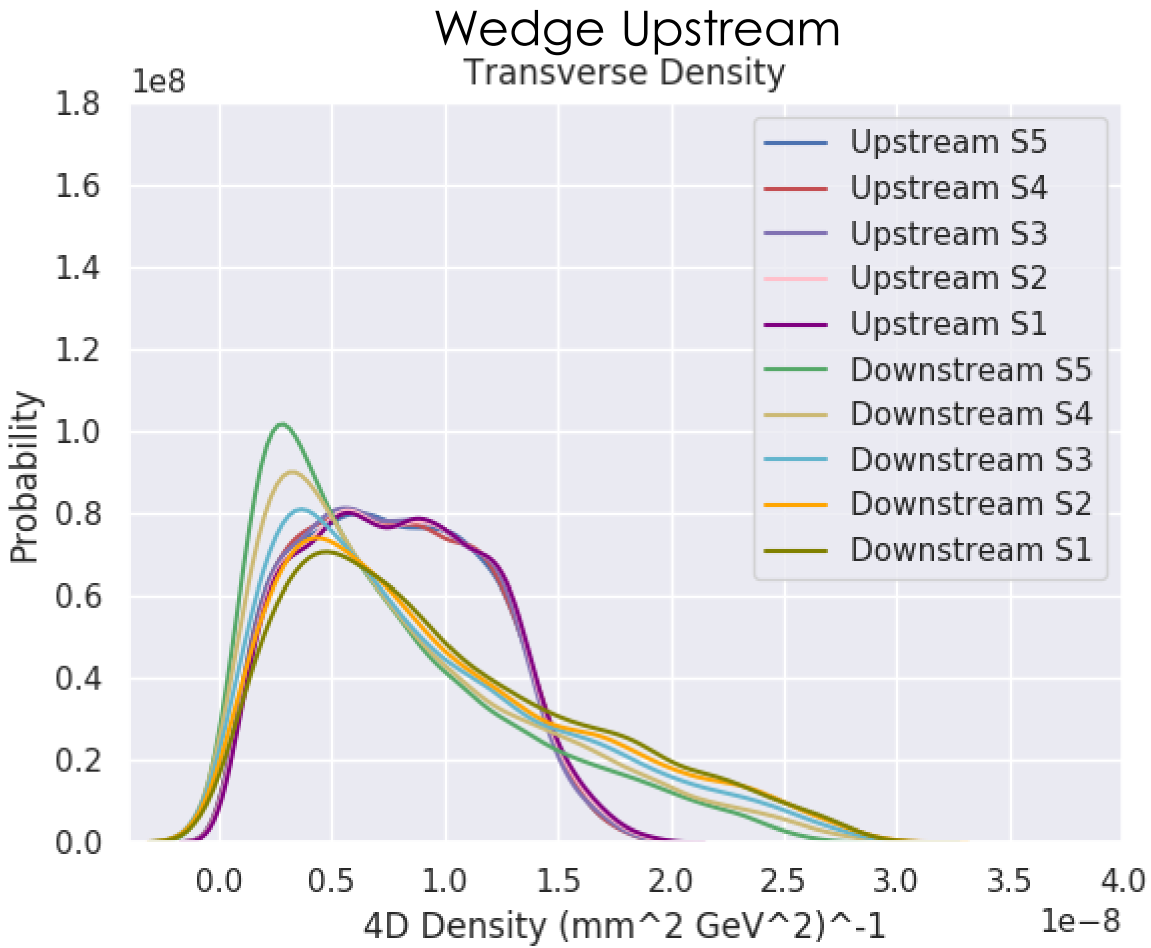
Phase Space Density Evolution Full sample





Phase Space Density Evolution
Only sample which makes it downstream





Phase Space Density Evolution

- Liouville: Phase Space Density doesn't change
- When look at phase space density of a selection of particles through the cooling channel when no absorber is present it remains constant bar for small changes due to the absorber windows
- The upstream sections of the wedge and no wedge case are not comparable as the selection has been biased by the wedge
- The wedge shows an increase in the phase space density for many particles. It also contains a significant number of particles that haven't gone through the wedge
- When look at full sample of particles at each station, there is a clear change between the upstream and downstream section as the particle distributions have changed in a non-random way

Conclusion

- KDE is not a poor estimator, for second-order kernels it has the same rate of convergence as for KNN
- The density calculated is driven by the particle selection
- The density is only conserved for that selection
- MICE has significant transmission losses. When comparing the Upstream and Downstream sections these transmission losses as well as scraping or scattering need to be accounted for as they bias the density calculation
- Without accounting for this, the absorber and no absorber cases can’t be compared
- When looking at the density of a particle selection through the cooling channel, it remains conserved for the no absorber case, and shows significant changes when the wedge is present

The End