Thesis Spring 2024 Review - Greta Hibbard

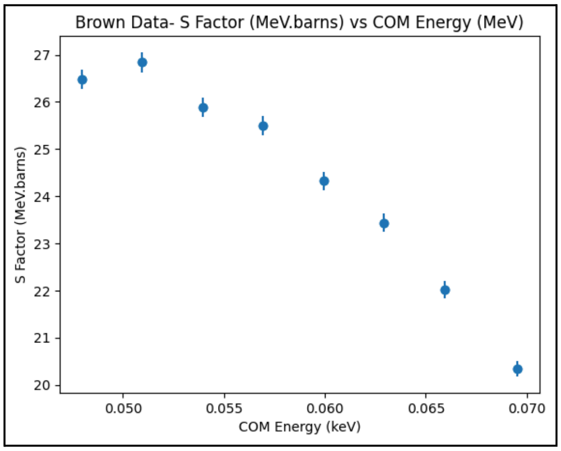
General

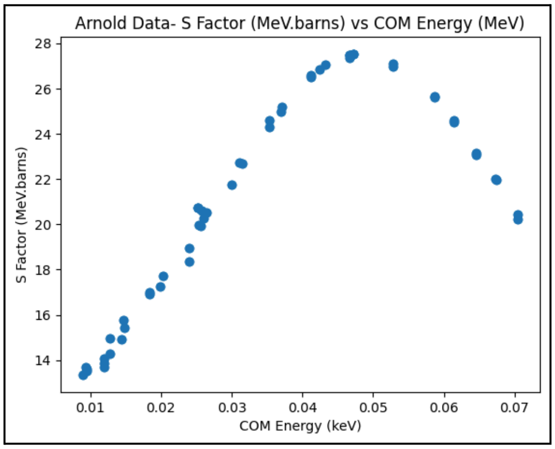
- Provide uncertainty quantification for light ion fusion reactions
  - Begin with D-T with the intent to apply the analysis to other reactions
    - Arnold, Conner, Brown, Kobzev, Jarmie data sets for D-T reaction
- A re-analysis of uncertainty quantification is necessitated
  - Last comprehensive analysis was the Bosch-Hale paper 30 years ago
    - New tools (Bayesian Analysis)
    - New data
    - New questions / projects
  - This research aims to provide an update to that paper using some of the new tools and new data available to answer new, relevant questions
- Additionally, the early weeks of the tutorial focused on writing skills preparing for the Goldwater submission at the end of January

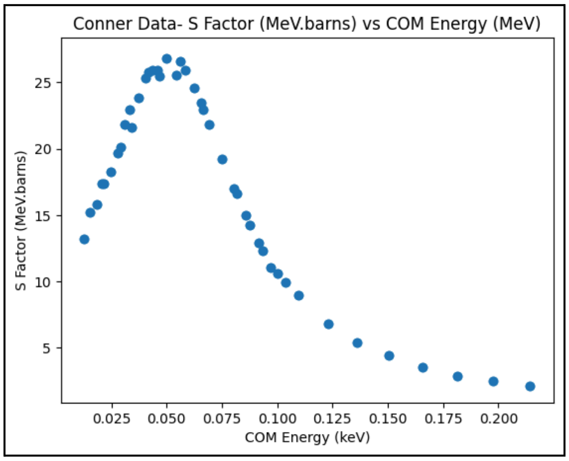
Tasks

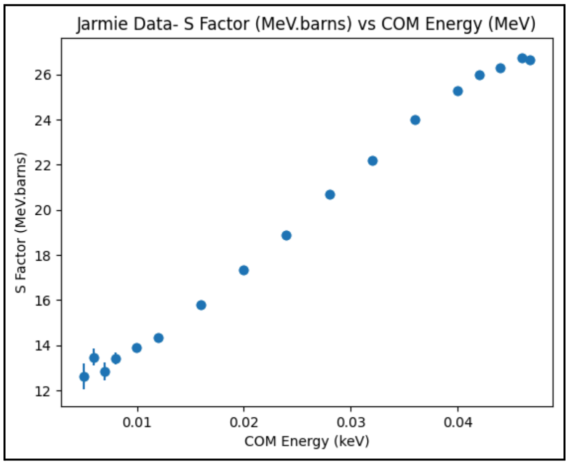
1. Plotting D-T data

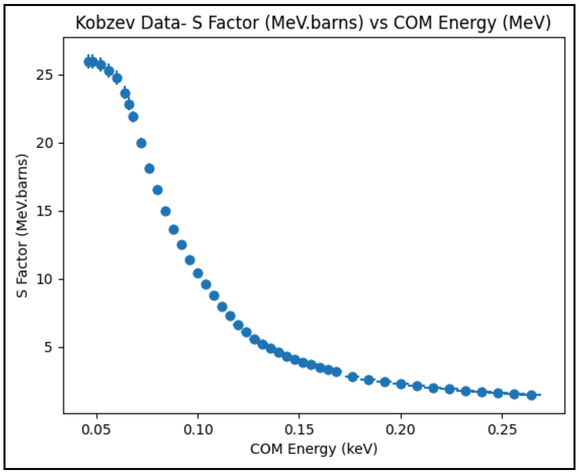
   I began by plotting the available data sets for deuterium tritium fusion reactions. For each of the 5 data sets, I plotted center of mass energy (MeV) vs S-factor (MeV*barns).



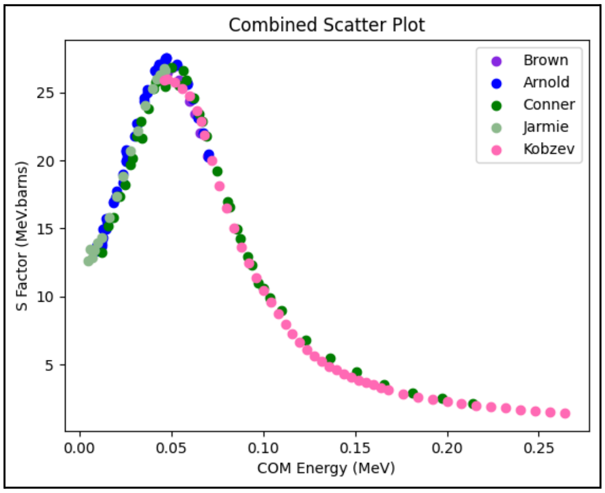








After plotting each individually, I combined the data sets onto the same plot.

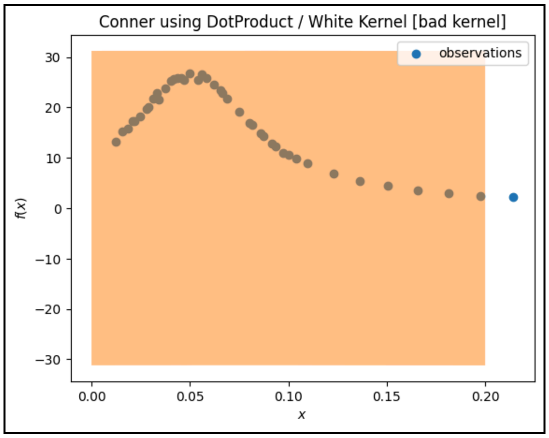


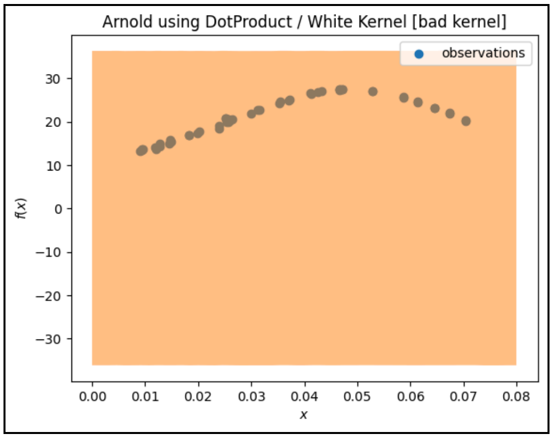
2. GPRs and Kernels

Now, with the data plotted, I began experimenting with different kernels and methods of plotting GPRS.

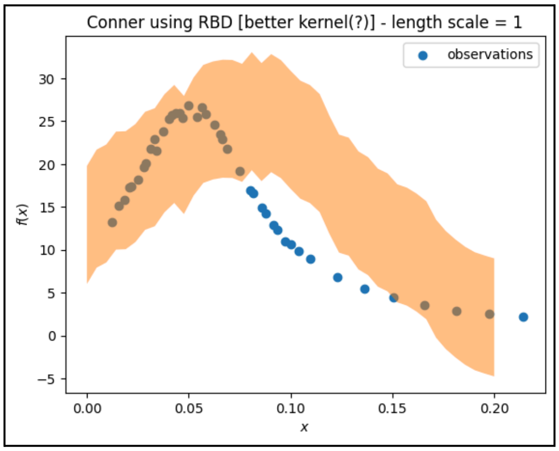
Before writing GP models, I had to learn to manipulate and separate the data based on what needed to be training / testing data. In the first examples, I overtrained the model by using all observations as training data. After finding the correct kernel, I begin to narrow down the amount of training points by eliminating as many points as possible while maintaining the reaction curve.

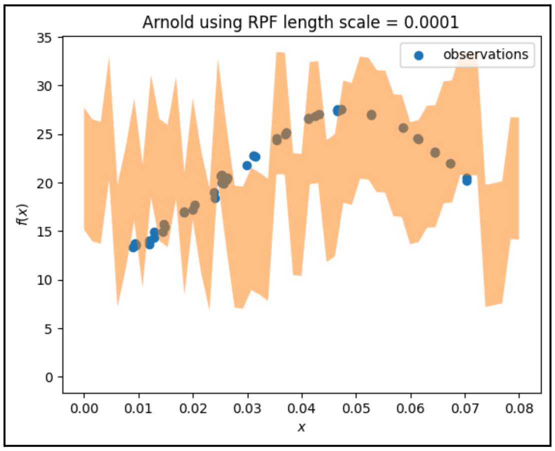
First, from following an example, I used the white kernel. Upon seeing the graphical results of the GP, I realized it was not a good kernel for my model. The main use-case of this kernel is as part of a sum-kernel where it explains the noise of the signal as independently and identically normally-distributed. Additionally, the length scale I used was too large, so standard deviation was too large and ultimately unhelpful.





I continued to experiment with different length scales and kernels until the standard deviations showed a reasonable prediction for the data point. I changed the kernel to a Radial Basis Kernel because I have used the RBF kernel in my previous research. I changed to length scale to a smaller value as well because the order of magnitude of the predicted points needs to be much smaller than 1.

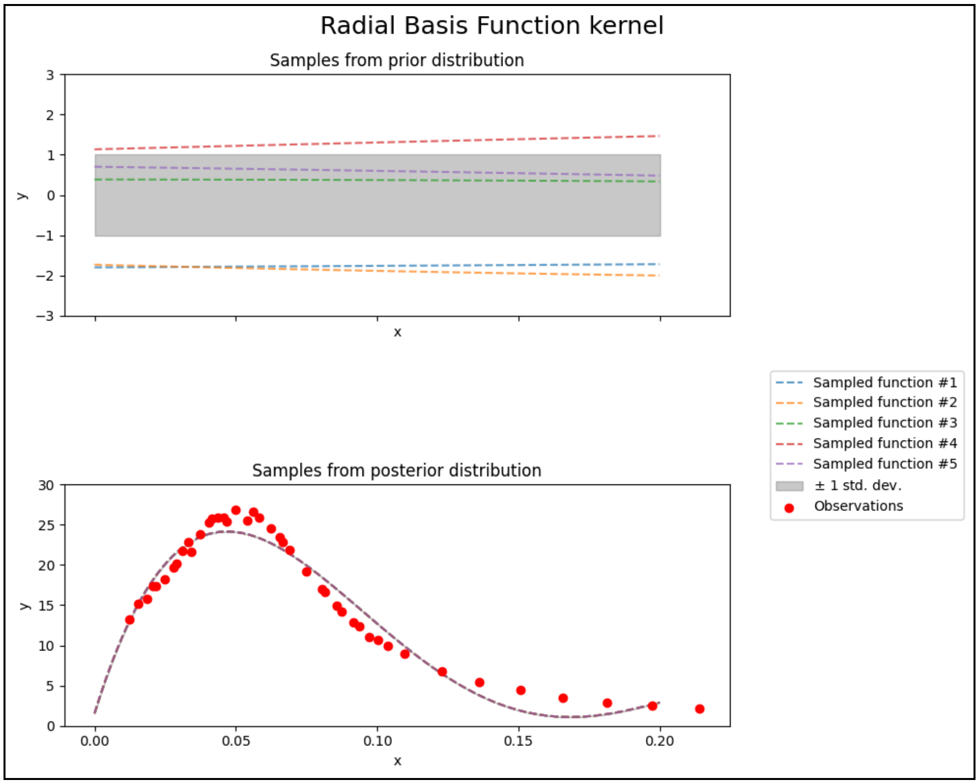




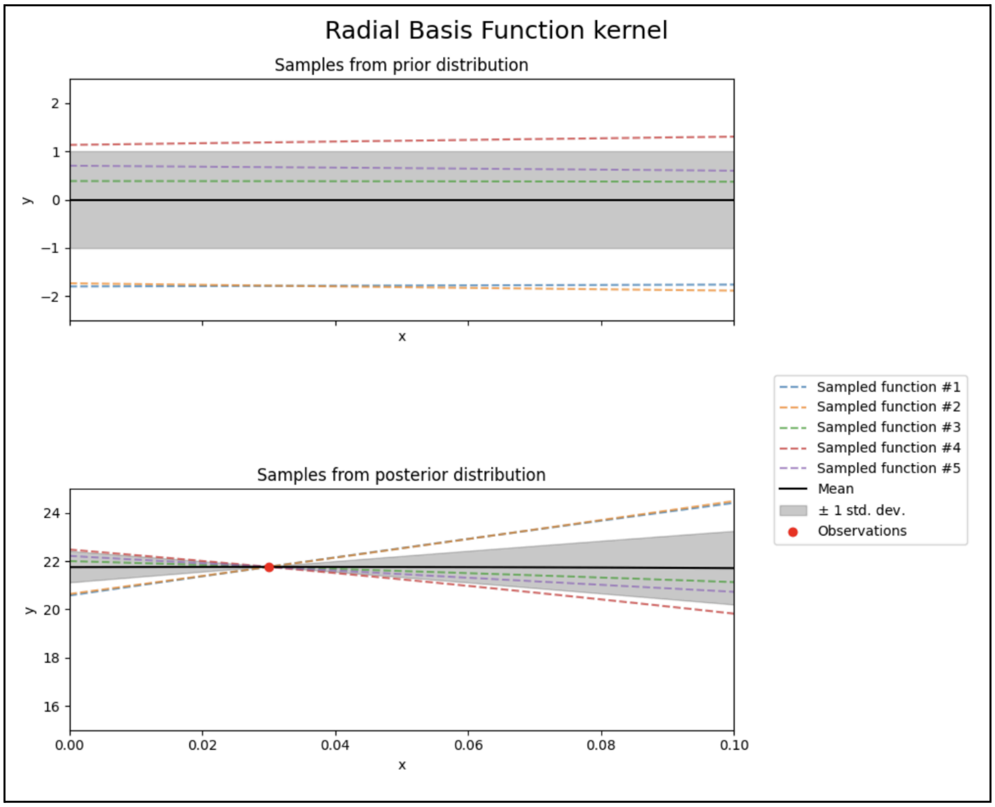
More noise can be observed when the length scale is reduced.

Next, I try following a different GP example for this data. The following example allows the length scale of the kernel to be optimized as the Gaussian Process is run. It plots the prior and posterior distributions of the GPs, including sampled functions and a mean function. This example allows for great visualization of the training process.

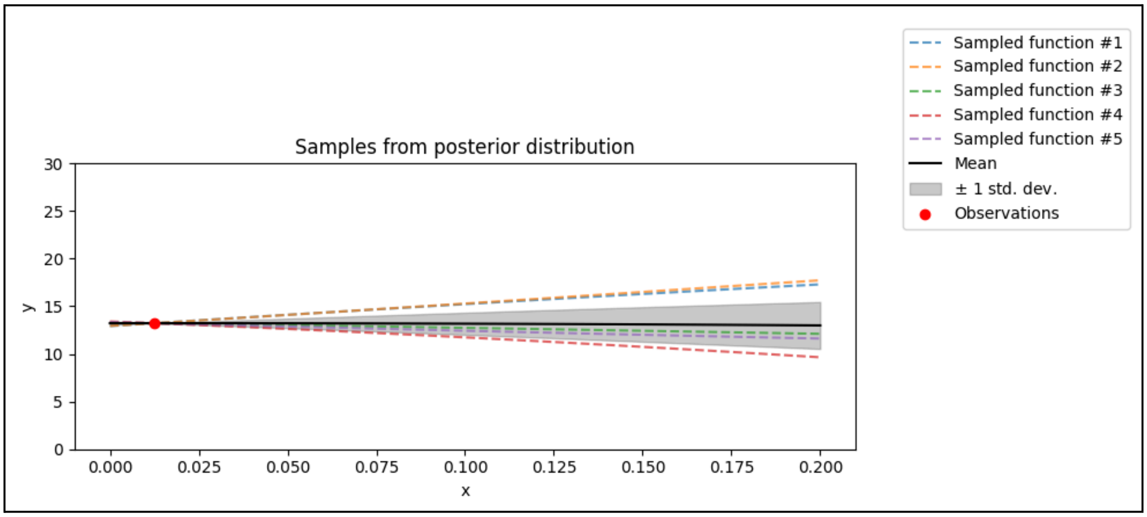
In this example, because there are an abundance of observations, the model gets overtrained. The sampled functions and mean function all hide behind one another and there is no observable standard deviation because the points are close enough together.

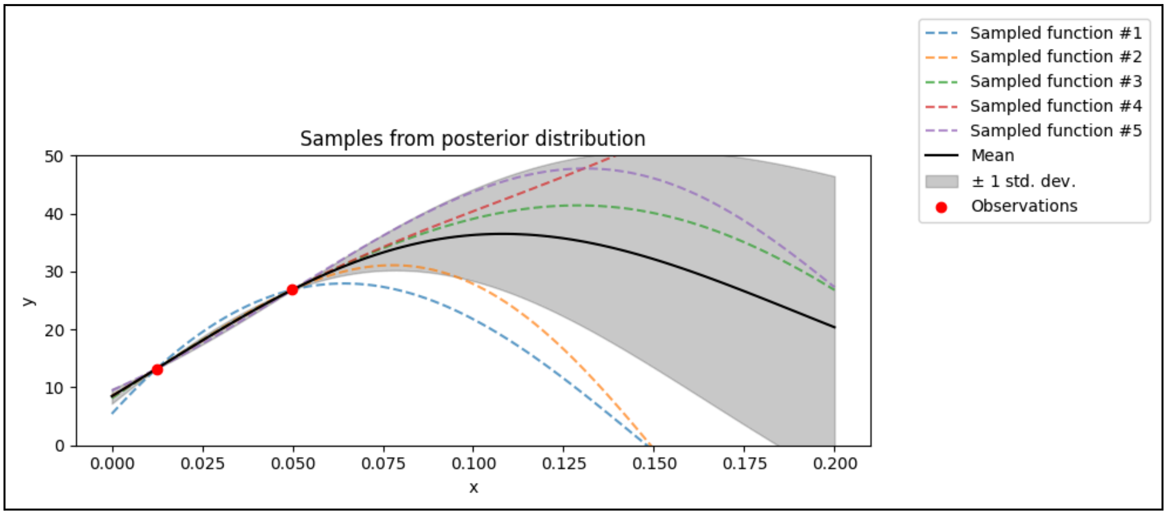


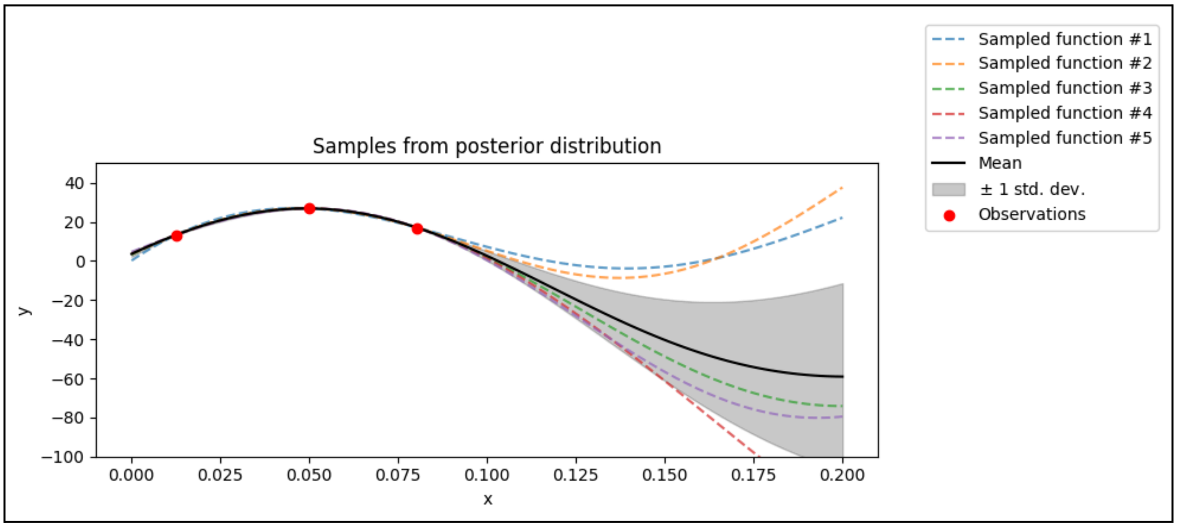
Now, I began reducing the amount of training points so as to not overtrain the model. I reduced observations until the standard deviation 'clouds' were visible. Starting with one observable, the reducing standards deviation around the observation becomes evident.

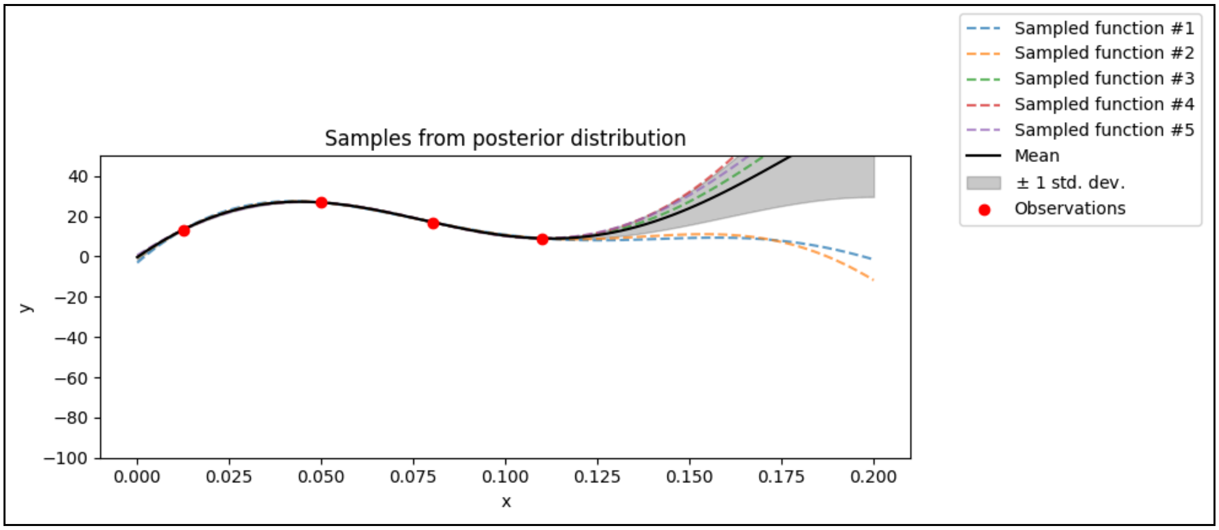


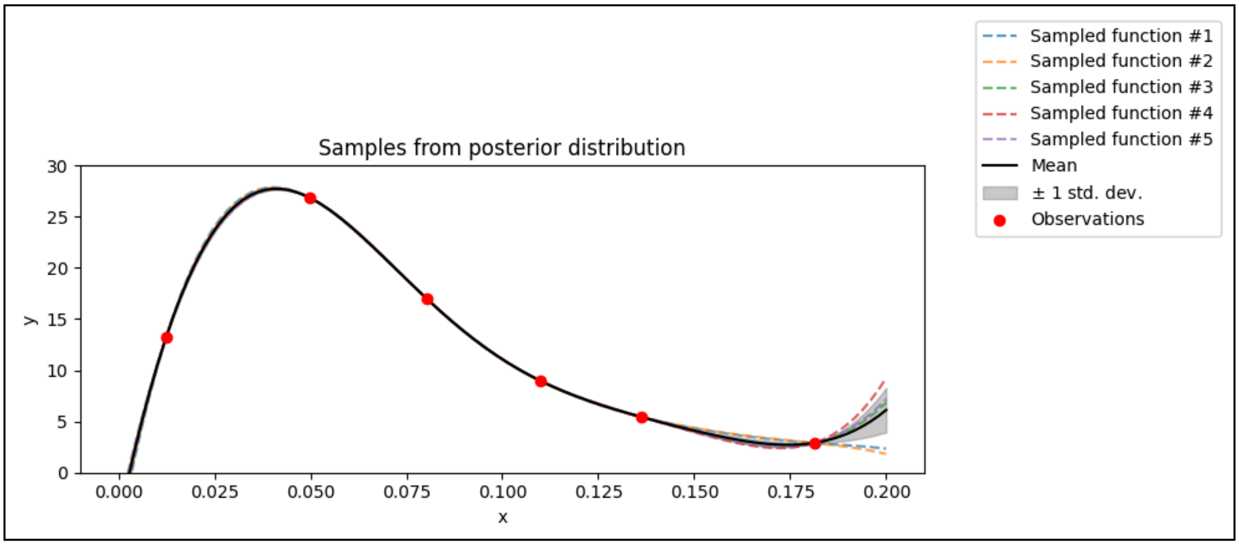
Following, is an illustration of the training process:











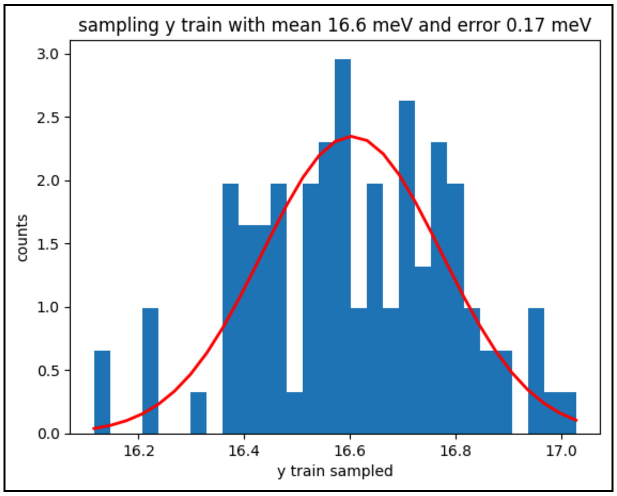
After adding six points, the data set is well trained without being overtrained. The kernel was trained to the following:

```
Kernel parameters before fit:
1**2 * RBF(length_scale=1))
Kernel parameters after fit:
252**2 * RBF(length_scale=0.1)
Log-likelihood: -29.07
```

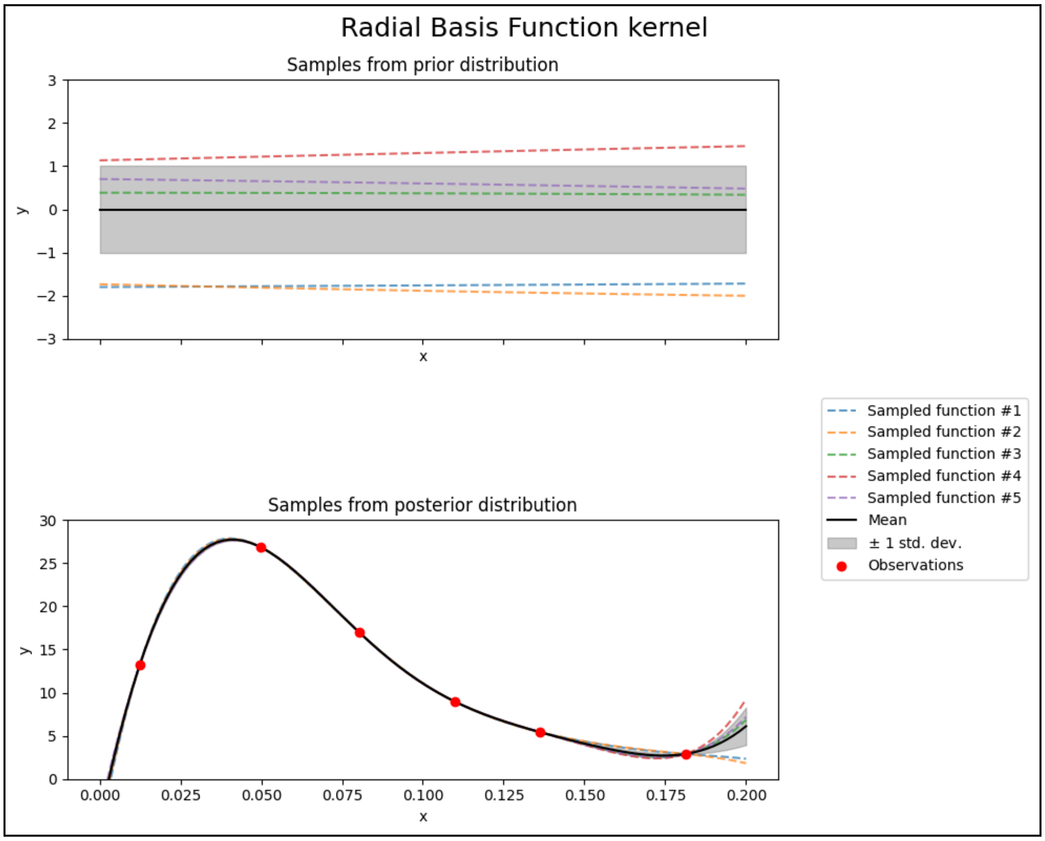
I proceeded to create GPs for each of the 5 data sets.

3. How good is my GP?

Next, I aimed to develop some metric to evaluate how likely the GP is to predict a given S factor value. I began by selecting one observation and sampling within the given error associated with the point. I sampled 100 points within a gaussian.

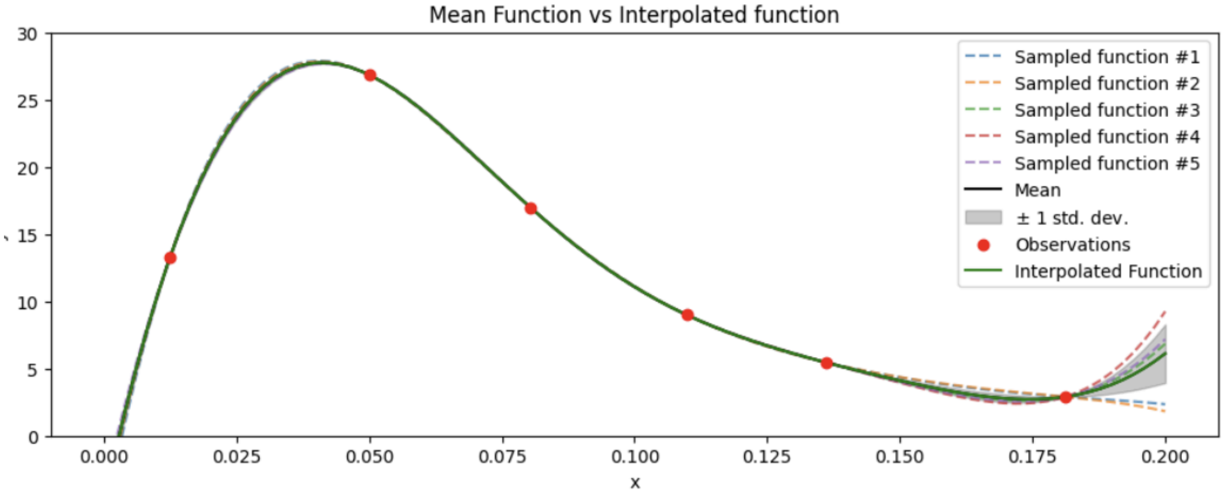


Next, I created a GP model to establish the posterior distribution function.



```
Kernel parameters before fit:
 1**2 * RBF(length_scale=1))
 Kernel parameters after fit:
252**2 * RBF(length_scale=0.1)
    Log-likelihood: -29.074
```

I used the data to interpolate a mean function of the PDF. Admittedly, I am not sure if this was the best way to get the mean function, but I could not find a way for sci-kit-learn to return the PDF directly. I plotted the interpolated mean function against the original mean function to evaluate accuracy, and it looked identical so I continued on using the interpolated function.


Mean Function vs Interpolated function

Now, selecting one point in the sampled distribution of the testing point:

- s = 16.49267033
- x_test = 8.1600E-02 sigma(x_test) is the s.d. of the GPR trained on your training data at the testing point x_test,
- sigma(x_test) = [0.00373278] mu(x_test) is its mean
- mu(x_test) = [16.551878]

The probability of predicting that value of s given the GPR is:

pr(s|GPR) \propto exp[-1/2 (s-mu(x_test)) 1/sigma(x_test)^2 (s-mu(x_test))]

Which yields an unlikely value of

2.3296064748574668e-55

Next, I generalized this calculation to all points in the sampled gaussian and computed the average likelihood of the GPR.

0.026588253937028697

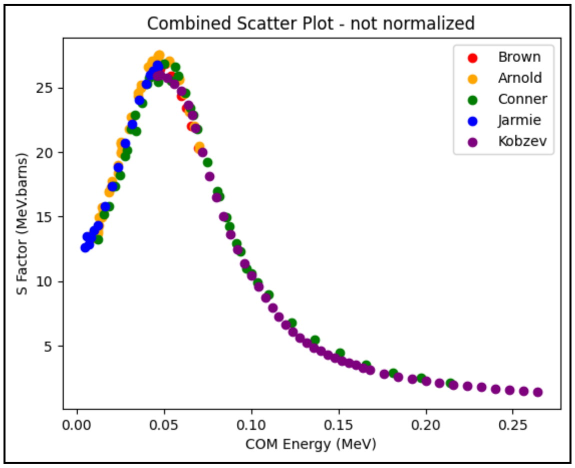
4. Normalization

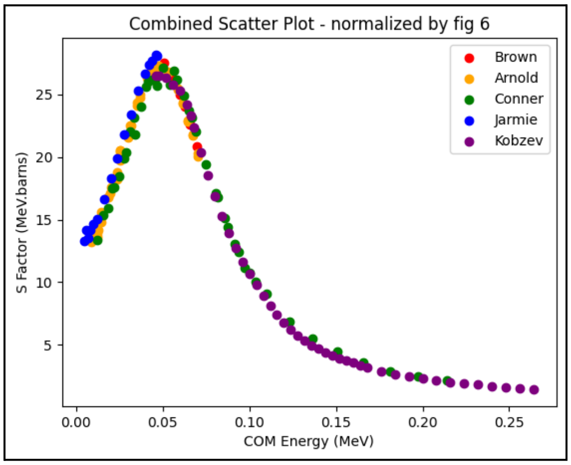
Next, I take a look at normalizing the data sets (Arnold, Brown, Conner, Jarmie, Kobzev). To do so, I utilized the normalization fractions presented in the Odell Brune Phillips paper.

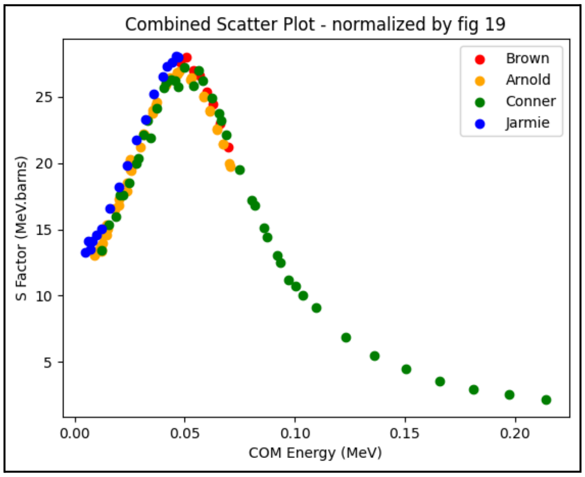
Data Set	Fraction by Fig. 6	Fraction by Fig. 19
Conner	0.99	0.997
Arnold	1.01	1.015
Kobzev	0.98	N/A

Jarmie	0.95	1.003
Brown	0.975	0.983

Allowing us to compare the following plots,





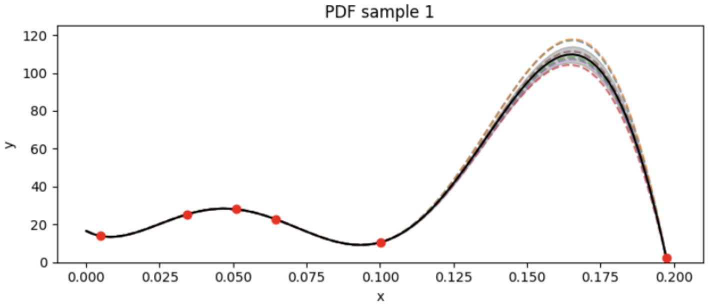


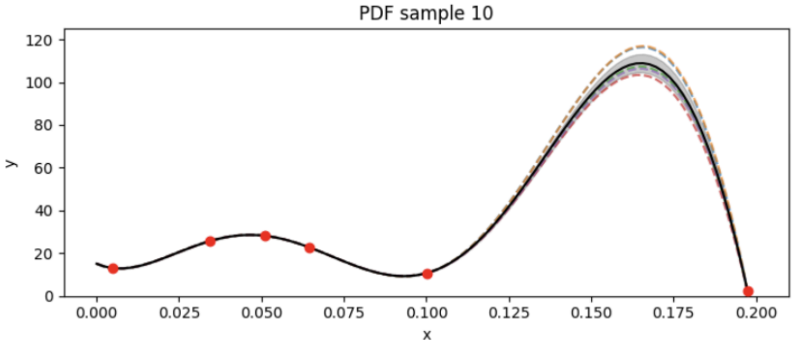
5. GP model with normalized data

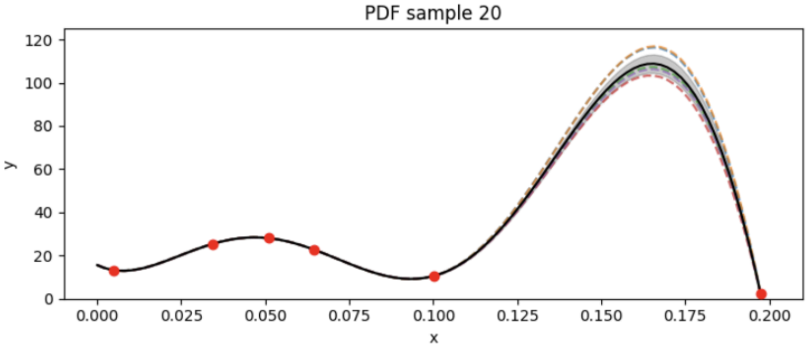
The final task completed during the spring semester was creating GP models with the normalized data. I selected points from the normalized data sets that had low uncertainties. I then sampled 100 points from each data sets from a normal distribution, as done in the process described in task 3. I did some manipulation to then form 100 6 training point arrays from the sampled original values like so…

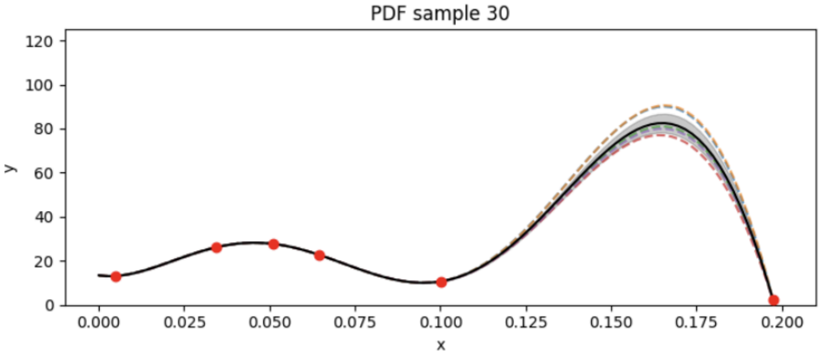
Sfactor sample set 1: [13.414127224752619, 25.895308188602684, 27.88562050310998, 22.66921050188562, 10.551001748860026, 2.5120293439780452]
Sfactor sample set 2: [13.370121682046351, 25.716289335646863, 27.861369832661158, 22.728216070768084, 10.514571616631175, 2.4768334674608847]
Sfactor sample set 3: [13.81152362914069, 25.909668707978668, 28.045112780841603, 22.691151086133576, 10.69640850224839, 2.5150947423676007]
…..
Sfactor sample set 98: [13.24375895621759, 25.59878189727652, 28.22472964700317, 22.682159709473346, 10.608843033564858, 2.4999387273241434]
Sfactor sample set 99: [12.986154028322296, 25.502444530160464, 27.600268437920164, 22.743182666508062, 10.299389582646421, 2.455001568949724]
Sfactor sample set 100: [12.784163227881429, 26.112265712632468, 28.005641693135622, 22.737976959418972, 10.720591474177493, 2.4900580499824416]

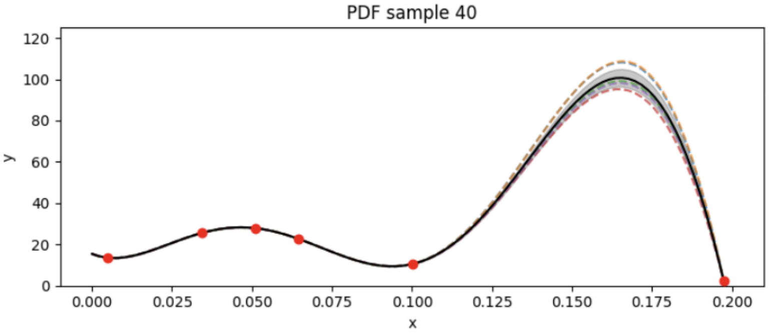
I then created GP models for all 100 sets of training data. As an example, I plotted every 10 GPs. Below, it is observed that I did not do a very good job of selecting points that spanned the entire x range due the the discrepancy around 0.17 MeV.

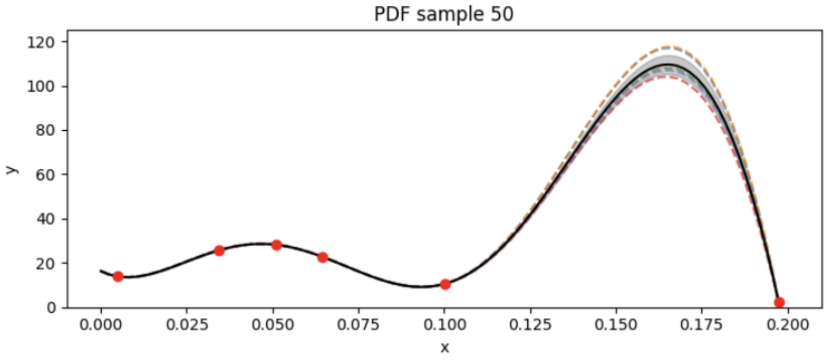


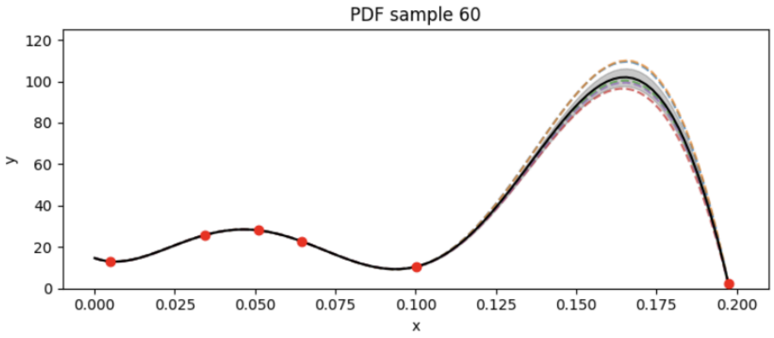

PDF sample 10
0
20
40
60
80
100
120
y
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
x

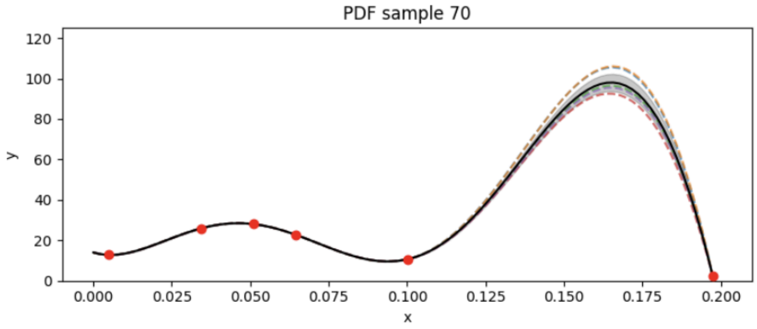

PDF sample 20
0
20
40
60
80
100
120
y
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
x

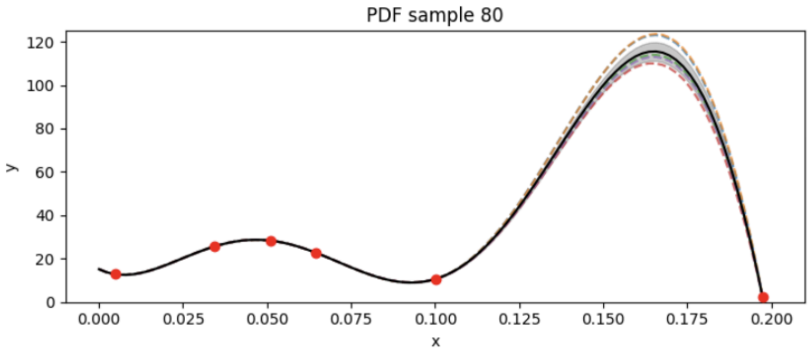

PDF sample 30
0
20
40
60
80
100
120
y
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
x

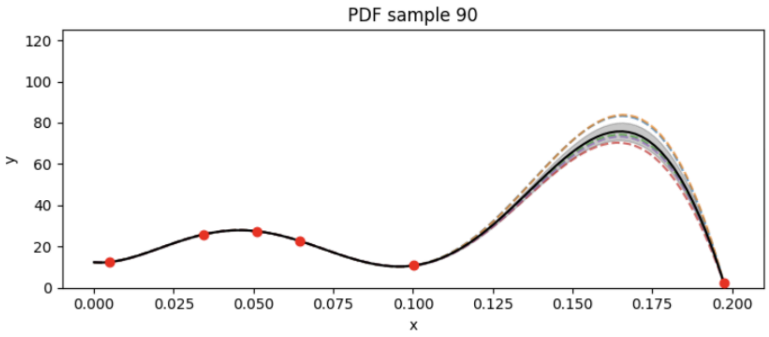

PDF sample 40
0
20
40
60
80
100
120
y
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
x

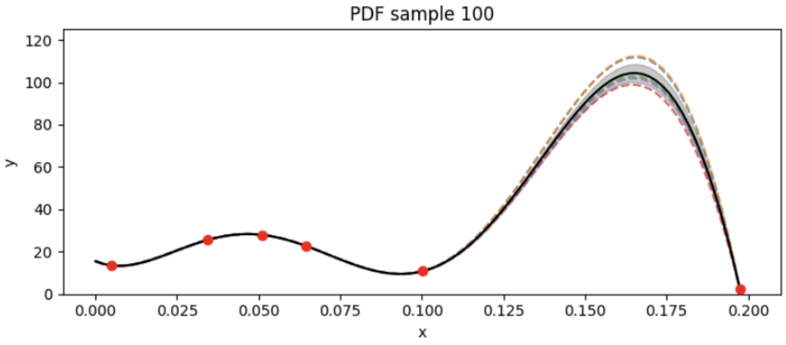

PDF sample 50
y
x
0
20
40
60
80
100
120
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200


PDF sample 60
y
x
0
20
40
60
80
100
120
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200


PDF sample 70
y
x
0
20
40
60
80
100
120
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200


PDF sample 80
y
x
0
20
40
60
80
100
120
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200





Looking Forward

Summer 2024 - Write prospectus