Lab Report 1

Exploring OriginPro and modeling CMBR and high energy emissions of neutron stars

Mohammad Mahmoud Ibrahim **202200438**

PEU 327 (Astronomy Laboratory)

Contents

1 Experiment (I): Scientific Graphing and Data Analysis Using OriginLab

1.1 Task (I): Statistical Analysis

(a) *Descriptive Statistics*
Figure 1 contains the descriptive statistics of the heights and the weights of 40 people sorted by gender then age.

Descriptive Statistics

			N total	Mean	Standard Deviation	Sum	Minimum	Median	Maximum
height	F	12	5.0	144.6	13.4	723.0	128.0	146.0	163.0
		13	3.0	145.3	6.4	436.0	138.0	148.0	150.0
		14	5.0	154.2	3.7	771.0	150.0	153.0	160.0
		15	2.0	155.5	3.5	311.0	153.0	155.5	158.0
		16	2.0	154.0	8.5	308.0	148.0	154.0	160.0
		17	1.0	153.0	--	153.0	153.0	153.0	153.0
	M	12	3.0	141.3	13.3	424.0	126.0	148.0	150.0
		13	4.0	151.0	7.9	604.0	143.0	150.5	160.0
		14	7.0	161.0	5.7	1127.0	155.0	158.0	170.0
		15	5.0	160.8	4.7	804.0	153.0	163.0	165.0
		16	1.0	168.0	--	168.0	168.0	168.0	168.0
		17	2.0	170.5	3.5	341.0	168.0	170.5	173.0
weight	F	12	5.0	44.9	15.2	224.5	28.7	42.2	65.0
		13	3.0	42.9	11.1	128.6	30.1	47.9	50.6
		14	5.0	43.6	11.6	217.8	36.8	38.5	64.1
		15	2.0	45.9	5.8	91.8	41.8	45.9	50.0
		16	2.0	51.0	1.5	101.9	49.9	51.0	52.0
		17	1.0	52.4	--	52.4	52.4	52.4	52.4
	M	12	3.0	44.0	12.2	131.9	35.9	38.0	58.0
		13	4.0	42.3	5.0	169.2	35.2	43.6	46.9
		14	7.0	46.5	4.7	325.8	41.3	44.3	53.6
		15	5.0	49.7	4.4	248.3	46.3	47.4	57.1
		16	1.0	57.8	--	57.8	57.8	57.8	57.8
		17	2.0	66.6	8.8	133.1	60.3	66.6	72.8

Figure 1: Descriptive Statistics of Heights and Weights

A standard deviation isn't available for any of the groups of sample size 1, since it would be nonsensical. We can see that the standard deviation for height and weight is largest in younger individuals, since they are still developing, the variablity is large. For males, the mean height is largest at age 17 (largest age), while for females it is at 15 (third largest age). This result does not make much sense, but could be interpreted as a lack of reliability of the sample (not representative enough, and small sample size=40). The mean weight is largest for males at age 17, and for females also at age 17, which makes sense.

(b) *Normality*
We generated a data set of 100 numbers that have a mean of 100 and standard deviation of 20 using the formula nint(100+20*normal(100)). We ran a Shapiro-Wilk test to check normality of the generated data set, which we naturally expect to be normally distributed.

NormalityTest

Shapiro-Wilk

	DF	Statistic	p-value	Decision at level(5%)
A	100	0.97543	0.05829	Can't reject normality

A: At the 0.05 level, the data was significantly drawn from a normally distributed population.

Figure 2: Normality Test of Generated Data Set

As expected, the test returns that with a confidence level of 95%, the data set cannot be rejected as being normally distributed.

(c) *Frequency Counts*
This is the histogram of the previously generated normally distributed data set.

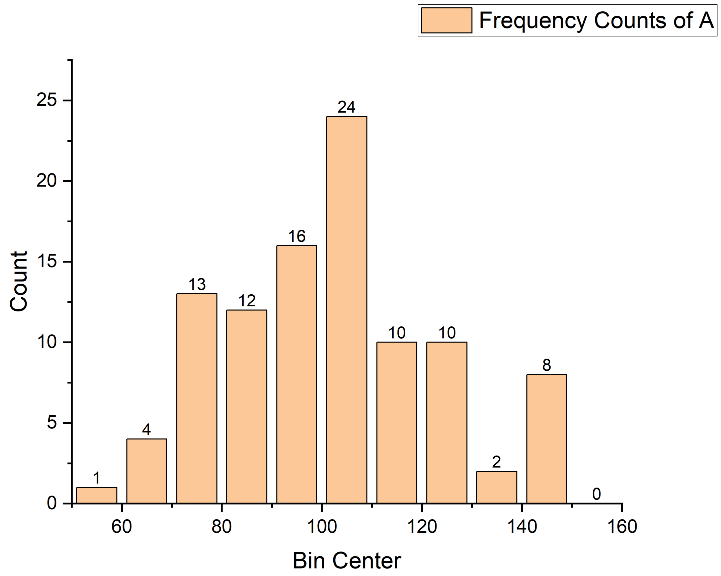


Figure 3: Histogram of Generated Data Set

We can see that the histogram looks relatively normally distributed (bell shape).

(d) *One Way ANOVA*
A mean comparison test is done (Tukey) to test the null hypothesis that the means of all nitrogen content of all 4 plant groups' populations are the same.

We find that with a 95% confidence level, we can reject the null hypothesis that the means of all 4 plant groups are the same.
This is more obvious if we look at the plot of mean differences.

Overall ANOVA

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	1996.36652	665.45551	12.86214	<0.0001
Error	76	3932.05317	51.73754		
Total	79	5928.41969			

Null Hypothesis: The means of all levels are equal.
Alternative Hypothesis: The means of one or more levels are different.
At the 0.05 level, the population means are significantly different.

Figure 4: One Way ANOVA of Plant Groups

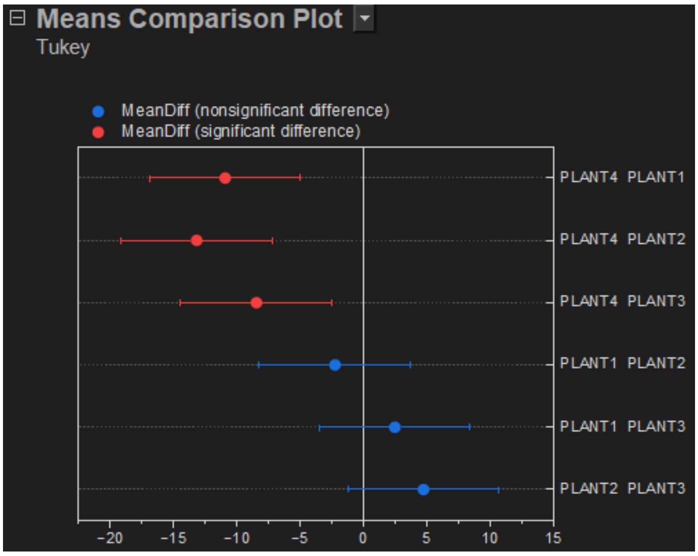


Figure 5: Mean Differences of Plant Groups

We can see that plant group 4 is further than all the others (because it has a much lower mean of 7.96), but the other 3 are close enough that to our significance level we could consider their populations to have equal means.

1.2 Task (II): Graphing

(a) *Batch Plotting*
We have 3 data sets (representing different systems), each with 3 parametrs (T, B, x). We want to plot the data sets in 3 different graphs, each with time on the x-axis and the y-axes representing the 3 parameters. We use the batch plotting feature to do this.

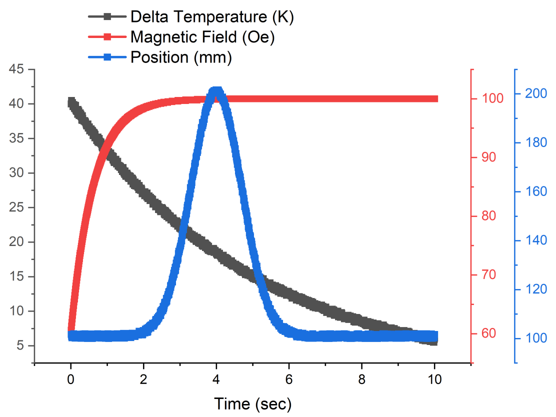


Figure 6: Graph 1

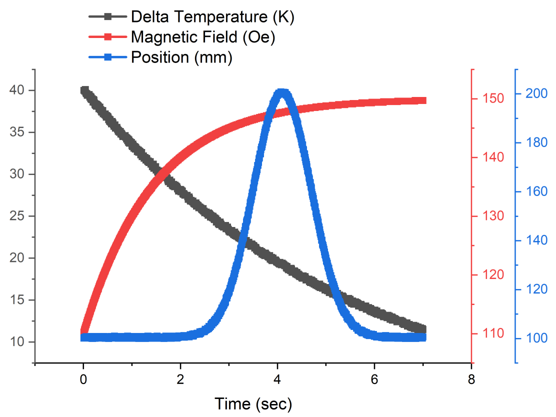


Figure 7: Graph 2

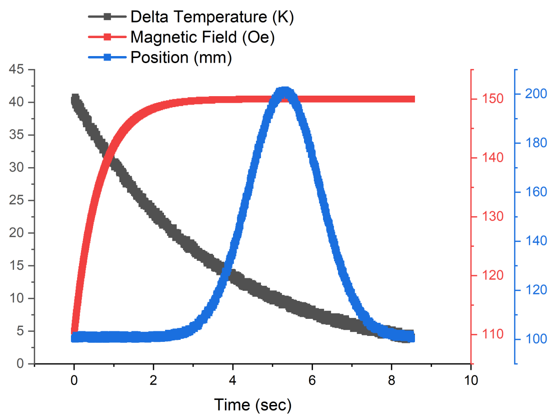


Figure 8: Graph 3

(b) *3D Surface*
We have a data set of 100 points, each with 3 values in 3 columns (A, B, C) where $C = me^{-n(A^2+B^2)}$ (Gaussian). We want to plot a 3D surface colormapped plot of the data set, where the colors reflect the C column values.

We can see that the plot in figure 9 looks like a Gaussian (bell shape).

(c) *Contour and Surface Plots from Virtual Matrix*
Here in figure 10 is a colormapped surface plot of a dataset, with its projection right above it.

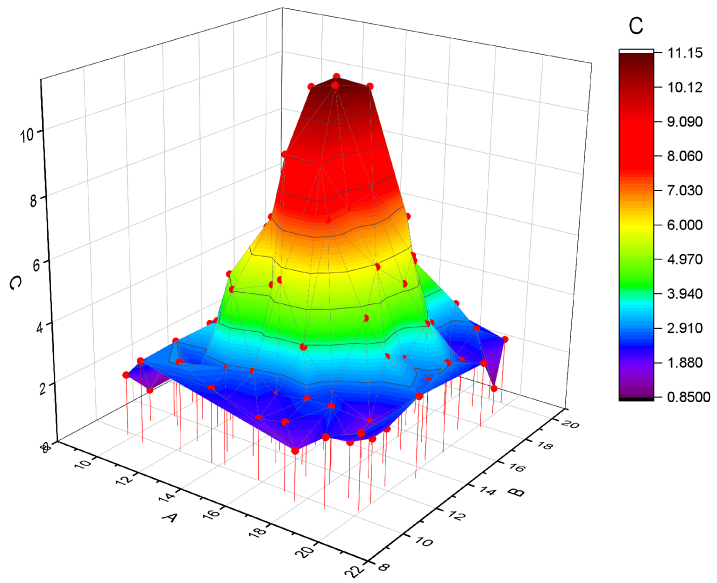


Figure 9: 3D Surface Plot of Data Set

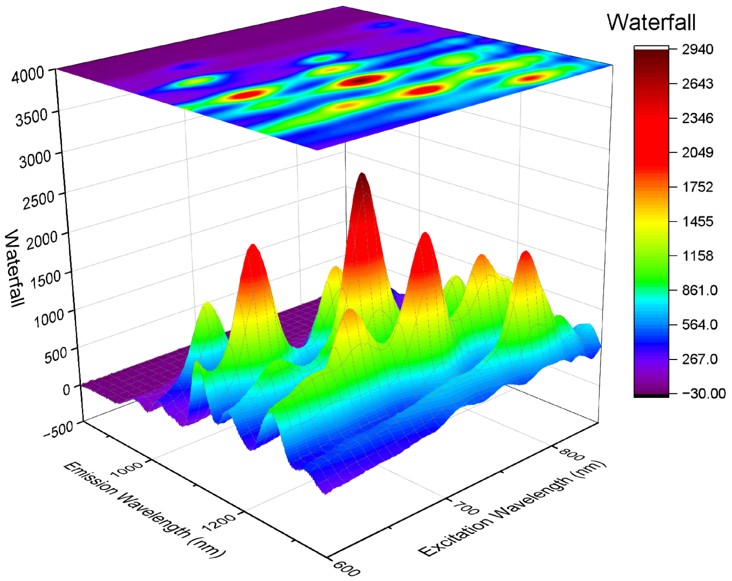


Figure 10: Contour and Surface Plots of Data Set

1.3 Task (III): 3D Visualization of the Hydrogen Atom d-Orbitals

(a) *3D Surface plot with colormap for data in virtual matrix*
We visualize an ellipsoid from a virtual matrix of data in figure 11, with the colormap reflecting the values of the data. We use the sph2cart function since the values are stored as spherical coordinates.

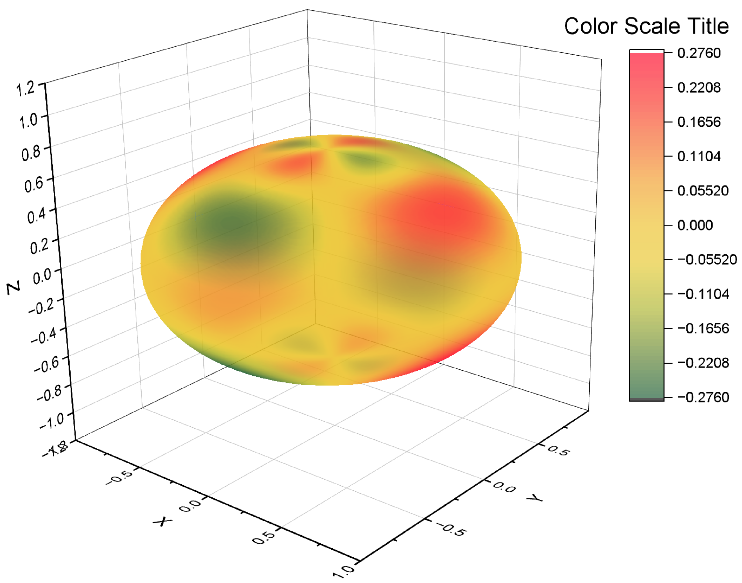


Figure 11: 3D Surface Plot of Ellipsoid

(b) *Convert data in a matrix object and make a 3D surface plot*
We plot the d-orbitals of the hydrogen atom ($l = 2$ orbitals) in figure 12. The angular variation of these orbitals are represented by the spherical harmonics Y_2^m. We take a linear combination of $\frac{1}{\sqrt{2}}(Y_2^2 + Y_2^{-2})$.

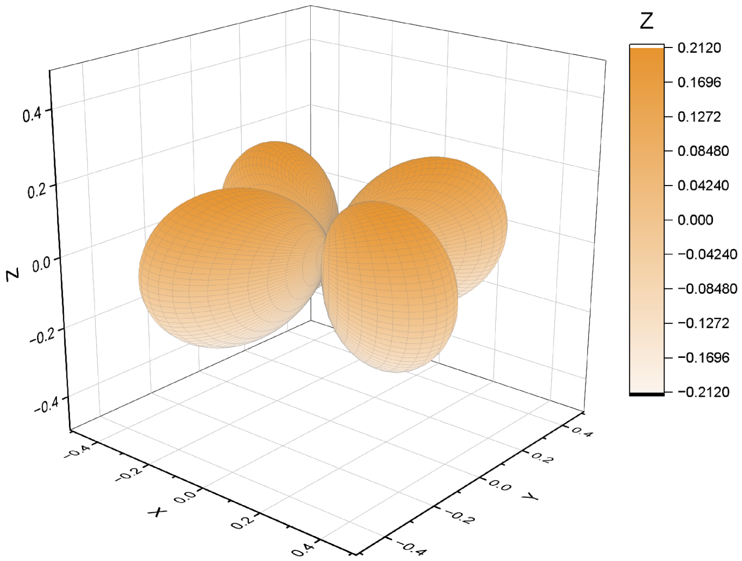


Figure 12: 3D Surface Plot of d-Orbitals

1.4 Task (IV): Linear and Polynomial Regression

(a) *Linear Regression and Outlier Removal*
Here in figure 13, we have the linear fit for the data with the outlier

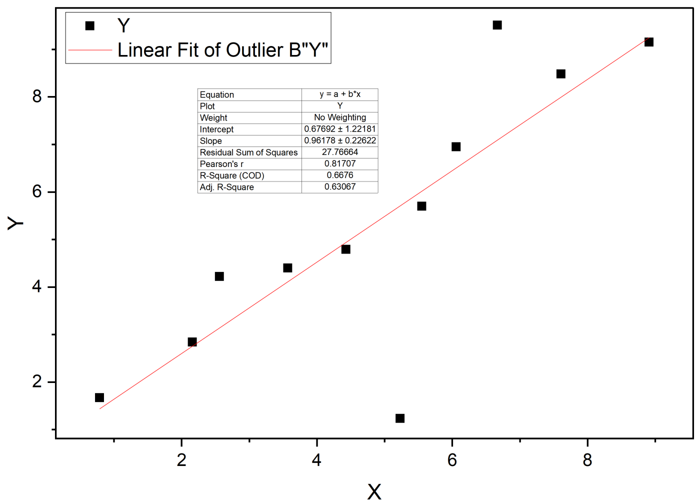


Figure 13: Linear Regression with Outlier

In figure 14, we have the linear fit without the outlier point with the standardized residual of -2.54889.

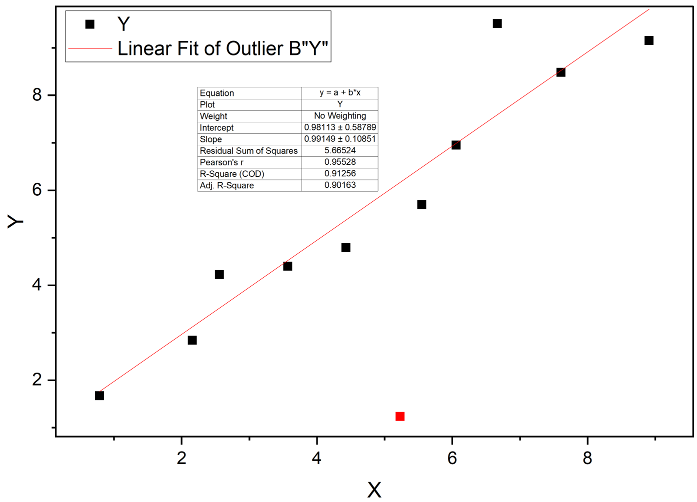


Figure 14: Linear Regression without Outlier

(b) *Linear fit for Langmuir model*

First, we plot the linear fitting for the data when transformed to the traditional linear Langmuir equation.

$$y = -\frac{1}{k}\frac{y}{x} + y_m$$

From here we can find that slope $= -\frac{1}{k} \approx -0.20384 \implies k \approx 4.9058$, and intercept $= y_m \approx 3.0053$.

Then, we plot the linear fitting for the data when transformed to the double reciprocal Langmuir equation.

$$\frac{1}{y} = \frac{1}{y_m k} + \frac{1}{y_m}$$

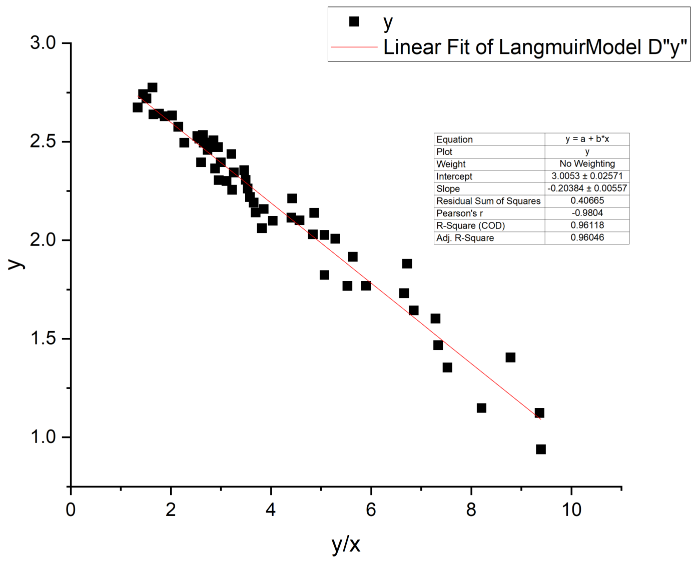


Figure 15: Linear Fit for Traditional Langmuir Model

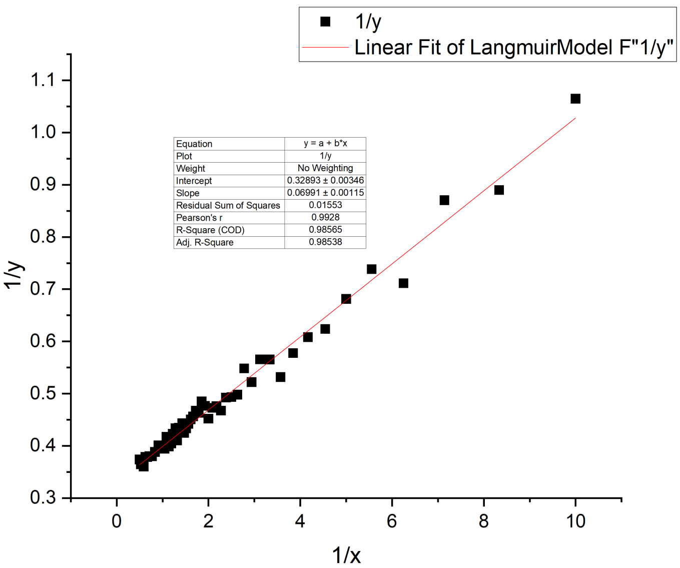


Figure 16: Linear Fit for Double Reciprocal Langmuir Model

Similary, we can find that the intercept $= \frac{1}{y_m} \approx 0.32893 \implies y_m \approx 3.0402$, and slope $= \frac{1}{y_m k} \approx 0.06991 \implies k \approx 4.705$

The results are similar, but not the same. The differences are due to rounding errors/precision in divisons and multiplications. We will now show the graph as in figure 16, except that is plotted by the apparent fit function in OriginLab.

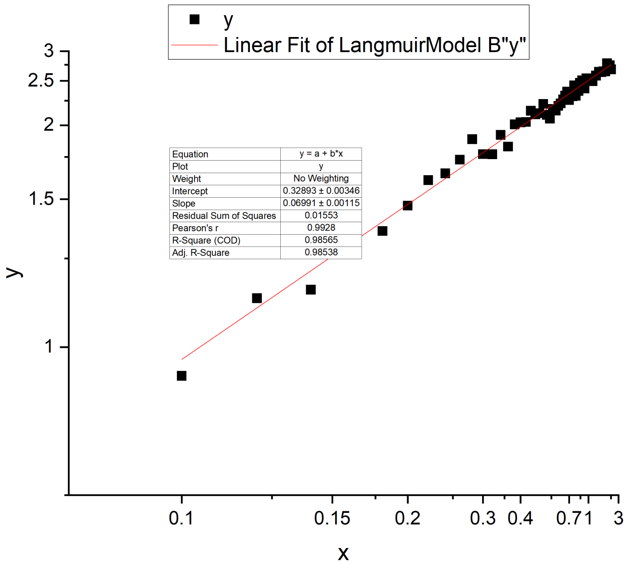


Figure 17: Apparent Linear Fit for Double Reciprocal Langmuir Model

1.5 Task (V): Nonlinear Fitting with Built-in and User-defined Functions

(a) *Use a built-in function to fit the data*
We will fit the given data set in figure 18 using a Gaussian function.

After fitting to a gaussian with 5 iterations, we get the following (figure 19 and 20) fitting curve and parameters.

We fit again but with x_c set fixed to 25, to get a standard error of 25. (figure 21)

(b) *Define and fit with a user-defined function*
We define the function $y = y_0 + ae^{-bx}$ to fit our data set to. After 9 iterations, we get the following fitting curve and parameters (figures 22 and 23).

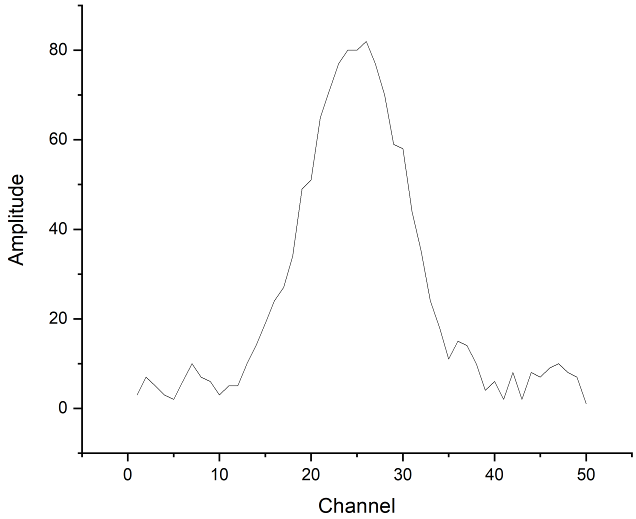


Figure 18: Data Set

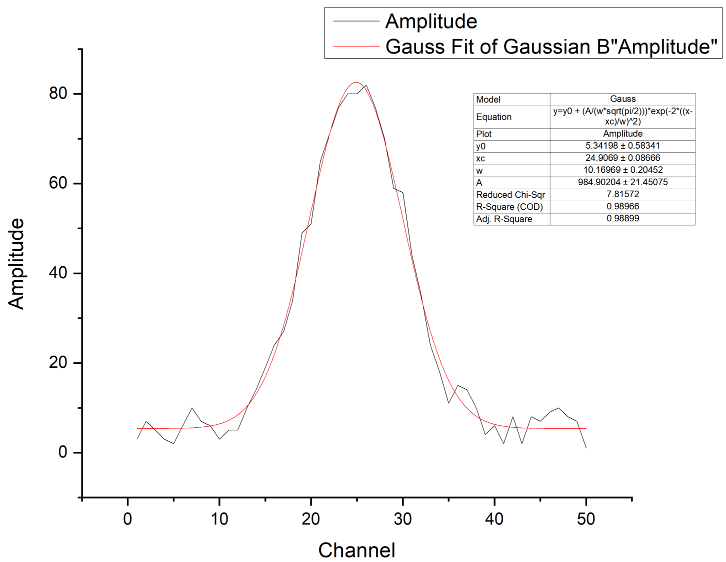


Figure 19: Gaussian Fit for Data Set

Nonlinear Curve Fit (Gauss) (21/2/2025 19:04:39)

⊞ Notes

⊞ Input Data

⊟ Parameters

		Value	Standard Error	t-Value	Prob>|t|	Dependency
Amplitude	y0	5.34198	0.58341	9.15655	6.21151E-12	0.54074
	xc	24.9069	0.08666	287.40475	1.60949E-76	1.07265E-11
	w	10.16969	0.20452	49.72362	1.22804E-41	0.52123
	A	984.90204	21.45075	45.91457	4.47306E-40	0.69383
	sigma	5.08485	0.10226			
	FWHM	11.9739	0.24081			
	Height	77.27256	1.21284			

Reduced Chi-sqr = 7.81572174469
COD(R^2) = 0.98966112534176
Iterations Performed = 5
Total Iterations in Session = 5
Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Figure 20: Parameters of initial Gaussian Fit

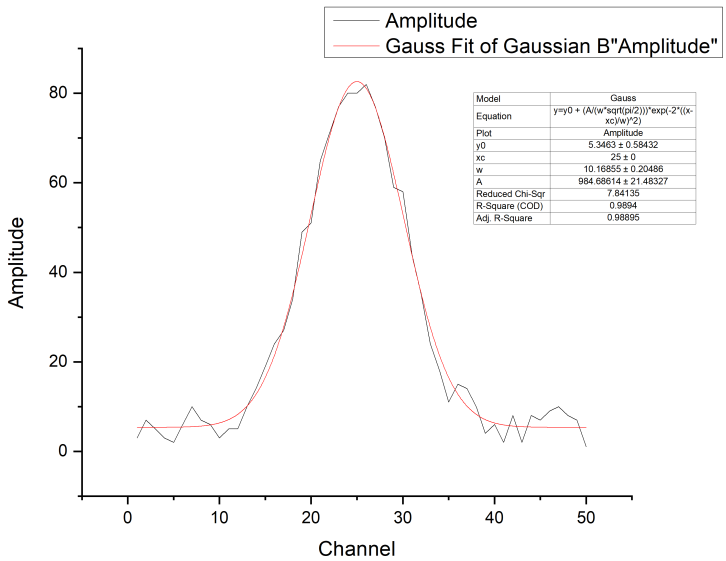


Figure 21: Second Gaussian Fit for Data Set

1.6 Task (VI): Peak Analysis with Deconvolution and Baseline

(a) *Multiple Peak Fit with Deconvolution*
We fit the peaks with different gaussian functions (7 gaussians for 7 peak). In figure 24, we show that gaussianss fitted to the original curve. In figure 25, we show the statistics of the fitting.

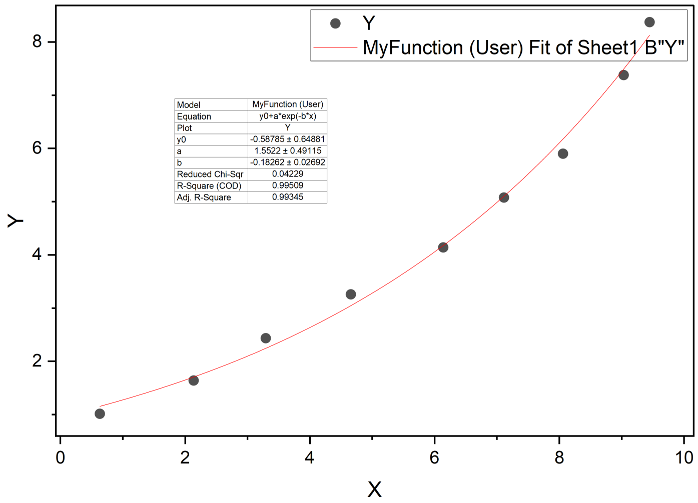
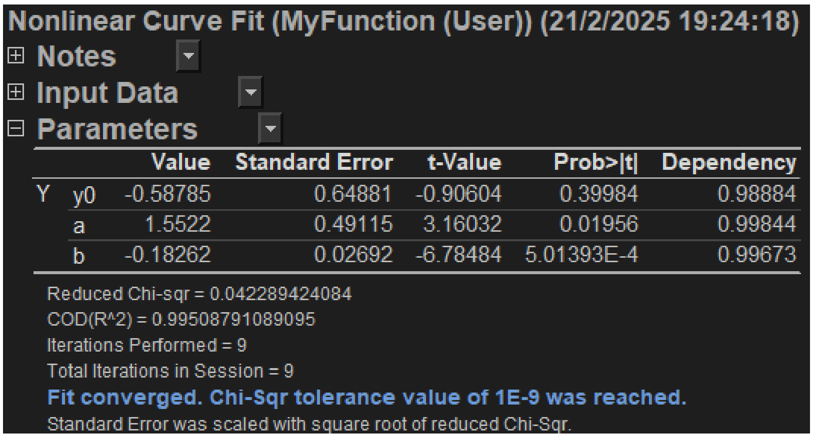


Figure 22: User-defined Fit for Data Set



Nonlinear Curve Fit (MyFunction (User)) (21/2/2025 19:24:18)

⊞ Notes

⊞ Input Data

⊟ Parameters

		Value	Standard Error	t-Value	Prob>|t|	Dependency
Y	y0	-0.58785	0.64881	-0.90604	0.39984	0.98884
	a	1.5522	0.49115	3.16032	0.01956	0.99844
	b	-0.18262	0.02692	-6.78484	5.01393E-4	0.99673

Reduced Chi-sqr = 0.042289424084
COD(R^2) = 0.99508791089095
Iterations Performed = 9
Total Iterations in Session = 9
Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.
Standard Error was scaled with square root of reduced Chi-Sqr.

Figure 23: Parameters of User-defined Fit

We see that the reduced Chi-Sqr value is extremely tiny ($\sim 10^{-12}$), while the adjusted R-squared value is 1, indicating that the fit is extremely good.

(b) *Fit Peaks with Baseline*
We fit the peaks with a baseline that we make, and show the fitting in figure 26. We also show the statistics of the fitting in figure 27.

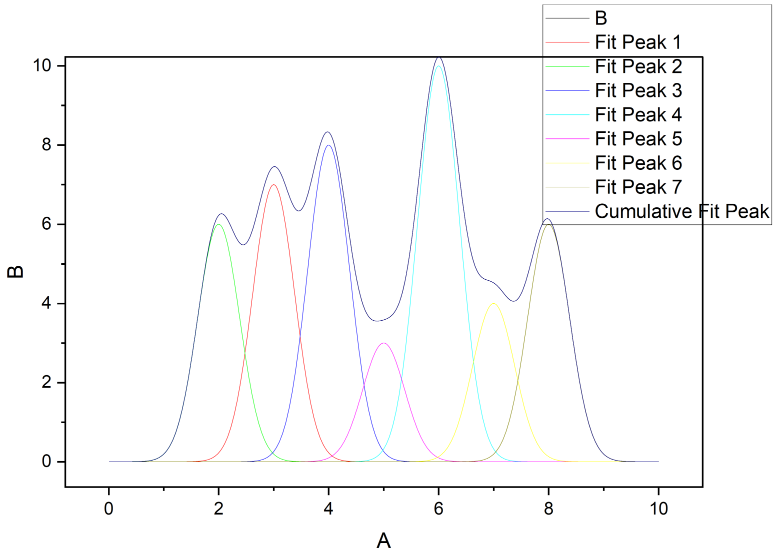
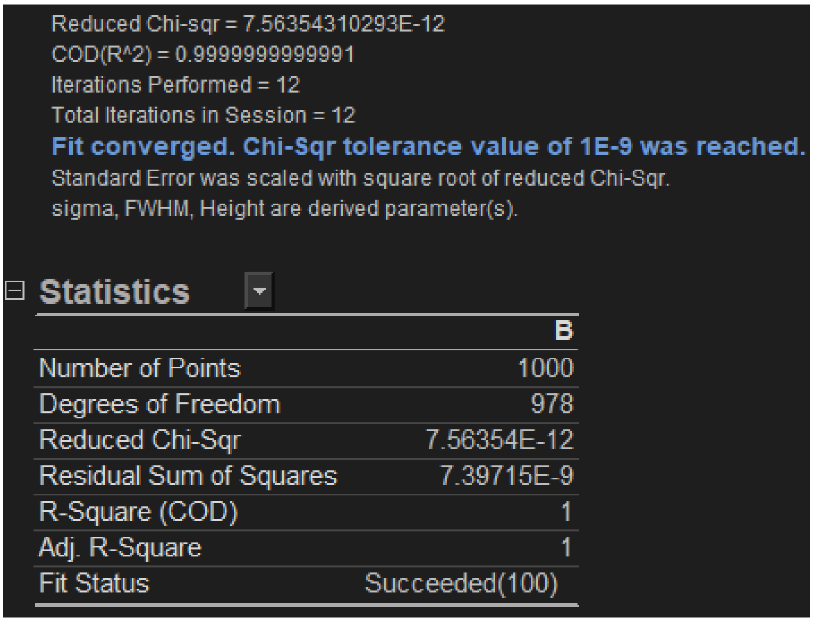


Figure 24: Multiple Peak Fit with Deconvolution



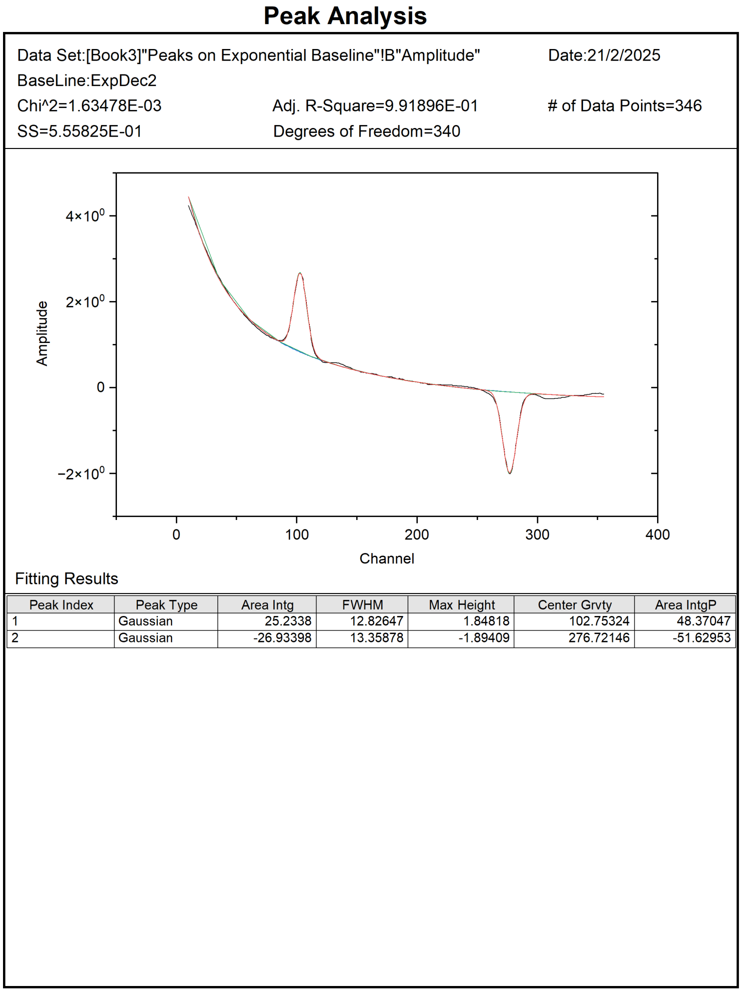
Reduced Chi-sqr = 7.56354310293E-12
COD(R^2) = 0.9999999999991
Iterations Performed = 12
Total Iterations in Session = 12
Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.
Standard Error was scaled with square root of reduced Chi-Sqr.
sigma, FWHM, Height are derived parameter(s).

Statistics

	B
Number of Points	1000
Degrees of Freedom	978
Reduced Chi-Sqr	7.56354E-12
Residual Sum of Squares	7.39715E-9
R-Square (COD)	1
Adj. R-Square	1
Fit Status	Succeeded(100)

Figure 25: Statistics of Multiple Peak Fit

We see that the reduced Chi-Sqr value is very small ($\sim 10^{-3}$), while the adjusted R-squared



Peak Index	Peak Type	Area Intg	FWHM	Max Height	Center Grvty	Area IntgP
1	Gaussian	25.2338	12.82647	1.84818	102.75324	48.37047
2	Gaussian	-26.93398	13.35878	-1.89409	276.72146	-51.62953

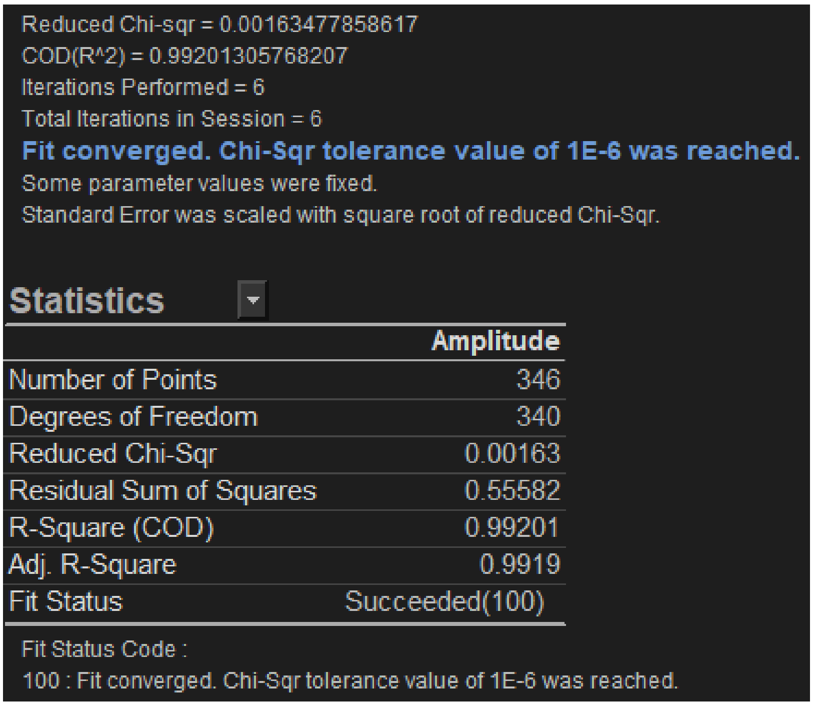
Figure 26: Fit Peaks with Baseline

value is 0.9919, indicating that the fit is really good.

1.7 Task (VII): Numerical integration using the integration Gadget

(a) *Integrate and output the quantities*
We integrate a single peak from the multiple peaks dataset. The results of the integration (area) as well as the maximum of the peak, and the full width at half maximum (FWHM). We show the results in figure 28.

This integration method is very useful if we have a large enough dataset such that we can


Reduced Chi-sqr = 0.00163477858617
COD(R^2) = 0.99201305768207
Iterations Performed = 6
Total Iterations in Session = 6
Fit converged. Chi-Sqr tolerance value of 1E-6 was reached.
Some parameter values were fixed.
Standard Error was scaled with square root of reduced Chi-Sqr.

Statistics

	Amplitude
Number of Points	346
Degrees of Freedom	340
Reduced Chi-Sqr	0.00163
Residual Sum of Squares	0.55582
R-Square (COD)	0.99201
Adj. R-Square	0.9919
Fit Status	Succeeded(100)

Fit Status Code :
100 : Fit converged. Chi-Sqr tolerance value of 1E-6 was reached.

Figure 27: Statistics of Fit Peaks with Baseline

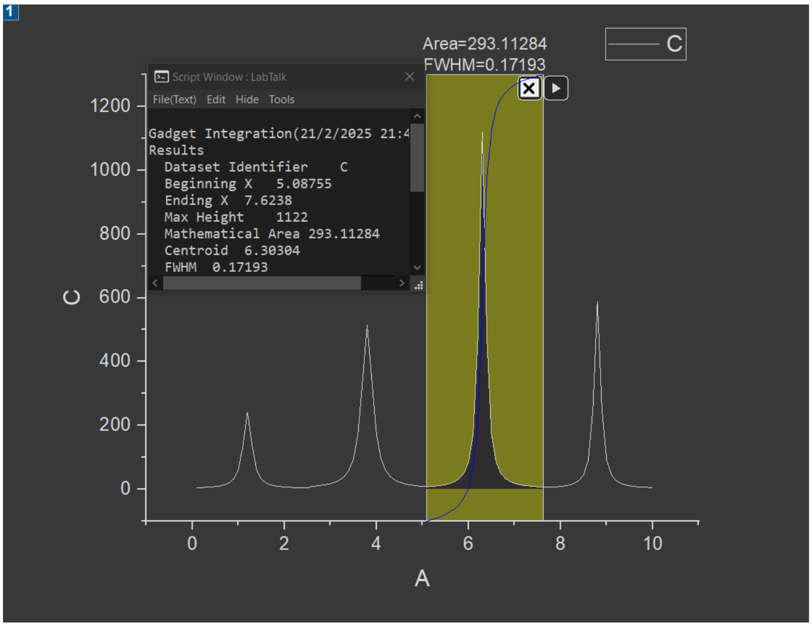


Figure 28: Integration Results

interpolate between them well enough to form a continous curve for which we can use this method on. Otherwise, if the graph has an insufficient data set, the interpolation of the data points may be a bad representation for the underlying phenomenon, such that the area will be inaccurate. A minor drawback, which can be easily resolved, is that setting the start and end points of the rectangle may not be very reliable when done using a mouse. This could be solved by just writing the point at which it should start integrating.

(b) *Integrate with a data plot baseline*
We now integrate a data plot with a baseline. Meaning that we do not get the area bounded by the x-axis, but by a baseline that we input in. We show the results in figure 29.

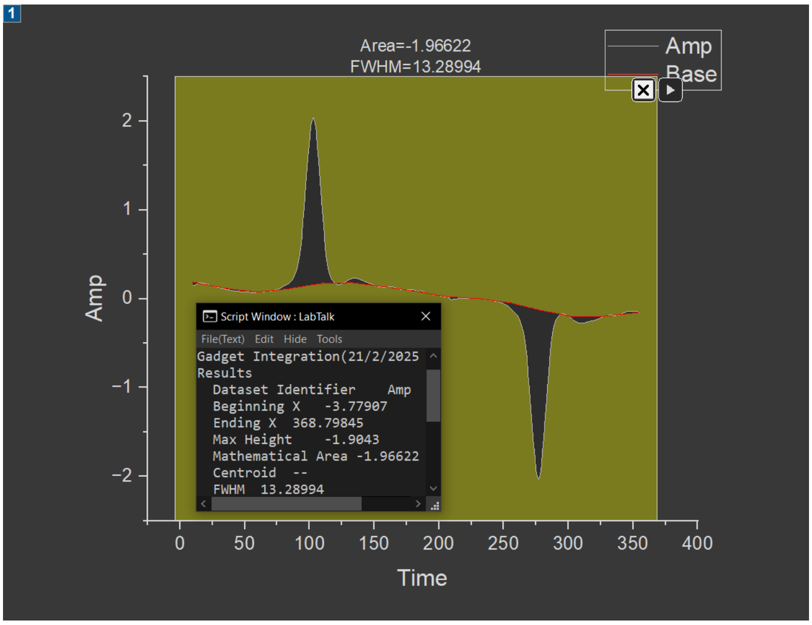


Figure 29: Integration with Baseline Results

This method is useful when we want to integrate a peak that is not bounded by the x-axis, but by some other line. This is commonly used in performing Ultraviolet-visible spectroscopy (UV-VIS). The accuracy of the results here will heavily depend on the accuracy of the readings of the baseline.

2 Experiment (II): Modeling Cosmic Background radiations and High Energy Emissions of Neutron Stars

2.1 Task (I): Modeling Cosmic Background Radiation

(a) *Analyzing the data (Descriptive statistics)*
We have 100 readings indicating the number of photons recieved per 10 seconds. These are their descriptive statistics:

Descriptive Statistics

	N total	Mean	Standard Deviation	Sum	Minimum	Median	Maximum
Data1	100	14.74	3.6725	1474	7	14.5	25

Figure 30: Descriptive Statistics of CMBR

The mean number of photons is 14.74, and the standard deviation is 3.6725. The median is 14.5, which aligns with the mean which suggets that the number of photons might be normally distributed. This would suggest that the dispersion is random. Further investigation must be done to make a call.

(b) *Analyzing the data (Histogram)*
We present the data set as a histogram in figure 31 by plotting the frequency of the number of photons recieved per 10 seconds.

We will now try to fit the data against many different models, to see which fits best.

(c) *Modeling the data*

i. **Gaussian**:
We fit the data to a Gaussian model, and show the fitting in figure 32, and show the statistics of the fitting (χ^2 value and adjusted R^2 value) in figure 33.

We see that the reduced χ^2 value is very large ≈ 10.68, while the adjusted R-squared value is somewhat small at 0.85611, indicating that the fit is not that good. We hope the other models will prove more succesful.

ii. **Poisson**:
We fit the data to a Poisson model, and show the fitting in figure 34, and show the statistics of the fitting (χ^2 value and adjusted R^2 value) in figure 35.

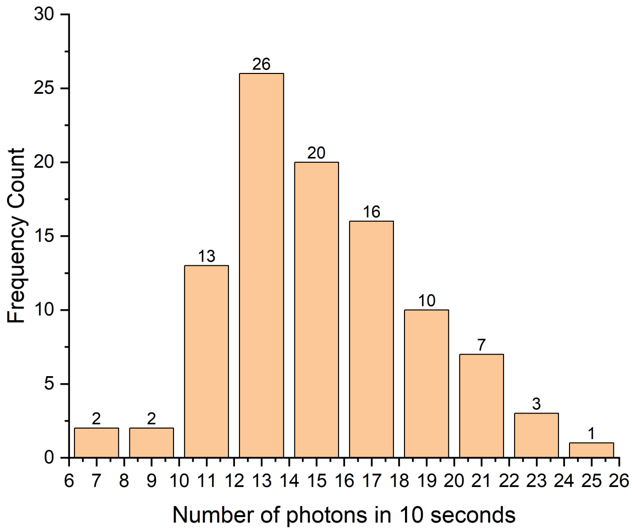


Figure 31: Histogram of CMBR

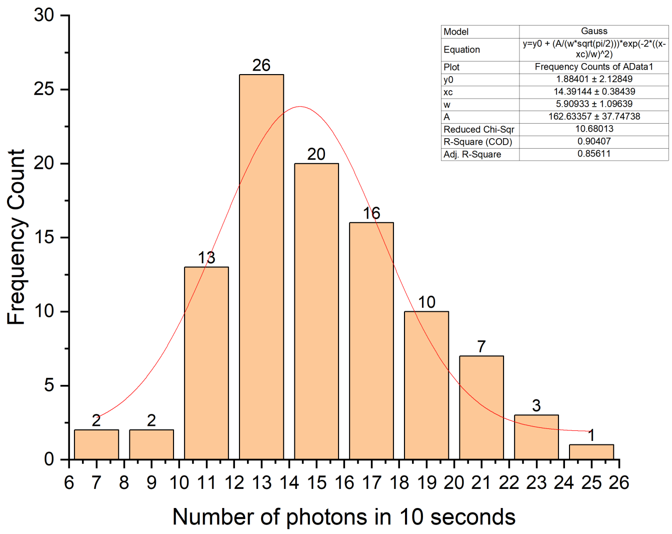
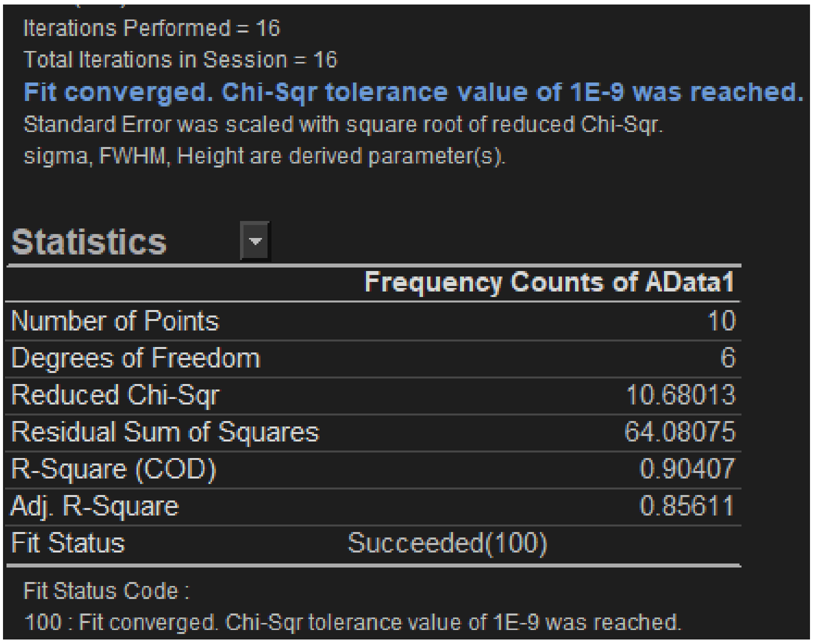


Figure 32: Gaussian Fit for CMBR

We see that the reduced χ^2 value is very large ≈ 83.5, while the adjusted R-squared value



Iterations Performed = 16
Total Iterations in Session = 16
Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.
Standard Error was scaled with square root of reduced Chi-Sqr.
sigma, FWHM, Height are derived parameter(s).

Statistics

	Frequency Counts of AData1
Number of Points	10
Degrees of Freedom	6
Reduced Chi-Sqr	10.68013
Residual Sum of Squares	64.08075
R-Square (COD)	0.90407
Adj. R-Square	0.85611
Fit Status	Succeeded(100)

Fit Status Code :
100 : Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Figure 33: Statistics of Gaussian Fit for CMBR

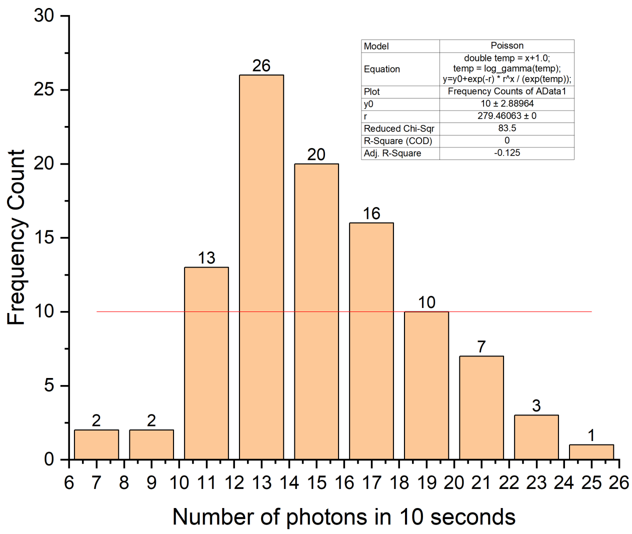


Figure 34: Poisson Fit for CMBR

is 0, indicating that the fit is terrible. We go ahead and move onto the next model.

iii. **Lorentzian**:

Total Iterations in Session = 3

Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Standard Error was scaled with square root of reduced Chi-Sqr.

Statistics

	Frequency Counts of AData1
Number of Points	10
Degrees of Freedom	8
Reduced Chi-Sqr	83.5
Residual Sum of Squares	668
R-Square (COD)	0
Adj. R-Square	-0.125
Fit Status	Succeeded(100)

Fit Status Code :

100 : Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Figure 35: Statistics of Poisson Fit for CMBR

We fit the data to a Lorentzian model, and show the fitting in figure 36, and show the statistics of the fitting (χ^2 value and adjusted R^2 value) in figure 37.

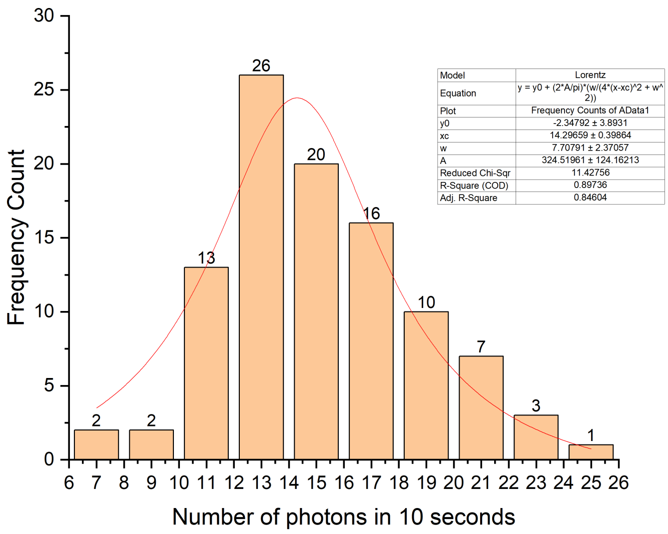


Figure 36: Lorentzian Fit for CMBR

Total Iterations in Session = 25

Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Standard Error was scaled with square root of reduced Chi-Sqr.

H are derived parameter(s).

Statistics

	Frequency Counts of AData1
Number of Points	10
Degrees of Freedom	6
Reduced Chi-Sqr	11.42756
Residual Sum of Squares	68.56535
R-Square (COD)	0.89736
Adj. R-Square	0.84604
Fit Status	Succeeded(100)

Fit Status Code :

100 : Fit converged. Chi-Sqr tolerance value of 1E-9 was reached.

Figure 37: Statistics of Lorentzian Fit for CMBR

We see that the reduced χ^2 value is larger than the gaussian ≈ 11.43, while the adjusted R-squared value is bigger than for the gaussian at 0.89736. The Gaussian and the Lorentzian are somewhat comparable in their reliability of representing the model. More data points are required, to differentiate which is more fit to represent the model.

2.2 Task (II): Modeling Magnetar Bursts

(a) *Distribution of Fitting Paramters*

Here we have the combined histograms of the frequency counts of magnetar emissions with a given photon index for long and short Bursts to test the PL+BB model. We fit both data sets to Lorentzians, since after trying Gaussians a lorentzian fit gave a smaller χ^2 value. We show the fitting in figure 38.

We see that the reduced χ^2 value for the short bursts is 0.10256, and the adjusted R^2 value is 0.98168. Which indicates that the lorentzian fit is an extremely good fit for the short bursts data.

However for the long bursts we have a reduced χ^2 value of 2.01958, and the adjusted R^2 value is 0.92115. Which indicates that the fit is really good, but not as fit as for the short bursts.

(b) *Analyzing Flux and Fluence emissions of Magnetars*

The all flux emissions (figure 39) fits a lorentzian model the best, however the χ^2 value is way too high at 58.55. It seems like it follows an exponential decay, however fitting it to an exponential decay using Origin did not converge. The flux ratio emissions (figure 40) fits a

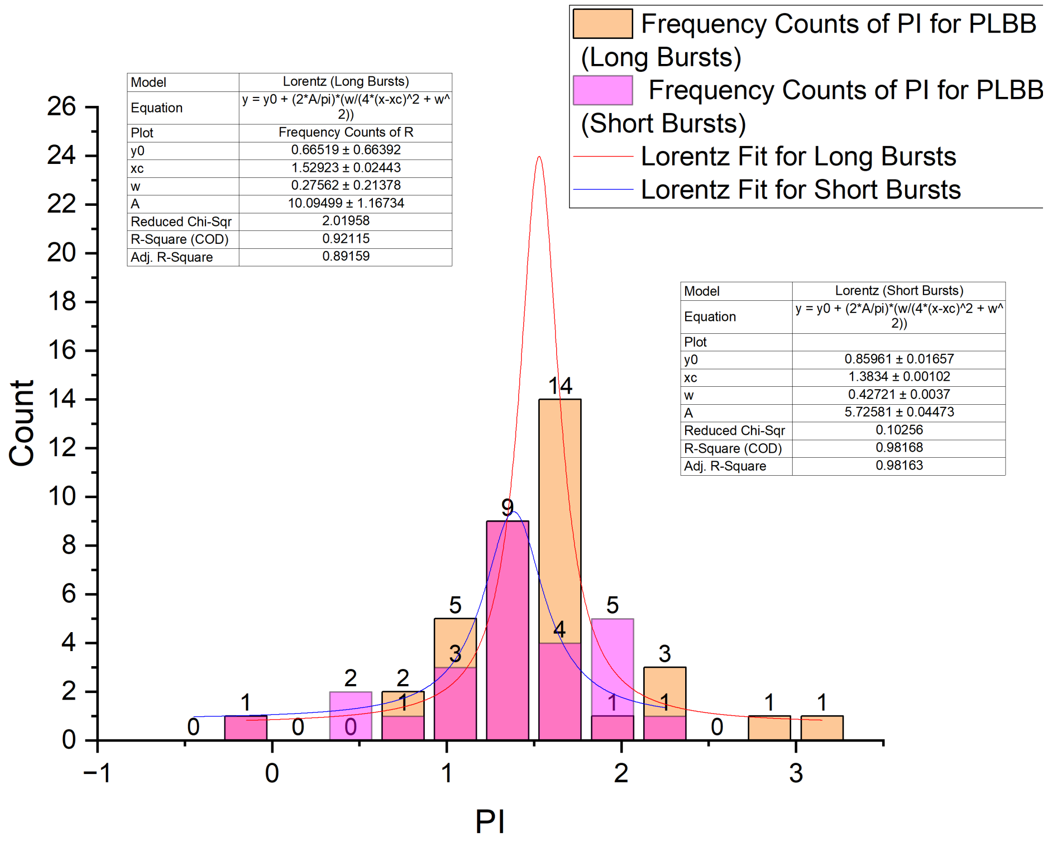


Figure 38: Lorentzian Fit for Magnetar Emissions

lorentzian model the best, with a reduced χ^2 value of 13.678, and an adjusted R^2 value of 0.83666. This indicates that the lorentzian fit might be a good model for flux ratio emissions data.

For the all fluence data (figure 41), it fit an exponential decay really well with a χ^2 value of 17.53, and an adjusted R^2 value of 0.97293. Suggesting, that an exponential model is a good fit.
For the fluence ratio emissions (figure 42), the lorentzian is a good fit since it has a χ^2 value of 4.93302, and an adjusted R^2 value of 0.97454.

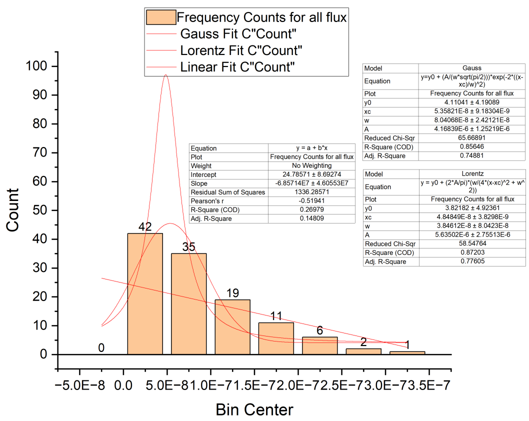


Figure 39: All Flux Emissions

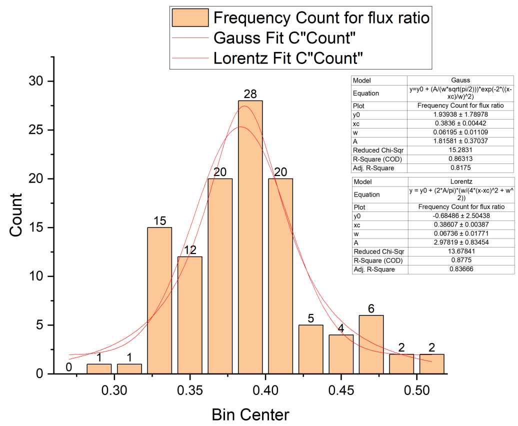


Figure 40: Flux Ratio Emissions

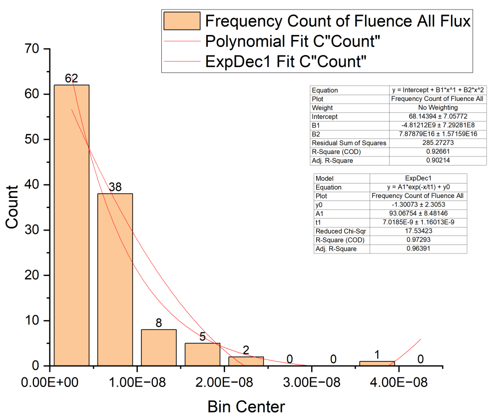


Figure 41: All Fluence Emissions

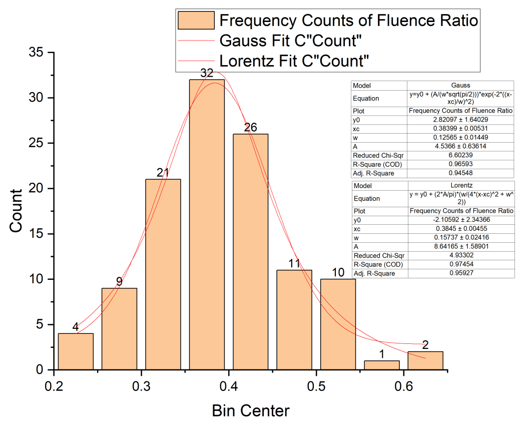


Figure 42: Fluence Ratio Emissions