An Analysis of Missing Transverse Momentum Triggers for Improving Efficiency at the ATLAS Experiment at CERN

by

Joseph Corrado

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Bachelors with Honors

in the
Department of Physics
Faculty of NYU Physics

NEW YORK UNIVERSITY
Spring 2019

Approval

Name:	**Joseph Corrado**
Degree:	**Bachelors with Honors (Physics)**
Title:	**An Analysis of Missing Transverse Momentum Triggers for Improving Efficiency at the ATLAS Experiment at CERN**
Examining Committee:	**Chair:** Allen Mincer Senior Supervisor Professor
	Kyle Cranmer Professor
	Daniel Zwanziger Dean of Undergraduate Studies Professor
Date Defended:	**May 20th, 2019**

Acknowledgements

Firstly, I would like to thank my principal investigator, Allen, for always being available to answer my questions, and to have seemingly infinite patience when I try to dig deep into my understanding of our research.

In addition, I would like to thank my dear friend and colleague, Don, for always being there for me, and specifically for always being immensely helpful whenever I wanted to bounce off any questions for my understanding and how to solve problems in our research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Large Hadron Collider

The Large Hadron Collider is the most powerful particle accelerator in the world, located in CERN on the France-Swiss border. The circumference of the LHC is 27km, and is located 100m underground. The maximum energy possible to accelerate the particles to in the LHC is directly dependent on the size of the LHC and the strength of the magnets used to accelerate the particles. In order to achieve the design energy of 7 TeV per proton, it was necessary to have a circumference of 27km, and to use some of the most powerful dipoles and radiofrequency cavities in existence. Inside the pipes where the protons travel, a very strong vacuum is required, and so the pressure in some parts is over $10^{-9}Pa$. The beams are made up of cylinder-like bunches. Using these bunches, the expected number of collisions is 10^{34} per cm^2 per second. If we take into account the cross section of the proton-proton interaction, roughly 100mBarns (1Barn = 10^{-24}cm^2), then the expected number of proton proton interactions is:

$$10^{34}\text{cm}^{-2}\text{s}^{-1}(100\text{mBarns})\frac{1\text{Barn}}{1000\text{mBarn}}\frac{10^{-24}\text{cm}^2}{1\text{Barn}} = 10^{9}\text{s}^{-1} \tag{1.1}$$

The time between bunches is about $25ns$, so in theory, there could be about 3550 bunches in the entire circumference of the LHC, but in practice there are only usually a maximum of 2556 bunches in either direction when the LHC is filled.

1.2 ATLAS Experiment

ATLAS (A Toroidal LHC ApparatuS) is one of seven particle detector experiments constructed at the Large Hadron Collider, a particle accelerator at CERN. When the LHC runs at full energy and intensity, there are bunch crossings every 25 nanoseconds, which mean there are about 25 million proton-proton collisions taking place every second inside the ATLAS detector. The amount of data collected for each event is around 1MB. For the ATLAS experiment, there is about a gigabyte of data collected each second. Because

this is significantly larger than any practical system can handle, there are **triggers** that are designed to reject uninteresting events and keep the interesting ones. For ATLAS, the trigger system is designed to run at about 1kHz or 1000 events per second. This means that ATLAS collects about 4 petabytes of data per year. There are 10^{11} protons in a bunch. The proton-proton interaction cross section is approximately $100mB$.

1.3 Trigger System

In particle physics, a trigger is a system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded. The trigger system is necessary because of limitations in terms of data storage capacity and rates. In general, the experiments typically search for "interesting" events (decays of rare particles) that occur at relatively low rates, so we need to have trigger systems that identify events that should be recorded for later analysis. The Large Hadron Collider has an event rate of approximately 1 GHz. The triggers are divided into levels so that each level selects the data that becomes an input for the next level, which has more time available and more information to make better decisions. There is the **Level-1 (L1)** system, which is based on custom electronics, and the **High Level Trigger (HLT)** system, that relies on commercial processors. The L1 system uses only coarsely segmented data from the calorimeter and muon detectors, while holding all the high-resolution data in pipeline memories in the electronics.

1.4 Missing Transverse Momentum

The transverse momentum is defined to be the momentum in the transverse plane to the beam axis. Because the protons in the beam pipe collide approximately head-on in opposite directions, we expect the produced particles to have approximately zero momentum. As a result, we know we must conserve transverse momentum. Longitudinal momentum is also conserved, but because it is very difficult to measure the momentum in the direction of the axis of the beam pipe, the detector is not able to measure it. The total transverse momentum is defined to be the vector sum of the transverse momenta of the particles produced in the beam pipe. Missing transverse momentum is defined as the vector sum of the transverse momenta of all invisible particles. The missing transverse momentum can be understood as the amount by which produced particles are failing to obey conservation of the three-momentum. Therefore, by only using muon-triggered events that passed a W transverse mass cut enriches the W and therefore real MET content.

1.5 Efficiency Curves

An efficiency curve illustrates the probability of an algorithm to classify the MET of an event as above or below a threshold value as a function of the MET as determined by

another, "true" algorithm. So we can ask, what is the efficiency of L1 $>$ 30 as a function of CELL MET. What this is means is we are taking the MET as determined by CELL to be the true MET, and we want to know how well (what the probability is) L1 does at classifying events as having MET above or below 30 at each value of the MET determined by CELL. The way one would read a plot of this efficiency is to pick a value of CELL MET (on the x-axis) and ask "of the number of events CELL determined had this MET, what fraction did L1 determine had MET greater than the threshold [30 GeV]". The fraction [of the total amount of events CELL determined was in that MET bin] that L1 determined was greater than the threshold would be the height of the efficiency curve at that value of CELL MET. A perfect efficiency curve would look like a step function centered at the threshold around which one is trying to classify the MET of events. Consider figure 1.1. This is a plot of the efficiency of $L1 > 60GeV$ as a function of L1 MET. It is a step function because if you take a point on the x axis lower than the threshold, you would expect there to be no events L1 said was greater than the threshold, thus yielding a zero efficiency. The same reasoning holds for points to the right of the threshold. The fact that efficiency curves in reality do not look like step functions can be understood in terms of Type I and Type II error. The step function for the efficiency curve would be centered on the threshold one is asking for the efficiency about. Consider figure 1.2. This is a plot of the efficiency of $L1 > 60GeV$, as a function of CELL MET. The fact that the efficiency curve immediately to the left of the threshold is not zero means that, of the events CELL determined had MET lower than the threshold, L1 determined some fraction of those events had an MET higher than the threshold. The fact that the efficiency curve immediately to the right of the threshold is not one means that, of the events CELL said had MET larger than the threshold, L1 said some fraction of those events had MET lower than the threshold. In this case, the fraction of events L1 determined had an MET higher than the threshold, given that CELL said the MET was higher than the threshold, is less than one.

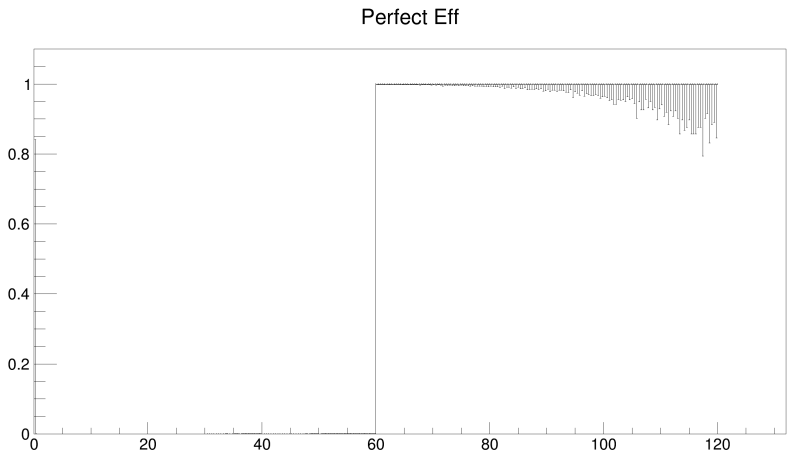

Figure 1.1: An example of a perfect efficiency curve for L1 > 60GeV

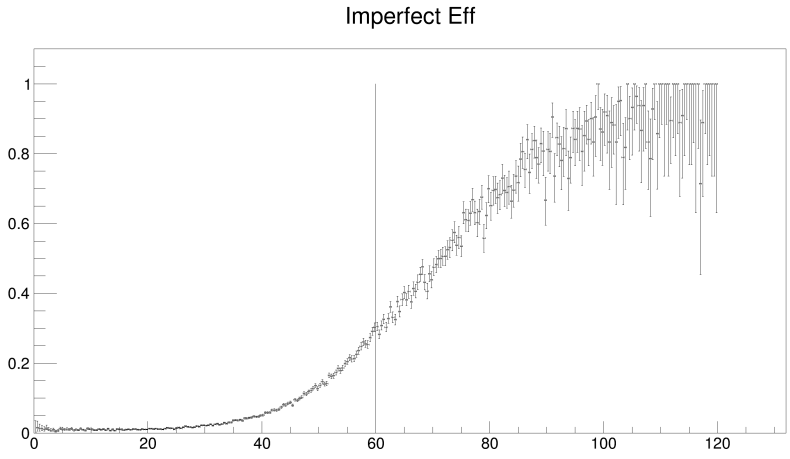

Figure 1.2: An example of an imperfect efficiency curve for L1 > 60GeV

Chapter 2

Improving Efficiency Using Combined Algorithms

For our project, we wanted to improve the efficiency of the algorithms in classifying the MET of a signal event as larger than some MET. The way we do this is by determining the thresholds needed to satisfy the trigger rate based on the fraction of background events kept by the trigger system, and then by using those thresholds to compute an efficiency for the same pair of algorithms to classify a signal event as larger than some value of MET. By signal events, we mean events for which the muon trigger fired, which correspond to events for which a muon was detected. In this assignment, we wanted to see if we could obtain an increase in efficiency by combining uncorrelated algorithms together subject to the constraint of the trigger rate. The issue is that there are many pairs of thresholds that when combine will keep the proper trigger rate. In the parameter space of all possible combinations of combined thresholds, we have two degrees of freedom in determining what the appropriate thresholds on a pair of algorithms should be. In order to simplify the problem, I imposed the constraint that the two trigger rates for the algorithms had to be the same. Therefore, I removed a degree of freedom in the parameter space by constraining my solution space to pairs of thresholds that satisfy the trigger rate and lie on the line $y = x$. The monotonicity of the fraction of events kept as you change either of the thresholds on the algorithms guarantees that there will be a unique solution to both of these constraints on the parameter space.

Once we added the constraint of keeping the same fraction individually, the problem essentially became one-dimensional, and I used the popular bisection root-finding method in order to compute the solution to the optimal pair of thresholds.

This consisted of guessing an individual fraction for each of the algorithms to keep, and then computing what the combined fraction of events kept was. We iterated this process until the combined fraction kept matched the trigger rate to within one bin. The level curve describing the set of pairs of thresholds such that the trigger rate constraint is satisfied is

given by the constraint:

$$f(\tau_\alpha, \tau_\beta) = C$$

for some C. Here, f is the function representing the fraction of events kept when the algorithms are used together at the same time. This C is determined by the fraction of background events that are kept by the trigger system. In order to compute C, we used the fraction of passnoalg data that passed an L1 MET cut of 50 GeV and a CELL MET cut of 100 GeV. For our analysis, C turned out to be 0.0059. So we needed to solve the equation $f(\tau_\alpha, \tau_\beta) = 0.0059$. However, because the parameter space is two-dimensional, and the evaluation of f takes a long time (fraction of events kept by both algorithms, and by each one individually), we introduced the constraint that the two individual fractions kept needed to be the same. The solution space for our problem is depicted in figure **??** In summary, we use the background events in order to compute what the thresholds on the algorithms need to be in order to keep the trigger rate, and then we determine the efficiencies of this combination of algorithms at those thresholds from the signal events.

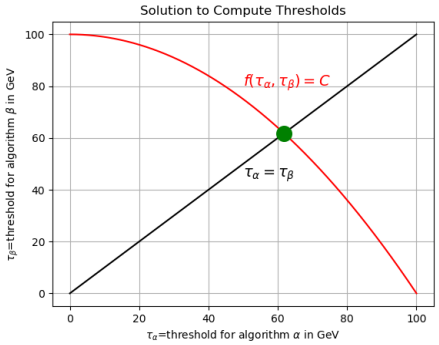

Figure 2.1: Solution to Bisection in Threshold Space

2.1 Signal Selection on Transverse Mass

In addition to the cuts on the various algorithms, we also needed to introduce a cut on the transverse mass that is detected to ensure we only keep events with a transverse mass close to that of the W boson ($80.379 \pm 0.012 GeV/c^2$). We compute the transverse mass using:

$$m_T = \sqrt{2P_\mu P_\nu (1 + \cos(\phi))}$$

In addition to the previously mentioned cuts, we also added a cut on the transverse mass for the range $40 \leq m_T \leq 100$. This was necessary in order to select signal events that are likely to have a high rate of weakly interacting particles produced.

2.2 Results

2.2.1 Algorithms that did not do better combined

For most pairs of algorithms, we found that combining algorithms such that they keep the correct trigger rate did not yield any gain in efficiency. As in figure 2.2, we see that the red curve lies underneath the green curve.

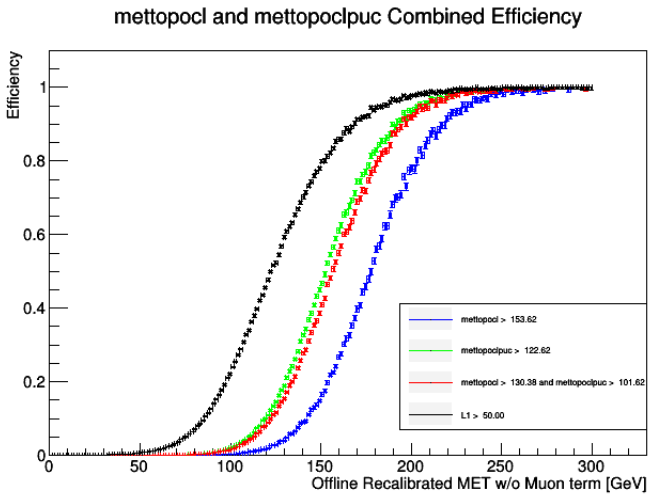

Figure 2.2: Efficiencies of METTOPOCL and METTOPOCLPUC

2.2.2 Algorithms that did do better combined

We found that we were able to achieve an increase in the overall efficiency for some of the pairs of algorithms considered.

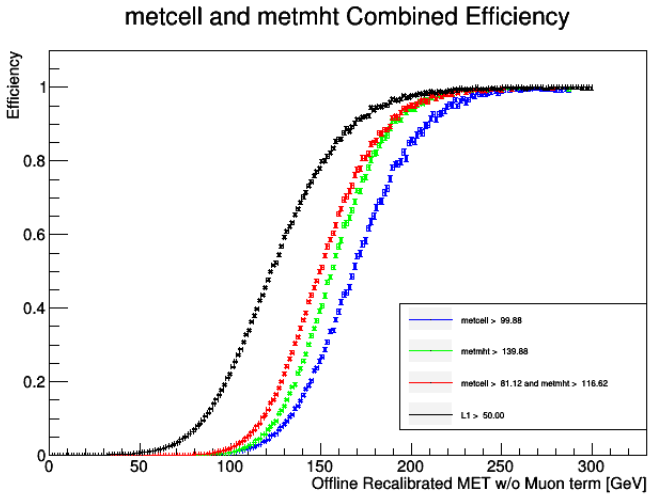

Figure 2.3: Efficiencies of METCELL and METMHT

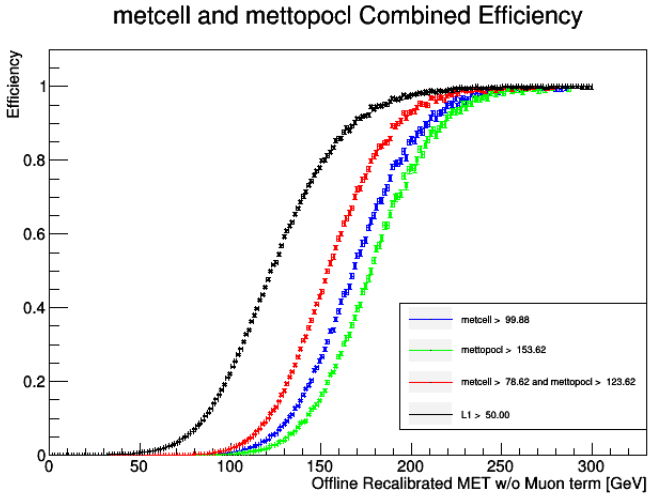

Figure 2.4: Efficiencies of METCELL and METTOPOCL

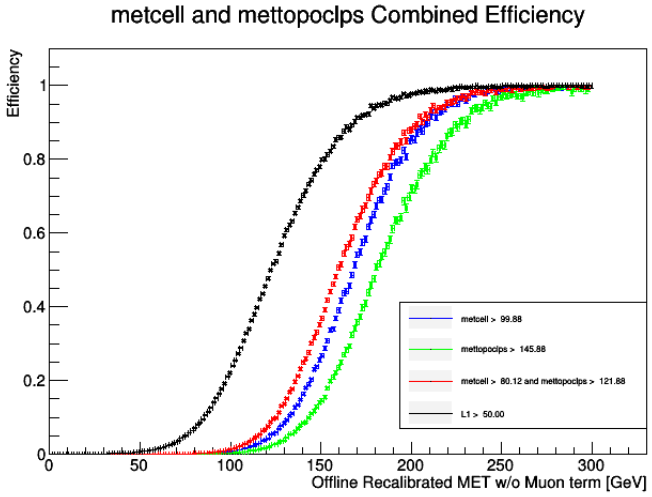

Figure 2.5: Efficiencies of METCELL and METTOPOCLPS

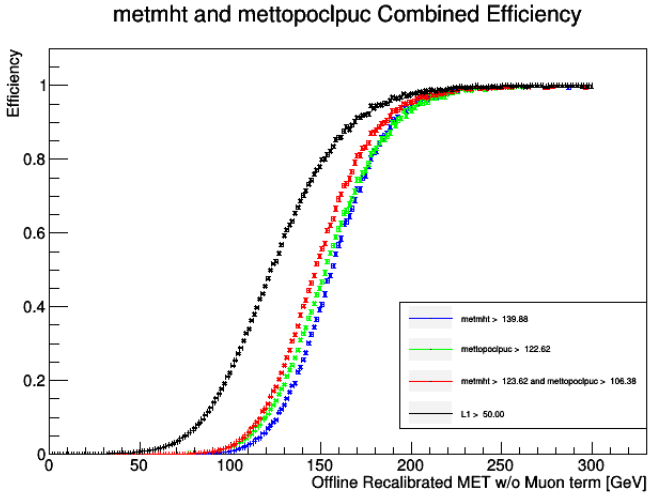

Figure 2.6: Efficiencies of METMHT and METTOPOCLPUC

2.2.3 Best Combined Algorithms and Best Individual Algorithms

During the summer of 2018, the pufit plus cell algorithm trigger was the main one that was actually used for the second part of Run 2. In figure 2.7, I've plotted the efficiencies of the best combined algorithms along with the best efficiencies of the individual algorithms. We see that the top curves on this plot are those for combined algorithms, as well as the individual mettopoclpuc and metmht curves.

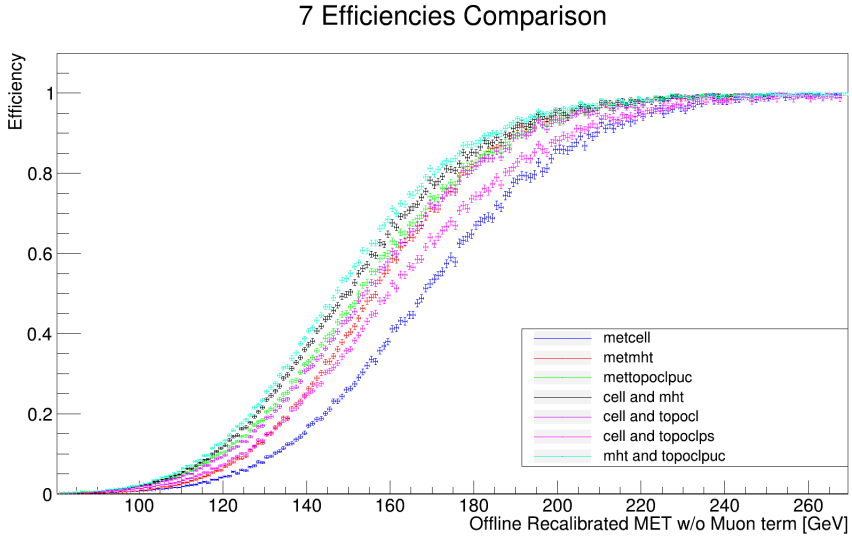

Figure 2.7: Best Combined Efficiency Algorithms Versus Best Individual Algorithms

Chapter 3

Predicting the unbiased MET Distribution at Higher Luminosity

3.1 Objective

Our goal is to empirically reconstruct the unbiased CELL MET distribution using `HLT_noalg_L1ZB`, `HLT_noalg_L1XE30` and `HLT_noalg_L1XE50` data. By this data, I mean the events for which these triggers fired. The `HLT_noalg_L1XExx` triggers are triggers that fire randomly, but only select events that have L1 MET greater than xx. As a result, the data selected by these triggers are biased with respect to L1 (except for the `HLT_noalg_L1ZB`). However, we want to use this data to determine the CELL MET distribution as a function of μ. Because the ZeroBias events only allow us to go up to about 80 GeV, we use the `HLTnoalg_L1XExx` triggered events in order to extend to higher MET. This is important because knowing the distribution of MET as determined by the CELL algorithm as a function of μ helps us to better understand the background rate of events and allows us to test models for this rate. This allows us to get a better handle on predicting what kind of rates of events we'll need to account for when designing the triggers and how the increasing luminosity could affect the efficiency of the trigger system. As a result of the bias with respect to L1, in order to get the unbiased distribution, we must correct the `HLTnoalg_L1XExx` data using the efficiency curves determined from lower threshold triggers back to ZeroBias. While performing this correction, it was important to propagate the errors due to the determination in the efficiency, and those due to statistical uncertainties. Performing the reconstruction involved several steps:

1. compute the efficiency of $L1 > 30$GeV for `HLT_ZB_L1ZB` data as a function of CELL MET

2. correct the `HLT_ZB_L1XE30` data back to the `HLT_ZB_L1ZB` distribution by multiplying by the prescale and dividing by the efficiency.

3. compute the efficiency of $L1 > 50GeV$ for `HLT_ZB_L1XE30` data as a function of CELL MET

4. correct the `HLT_ZB_L1XE50` data back to the `HLT_ZB_L1ZB` distribution by multiplying by the corresponding prescale, and dividing by both of the previously computed efficiencies.

For this project, we used the 2015, 2016 and 2017 combined `HLTnoalg_L1ZB` , `HLTnoalg_L1XE30` and `HLTnoalg_L1XE50` data produced by Jonathan Burr on 11/17/2017 from the zerobias and JETM10 trees. In addition, we removed the events from runs 330203, 331975 and 334487 because these events had large MET events without jets and the logbook says there were calorimeter noise problems in these runs.

3.2 Efficiency Fits

It was necessary to come up with a model that could be used to fit to our efficiency data. To create this model, we assumed that the conditional distribution of the value of MET as determined by L1, given the value of MET as determined by CELL was a gaussian:

$$\mathbb{P}(X = x|Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2\sigma^2} \tag{3.1}$$

Where we have defined:

$$z = y - (ax + b) \tag{3.2}$$

Then, in order to derive the expression for the model for the efficiency curves, we compute the probability that L1 determines as MET value higher than the value of the threshold. In order to do this, we integrate from the value of the threshold (T) to infinity:

$$\varepsilon(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_T^\infty e^{-z^2/2\sigma^2} dz \tag{3.3}$$

One may manipulate this expression using properties of integration, probability density functions and gaussians in order to write an expression for the efficiency in terms of the error function:

$$\text{Erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{3.4}$$

Finally, the expression we get is:

$$\varepsilon(x) = \frac{1}{2}\left(1 + \text{Erf}\left(\frac{ax + b - T}{\sigma\sqrt{2}}\right)\right) \tag{3.5}$$

We performed this fit for the distribution of MET in μ bins of $0 - 10, \ldots, 60 - 70$

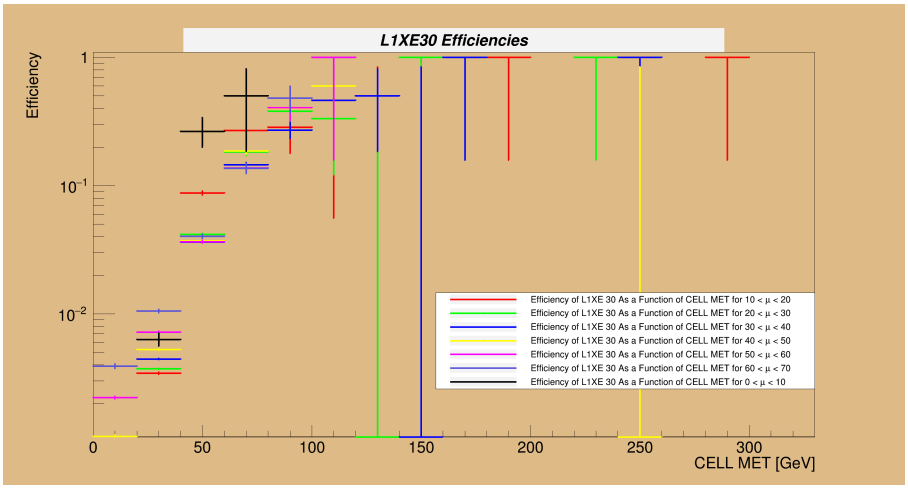

Figure 3.1: Efficiency Curves of L1> 30 on the `HLT_noalg_L1ZB` data

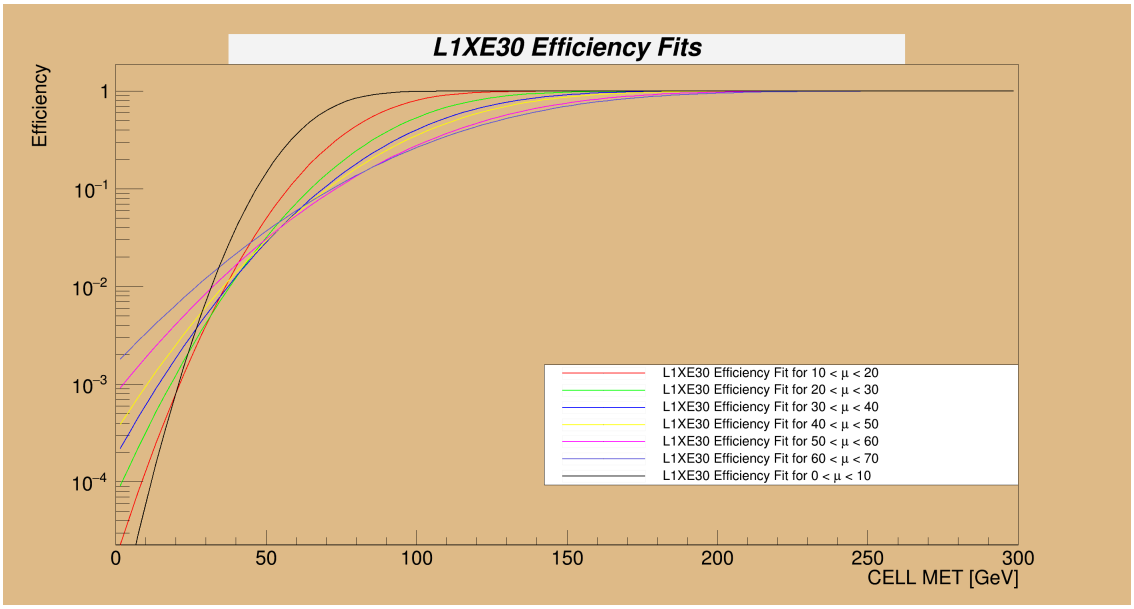

Figure 3.2: Efficiency Curves of L1> 30 on the `HLT_noalg_L1ZB` data

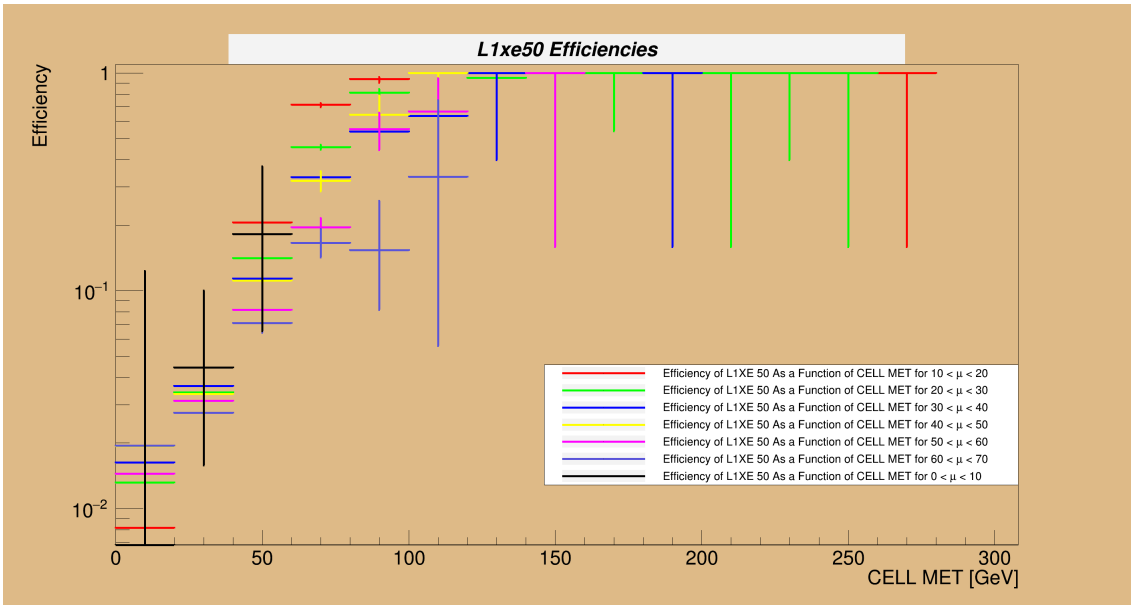

Figure 3.3: Efficiency Curves of L1> 50 on the `HLT_noalg_L1XE30` data

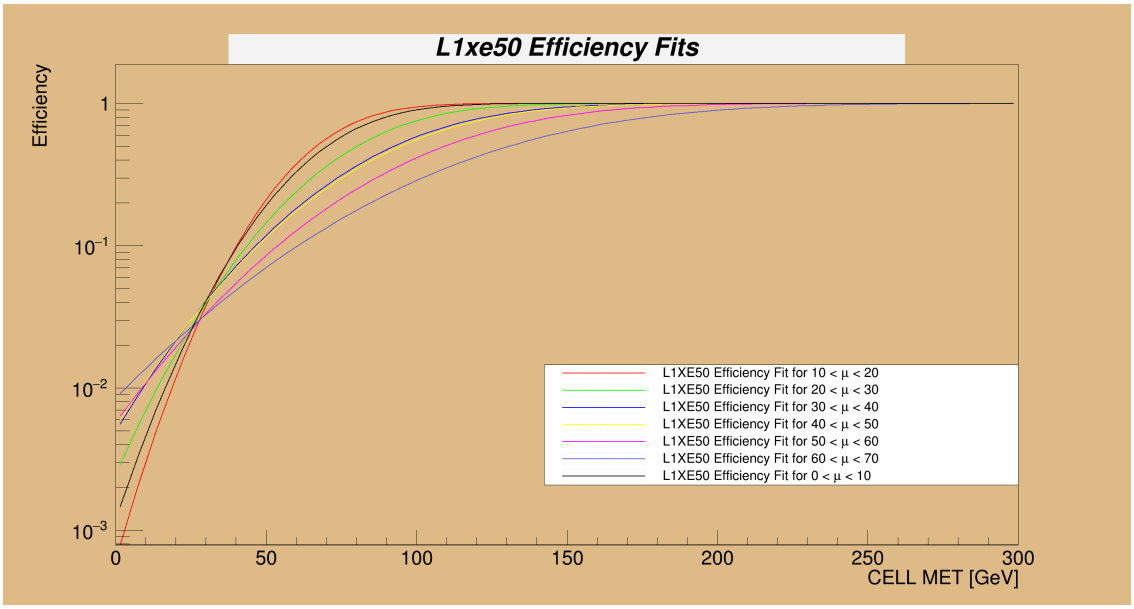

Figure 3.4: Efficiency Fits of L1> 50 on the `HLT_noalg_L1XE30` data

3.3 Error Propagation

In order to obtain accurate error bars on the bins of our reconstructed distribution, it was necessary to appropriately propagate errors due to the determination of the efficiency and the statistical error. Because the value of the prescale varies within a given bin, it was necessary to keep track of the error accumulated on a particular bin manually. In order to do this, the formula for propagation of fractional uncertainties was used. We have the fact that the expression for what the contribution from a single event to the current bin should look like is given by:

$$n_{ZB}^{30} = \frac{P_{30}}{\varepsilon_{ZB}^{30}} \cdot n \tag{3.6}$$

$$\left(\frac{\delta n_{ZB}^{30}}{n_{ZB}^{30}}\right)^2 = \left(\frac{\delta P_{30}}{P_{30}}\right)^2 + \left(\frac{\delta \varepsilon_{ZB}^{30}}{\varepsilon_{ZB}^{30}}\right)^2 + \left(\frac{\delta n}{n}\right)^2 \tag{3.7}$$

$$= \left(\frac{\delta \varepsilon_{ZB}^{30}}{\varepsilon_{ZB}^{30}}\right)^2 + 1 \tag{3.8}$$

Because we know the prescale exactly, we have that $\delta P_{30} = 0$ and $\delta P_{50} = 0$, so those terms drop out. Because this is a single event, we have that $n = 1$ and by the standard deviation of a poisson process $\delta n = \sqrt{n}$, which means that $\delta n = 1$. So the term due to statistical uncertainty contributes unity on the right hand side. In addition to computing the uncertainty on the contribution of a single event with given prescale and efficiency during the correction of the `HLT_noalg_L1XE30` data to the unbiased distribution, we need to also know the uncertainty on the correction of a given event from the `HLT_noalg_L1XE50` data to the unbiased distribution. In order to compute the unbiased distribution from the `HLT_noalg_L1XE50` events, it was necessary to multiply a given event by its recorded prescale (P_{50}), as well as divide by both the efficiency of the `HLT_noalg_L1XE50` data relative tot he `HLT_noalg_L1XE30` data, and the efficiency of the `HLT_noalg_L1XE30` data to the `HLT_noalg_L1ZB` data.

$$n_{ZB}^{50} = \frac{P_{50}}{\varepsilon_{ZB}^{30}\varepsilon_{ZB}^{50}} \cdot n \tag{3.9}$$

$$\left(\frac{\delta n_{ZB}^{50}}{n_{ZB}^{50}}\right)^2 = \left(\frac{\delta P_{50}}{P_{50}}\right)^2 + \left(\frac{\delta \varepsilon_{ZB}^{30}}{\varepsilon_{ZB}^{30}}\right)^2 + \left(\frac{\delta \varepsilon_{ZB}^{50}}{\varepsilon_{ZB}^{50}}\right)^2 + 1 \tag{3.10}$$

$$= \left(\frac{\delta \varepsilon_{ZB}^{30}}{\varepsilon_{ZB}^{30}}\right)^2 + \left(\frac{\delta \varepsilon_{ZB}^{50}}{\varepsilon_{ZB}^{50}}\right)^2 + 1 \tag{3.11}$$

I track of these uncertainties by incrementing the elements of an array that corresponded to the square error of that bin of the histogram, and taking the square root at the end to get

the numeric value for the uncertainty on that bin for both of the corrected distributions. In order to compute the factors ε_{ZB}^{30} and ε_{ZB}^{50}, it was necessary to propagate errors in quadrature. The expression for this computation is given by:

$$(\delta\varepsilon)^2 = \sum_{i=1}^{k} \left(\delta\theta_i \frac{\partial f}{\partial \theta_i} \right)^2 \tag{3.12}$$

where k is the number of parameters in the fitting function we used, f is the fitting function we used, and θ_i is the ith parameter. After some manipulation with partial derivatives and exponential integrals, one can derive that the expression for $\delta\varepsilon_{ZB}^{30}$ is given by:

$$\delta\varepsilon_{ZB}^{30} = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\left(\frac{ax+b-30}{\sqrt{2}\sigma} \right)^2 \right) \left[\delta a^2 x^2 + \delta b^2 + \delta\sigma^2 \left(\frac{ax+b-30}{\sigma} \right)^2 \right]^{1/2} \tag{3.13}$$

3.4 Relative Normalization

Because the error bars are larger at lower values of MET, it is necessary to also perform an overall relative normalization between the `HLT_noalg_L1XExx` data and the `HLT_noalg_L1ZB` data in addition to the normalization to unity. This was done in order to compare the shapes of the curves more easily. The relative normalization factor was computed by taking a weighted average of ratios sampled at various points in the tail of the distribution where the slopes of the distributions were approximately parallel. The expression for the relative normalization factor is given by:

$$\hat{f}_{MLE} = \frac{\sum_{i=1}^{k} \frac{1}{\sigma_i^2} f_i}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \tag{3.14}$$

where k is the number of samples taken from the tail of the distribution in the region where the slopes were approximately parallel. Using the expression for the propagation of fractional uncertainty, one can calculate the error in this expression for the relative normalization factor:

$$\sigma_i(f_i) = f_i \sqrt{\left(\frac{\sigma_{ZB}^{50}}{N_{ZB}^{50}} \right)^2 + \left(\frac{\sigma_{ZB}}{N_{ZB}} \right)^2} \tag{3.15}$$

Note that in this case, we are not propagating the error on a single event, but on a given ratio of bin values. The variables are: σ_{ZB}^{50} is the error on the given bin from the `HLT_noalg_L1XE50` distribution after being corrected to the unbiased distribution, the N_{ZB}^{50} is the value of that bin, and the σ_{ZB} is the error on the corresponding bin from the `HLT_noalg_L1ZB` distribution and N_{ZB} is the number of events in that bin.

3.5 Results

After computing the efficiency curves, their fits and performing the appropriate corrections, we found that we were able to replicate the unbiased distribution remarkably well.

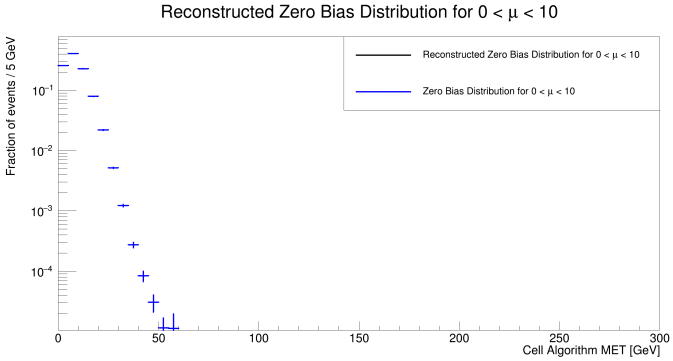

Figure 3.5: Reconstructed Zero Bias distribution for $\mu \in [0, 10)$

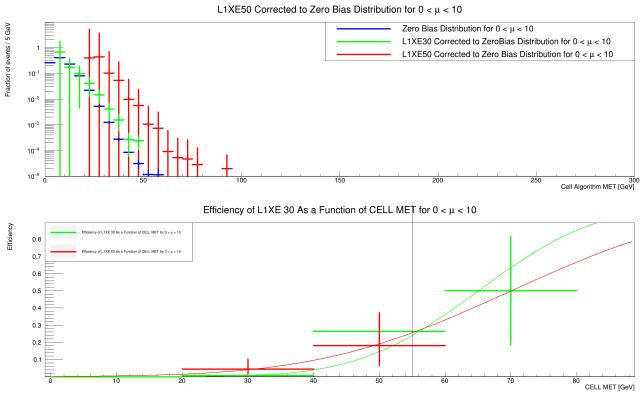

Figure 3.6: Corrected Distributions for $\mu \in [0, 10)$

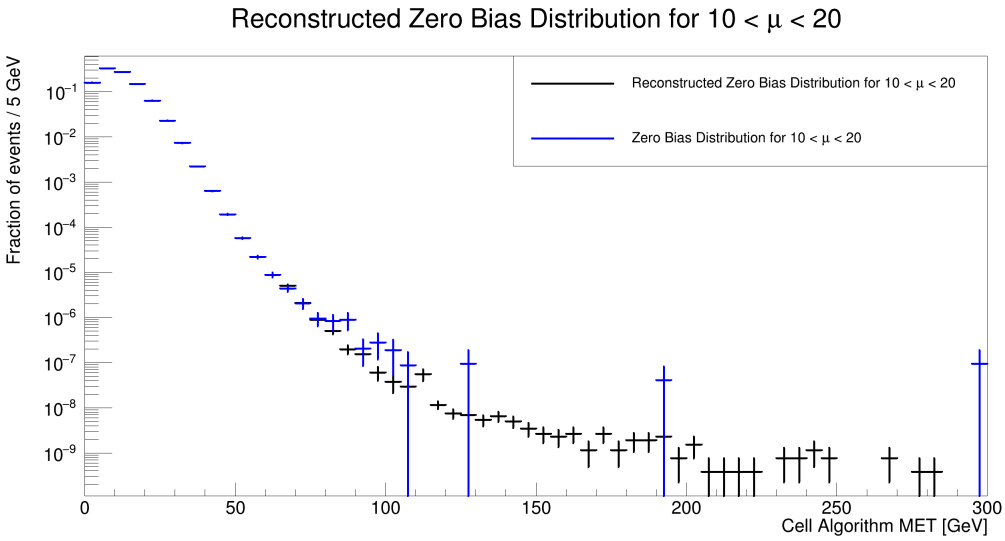

Figure 3.7: Reconstructed Zero Bias distribution for $\mu \in [10, 20)$

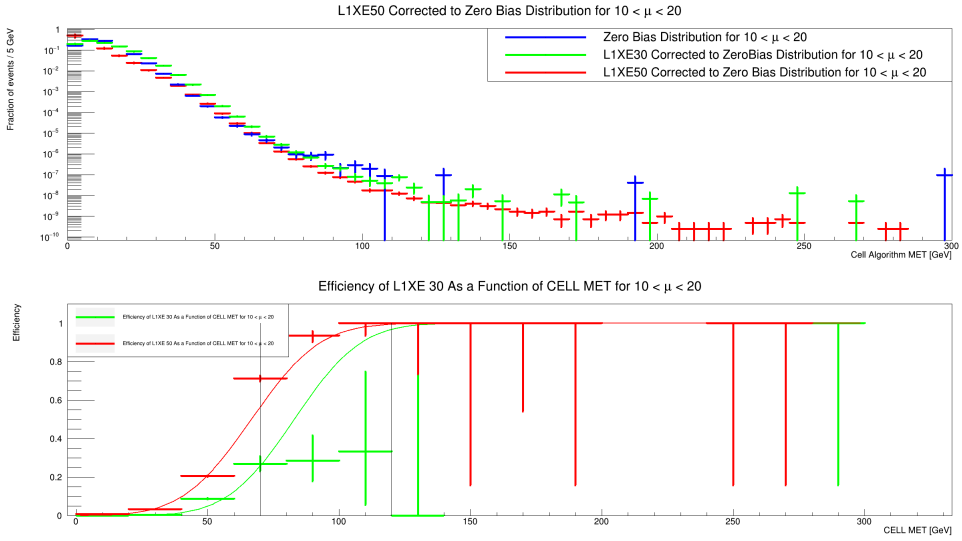

Figure 3.8: Corrected Distributions for $\mu \in [10, 20)$

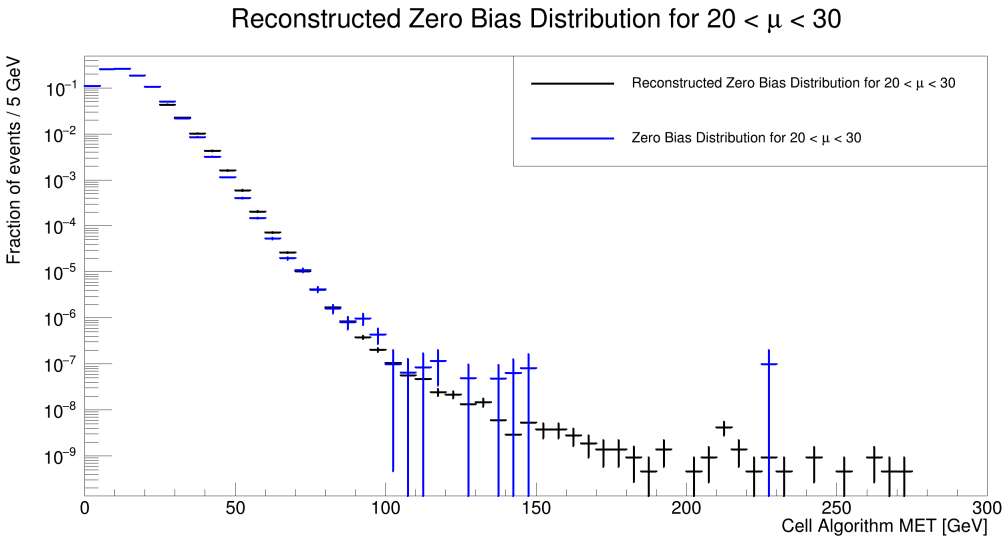

Figure 3.9: Reconstructed Zero Bias distribution for $\mu \in [20, 30)$

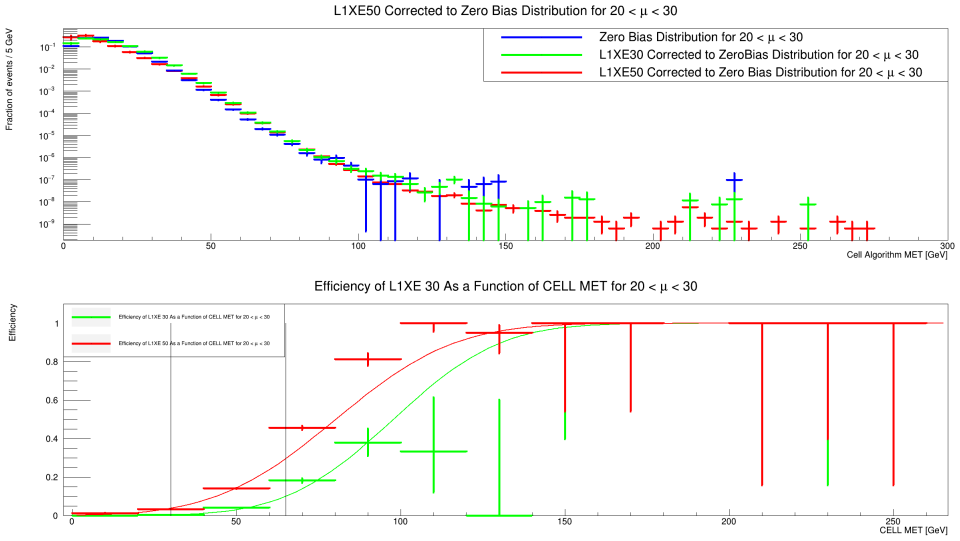

Figure 3.10: Corrected Distributions for $\mu \in [20, 30)$

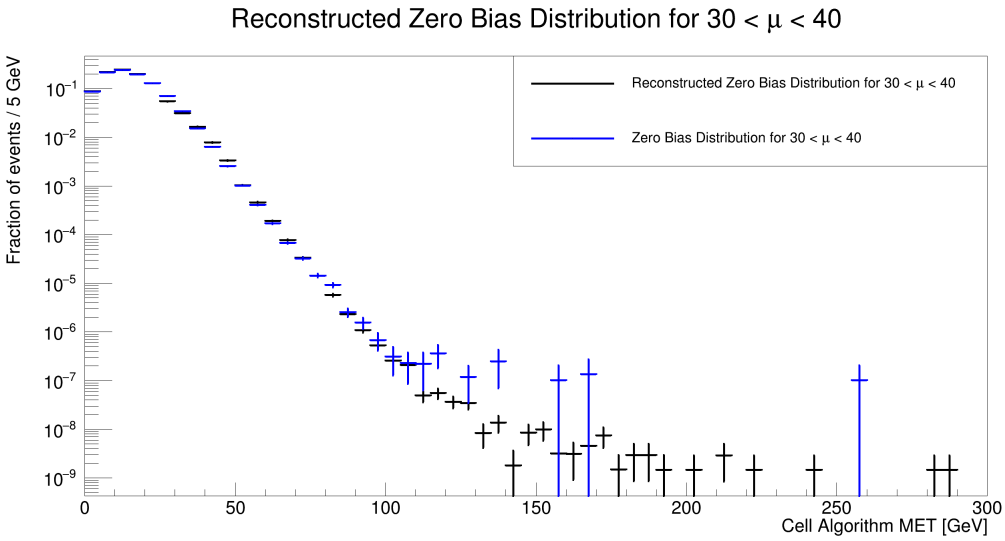

Figure 3.11: Reconstructed Zero Bias distribution for $\mu \in [30, 40)$

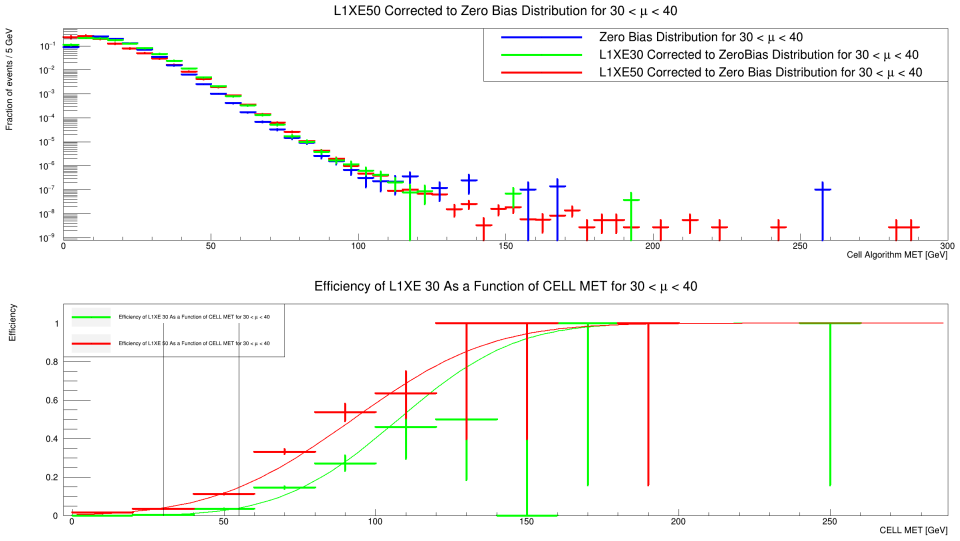

Figure 3.12: Corrected Distributions for $\mu \in [30, 40)$

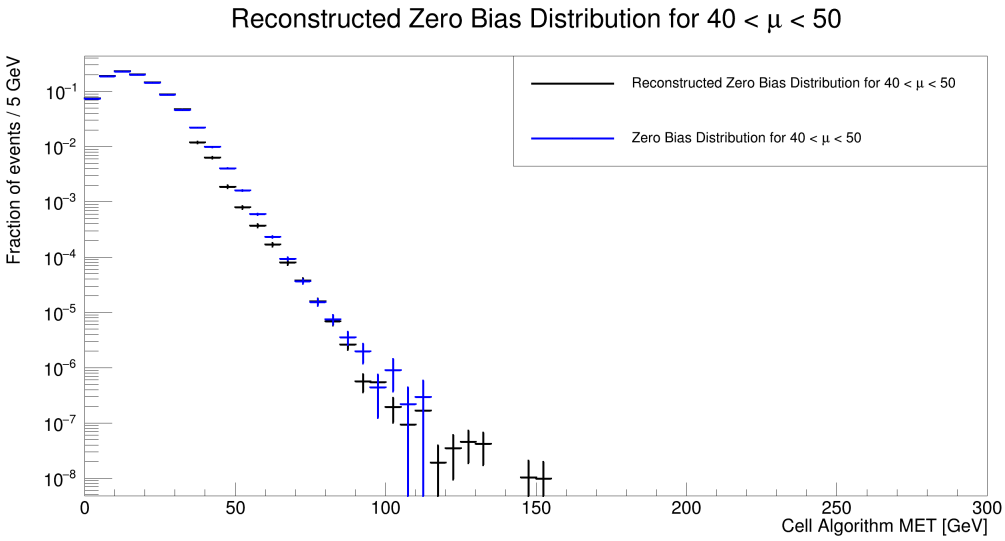

Figure 3.13: Reconstructed Zero Bias distribution for $\mu \in [40, 50)$

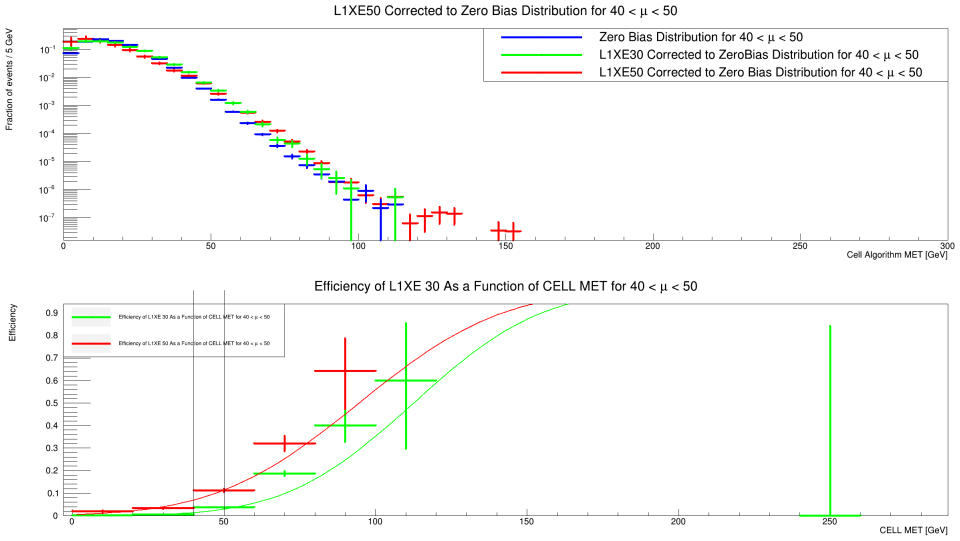

Figure 3.14: Corrected Distributions for $\mu \in [40, 50)$

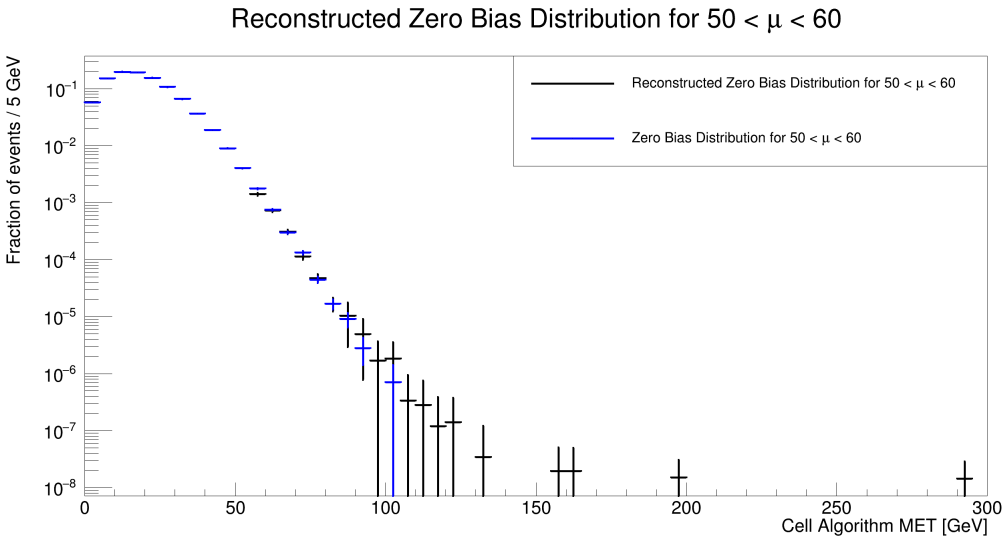

Figure 3.15: Reconstructed Zero Bias distribution for $\mu \in [50, 60)$

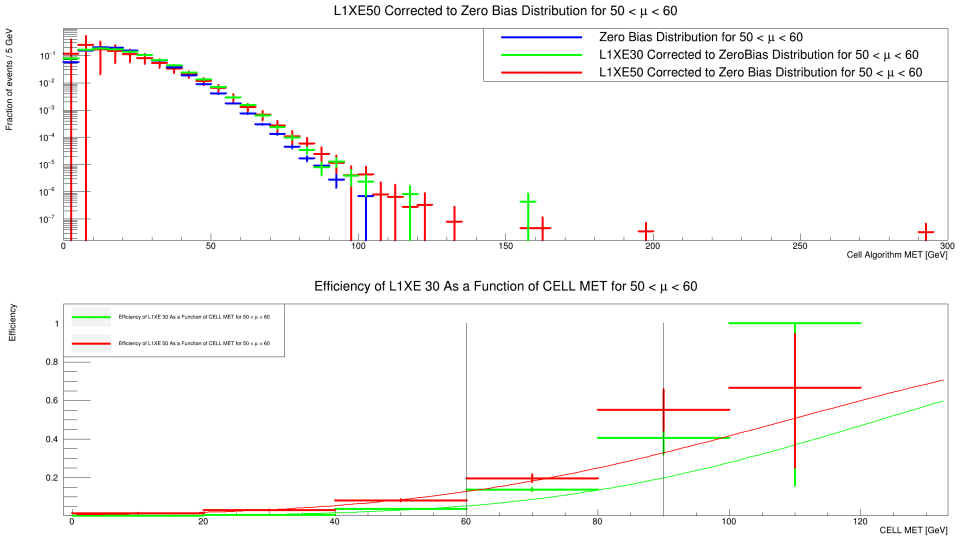

Figure 3.16: Corrected Distributions for $\mu \in [50, 60)$

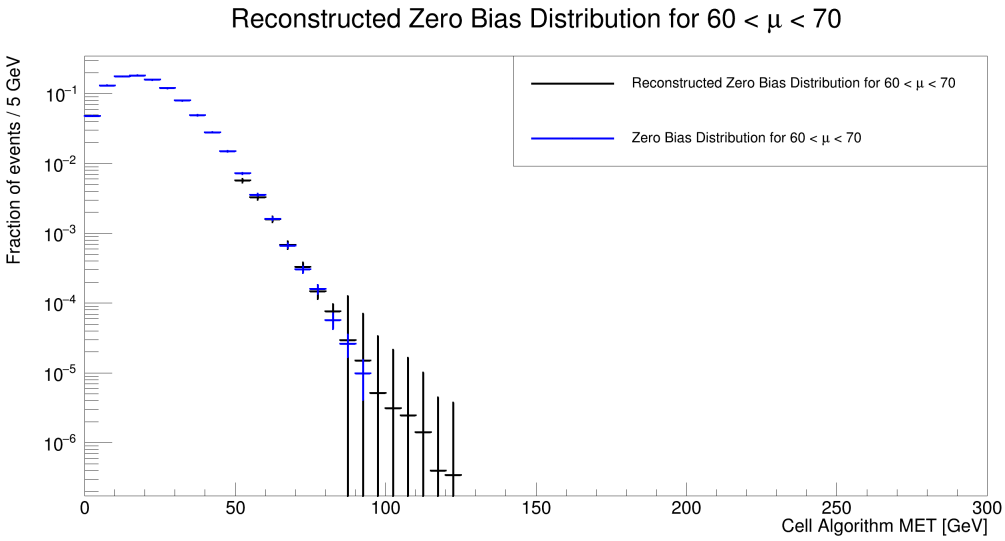

Figure 3.17: Reconstructed Zero Bias distribution for $\mu \in [60, 70)$

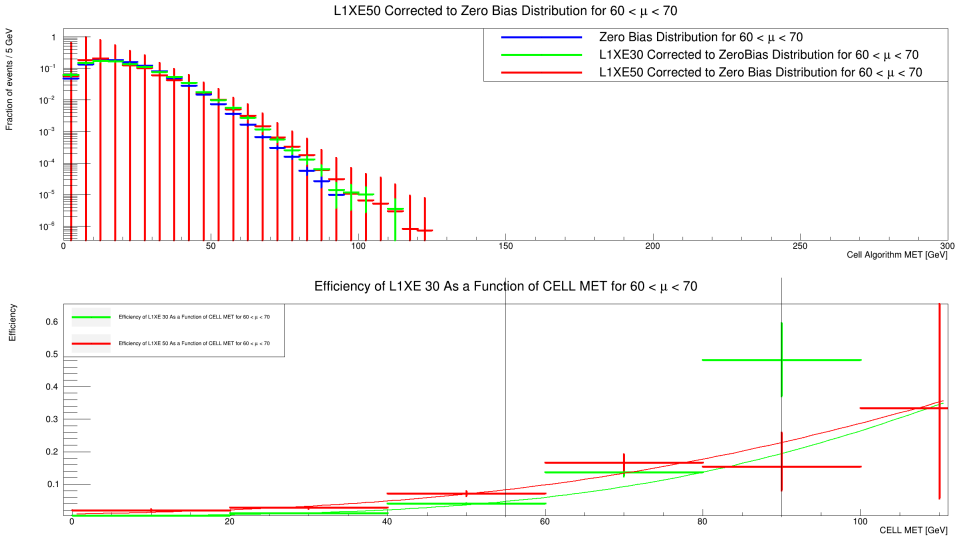

Figure 3.18: Corrected Distributions for $\mu \in [60, 70)$

Appendix A

Code

Here is the link to my github:
`https://github.com/jcorrado76/ATLAS_Research`

Appendix B

Table of Fitting Parameters for Efficiency Fits

a	b	σ	L1XE	μ bin
0.536043	-4.88401	7.63437	30	0
0.449818	18.4754	10.3507	50	0
0.40883	-3.87341	8.13195	30	1
0.505088	16.3944	10.3677	50	1
0.336915	-2.90115	8.63962	30	2
0.345437	22.3296	9.83129	50	2
0.29943	-2.17211	9.01473	30	3
0.277972	24.32	9.94887	50	3
0.281092	-1.63701	9.27598	30	4
0.289215	22.6488	10.6932	50	4
0.2487	-0.58147	9.68806	30	5
0.230607	24.8226	9.95501	50	5
0.231716	0.541431	9.99171	30	6
0.183126	26.0774	10.0148	50	6

Table B.1: Fit Parameter Table