CERN Summer Student Project:
Standard Method Dijet Balance

Author: Martin Friedl
Supervisor: Dr. Thomas Koffas
Carleton University - ATLAS Group

Sept 4th, 2012

Abstract

The forward jet response in the ATLAS detector requires an in situ correction to capture the effects that are not captured by the initial Monte Carlo (MC) derived calibration. Given that the detector's response to central jets is well modeled by Monte Carlo, an in situ analysis using dijet events is performed which probes the response in the forward regions of the detector by exploiting transverse momentum balance between a central and forward jet. From this analysis, response curves are generated giving the response of forward jets ($|\eta| > 0.8$) relative to central jets ($|\eta| < 0.8$) for both data and MC. An attempt is then made to implement pileup reweighting of MC data in order to normalize the MC pileup conditions to those observed in experimental data.

1 Introduction

In the ATLAS detector, two proton beams traveling in opposite directions interact with one another. In each beam, the protons travel in bunches with a separation of 50ns. When two bunches of protons collide (called a bunch crossing), often the constituents of the two protons (partons) will hit each other with enough energy that they hadronize, causing a parton shower. A parton shower is simply a collection of partons that get ejected from the beam axis. Once ejected, this cluster of partons will hadronize, forming a collection of more stable particles which are called a jet. This jet is then detected by the ATLAS calorimeters as a deposit of energy. Fig. (1.1) illustrates a two proton beam interaction resulting in two jets. This type of interaction is also called a dijet interaction.

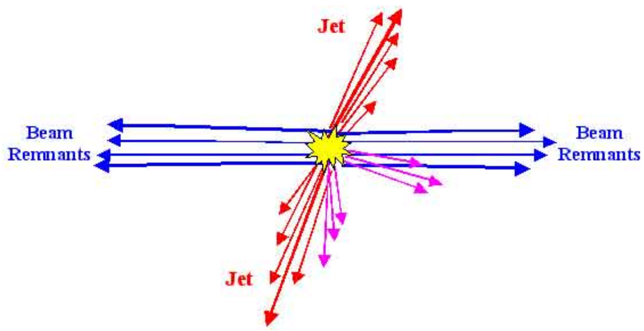

Figure 1.1: Diagram of a proton bunch crossing resulting in a dijet. [4]

1.1 The ATLAS Coordinate System

ATLAS uses a standard, right-handed, coordinate system to define the position and direction of travel of a particle within the detector. In this coordinate system, the z-axis points along the direction of the beam line, the x-axis points towards the centre of the accelerator and the y-axis points upwards. The polar angle θ is then measured from the positive z-axis to the jet vector. Depending on how the initial protons interact, jets are produced that travel in various directions. In order to measure which direction a jet is travelling with respect to the beam axis, an angular dependent term called pseudorapidity η is defined. Pseudorapidity is related to the polar angle θ by Eq. (1.1).

$$\eta = -\log(\tan(\theta/2)) \tag{1.1}$$

In Fig.(1.2), a diagram shows the relationship between angle θ and pseudorapidity η.

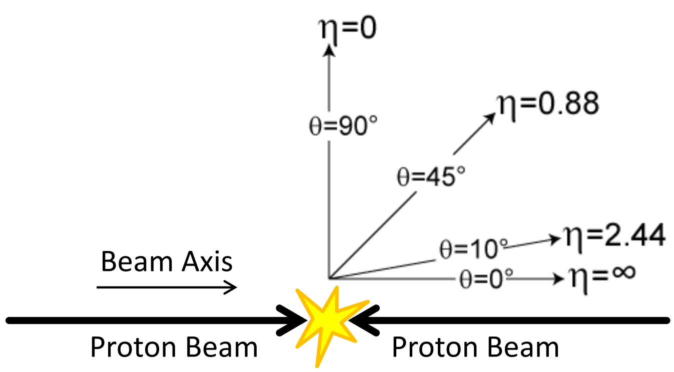

Figure 1.2: Diagram of pseudorapidity η measurement. [5]

1.2 Central and Forward Jet Definition

If a particle is ejected almost perpendicular to the beam axis ($\eta \approx 0$), it will land in the central part of the detector and is therefore called a central jet. On the other hand, a jet that travels very far along the beam axis before being absorbed

($|\eta| >> 0$) will land in the forward part of the detector and is for this reason called a forward jet. As seen in the diagram in Fig.(1.2), a forward jet will have a large $|\eta|$, whereas a central jet will have an $|\eta|$ close to zero. In this analysis, central jets are defined as having $|\eta| < 0.8$, corresponding to the central region of the barrel calorimeter in ATLAS. Any jets with $|\eta| > 0.8$ are considered to be forward jets [1]. The central jets are then used as a reference in order to probe the rest of the detector. The details of this method are further discussed in Section 2 of this report.

The calorimeters in the forward region of the ATLAS detector do not properly reconstruct the exact transverse momentum p_T of forward jets due to factors such as: differing detector technologies, dead mass, and the small angle at which jets enter the forward regions of the detector. On the other hand, the central region of the barrel calorimeter in ATLAS has been well defined and modeled such that detected particles in the central region are measured to have a direction and energy that is very close to Monte Carlo (MC) simulation. In fact, the uncertainty on the measured energy for central jets has been determined to be about 1-3% [3].

1.3 Experimental Data

The purpose of this project is to study the detector jet response for forward jets relative to central jets. The jet response is studied using jets reconstructed from experimental data and also using jets reconstructed from MC simulation. The results from data and MC are then compared for consistency.

In the case of experimental events, when an interaction occurs, the particles ejected from the event are absorbed by the liquid argon calorimeters inside the ATLAS detector. The calorimeters register a deposit of energy which is sent as an electrical signal to the level 1 triggers which are calorimeter-based hardware triggers. ATLAS has many levels of hardware and software triggers to reduce the stored data rate to a manageable amount. An Anti-Kt jet-finding algorithm [6] is then run on all the events that pass the level 1 trigger in order reconstruct the jets. To further reduce the amount of data collected so that it can be processed in a reasonable amount of time, the jet objects are sent to higher-level, software-based, triggers called event filter triggers. After having gone through all the different trigger levels, the data rate is reduced to a manageable amount which can then be stored and analyzed.

For this analysis, high level software-based event filter triggers were used. The jet triggers are divided into two categories, forward and central, corresponding to $|\eta| > 2.8$ and $|\eta| < 2.8$ respectively. Forward and central triggers are further divided and categorized by the jet p_T threshold required to fire the trigger. Thus, each event is categorized as being triggered by either a forward or central jet corresponding to a particular jet p_T.

1.4 Generating Monte Carlo Data

MC data is generated based on theoretical models that give each outcome of a collision a probability. After the collisions are generated, they are put through a simulated model of the ATLAS detector which turns these generated collisions into signals. These signals are also then fed into an Anti-Kt jet-finding algorithm in order to reconstruct the jets, much like with experimental data.

2 Standard Method Di-Jet Analysis

In an interaction, the initial protons have negligible transverse momentum p_T since they are traveling along the beam axis. However, after the interaction, jets are produced that have a nonzero transverse momentum. By conservation of momentum the total vector sum of the jets' p_T is zero. Thus, when the measured p_T of two jets from a dijet interaction are compared, it is expected that they are equal. If events are chosen in which at least one jet falls in the central region, by looking at the p_T balance between the two jets, various regions of the detector can be probed in order to get the detector jet response in all other regions of the detector.

For standard method dijet analysis, standard dijet event selections are first applied to the data in order to ensure that all remaining events have interactions with the previously mentioned two parton to two jet topology, with one jet falling in the central region of the detector. The two jets are referred to as either the reference jet or the probe jet. The reference jet falls within the central region of the detector where the detector response is well modelled and therefore the reference jet is considered to have its p_T reconstructed very accurately. However, this is not always the case for jets at higher pseudorapidity ranges. After the reference jet has been defined, the probe jet is allowed to fall anywhere inside the detector.

Events are binned together by their p_T and their pseudorapidity η (eta). The events are binned in p_T because the response of the detector can change with varying p_T. Therefore, in order to get a detailed response profile, the events are broken up into separate p_T bins. The p_T bins are defined by the previously mentioned ATLAS jet triggers, with the lower edge of each p_T bin corresponding to the >99% efficiency point for a given trigger. In addition, each p_T bin is further sub-divided into η bins in order to observe the response change over the whole detector. The η binning is defined by the physical detector transition regions to determine how the response changes across the transition regions in the detector.

2.1 Asymmetry Distributions

For each p_T and η bin, an asymmetry value (A) is calculated according to Eq. (2.1).

$$A = \frac{p_T^{probe} - p_T^{ref}}{p_T^{avg}} \tag{2.1}$$

A is a measure of the difference in p_T between the two jets, normalized to the p_T^{avg} of the two jets. Ideally, it is expected that both p_Ts would be equal which would make A zero. Though, due to several of the previously mentioned inefficiencies, A is often slightly more or less than zero. The A values of all the events in a given p_T and η bin are added together to give an asymmetry distribution as seen in Fig. (2.1) . From this distribution, the mean asymmetry for that individual bin is calculated.

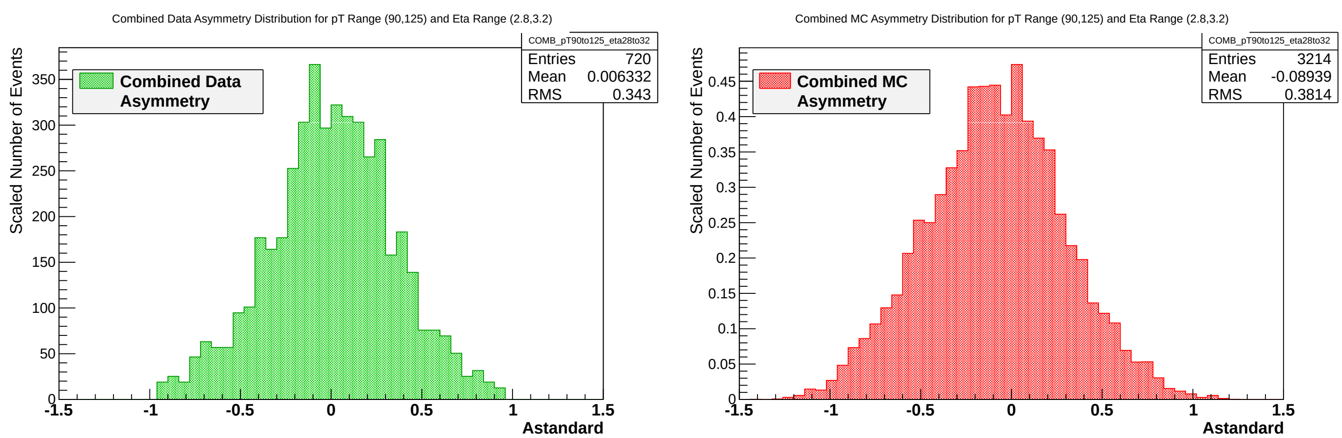

Figure 2.1: Asymmetry distributions for Monte Carlo and experimental data.

2.2 Data Scaling

In the case of experimental data, in order to get the proper asymmetry distribution, the events captured by the forward p_T trigger are added to the events captured by the central p_T trigger. However, because of a difference in luminosity (number of particles per unit area per unit time) collected by each, the two distributions cannot be added directly. Instead, a luminosity scaling factor must first be applied to each of the two trigger distributions in order to compensate for the two different trigger luminosities.

In Fig. (2.2) it can be seen how initially the central and forward trigger asymmetry distributions are comparable. After the luminosity scaling factor is applied, the forward trigger becomes much less significant compared to the central trigger.

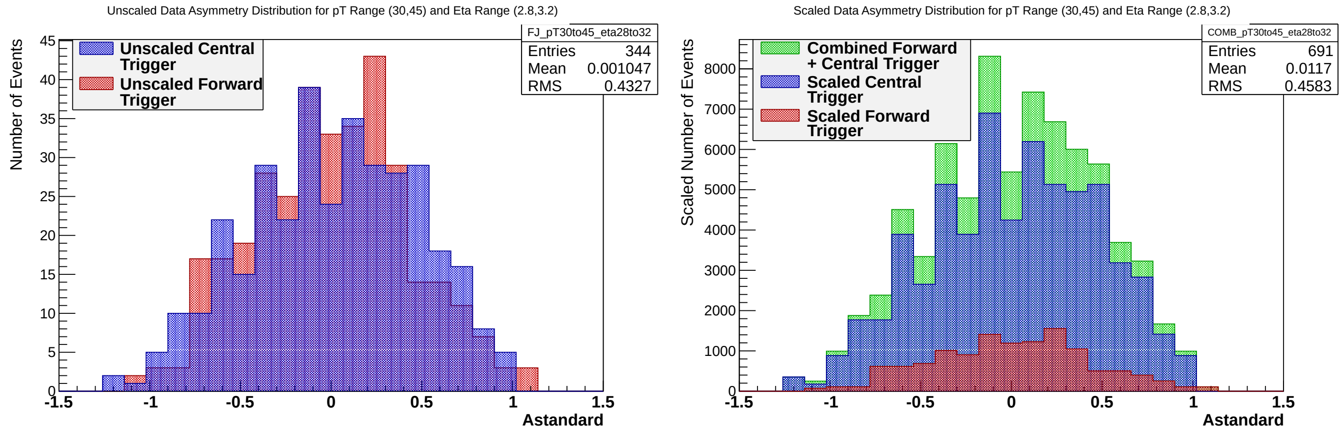

Figure 2.2: Unscaled and scaled asymmetry distributions for experimental data.

2.3 MC Scaling

For MC, the total data set is constructed from 8 data subsets called JX samples that vary from J0 up to J7. Each sample is generated from a Poisson distribution centered at a certain p_T which increases as you go up in number of the JX sample. Initially, all JX samples have the same number of events, but because high p_T events are less likely to occur than low p_T events (smaller cross section), the higher JX samples must be scaled by a smaller weight than the low JX samples. These applied weights are called the cross-section weights. In addition, each event is also weighed with a filter efficiency weight, a luminosity weight and an event weight. The overall scaling factor applied to each MC event is shown in Eq. (2.2). For each p_T and η bin, all of the JX data subsets must first be scaled by their corresponding W and added together make up the asymmetry distribution for that bin.

$$W = \frac{W_{FilterEff} W_{CrossSection} W_{Lumi}}{W_{Event}} \tag{2.2}$$

In MC, as seen in Fig. (2.3), the J1 sample is made much more statistically significant by the W scaling. Since the J1 sample has a larger cross-section than the J2 sample it is scaled by a larger number than the J2 sample. Therefore, after scaling, the J1 sample is more statistically significant than the J2 sample. After scaling all of the JX samples by their W, they may be added together to create the overall asymmetry distribution for each p_T and η bin. Note that the rest of the JX samples (J0, J3-J7) were not included in Fig. (2.3) because they do not contribute any events to this particular bin.

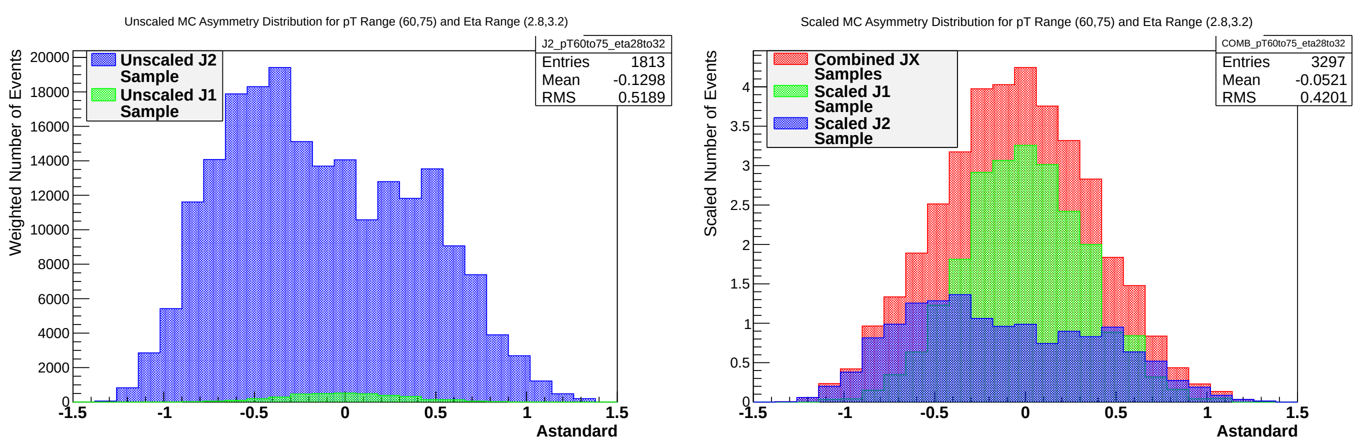

Figure 2.3: Unscaled and scaled JX asymmetry distributions for Monte Carlo data.

2.4 Generating the Response Curves

Once the asymmetry distributions for all η bins within a certain p_T bin were generated, the mean asymmetry was calculated for each distribution. From the mean asymmetry, the jet response R was then calculated using Eq. (2.3) for each p_T and η bin. By construction, this method defines the R in the central region to be 1, so if the probe jet lands in the central region of the detector, R will be 1. If the probe jet lands outside of the central region, R will deviate from the expected value of 1 due to the different p_T response for forward jets.

$$R = \frac{p_T^{probe}}{p_T^{ref}} = \frac{2+A}{2-A} \tag{2.3}$$

The R values from each η bin within a p_T range were plotted in order to give a response curve over all η for a given p_T range. This was repeated for all p_T ranges, resulting in multiple response curves, each for a given p_T. An illustration of this process can be seen in Fig. (2.4). In this illustration, it can be seen how the A_{avg} is taken from each A distribution, then R_{avg} is calculated and finally plotted on the response curve as a function of η bins.

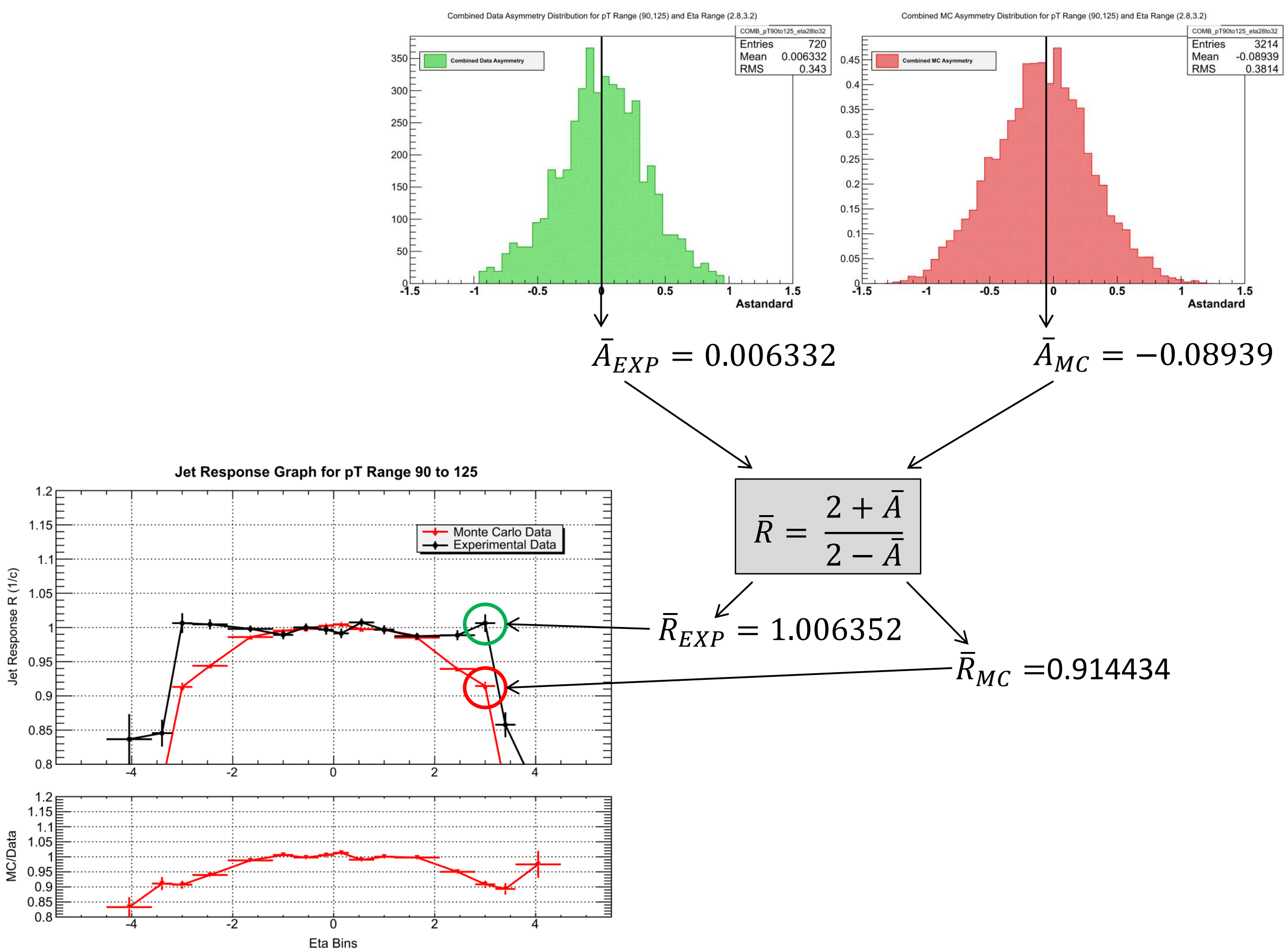

Figure 2.4: Illustration showing how the response curves were created.

2.5 Results

The response curves were calculated for each p_T bin with MC data and experimental data both plotted on the same graph for comparison. An additional graph was added below the response curves showing the ratio of R_{MC}/R_{Data} vs. η in order to better display the discrepancy between MC data and experimental data. The response curves for a few p_T bins can be seen in Fig. (2.5).

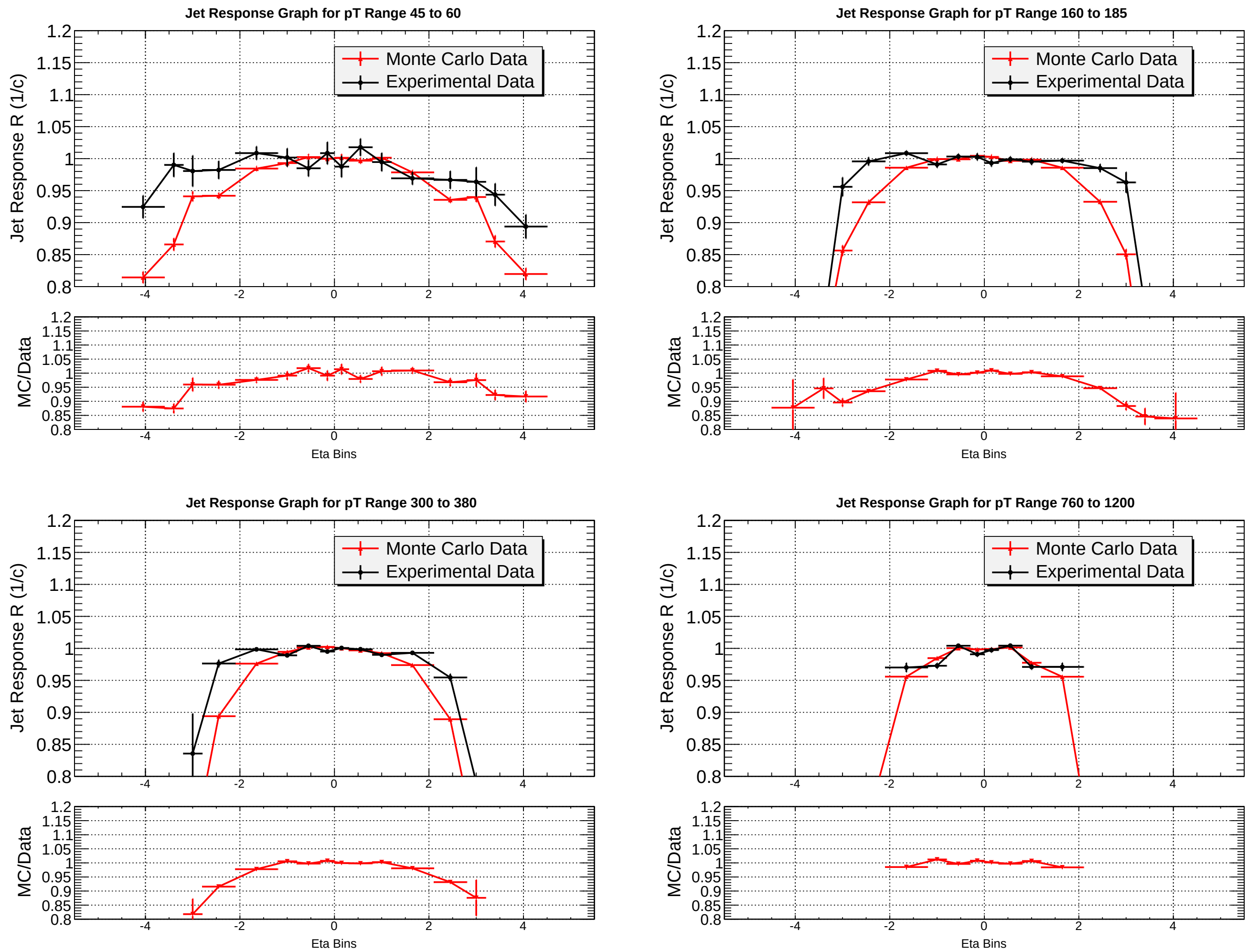

Figure 2.5: Monte Carlo and experimental data response curves for various p_T ranges.

In Fig. (2.5) it can be seen how the response curve η range decreases as p_T is increased. This is by explained by the fact that jets only have a finite total momentum. At very high p_T, most of the jet's momentum will be in the transverse direction and only a small fraction of the total momentum will then be in the forward direction. This kinematic limit makes it impossible for high p_T jets to land in the very forward regions of the detector since these jets will be ejected nearly perpendicular to the beam axis and therefore land closer to the central region of the detector.

3 Pileup Re-Weighting

In order to improve the accuracy of the response curves, the MC data must be also re-weighted to the pileup conditions at which the data was taken. Pileup is a term used to describe the process where a large number of simultaneously occurring interactions cause a “pileup” of signals in the detector. There are two types of pileup: in-time pileup and out-of-time pileup. In-time pileup is caused by multiple interactions in the same bunch crossing. Out-of-time pileup is caused by multiple events overlapping from consecutive bunch crossings and arises due to the long integration time ($\sim 400ns$) of the calorimeters. In 2012, due to the smaller times in between proton bunch crossings (currently 50ns), there is a larger amount of out-of-time pileup. This means that the detector sometimes does not have enough time to “reset” itself before the next bunch crossing and therefore registers a signal from the previous bunch crossing.

In order to characterize pileup, the average number of proton interactions (collisions) per bunch crossing μ are recorded while data is being collected in ATLAS. Since pileup is directly related to the number of interactions per bunch crossing, a μ distribution is a good estimator of the pileup conditions in a given data set [2].

The pileup conditions in MC data and collected data are not initially the same because MC data is generated with a default set of pileup conditions. Experimental data, on the other hand, is captured in the ATLAS detector where the average number of interactions per bunch crossing μ is constantly measured and recorded. Both the data and MC μ distributions can be seen in Fig. (3.1).

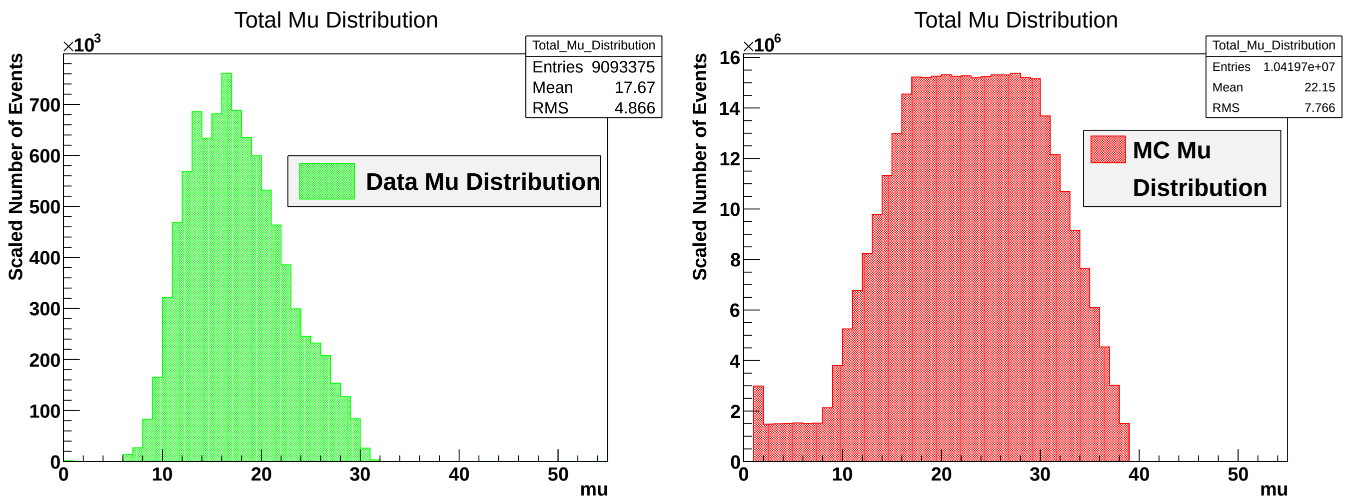

Figure 3.1: Total μ distributions for experimental data and Monte Carlo data before pileup reweighting.

From Fig. (3.1) it can be seen that the unweighted μ distribution for MC is not of the same order as the distribution for experimental data. Therefore, in order to compare the two, a normalizing factor is applied to both which normalizes the integral of each distribution to 1. Both distributions can then be plotted on the same axes in order to compare their shapes. In Fig. (3.2) this comparison is shown and it can be seen that the experimental μ and the MC μ do not agree. This proves that the MC data must be re-weighted in order for it to match the pileup conditions of the collected data.

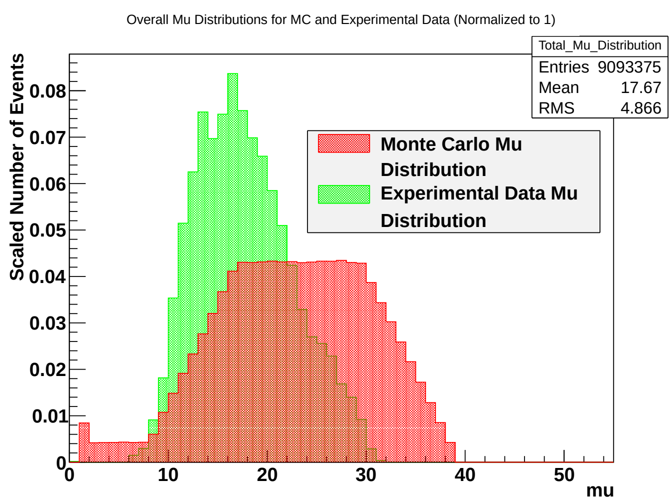

Figure 3.2: Total normalized μ distributions for experimental data and Monte Carlo data before pileup reweighting.

A tool has been developed called the Pileup Reweighting Tool which reads in luminosity data files that contain $<\mu>$ for a specific time block in the data collection run. This tool generates weights for each MC event in order to normalize the MC μ distribution to the experimental μ distribution.

The reason for applying these weights is that, for example, if the experimental data set only has events with $\mu < 30$, comparing this data to an MC data set that has events with $\mu > 30$ would result in inconsistencies. The inconsistencies would arise due to the fact that the two data sets were produced under different pileup conditions. Instead, using the Pileup Reweighting Tool, weights are generated and applied to the MC data which has the effect of normalizing the MC μ distribution to match that of data. In the end, the μ distribution of the MC data is brought down by the weights to correspond exactly with the μ distribution of the collected data set, as shown in (3.3).

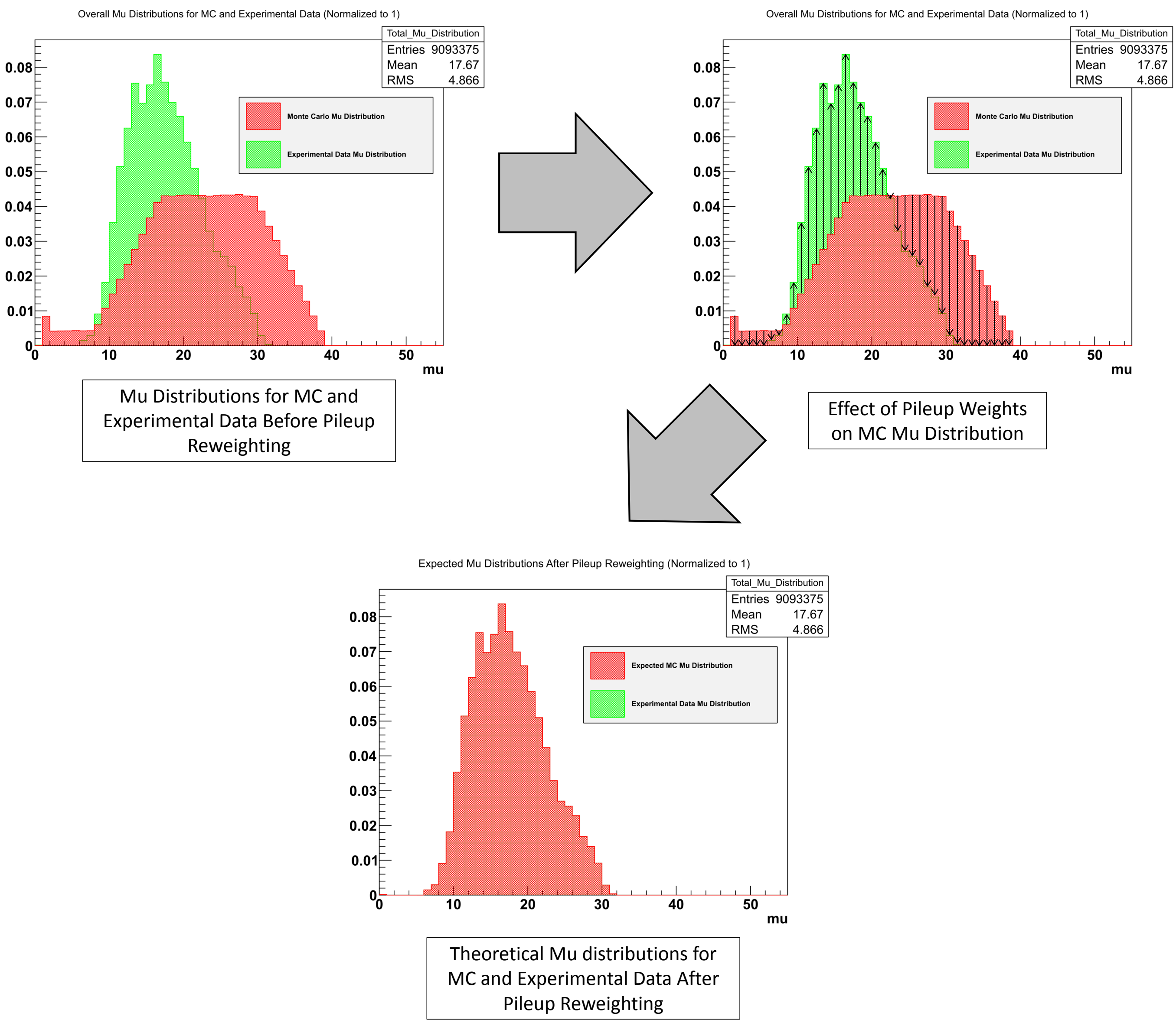

Figure 3.3: Flow chart showing the expected effect of introducing pileup weights to Monte Carlo data μ distributions.

In Fig. (3.3) the expected effect that pileup reweighting should have on the MC distribution is shown. After pileup reweighting, it is expected that the MC μ distribution gets reweighted to perfectly fit the data μ distribution. By confirming that the two μ distributions coincide, we can be certain that the pileup weights were applied properly and that the MC data is now normalized to the same pileup conditions as the experimental data.

3.1 Pileup Reweighting Results

The results of the pileup reweighting being applied to the MC data can be seen in Fig. (3.4). It is clear from the plot that the pileup reweighting did not properly reweigh the MC μ distribution. It was expected that the MC μ distribution would be scaled perfectly in order to coincide with the experimental data μ distribution, but this was not the case. Instead, in Fig. (3.4), it appears as if there is a shift by one bin to the right in the reweighted MC data sample, though only on the left side of the distribution. On the right side of this same distribution, near $\mu = 30$, there no longer appears to be any shift. A possible explanation for this discrepancy could be that the luminosity files that were given to generate the weights were inconsistent with the actual data that was used for the comparison. Therefore, the weights were generated using an assumed data set, but when these weights were applied to the actual data (which was different from the assumed data set) it caused the discrepancy between MC and data seen in Fig. (3.4).

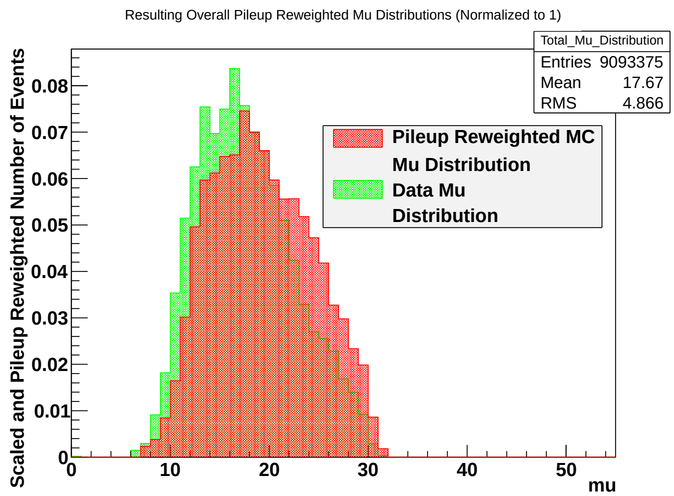

Figure 3.4: μ distribution of the actual result of introducing pileup weights to the Monte Carlo μ distributions.

4 Conclusion

The standard dijet analysis method was used in order to probe the jet response of the forward regions of the ATLAS detector. The p_T of forward and central jets were compared in order generate response curves as a function of η for various p_T. It was seen that the detector response falls off as $|\eta|$ increases but is generally in agreement to within 10% with MC data. However, it is important to note that the standard method dijet p_T balance has the disadvantage that it requires at least one of the two jets to fall in the central region of the detector, resulting in limited statistics. Dijet events where both jets fall outside the reference region are not used in the analysis. As an improvement, there exists a matrix dijet balance method which does not have this restriction. In this method, the reference and probe jets are instead simply replaced with left and right jets. A minimizing matrix of linear equations is then constructed in order to solve for a calibration factor, from which the response curves are obtained [1]. This method is now used as the default dijet intercalibration method because it is able to produce more accurate response graphs with smaller error bars due to larger statistics.

After generating the response curves, an attempt was made in order to correct for the difference in pileup conditions between MC generated data and experimental data. The Pileup Reweighting Tool was implemented to generate pileup weights which were then applied to the events from MC data. The resulting weighted μ distribution was compared to that from experimental data to check that the reweighting was successful. It was found that discrepancies still existed between the MC data and the experimental data. This suggests that perhaps the data set given to the Pileup Reweighting Tool was not consistent with the data set used to compare with MC after pileup reweighting. The cause of the improper weighting has to be determined and corrected before response curves can be produced using the pileup weights. Once a solution is found, new response curves can be produced and compared with the response curves generated before the pileup reweighting in order determine the effect of the pileup correction on the measured response of the detector.

5 Acknowledgements

I would like to give my thanks to my supervisor Professor Thomas Koffas from Carleton University for his help and advice this summer. I would like to thank James Lacey from Carleton University for tirelessly answering all of my questions during my time at CERN. I would also like to thank Professor Gerald Oakham from Carleton University for being my supervisor while I was working with thin gap chambers during my first two months this summer at Carleton. Finally, I would also like to thank Professor William Trischuk of the University of Toronto for offering me this amazing opportunity to come to CERN as a summer student and being there to help every step of the way.

References

[1] The ATLAS Collaboration. In-situ pseudorapidity intercalibration to evaluate jet energy scale uncertainty and calorimeter performance in the forward region. *ATLAS-CONF-2010-055*, 2010.

[2] The ATLAS Collaboration. Pile-up corrections for jets from proton-proton collisions at sqrt(s) = 7 tev in atlas in 2011. *ATLAS-COM-CONF-2012-078*, 2012.

[3] The ATLAS Collaboration. Single hadron response measurement and calorimeter jet energy scale uncertainty with the atlas detector at the lhc. *CERN-PH-EP-2012-005*, 2012.

[4] The CDF Collaboration. Charged jet evolution and the underlying event in proton-antiproton collisions, December 2001.

[5] JabberWok. Pseudorapidity2.png, May 2007.

[6] Gregory Soyez Matteo Cacciari, Gavin P. Salam. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008.