Measurement of dijet production at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector

A thesis submitted to the University of Manchester for the degree of

PhD

in the Faculty of Engineering and Physical Sciences

2011

Graham Jones

School of Physics and Astronomy

Contents

Total Word Count: 22,411

List of Figures

List of Tables

Abstract

A measurement using the ATLAS detector has been made of the fraction of events in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ that do not contain additional jets in the rapidity region bounded by a di-jet system. This provided a strong test of perturbative QCD in the new energy regime of the Large Hadron Collider. Additional measurements have been made of the mean jet multiplicity in the rapidity bounded region and using a variety of different event selections. These observation were compared to recent next to leading order dijet predictions produced by POWHEG and a variety of leading order generators.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Graham Jones
School of Physics and Astronomy
University of Manchester
Oxford Road
Manchester
M13 9PL
United Kingdom
September 2011

Copyright

//www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Acknowledgments

I would like to thank Fred and Andy for all their time and help without which this thesis would surely have sunk faster than the Titanic. Sabah probably deserves a pay rise for all the time he spent helping me.

Manchester is quite a marvelous place to be as a student where barely a day goes by that there is not a cake to be enjoyed after a hard afternoon of exploring physics or more commonly the ROOT support forums.

Spending two years at CERN was a great experience. To see decades of work by thousands of scientists and engineers come to fruition and be part of it in some small way made the last four years quite fullfilling.

I am also grateful for friends who managed to put up with me and hope they continue to keep doing so.

It would of course have not been possible without the love and support my family, whom I owe everything.

Chapter 1

Introduction

For over a year the LHC has operated at a centre of mass energy of 7 TeV. During this initial period of running it has been very important for all the experiments to re-discover the Standard Model of particle physics. Key to the Standard Model at any hadron-hadron collider is the understanding of QCD, as the dominantly produced physics object in the ATLAS detector is the jet.

An area of difficulty expected for the standard QCD predictions is the production of wide angle radiation, whose cross section is highly uncertain theoretically. Previous to this work a measurement of these effects had not been made at the new energies which the LHC operates at. A measurement has been made using a central jet veto to study QCD radiation at the new energies available at the LHC. Using results from this thesis a paper has been published by ATLAS for the data collected in the 2010 period of proton-proton collisions.

For the purposes of theoretical comparison a number of different leading order Monte Carlo generators were used to generate predictions. In addition to comparison between the generators, these samples were also used for the derivation of data corrections that allowed the removal of residual effects the detector may have upon the data. For the purposes of this analysis it was found that the default method of generation was not sufficient for producing samples for data correction. Two new schemes for weighting generated events were developed to greatly improve the precision of the data corrections applied.

NLO dijet predictions have been made using the recently developed POWHEG framework. Uncertainties related to the predictions have been investigated with considerations of both the parton distribution functions and scales chosen for renormalization and factorization.

Chapter 2

Theory

In this chapter a brief outline of the theory of quantum chromodynamics is given with some discussion of proton-proton collisions. The jet physics object and the method of its construction is described in detail. Finally the important variables and observations that will be investigated in this thesis are identified.

2.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics describes the interactions of the fundamental particles through the exchange of gauge field particles [1]. These force-matter interactions are described by Lagrangians in field theory, where the Lagrangian is classically considered to be the difference between the kinetic and potential energy.

The particles described by the SM are the four vector bosons, which mediate the three forces described in the SM; and twelve fermions, which are the particles that make up the matter within our universe. The matter particles are split into two types, the leptons and quarks. Lastly a scalar boson, known as the Higgs, is thought to provide the mechanism through which particles acquire mass.

The three forces of nature described within the SM are the electromagnetic (EM), weak and strong nuclear force. A fourth force, gravity, the weakest of the four fundamental forces

is not yet described by the SM, however its effect on particles is very small. The force carrier for the EM interaction is the photon. The photon is a massless boson which transmits the EM force between particles with charge.

The weak force allows the interaction of both quarks and leptons. The gauge invariance of the theory is maintained by the introduction of the three gauge bosons $W^{\pm}$ and Z. It is described by a non-albelian theory, where the the gauge bosons are able to interact with themselves. The relative weakness of the force can be explained by the massive nature of the gauge field particles, which limits the range of interactions. The last force to be discussed here is the strong nuclear force, and is also described by a non-albelian theory, where the gauge boson mediating the force is the gluon. The fermions that interact through this force are limited to the quarks as the leptons do not have a colour, where the colour is the equivalent of charge in the EM case. This colourful force is of detailed interest within this thesis and will be described further in the next section.

The strength of the interactions are described by the coupling constants for each of the forces. The coupling constants vary as a function of the interaction energy, or equivalently the distance scale, because of charge screening. In the case of non-albelian theories, such as in the strong or weak cases, the coupling decreases with energy because of the possibility of self interaction by the gauge bosons. Whilst for albelian theories, such as that which describes EM interactions, see the coupling increase with energy.

The goal of research today is to test the predictive nature of the SM and to look for physics that lies beyond well tested theory. Investigations include testing for structure in quarks and leptons, the unification of the forces, matter-antimatter asymmetry and the incorporation of the force of gravity to name a few.

2.2 Quantum Chromodynamics and Hadron-Hadron Interactions

Quantum chromodynamics (QCD) describes the strong interactions between coloured quarks and gluons. Within the SM there are three generations of quarks

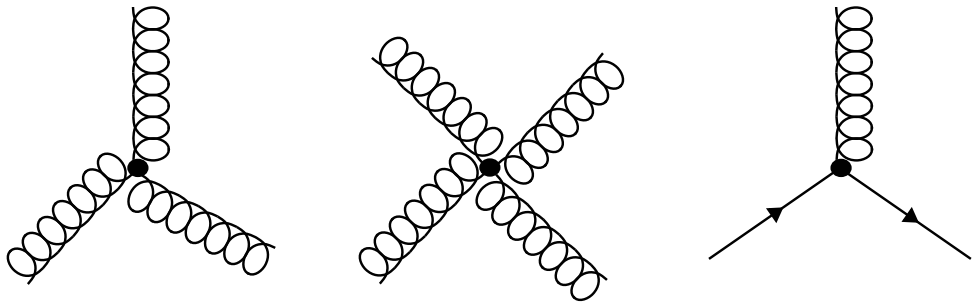
$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

where u, d, c, s, t, b are the up, down, charm, strange, top, and bottom quarks respectively. The up, charm and top quarks all have a charge $+\frac{2}{3}$. The down, strange and bottom quarks all have charge $-\frac{1}{3}$. The force carrier of the strong interaction is the gluon. The SM Lagrangian of QCD is given by

$$\mathcal{L} = \sum_q \overline{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} A^C_\mu - m_q \delta_{ab})\psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu}, \tag{2.1}$$

where $\psi_{q,a}$ are the quark-field spinors for a quark of flavour q and mass m; γ^μ are the Dirac γ-matrices; g_s is the strong coupling and is related to α_s by $g_s^2 = 4\pi\alpha_s$; A^C_μ is the gluon field and $F^a_{\mu\nu}$ is the gluon field strength tensor for a gluon colour index a [2]. The first term in the equation handles the propagation and interaction of quarks and the second handles the gluon propagation and self-interaction.

The interaction vertices for QCD, which are encapsulated in the Lagrangian above, are shown in figure 2.1. QCD has the property of asymptotic freedom, where as the interaction energy increases the strong coupling decreases. It is in this high energy limit that perturbation theory becomes applicable. The converse is true; as quarks move apart the force becomes stronger, requiring an infinite amount of energy to separate two quarks. This confining nature of the strong interaction leads to the absence of free quarks and in general why we only observe colourless final states.



(a) (b) (c)

Figure 2.1: Interaction vertices of QCD where the curly line represents the gluon and straight lines fermions. Taken from [3]

The main tool for solving the QCD Lagrangian is perturbative quantum field theory. The definition of the partonic scattering cross section is given by

$$d\hat{\sigma}_{ab\to F} = \frac{1}{2\hat{s}_{ab}}|\mathcal{M}_{ab\to F}|^2(\Phi_F;\mu_F,\mu_R), \tag{2.2}$$

where $|\mathcal{M}|^2$ are the matrix element squared of the process summed and averaged over the helicities and colours. μ_F and μ_R are the factorization and renormalization scales respectively. The factorization scale separates long and short distance physics, below the scale the parton is absorbed into hadron structure. The leading order (LO) Feynman diagrams for jet production are shown in figure 2.2. Also shown in figure 2.3 are some of the next to leading order (NLO) Feynman diagrams, which show additional real emissions and virtual loop corrections that are important at this level.

The differential cross section for an observable O can be computed with

$$\frac{\mathrm{d}\sigma}{\mathrm{d}O} = \sum_{a,b}\int_0^1 \mathrm{d}x_a\mathrm{d}x_b \sum_F \int \mathrm{d}\Phi_F f_a^{h1}(x_a,\mu_F) f_b^{h2}(x_b,\mu_F)\frac{d\hat{\sigma}_{ab\to F}}{d\hat{O}} D_F(\hat{O}\to O,\mu_F), \tag{2.3}$$

a sum over all the constituents of the interacting protons and all the possible final states F [2]. The integral shows the separation applied to the physics into short and long timescales, where

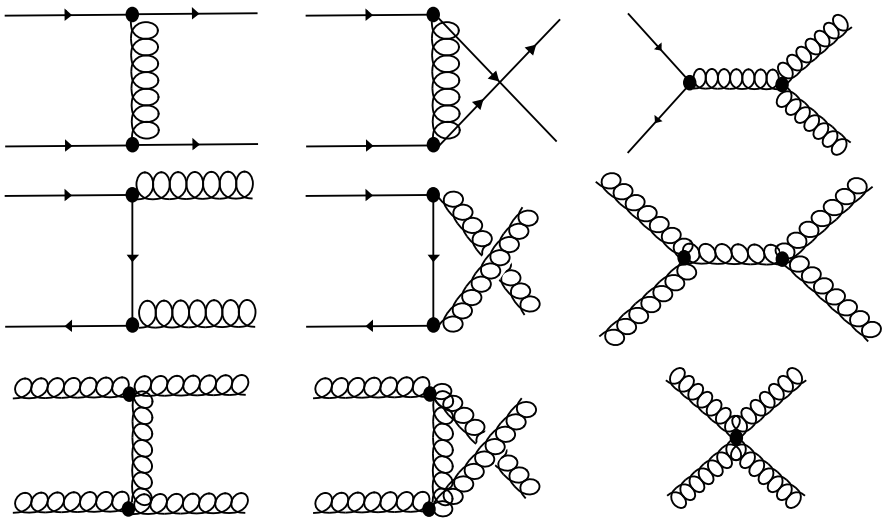


Figure 2.2: Examples of some of the diagrams for jet production processes at LO level. The curly line represents the gluon and straight lines fermions. Taken from [3].

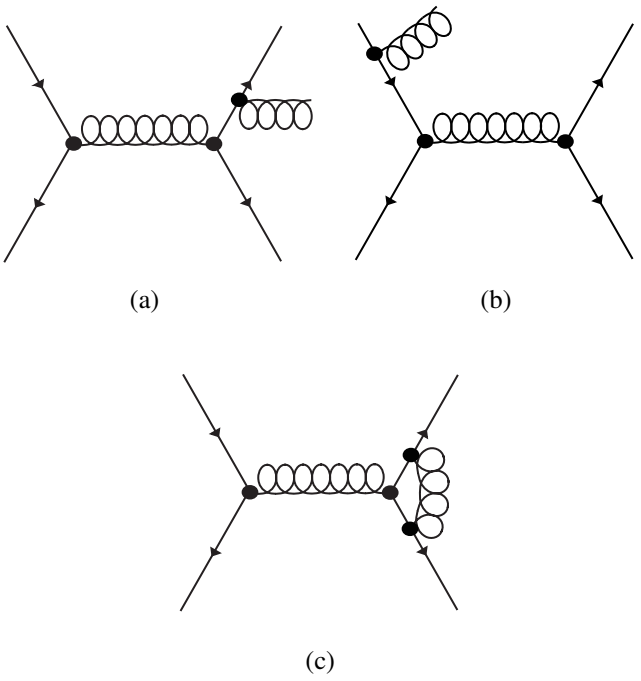


Figure 2.3: Examples of some of the diagrams for QCD processes at NLO level. The curly line represents the gluon and straight lines fermions. Taken from [3].

the parton distribution functions $f_{a,b}$ give the probability of finding a parton of momentum $x_{a,b}$ with a flavour a, b within a proton when it is probed by an interaction of scale μ_F. A long time after the parton interaction the transition of the produced partons to colourless hadrons

is handled by the fragmentation function D_F.

Underlying Event

The underlying event is the energy which is attributed to the primary proton interaction but not originating from the hardest parton interaction. There is a non-negligible probability that there will be more than one interaction between pairs of partons within the same proton-proton collision, this is known as multiple partonic interactions. Additionally if after the collision the protons disassociate then their remnants will add to the underlying event.

2.3 Jets

Jets are collimated sprays of hadrons carrying the momentum from the interaction. It is useful to describe the kinematic properties of jets in terms of Lorentz invariant quantities. The rapidity is defined as

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right), \tag{2.4}$$

where E is the energy and p_z is the momentum along the direction of the beam. The transverse momentum of a jet is defined as

$$p_T = \sqrt{p_{T\,x}^2 + p_{T\,y}^2} \tag{2.5}$$

where p_x and p_y are the transverse momenta in the x and y directions respectively.

2.3.1 Important Properties

A number of important properties for jet algorithms have been identified in [4], including invariance under boosts, order independence, infrared safety and collinear safety. Invariance under boosts requires that jet properties do not change when observed in a different frame. Order independence states that the jets identified at the parton level should be the same as

those identified at hadron level and also after measurement by a detector.

Infrared Safety

Proton-proton collisions will typically contain a large number of low p_{T} particles in each event. Jet algorithms must avoid being sensitive to these particles. Jet algorithms which use seeds can be especially sensitive to these problems, as shown in figure 2.4 where if there is a low p_{T} object present between the high p_{T} objects then they may all be merged into a single jet. This would leave jets sensitive to large uncertainties in both measuring these low p_{T} objects and also predicting them.

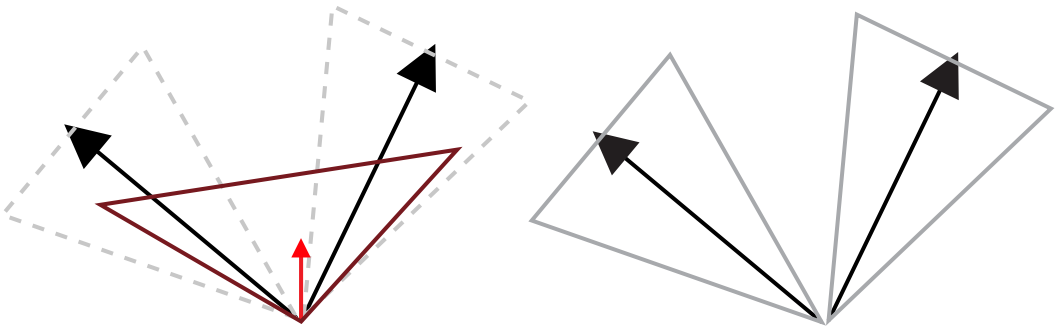


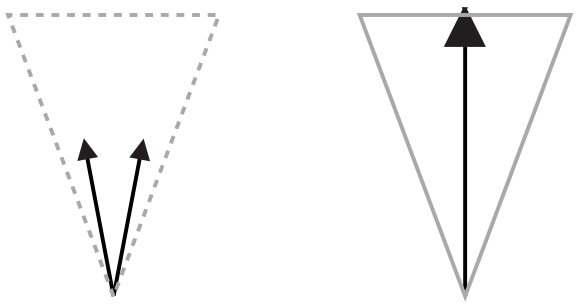
Figure 2.4: Examples of clustering of entities into jets using an infrared unsafe algorithm with (a) and without (b) an additional soft entity.

Collinear Safety

The jet algorithm should also not be sensitive to the case where a particle is replaced with two collinear particles. In figure 2.5 a jet algorithm which uses seeds may fail to identify the jet if the parton is split.

2.3.2 Anti-k_t Jets

The standard method of creating jet objects in ATLAS uses the anti-k_t jet algorithm [5], which is both infrared and collinear safe. The anti-k_t jet algorithm constructs jets through



(a) (b)

Figure 2.5: Examples of jet finding with (a) and without (b) collinear splitting of the parton.

iterative recombination of entities. Examples of entities used by ATLAS are the clusters of energy measured in the calorimeter, particles measured by the tracker and truth particles from the Monte Carlo event record. The decision to cluster entities is handled through the distance metrics

$$d_{ij} = \min(p_{T\,i}^{-2}, p_{T\,j}^{-2})\frac{\Delta_{ij}^2}{R^2}, \tag{2.6}$$

$$d_{iB} = p_{T\,i}^{-2}, \tag{2.7}$$

where $p_{T\,i}$ is the transverse momentum of the i-th entity, R is the recombination parameter and Δ_{ij} is the distance between the i-th and j-th entity defined as

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}, \tag{2.8}$$

where y_i and ϕ_i are the rapidity and polar angle of the i-th entity respectively. For each pair of entities d_{ij} and d_{iB} are calculated and then sorted from smallest to largest. If the smallest effective distance is the d_{ij} case then the i-th and j-th entities are combined, in the d_{iB} case the i-th entity is removed from the list and it becomes a jet. This process repeats until all the

entities have formed jets.

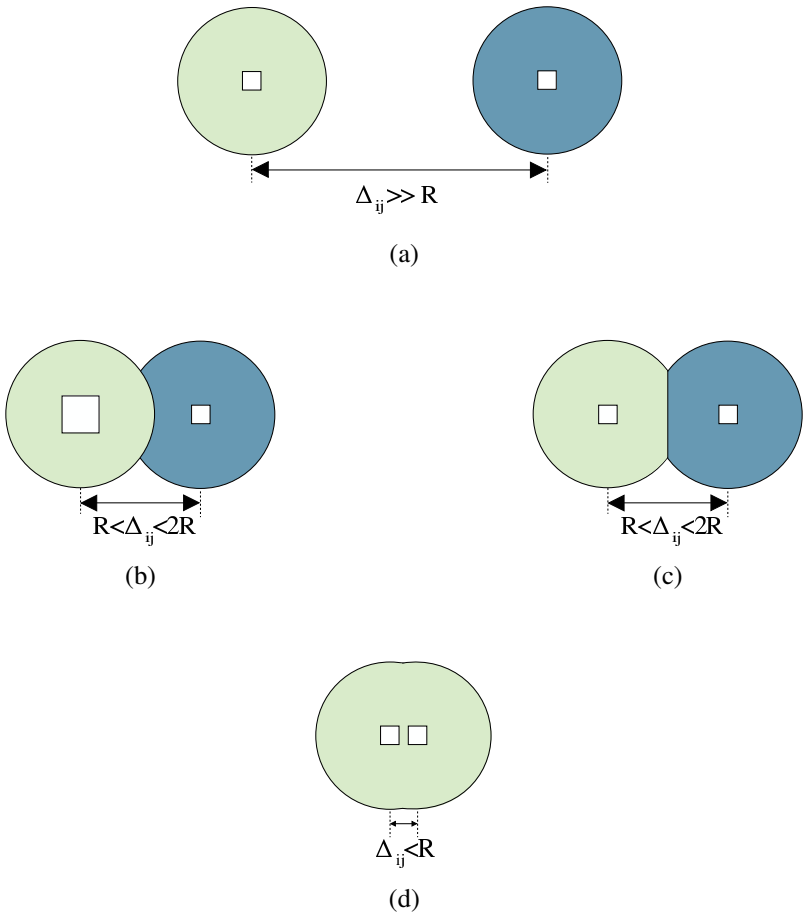


Figure 2.6: Catchment areas of jets formed around boxes that represent high p_T entities. Two equally high p_T entities are considered in (a), (c) and (d). The asymmetric energy case is considered in (b). Δ_{ij} is the distance between the original high p_T entities.

The result of the clustering procedure is that low p_T entities will tend to cluster with the high p_T entities first. This will cause the formation of cone shaped jets around the high p_T entities. So for the case where two high p_T entities are separated by more than double the jet radius parameter, figure 2.6(a), two cone shaped jets will be formed. As the distance between the high p_T entities falls then shape of the jets will change depending upon the relative p_T. For the case of asymmetric p_T, shown in figure 2.6(b), the higher p_T entity will still form a cone shaped jet whilst the other will form a crescent shaped jet. In the completely symmetric case, in figure 2.6(c), the boundary between the jets will be a straight line. Finally if the two entities are sufficiently close, figure 2.6(d), they will be merged into one jet with a possibly

complex shape.

The clustering properties of the anti-k_t algorithm result in jets are less sensitive to soft energy including pile-up, which is energy from additional proton-proton interactions; and the underlying event. The reduced sensitivity to soft energy has the downside that the algorithm will also be less adaptive to soft and collinear emission expected to occur for the energetic particles coming from the hard interaction.

2.4 Inclusive Jet and Dijet Cross Sections

The measurements of the inclusive jet and dijet cross sections are a strong test of perturbative QCD and so form an important component of the benchmark processes that the LHC must measure to rediscover the Standard Model. In addition, the increase in the centre-of-mass energy at the LHC will allow much higher jet energies to be probed, possibly uncovering new phenomena such as excited quarks, axigluons and colour octet scalar resonances. Already ATLAS has searched for resonances from new phenomena in the dijet mass spectrum, finding no evidence up to a mass of $\sim 4\,\text{TeV}$ [6]. Jets at hadron colliders also form an important background to many searches in other channels, further necessitating precision measurement to increase the sensitivity in those searches as well.

2.4.1 Variables of Interest

The first distribution of interest is the inclusive single jet cross section

$$\frac{d^2\sigma}{d|y|dp_T} \tag{2.9}$$

where $|y|$ and p_T are the absolute rapidity and the transverse momentum of the jet. The minimum p_{T} considered here was 20 GeV and the maximum rapidity 4.4. The second distribution investigated is the double differential cross section

$$\frac{d^2\sigma}{d\mathrm{y}^* dm_{jj}} \tag{2.10}$$

where $y^* = |y_1 - y_2|/2$ is the corresponding rapidity of the jets in their mutual centre of mass frame and m_{jj} the dijet mass. To allow for comparison with fixed order perturbative QCD it is necessary to also apply a cut requiring that

$$p_{T\,1} \geq p_{T\,2} + \Delta, \tag{2.11}$$

where $p_{T\,1}$ and $p_{T\,2}$ are the minimum transverse momenta of the two leading jets, and Δ is a non-zero constant. As $\Delta \to 0$ the fixed order predictions become unstable [7]. The instability is caused by the vetoing of soft radiation leading to large uncanceled virtual corrections. In this thesis $p_{T\,1} = 30\,\mathrm{GeV}$ and $p_{T\,2} = 20\,\mathrm{GeV}$ are considered for these distributions.

2.5 Dijet With a Central Jet Veto

The main analysis goal of this thesis was to make a measurement of the jet environment restricted to a rapidity region between a chosen dijet system. As a simplification of a typical event one can consider the energy originating from the interaction flowing through a cylinder, which behaves as an ideal detector. In figure 2.7 a typical layout of an event structure is projected onto the surface of a cylinder, whose centre is the position of interaction. The circles on the surface of the cylinder represent areas where energy from the interaction intersects the cylinder. The yellow region is defined to be the interval within which a test is made for additional energy.

2.5.1 Variables of Interest

The primary test of the central jet veto (CJV) study is the fraction of events which do not contain additional radiation above a certain minimum threshold in the defined interval region. This quantity shall be referred to as the gap fraction and is defined as

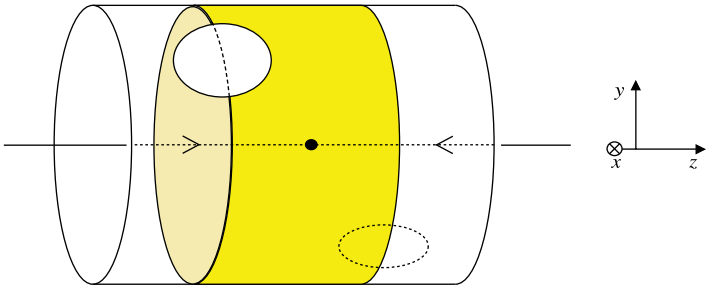


Figure 2.7: Visualization of the energy flow from an interaction through a cylindrical surface. The yellow region indicates the area to be tested for additional radiation.

$$\text{Gap Fraction} = \frac{\sigma_{empty}}{\sigma_{total}}, \tag{2.12}$$

where σ_{empty} is the cross section for events that satisfy a veto requirement of no additional radiation in the interval region and σ_{total} is the total cross section. The gap fraction was studied as a function of both the size of the interval between the dijet system, $|\Delta y|$, and the average transverse momentum, $\overline{p_T}$, of the jets which defined the interval. As $|\Delta y|$ increases in size the available phase space available for the emission of radiation increases and so the gap fraction should fall. However as both $|\Delta y|$ and $\overline{p_T}$ increase conservation of energy will prevent the production of radiation and so the gap fraction should slow its descent and eventually start to rise again.

The explicit requirement for the additional radiation is that there should be no jet found in the interval region with

$$p_{T\,3} > Q_0, \tag{2.13}$$

where Q_0 is a scale chosen to be as low as minimally possible for the detector to measure accurately. The dependence of the gap fraction as the veto scale Q_0 is varied is also of interest. By varying the Q_0 the dependence of the soft underlying physics can be tested. If a high value of Q_0 is chosen then there should be little effect from the underlying event or pile-up.

Another test of the CJV is the average multiplicity of jets, $\overline{N_{jet}}$ within the interval region

above the veto scale. The behavior is opposite to the gap fraction, where $\overline{N_{jet}}$ will rise with increasing |Δy|. Again this quantity will be investigated as a function of both |Δy| and $\overline{p_{\mathrm{T}}}$.

2.5.2 Kinematic Selection

There were three kinematic selections considered in the dijet with CJV analysis, these shall be referred to as selection A, B and C. By default the jets used in the selection definitions were found using the anti-k_t jet algorithm with $R = 0.6$. The only deviation from these kinematic selections occurs in chapter 7 where anti-k_t jet algorithm with $R = 0.4$ jets were also considered and the veto threshold in selection A is varied.

Selection A

1. All considered jets must have $|y| < 4.4$.
2. Interval defining jets are the two highest p_{T} jets.
3. The interval defining jets must both have $p_T > 20\,\mathrm{GeV}$ and $\overline{p_T} = (p_{\mathrm{T1}} + p_{\mathrm{T2}})/2 > 50\,\mathrm{GeV}$.
4. The veto jet transverse momentum threshold is $Q_0 = 20\,\mathrm{GeV}$.

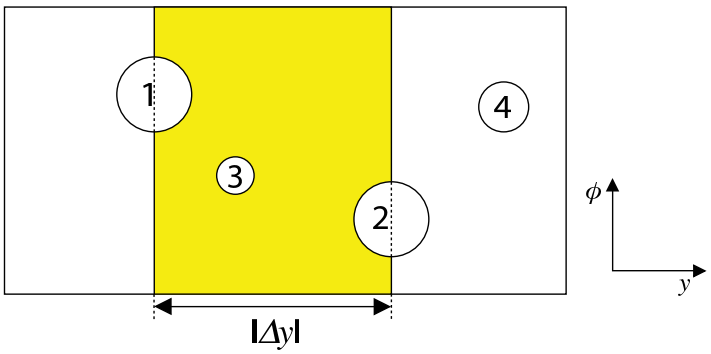


Figure 2.8: Diagram of the selection A in a typical event layout.

In figure 2.8 a typical layout of the event structure unrolled into the rapidity-ϕ plane is shown for selection A. The yellow region indicates the area of the detector where the jets are tested for the veto requirement, the extent of this region is between the centres of the objects identified as 1 and 2. If object 3 of the event meets the veto requirement then the interval region

will be defined as filled. Object 4 does not play a role in this event as it neither defines the interval region nor resides in a position where it would be considered as a possible veto.

Selection B

1. All considered jets must have $|y| < 4.4$.
2. The two interval defining jets are the most forward and backward jets in rapidity with $p_T > 20\,\text{GeV}$
3. The interval defining jets must have $\overline{p_T} = (p_{\text{T1}} + p_{\text{T2}})/2 > 50\,\text{GeV}$.
4. The veto jet transverse momentum threshold is $Q_0 = 20\,\text{GeV}$.

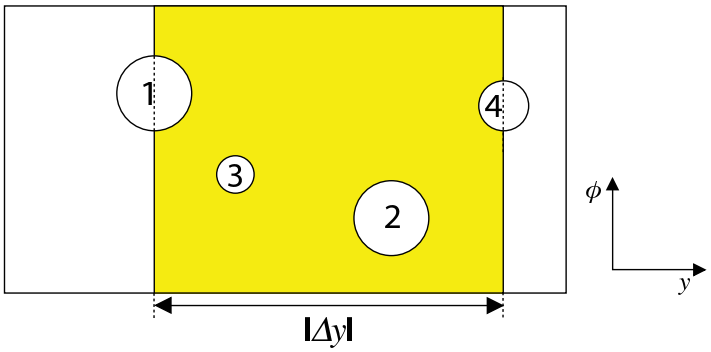


Figure 2.9: Diagram of the selection B in a typical event layout.

In figure 2.9 a typical layout of the event structure unrolled into the rapidity-ϕ plane is shown for selection B. As in the previous selection the yellow region indicates the area of the detector where the jets are tested for the veto requirement, the extent of this region is now defined between the centres of the objects identified as 1 and 4 as these jets are the most forward and backward jets that satisfy the minimum p_{T} threshold for interval definition jets. There are now two jets inside the interval region that must have their energy tested. By making this selection the $\overline{p_{\text{T}}}$ of events will be typically smaller than selection A.

Selection C

1. All considered jets must have $|y| < 4.4$.
2. The two interval defining jets are most forward and backward jets in rapidity with $p_T > 20\,\text{GeV}$
3. The interval defining jets must have $\overline{p_T} = (p_{\text{T1}} + p_{\text{T2}})/2 > 50\,\text{GeV}$.
4. The veto jet transverse momentum threshold is $Q_0 = \overline{p_T}$.

The only difference between selection B and C is the choice of veto definition. In selection C the veto threshold equal to $\overline{p_{\text{T}}}$, therefore adjusts on an event-by-event basis.

Bibliography

[1] I. J. R. Aitchison and A. J. G. Hey. *Gauge Theories in Particle Physics*, volume 1. Taylor and Francis, 2003.

[2] P. Skands. QCD for Collider Physics. *ArXiv e-prints*, April 2011.

[3] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and Collider Physics*, volume 8. Cambridge University Press, 1996.

[4] G. C. Blazey et al. Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop. *ArXiv High Energy Physics - Experiment e-prints*, May 2000.

[5] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k$_t$ jet clustering algorithm. *JHEP*, 0804:063, 2008.

[6] Georges Aad et al. Search for New Physics in the Dijet Mass Distribution using 1 fb^{-1} of pp Collision Data at $\sqrt{s} = 7$ TeV collected by the ATLAS Detector. 2011.

[7] Andrea Banfi and Mrinal Dasgupta. Dijet rates with symmetric E$_t$ cuts. *JHEP*, 01:027, 2004.

Chapter 3

The LHC and ATLAS

In this chapter the Large Hadron Collider is described. This includes a brief description of the accelerator complex, the evolution of the collider energy and a discussion of the machine luminosity. The ATLAS detector including the inner detector tracker, calorimeters, magnet system and trigger system will also be described. The chapter finishes with a description of the reconstruction and calibration of jets.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) [1][2][3], situated on the border of France and Switzerland, is the largest and most powerful particle physics facility constructed to date. The LHC accelerator is housed in a 26.7 km ring in which two proton beams rotating in opposite directions are brought into collision at four interaction points. The sites of those collisions are the experiments ATLAS, ALICE, CMS and LHCb. ATLAS and CMS are general purpose detectors whilst ALICE and LHCb specialise in the areas of heavy ion and B-physics respectively.

3.1.1 Accelerator Complex

The LHC uses almost all of the accelerators present at the CERN facility to to produce the proton beams in the correct configuration and energy. The configuration of the beams includes the number of protons per bunch, the number of bunches of protons and the spacing between the bunches. The protons are successively accelerated by Linac2, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before they are injected into the main LHC ring at an energy of 450 GeV. Figure 3.1 shows the layout of the accelerators.

In the main ring, the two beams are accelerated using the same superconducting magnets but in separate beam pipes. There are 9300 superconducting magnets in the main ring which operate at temperatures below 1.9 K.

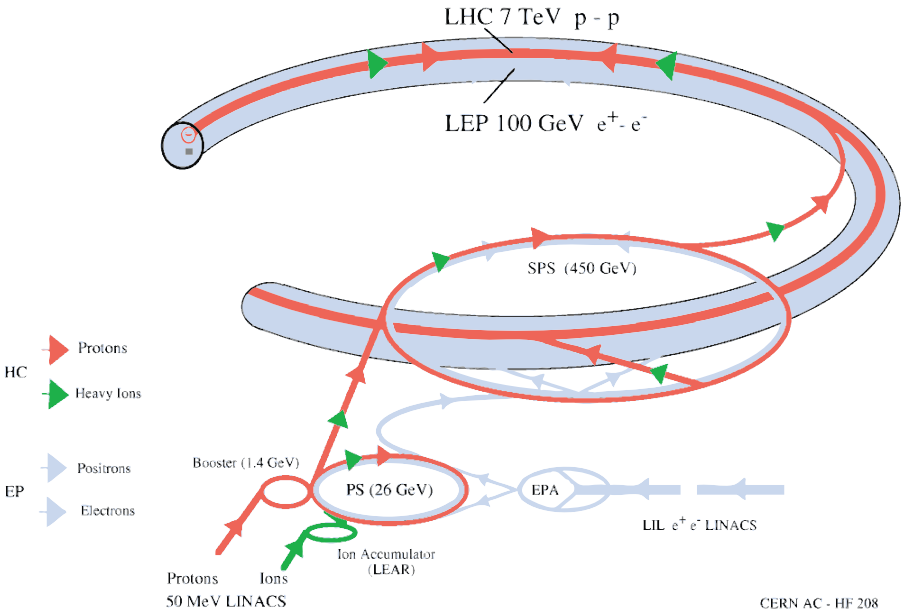


Figure 3.1: The LHC accelerator layout [4].

Collision Point

There are long straight sections of the accelerator at the four interaction points. Dipole and quadrupole magnets are used to steer and focus the beams towards a collision point at a non-zero crossing angle. The beams are kept apart as long as possible to prevent the

occurrence of parasitic collisions, where there is a collision between two protons outside of the experimental detectors.

3.1.2 Beam Energy

Early Problems

The design beam energy of the LHC is 7 TeV for each beam. Issues arising from the preparation of the magnets resulted in the planned beam energy being lowered to 5 TeV in 2008. Unfortunately an incident involving the connection between two magnets resulted in an explosion which caused damage to the accelerator [5]. This incident resulted in a delay to the start of LHC operations whilst magnets were repaired and work was carried out to prevent the same incident from happening again. In late 2009, the first collisions occurred at an energy of 450 GeV per beam and reached 1.18 TeV per beam before Christmas.

Physics Operation

The beam energy of 3.5 TeV was reached in early 2010 and is still the highest operating energy. A short period of proton-proton collisions were taken in 2011 with beams of energy 2.76 TeV. It is not expected that the design energy of 7 TeV will be reached until 2013 as a number of technical improvements are needed [6], which are planned for the long shutdown.

3.1.3 Machine Luminosity

The luminosity of the machine can be described by the formula

$$L = \frac{kN^2 f}{4\pi\sigma_x\sigma_y} \tag{3.1}$$

where k is the number of bunches, N is the number of protons in each bunch, f is the revolution frequency, σ_x and σ_y are the beam sizes at the collision point. The luminosity can be increased by decreasing the area of the beams at the collision point, increasing the number

of protons, or increasing the number of bunches. At peak luminosity operation in 2010, the operating conditions consisted of 348 bunches containing 1.2×10^{11} protons producing a luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The design luminosity for the LHC is $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, where the luminosity times the cross section is the collision rate. This was expected to be achieved with 2808 bunches containing 1.15×10^{11} protons each.

3.2 ATLAS

The ATLAS experiment (**A** **T**oroidal **L**HC **A**pparatu**S**) [7] is situated at Point 1 on the LHC ring. ATLAS is a general purpose experimental apparatus. ATLAS has been designed to investigate a wide range of physics including QCD, gauge bosons, B final states, top, heavy ion collisions and a range of exotic physics. The subdetectors are arranged in concentric layers surrounding the interaction point. Closest to the interaction is the inner tracking detector, which reconstructs charged particles. Outside this layer the calorimeters reside, which provide granular energy measurements of energy deposited by jets. The outermost layers are dedicated to the measurement of the properties of muons. The diagram showing the main subdetectors of the ATLAS experiment is shown in figure 3.2. The massive scale of the experiment is illustrated by the size of the human figures within the diagram. Further details of these subdetectors will be given in the following sections.

3.2.1 Conventions

There are a number of conventions and definitions used by the ATLAS collaboration: -

1. The positive x direction points towards the centre of the ring and the positive y direction vertically upwards. Using a right handed coordinate system the positive z direction is parallel with the beam towards LHCb.

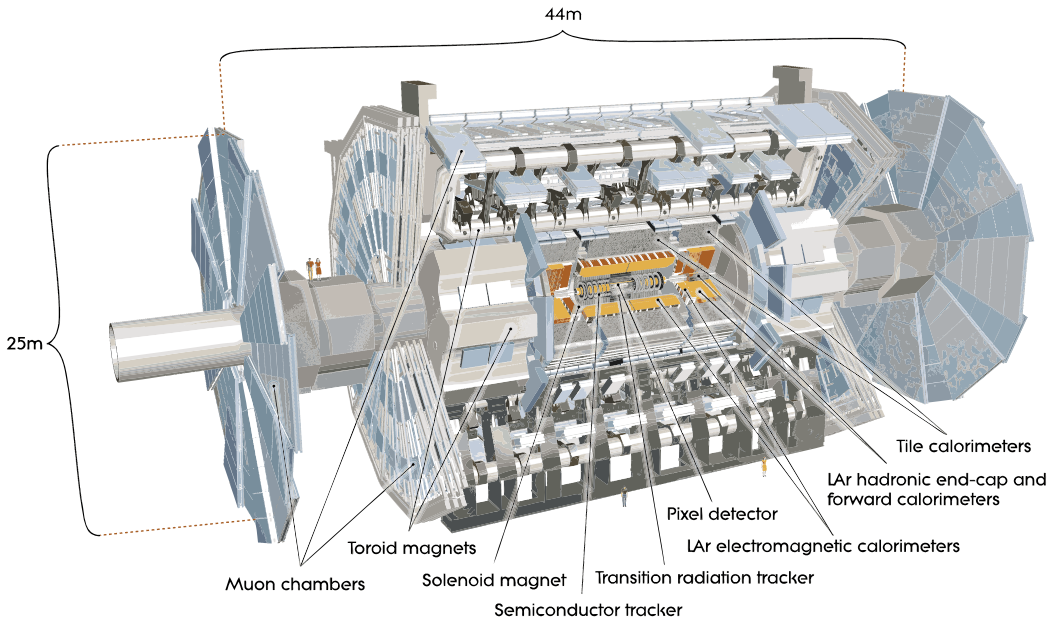


Figure 3.2: The ATLAS detector [7].

2. The polar angle θ is defined from the beam axis

$$cot(\theta) = \frac{z}{\sqrt{x^2 + y^2}} \tag{3.2}$$

3. The azimuthal angle is defined around the beam axis

$$tan(\phi) = \frac{p_y}{p_x} \tag{3.3}$$

where p_x and p_y are the momenta in the x-direction and y-direction.

4. The pseudorapidity is defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{3.4}$$

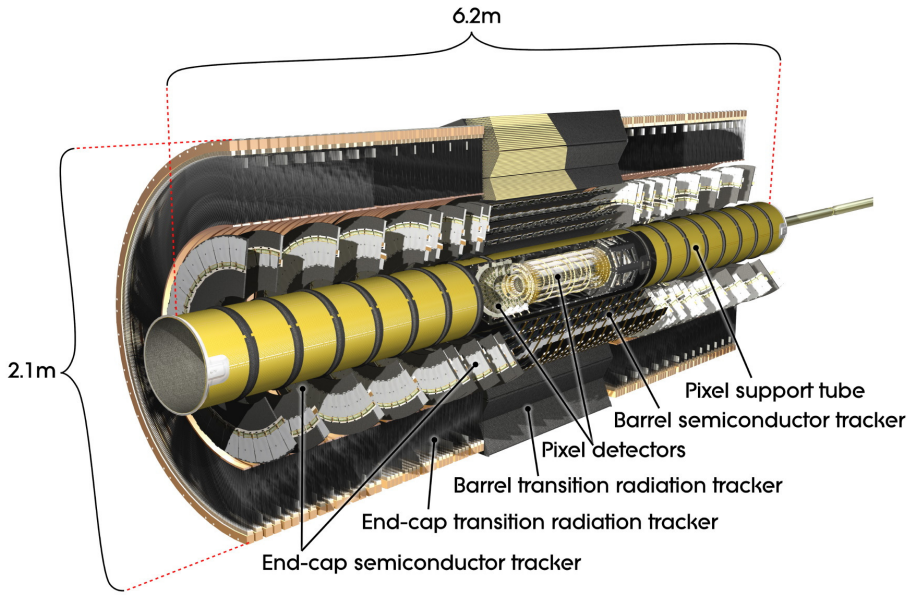


Figure 3.3: The inner detector layout.

3.2.2 Inner Detector Tracker

The inner detector as depicted in figure 3.3 is constructed out of a number of sub detectors which lie close to the interaction point. The inner detector allows the measurement of the sign and momentum of charged particles using the curvature of the track in the 2 T magnet field. A number of layers are needed to measure the position of particles along their track with high precision.

The inner detector allows the measurement of the interaction vertices in each event to be made with high efficiency. The efficiency for reconstructing a vertex is greater than 99% when the number of tracks with $p_T > 100$ MeV is greater than 3. The position resolution of vertices in events with more than 70 tracks or $\sqrt{\sum_{trk} p_T^2} > 8$ GeV is 30μm in the transverse plane and 50μm in the direction of the beam [8]. The reconstruction efficiency for tracks has been observed to be greater than 80% for tracks with $p_T > 500$ MeV and $|\eta| < 1$ [9]. The track reconstruction efficiency falls to 60% at $|\eta| = 2.5$, the edge of the inner detectors acceptance. The fall of the efficiency at high $|\eta|$ is due to the increased amount of material in this region.

Pixel Detectors

The innermost layers of the inner detector are composed of the silicon pixel detectors [10], consisting of three barrel layers and three disks at each side of the forward region. The silicon pixel detectors provide high granularity measurements close to the interaction point, and so contribute greatly to the vertex measurement. In the barrel the R-ϕ accuracy is 10 μm and z accuracy is 115 μm. In the end-caps the R-ϕ accuracy is 10 μm and R accuracy is 115 μm.

Semiconductor Tracker

The next sub-detector that particles pass through is the semiconductor tracker (SCT) [11], which is constructed from Silicon Strip Modules. Pairs of modules are glued back to back with a 40 mrad stereo angle between the sensors. These are arranged in a central barrel with four cylindrical layers and nine disks at either side of the forward region. In the barrel the R-ϕ accuracy is 17 μm and z accuracy is 580 μm. In the end-caps the R-ϕ accuracy is 17 μm and R accuracy is 580 μm.

Transition Radiation Tracker

The final sub-detector important for tracking in the inner detector is the Transition Radiation Tracker (TRT) which consists of around 300,000 proportional drift tubes [11]. As particles pass through the tubes the gas is ionised, the electrons then drift to the wire where a signal is produced. The R-ϕ accuracy is 130 μm. The TRT also provides particle identification as the emission of transition radiation is different for hadrons and electrons.

3.2.3 Calorimetry

The goal of the calorimeters is to measure the energy of the photons, electrons, jets and the missing transverse momentum. They can also provide information for particle identification. The ATLAS calorimeter is hermetic; it provides large angle coverage for $\eta \leq 4.9$ and complete ϕ coverage [12]. The fractional energy resolution for $p_T = 20$ GeV jets was evaluated

from MC to be 18%, whilst for $p_T > 100$ GeV the resolution was $< 9\%$. A number of in situ techniques have been investigated by ATLAS to test the reliability of MC prediction for the reconstruction efficiency and energy resolution of measured jets [13]. The selection efficiency, determined through a tag and probe technique using jets composed of of tracks, was seen to be greater than 98% for jets with $p_T > 20$ GeV. The jet energy resolution was measured using two techniques, the di-jet balance method and the bi-sector method. In both cases it was seen that the MC correctly describes the jet energy resolution within 14%.

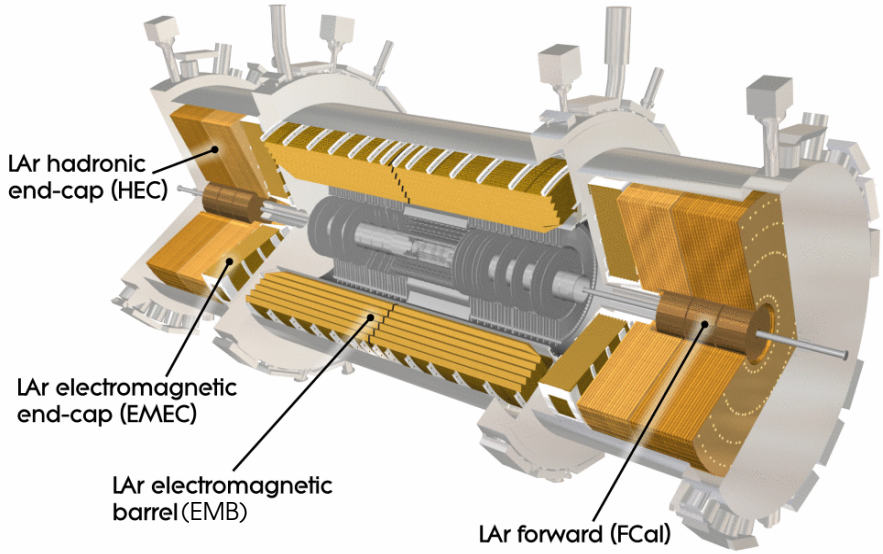


Figure 3.4: The liquid argon calorimeters.

Electromagnetic Calorimeter

The electromagnetic (EM) calorimeter is able to reconstruct with good resolution the energy of electrons and photons over a wide range of energies. It is important for the ability to distinguish electrons, photons and jets. The primary interaction which occurs when a high energy electron interacts with matter is bremsstrahlung, the emission of a photon. Whilst a high energy photon when interacting with matter will produce an electron and positron

pair. In this way a shower of particles will develop, where the path length through the material controls the number of interactions and hence the number of electrons produced. The calorimeter alternates between an absorber material which develops the shower through the EM interactions and electrodes which measure the current of electrons drifting in the gaps. The EM calorimeter uses liquid Argon (LAr) as an active medium and Lead as the absorber. The LAr EM barrel calorimeter is found in the central region $|\eta| < 1.475$, and the LAr EM end-cap calorimeter is present in the range $1.375 < |\eta| < 3.2$. The LAr calorimeter can be seen in figure 3.4.

Hadronic Calorimeter

The task of the hadronic calorimeter is to reconstruct jets and measure the missing transverse energy. It has low noise, high granularity and must be of sufficient thickness for the shower to be fully encapsulated in the calorimeter. The high energy hadrons interact with the absorber material through the strong interaction. The result of the interaction is the production of several lower energy hadrons. This process repeats until the hadrons stop or are absorbed. The hadronic LAr calorimeters are split into two components; the hadronic end-cap calorimeter covers the region $1.5 < |\eta| < 3.2$, and the forward calorimeter $3.1 < |\eta| < 4.9$. The LAr hadronic end-cap uses Copper as an absorber whilst the forward calorimeter uses both Tungsten and Copper. The hadronic tile calorimeter, which is a sampling calorimeter made from steel plates interspersed with scintillators, provides coverage in the central pseudorapidity region $|\eta| < 1.7$.

3.2.4 Magnet System

The ATLAS magnet system [14] is composed of the Central Solenoid together with the Barrel and End-Cap Toroids. The Central Solenoid operates at a current of 7.7 kA to produce a 2 T axial magnetic field within the inner detector. The Toroid magnets are installed outside of the calorimeter and operate at 20.4 kA to produce a 1 T tangential magnetic field for the

muon detectors.

3.2.5 Muon System

The muon system is composed of four different detector chamber technologies that lie outermost from the interaction point, where few particles other than neutrinos and muons can reach. The four systems are the Resistive Plate Chambers (RPCs), Monitored Drift Tube (MDT) chambers, Cathode Strip Chambers (CSCs) and the Thin Gap Chambers (TGCs) [7]. To measure the track properties three points on muons track must be identified, as with the inner detectors measurement of tracks. The RPCs form the first layer encountered within the barrel region, $|\eta| < 1$. The RPC is a simple gaseous detector which provide trigger information. Next to be encountered are the MDTs, which provide the precision track measurement.

In the forward most region of instrumentation for detecting muons, $2 < |\eta| < 2.7$, measurement of the muon tracks is made by the CSCs. The CSCs are multiwire proportional chambers which measure the passage of a muon by the formation of an avalanche on the anode wire. Here the TGCs provide the triggering information and additional track measurements.

3.2.6 Trigger System

The expected rate of interactions is too high to record the detector information for all the events observed so a selection must be made. This selection is carried out by the trigger system. The trigger system is made up of 3 levels; Level 1, Level 2 and the Event Filter. The Level 1 trigger is limited to a subset of detector information and the decision must be made within 2.5 μs to handle the design interaction rate of 40 MHz. At maximum, the Level 1 trigger can accept an event rate of 75 kHz but during 2010 the maximum rate kept was 30 kHz [15]. If accepted by the Level 1 trigger the event is transferred to Level 2 where regions of interest identified in the first level are investigated. Level 2 has complete access to the data but is still limited by time restrictions. The rate is lowered by Level 2 to approximately

3 kHz.

The final stage of the trigger system is the Event Filter which can run reconstruction algorithms of greater complexity because it has more time to make its decision. This results in improved efficiency and the ability to set tighter thresholds. The final maximum event rate recordable to disk was 200 Hz.

Minimum Bias Triggers

The Minimum Bias Trigger Scintillators (MBTS), which also lies within the inner detector, and the silicon tracking detectors provide the minimum bias triggers. The MBTS is divided into two rings and eight ϕ sections that cover the region $2.09 < |\eta| < 3.84$. The efficiency for triggering on an event when there is a track incident on the MBTS with $p_T > 100$ MeV is above 97% [16]. Given this high efficiency it makes the MBTS an invaluable tool for evaluating the absolute trigger efficiency for the jet triggers.

Jet Triggers

The jet triggers select events which contain high transverse momentum deposits in the Liquid Argon and Tile calorimeters [11]. For $|\eta| < 3.2$ at Level 1 a $\Delta\eta$-$\Delta\phi$ sliding window of dimensions 0.4×0.4, 0.6×0.6 and 0.8×0.8 is used to search for energy deposits above a chosen threshold. In the Forward Calorimeter there is no η granularity at Level 1. The Level 2 trigger uses a simplified cone algorithm to find jets whilst the Event Filter uses the same jet algorithms as physics analysis. The Event Filter for the jet triggers was not enabled for data collected in 2010. The performance of the jet triggers were seen to be very good with efficiencies > 99% observed for events containing jets with $p_T > 60$ GeV within the region of $|\eta| < 3.2$ [17].

3.3 Jet Reconstruction

Jets in ATLAS are reconstructed from energy deposited in the calorimeter system. Clusters of energy are built from cells in all of the calorimeter layers, and these clusters are then used as input to the jet algorithms. The cells from which the clusters are built are calibrated to the electromagnetic (EM) scale. The EM scale is defined to be the calibration where the energy of electrons and photons are measured correctly by the calorimeters. After cluster building and jet finding the jets are calibrated through the application of the jet energy scale (JES). The jet algorithm used to reconstruct the jets was anti-k_t with radius parameters R = 0.6 and R = 0.4.

Truth jets are constructed from running the same anti-k_t jet algorithms over the stable particles in the Monte Carlo event record, but not including either muons or neutrinos.

3.3.1 Initial Calibration

A calibration is necessary to convert the calorimeter signals measured in picoColoumbs to the energy that the particles have deposited. In ATLAS it was chosen that all energies should be first reconstructed to the electromagnetic energy (EM) scale. These calibration factors have been obtained by exposing the calorimeters to test beams of electrons and muons with a range of known energies [18].

3.3.2 Cluster building

Clusters are built from neighbouring cells in the calorimeters, where the neighbours include both the surrounding eight cells within each layer and those cells in the surrounding layers [19]. Varying granularity of each layer can result in more than ten neighbouring cells. Given that the clusters lie across multiple layers of calorimeter, they are known as 3d topological clusters. These exploit the fine granularity in the calorimeter system to give a detailed picture of the shower development. The steps to build a 3d topological cluster are:

1. Identify calorimeter seed cells which have a large signal to noise ratio, $|E| > 4\sigma$ where E is the energy of the cell and σ is the noise level of the cell;

2. Iteratively combine seed cells with neighbouring cells that satisfy $|E| > 2\sigma$ to form protoclusters. During this process neighbouring seeds may be merged;

3. All the direct neighbours with $|E| > 0$ are added to each of the protoclusters.

After the protoclusters have been built a splitting procedure is applied. Within the protoclusters local maxima are found. For these maxima, cells must satisfy the requirements of $E > 500\,\text{MeV}$, energy greater than any neighbouring cells and there must be more than three cells within the protocluster with $|E| > 4\sigma$. Clusters are re-grown around these local maxima without any noise requirements or merging of protoclusters. This helps to prevent the formation of clusters covering large fractions of the detector. A typical value of the noise in the EM calorimeter was seen to be 10-40 MeV, with higher levels of noise seen in the forward region.

3.3.3 Calibration to hadron level

The EM scale calibration previously applied does not account for a number of effects including the different fraction of energy that hadrons deposit, inactive material in the detector, particles not contained in the calorimeter, particles that fall outside the reconstructed jet and inefficiencies of clustering and reconstruction. To account for these a jet energy scale correction is applied on a jet-by-jet basis depending on the η and p_{T} of the jet [20]. Through the matching of MC Truth jets and jets at the EM scale the response can be determined, $R^{\text{EM}}(p_T^{\text{jet,EM}}, \eta)$, and so the calibrated momentum of the jets are

$$p_T^{jet} = [1/R^{EM}(p_T^{\text{jet,EM}}, \eta)] \cdot p_T^{\text{jet,EM}}, \tag{3.5}$$

this results in no directional change for the jet. This calibration was shown to correct isolated jet energy to within 2% [20]. The jets are now calibrated to what is known as the EM+JES

level. This was the final calibration applied before the jets were used in the analysis. A number of crosschecks using data have been performed by ATLAS to validate the JES calibration derived, one such tool being the jet-photon balance [21].

Pile-up correction

The default jet energy scale that was applied to the EM+JES calibrated topological cluster jets contains a simple correction for additional energy originating from pile-up [22], which is considered to be uncorrelated with the primary scatter and its jets. This correction does not attempt to remove energy from other sources of underlying event, such as multiple partonic interactions.

The energy offset was derived through the measurement of the energy density in the calorimeter towers using events where one primary vertex was required. This offset was translated to the topological cluster jets by using the average calorimeter tower multiplicity per jet. The correction required can be considered to be a function of the η, the number of primary vertices and the bunch spacing, although the bunch spacing had not yet become important for the 2010 data.

Jet energy scale uncertainty

There are a number of sources of uncertainty for the JES calibration including an uncertainty from single particle propagation, variation of the noise levels in the calorimeters, the MC tune and model used in the evaluation and pile-up. The largest source of uncertainty was the single particle propagation using both in situ techniques and also test-beam measurements [20]. The residual bias after the pile-up correction of ≈ 0.5GeV is estimated to be ≈ 160MeV per additional vertex for jets with $|\eta| < 1.9$. Higher uncertainties in the forward region should have minimal effect for the higher p_{T} interval defining jets and the central veto jet is typically found at low values of $|\eta|$.

Bibliography

[1] Thomas Sven Pettersson and P Lefvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05 LHC, CERN, Geneva, Oct 1995.

[2] G Brianti. The Large Hadron Collider in the LEP tunnel, CERN-LHC-Note-134. page 14 p, Sep 1990.

[3] Oliver Sim Bruning et al. *LHC Design Report*. CERN, Geneva, 2004.

[4] Jean-Luc Caron. The LHC injection complex. L'ensemble d'injection du LHC. AC Collection. Legacy of AC. Pictures from 1992 to 2002., May 1993.

[5] M Bajko et al. Report of the Task Force on the Incident of 19th September 2008 at the LHC. Technical Report LHC-PROJECT-Report-1168. CERN-LHC-PROJECT-Report-1168, CERN, Geneva, Mar 2009.

[6] F Bertinelli et al. Towards a Consolidation of LHC Superconducting Splices for 7 TeV Operation. *CERN-ATS-2010-144*, page 4 p, Jun 2010.

[7] T Lagouri. ATLAS : A General Purpose pp Experiment at the Large Hadron Collider at CERN. Technical Report MPI-PhE-99-19, Max-Planck Inst., Mnchen, Nov 1999.

[8] Performance of primary vertex reconstruction in pp collisions at $\sqrt{s}$ = 7 TeV in the ATLAS experiment. Technical Report ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010.

[9] G. Aad et al. Charged-particle multiplicities in pp interactions at $\sqrt{s}$ = 900 GeV measured with the ATLAS detector at the LHC. *Phys.Lett.*, B688:21–42, 2010.

[10] G Aad et al. Atlas pixel detector electronics and sensors. *Journal of Instrumentation*, 3(07):P07007, 2008.

[11] G Aad et al. Expected performance of the ATLAS experiment: detector, trigger and physics. 2009.

[12] ATLAS calorimeter performance: Technical Design Report. 1996.

[13] Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $\sqrt{s}$ = 7 TeV. Technical Report ATLAS-CONF-2010-054, CERN, Geneva, Jul 2010.

[14] H H J ten Kate. ATLAS superconducting toroids and solenoid. *IEEE Trans. Appl. Supercond.*, 15(2 pt.2):1267–1270, 2005.

[15] Georges Aad et al. Performance of the ATLAS Trigger System in 2010. 2011.

[16] L Tompkins. Performance of the ATLAS Minimum Bias Trigger in pp collisions at the LHC. Technical Report ATL-DAQ-PROC-2010-033, CERN, Geneva, Sep 2010.

[17] Performance of the ATLAS Jet Trigger in the Early $\sqrt{s}$ = 7 TeV Data. Technical Report ATLAS-CONF-2010-094, CERN, Geneva, Oct 2010.

[18] Yu A Kulchitskii et al. Electromagnetic Cell Level Calibration for ATLAS Tile Calorimeter Modules. Technical Report ATL-TILECAL-PUB-2007-001. ATL-COM-TILECAL-2006-013. CERN-ATL-TILECAL-PUB-2007-001, CERN, Geneva, Dec 2006.

[19] W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.

[20] Jet energy scale and its systematic uncertainty for jets produced in pp collisions at $\sqrt{s}$ = 7 TeV and measured with the ATLAS detector. Technical Report ATLAS-CONF-2010-056, CERN, Geneva, Jul 2010.

[21] Determination of the atlas jet energy measurement uncertainty using photon-jet events in proton-proton collisions at $\sqrt{s}$ = 7 TeV. Technical Report ATLAS-CONF-2011-031, CERN, Geneva, Mar 2011.

[22] Jet energy scale and its systematic uncertainty in pp collisions at $\sqrt{s}$ = 7 TeV in ATLAS 2010 data. Technical Report ATLAS-CONF-2011-032, CERN, Geneva, Mar 2011.

Chapter 4

Monte Carlo Generation

In this chapter the basic process of Monte Carlo event generation is described and details about the leading order Monte Carlo generators used are given. The predictions made for the dijet with jet veto measurement are shown for these Monte Carlo generators. A short description of the computer simulation of the ATLAS detector follows. Finally two methods for producing weighted Monte Carlo event samples are described.

4.1 Monte Carlo Generator Theory

Monte Carlo (MC) event generation provides a way to simulate the type of events that are expected to be produced in proton-proton collisions at the LHC.

4.1.1 General Program Flow

1. The generation of the hard process takes place using perturbation theory.

2. Additional partonic activity is generated. Including initial and final state radiation, multiple partonic interactions and beam remnants.

3. The mostly non-perturbative physics is modeled. The fragmentation, hadronisation and decay to stable particles is handled.

4.1.2 PYTHIA

PYTHIA 6 [1] is a Monte Carlo generator capable of simulating the QCD dijet process events at leading order accuracy. The parton shower used in the generation is ordered in transverse momentum of the radiation. These samples cover the production of $2 \rightarrow 2$ parton scatters within a range of $17 < p_T < 1120\,\text{GeV}$, this is accomplished through the use of multiple slices with different parton $p_{T\,min}$ and $p_{T\,max}$. A subset of parameters is listed in table 4.1 for the samples considered in this chapter for standard PYTHIA.

Parton $p_{T\,min}$ [GeV]	Parton $p_{T\,max}$ [GeV]	Cross Section [nb]
17	35	6.78×10^{5}
35	70	4.10×10^{4}
70	140	2.19×10^{3}
140	280	8.77×10^{1}
280	560	2.35×10^{0}
560	1120	3.36×10^{-2}

Table 4.1: Parameters for standard PYTHIA dijet samples. Each sample contains 1.4million events.

The tune used for the FORTRAN PYTHIA samples is AMBT1 [2]. The parameters considered for this tune consisted of MPI, colour reconnection, initial and final state radiation variables. The data used in the making of this tune included the CDF runs 1 and 2; and ATLAS 0.9 TeV and early 7 TeV data. The ATLAS data received a higher weighting. The PDF used for this tune was MRST 2007LO* [3]. The LO* PDF is a leading order PDF with some modifications to make it behave more like a NLO PDF, including the relaxation of the momentum sum rule and the use of the NLO strong coupling definition. In a range of standard model processes the LO* PDF set combined with a LO matrix element process was observed to provide the best agreement with NLO PDF combined with a NLO matrix element process [3]. Thus when only the LO matrix element is available for a process the LO* PDF would be expected to provide the prediction closest to the truth.

4.1.3 Herwig++

Herwig++ [4] is another general purpose event generator which provides leading order dijet production with parton showering, hadronization and multiple interactions. The parton shower used is angular ordered, and so the largest angle emissions are handled first. Herwig++ events were generated with the LHC-UE7-1 tune, which also uses the MRST 2007LO* PDF.

4.1.4 Alpgen

Alpgen [5] provides the generation of leading order processes of up to six outgoing parton legs. The subsequent evolution of the event is handled by HERWIG+Jimmy [6, 7], which showers, hadronizes and produces the multiple partonic interactions. The PDF used for the production of Alpgen events was CTEQ6l1 [8]. The ALPGEN samples used where generated by the ATLAS collaboration and have limited event numbers.

4.2 Monte Carlo Predictions

A set of predictions for the dijet with jet veto analysis are shown in figure 4.1 for the case where the interval is defined by the two highest p_{T} jets. In figure 4.1a the predictions for the gap fraction diverge at large $|\Delta y|$; both Herwig++ and Alpgen gap fractions fall relative to PYTHIA 6. This is consistent with figure 4.1c where the average number of jets in the interval region is higher in the Alpgen and Herwig++ cases. Alpgen is seen to produce significantly more jets than the other generators.

In a variety of regions within figure 4.1 differences of up to 50% are seen between the MC predictions. The variation seen in the predictions can be considered to be representative of the likely large instrinsic theoretical uncertainties, which are expected for these variables. A more detailed evaluation of theoretical uncertainties can be found in [9], where uncertainties of a similar scale to the variation in figure 4.1 can be seen. These substantial uncertainties are

the dominant motivation for the investigation of other possibly more accurate MC programs.

4.3 Detector Simulation

After events had been generated in one of the MC generators it was possible to apply the full detector simulation to the final stable states of the MC event. ATLAS has developed a comprehensive software framework called Athena [10] within which the simulation of the ATLAS detector takes place. The GEANT4 [11] framework is used for the simulation of particles interacting with the detector volume. The reconstruction of these simulated events is made using the same algorithms that were used for data collected by the detector. The precise geometry of the ATLAS detector needs to be described, including all the material with which particles may interact. Dead material in which energy can be lost is considered to be important for the accurate simulation of the detector. The conditions of the detector during operation such as the alignment of detectors, mapping of the magnetic fields and the temperature distribution were also considered.

4.4 Event Weighting

Inside the ATLAS software framework it is possible to apply filtering algorithms, which allows a specific event to be retained for the full simulation stage. Application of these algorithms is applied after the event has been generated by the Monte Carlo generator. This makes it a generally applicable method for any of the Monte Carlo generators inside the Athena framework. Using the information present in the truth event record and derived truth jets allows events of interest to be selected. In this case we would like to increase the yield of events with large intervals between the leading jets.

The standard PYTHIA dijet samples that the ATLAS collaboration produce are unweighted. Thus to cover the entire phase space for dijet production requires that the events are produced in sets with parton p_{T} in different ranges. A deficiency of these samples is

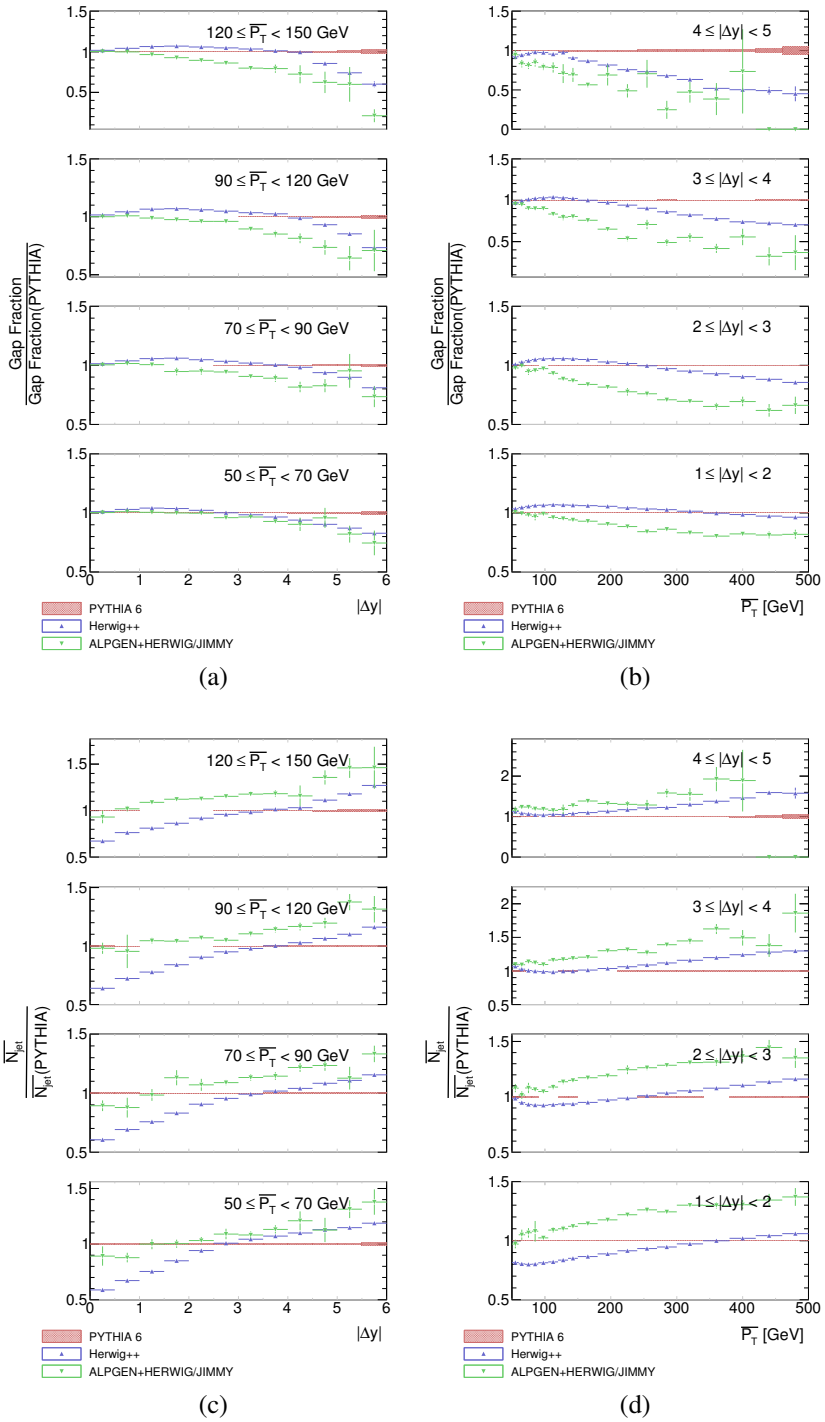


Figure 4.1: Comparison of three MC predictions PYTHIA 6 (AMBT1), Herwig++ (LHC-UE7-1) and ALPGEN+HERWIG/JIMMY (AUET1). The gap fraction as a function of |Δy| is shown for a number of $\overline{p_{\mathrm{T}}}$ regions (a). The gap fraction as a function of $\overline{p_{\mathrm{T}}}$ is shown for a number of |Δy| regions (b). The average number of jets in the event as a function of |Δy| is shown for a number of $\overline{p_{\mathrm{T}}}$ regions (c). The average number of jets in the event as a function of $\overline{p_{\mathrm{T}}}$ is shown for a number of |Δy| regions (d).

that the region of large rapidity separation is poorly sampled. Figure 4.2 shows the number of events generated as a function of the average transverse momentum versus the rapidity separation of the leading two jets. The deficiency in the large |Δy| region is clear; large |Δy| regions have few, if any, events. This could have a detrimental impact on using these events to compare data to MC, in particular the MC will be used for systematic uncertainty determination.

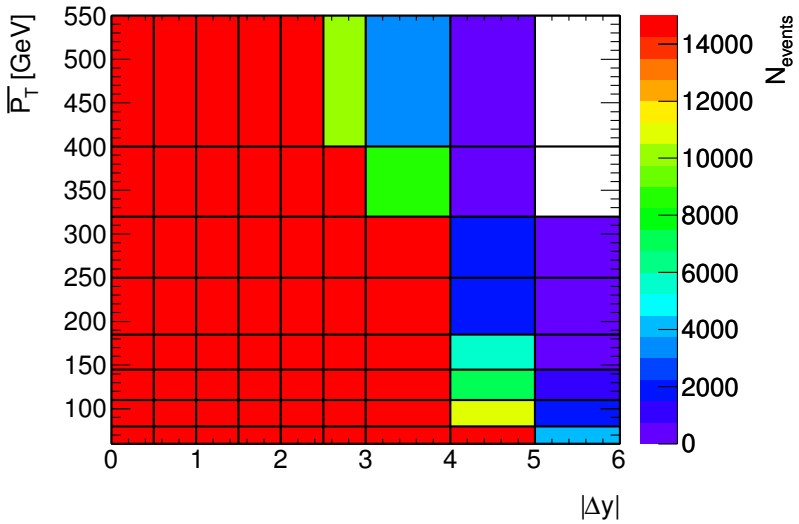


Figure 4.2: The number of events found in the standard PYTHIA event sample as a function of both $\overline{p_T}$ and |Δy| of the leading jets.

Although simply increasing the sample sizes could overcome the problems it would require simulating a factor of 100-1000 more events. At peak production of fully simulated events, ATLAS is able to produce ≈ 10 million events per day. To obtain the necessary event numbers it would take somewhere between 60 and 600 days, if ATLAS were to dedicate all its resources to this analysis. So the use of unweighted events is simply not possible with the relatively time consuming ATLAS full simulation. To overcome these shortcomings filtered and weighted events needed to be generated, which target the underpopulated phase space regions

Two weighting schemes were developed to compensate for the rapidly falling dijet cross section at large rapidity separations. The first involves simply weighting events by the ra-

pidity separation of the leading jets, and the second by both the separation and the p_{T} of the lead jet in the event. Both allow the use of greater levels of weighting to access the largest values of |Δy| which are kinematically allowed at the LHC.

4.4.1 Rapidity separation weighted samples

Weighting Function

The weighting is carried out using an acceptance-rejection method whose weight will vary as a function of |Δy|. A Gaussian function is used to represent the |Δy| distribution. An example of this can be seen in figure 4.3, where there is seen to be reasonable agreement even for such a simple model. The weight is then given by the expression

$$W = \frac{\exp\left(-0.5\left(\frac{|\Delta y|-\mu}{\sigma}\right)^2\right)}{\exp\left(-0.5\left(\frac{|\Delta y|_{max}-\mu}{\sigma}\right)^2\right)} \text{ for } |\Delta y| < |\Delta y|_{max} \tag{4.1}$$

where |Δy| is the interval size between the leading jets, σ is the width of the fitted Gaussian distribution, μ is the mean of the fitted Gaussian distribution and $|\Delta y|_{max}$ is the maximum interval size which is to be weighted, referred to in text as the control point. Above the control point the events are all kept with unit weight.

The parameters used for the weighting function are given in table 4.2. The cross section falls as the parton p_{T} increases, which is expected because it requires more energetic jets to produce the same p_{T} jets at higher rapidity values. The inverse of the efficiency numbers gives a rule of thumb for how many events need to be generated on average for one to be kept by the filter. For example approximately 625 events need to be generated for one event to pass the filtering.

Weighted Generation

The steps carried out to generate rapidity separation weighted samples are

1. Apply jet finding to the stable final state particles in the event record. The chosen jet

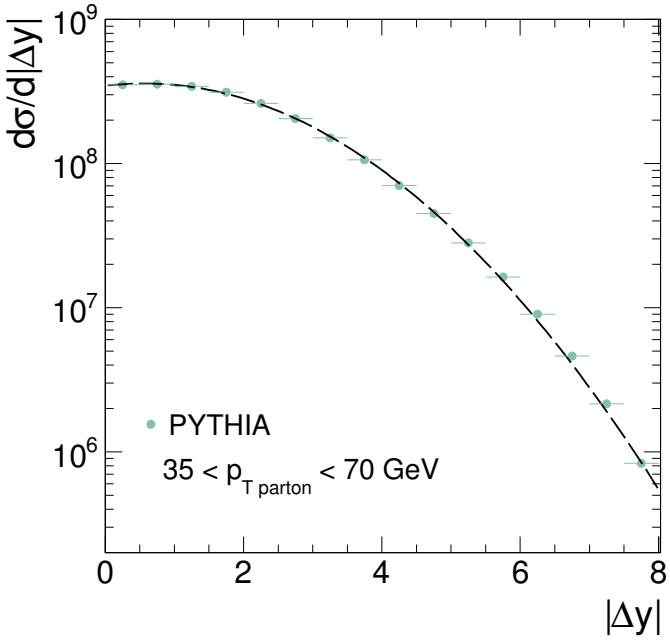


Figure 4.3: Gaussian fit (dashed line) of the interval size distribution, |Δy|, for the unweighted subsample of PYTHIA dijet events. ROOT [12] is used to carry out the fitting.

$p_{T\,min}$ [GeV]	$p_{T\,max}$ [GeV]	μ	σ	$\lvert\Delta y\rvert_{max}$	Efficiency	Event Number
17	35	0.58	2.10	8.0	4.9×10^{-3}	1,400,000
35	70	0.58	1.90	7.0	8.8×10^{-3}	1,400,000
70	140	0.59	1.60	6.0	9.3×10^{-3}	500,000
140	280	0.55	1.41	6.0	1.6×10^{-3}	500,000
280	560	0.49	1.20	5.0	2.1×10^{-3}	500,000
560	1120	0.46	0.95	4.0	2.3×10^{-3}	500,000

Table 4.2: Parameters for generation of the |Δy| weighted PYTHIA dijet samples.

algorithm is anti-k_t and uses a radius parameter of $R = 0.6$. This is the same choice of jet algorithm primarily used in the main analysis presented in this thesis.

2. Events are filtered by requiring at least two jets with $p_{\mathrm{T}} > 12\,\mathrm{GeV}$ in the event.

3. The |Δy| between the two highest p_{T} jets in the event is calculated.

4. The event weight is calculated using the formula 4.1 with the events interval size.

5. A random number is generated between zero and one. The event is kept if the random number is less than the inverse of the event weight.

Results

It can be seen in figure 4.4 that the coverage of the new weighted sample is greatly improved over the standard events in figure 4.2. Even the high $\overline{p_{\mathrm{T}}}$ region, where the weighted samples are significantly smaller, still shows great improvement.

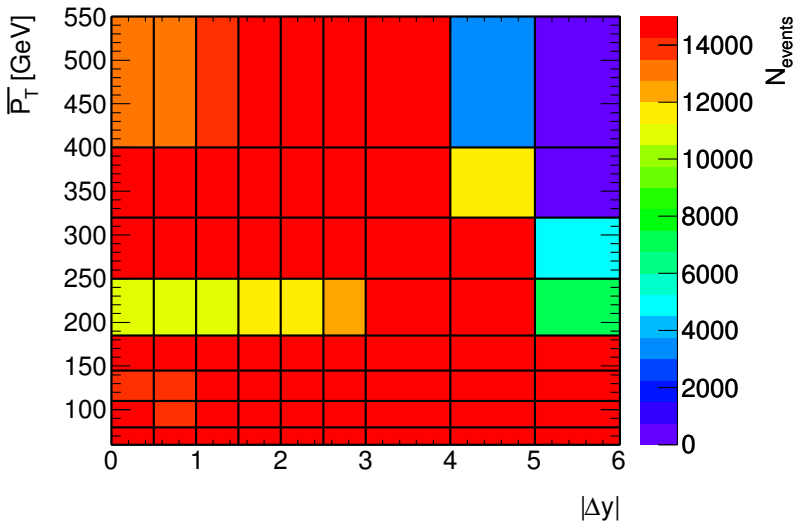


Figure 4.4: The number of events found in the |Δy| weighted PYTHIA event sample as a function of |Δy| of the leading jets.

4.4.2 p_{T} and |Δy| weighted samples

A deficiency found in the previous weighted samples was that only selecting events based on the |Δy| value would reduce the fraction of events which had both high |Δy| and $\overline{p_{\mathrm{T}}}$ within each individual parton p_{T} slice. This problem can be seen in figure 4.4 where there is a lack of events in the region $180 < \overline{p_T} < 250$ GeV. To counteract this, a new improved weighting method was developed.

Weighted Function

A two dimensional function considering both the leading jet p_{T} and also the |Δy| is used. The functional form used to represent the unweighted distribution is a mixture model composed of a set of bivariate distributions represented by equation (4.2)

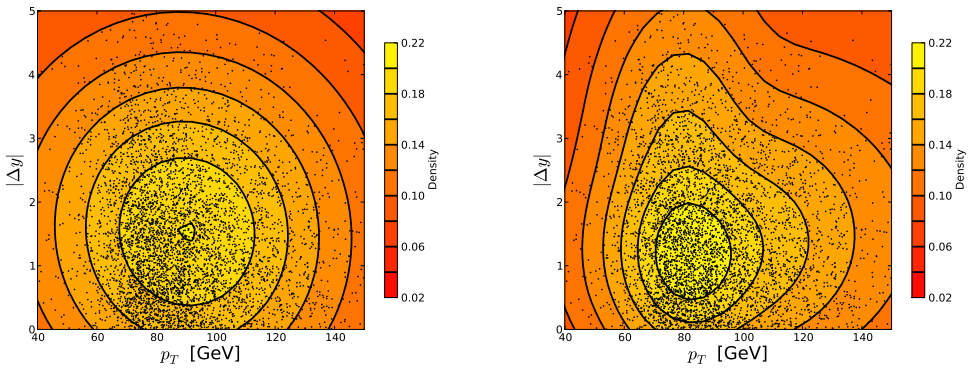
$$F(|\Delta y|, p_T) = \prod_{i=1}^{N} w_i f_i(|\Delta y|, p_T) \tag{4.2}$$

where w_i is the weight for each bivariate distribution, f_i, which is defined in equation

$$f_i(|\Delta y|, p_T) = \frac{1}{2\pi\sigma_{|\Delta y|,i}\sigma_{p_T,i}\sqrt{1-\rho_i^2}} \exp\left(-\frac{1}{2(1-\rho_i^2)}\left[\frac{(|\Delta y| - \mu^i_{|\Delta y|})^2}{\sigma^2_{|\Delta y|,i}} + \frac{(p_T - \mu^i_{p_T})^2}{\sigma^2_{p_T,i}} - \frac{2\rho_i(|\Delta y| - \mu^i_{|\Delta y|})(p_T - \mu^i_{p_T})}{\sigma_{|\Delta y|,i}\sigma_{p_T,i}}\right]\right) \tag{4.3}$$

where ρ is the correlation between the |Δy| and p_{T} variables, μ_x is the mean value of variable x and σ_x is the width of variable x. The increased complexity of the model means that it is no longer possible to use ROOT to perform the fit. Instead a package present in the SciPy machine learning framework [13] was used. This uses an expectation maximization routine to iteratively improve the model's fit from an initial guess. The number of bivariate distributions used in the mixture model is a free parameter. The use of five distributions was

found sufficient for the purpose of representing the density of the events. It should however be possible to optimize the choice of the number of bivariate distributions. Over-fitting is not necessarily a problem in this case as the density of events will still be described well, however introducing too many parameters will result in the fit taking a very long time. Figure 4.5 shows two example fits using mixture models containing either one or five gaussian distributions. The case with only one gaussian distribution is not a good fit of the sample, whilst the second example shows a good fit can be achieved with five distributions. It should be noted that the fit in the low density regions will not be very good, but the aim of the weighting is to suppress the regions which already contain a high density of events, thus those areas will have good determination of density.



(a) (b)

Figure 4.5: The contour lines show the density of the mixture model which had been fitted to the unweighted PYTHIA sample with $70 < p_{T\,lead\ parton} < 140$ GeV. Subfigure (a) used a single Gaussian distribution and (b) used five Gaussian distributions in the mixture model. The axis on the right-hand side shows the density of events. The points illustrate the distribution of Monte Carlo events for a small fraction of the sample used to produce the fit.

Weighted Generation

In the filtering algorithm for this method the same requirement is made that there be at least two jets with $p_T > 12$ GeV. The weight is now derived by evaluating equation 4.2 for newly generated events and comparing it to the value obtained at a control point. If the probability density is lower than the control point density then the event weight used will be one. Again

the same rejection technique is used by comparing a random number to the inverse of the weight.

$p_{T\,min}$ [GeV]	$p_{T\,max}$ [GeV]	Efficiency	Event Number
17	35	2.9×10^{-4}	1,000,000
35	70	1.1×10^{-4}	5,000,000
70	140	1.8×10^{-4}	5,000,000
140	280	3.5×10^{-4}	1,000,000
280	560	4.3×10^{-4}	1,000,000
560	1120	5.7×10^{-4}	1,000,000

Table 4.3: Sample statistics for generation of the p_T−|Δy| weighted PYTHIA dijet samples.

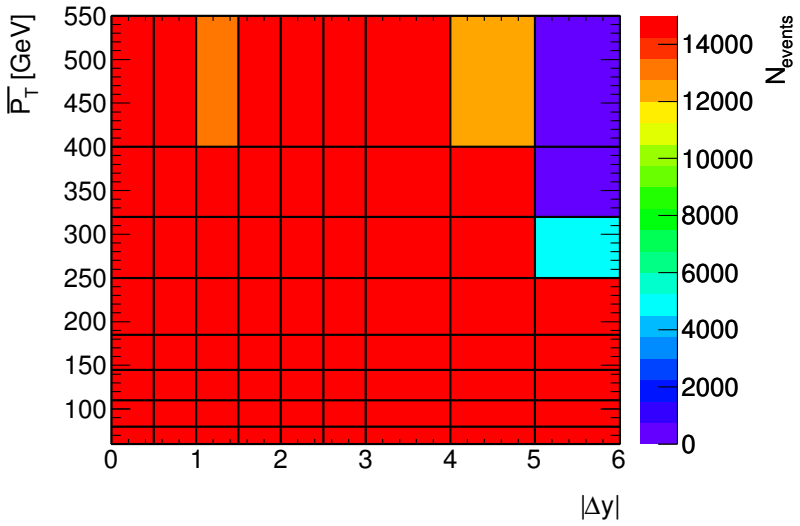


Figure 4.6: The number of events found in the |Δy| weighted PYTHIA event sample as a function of both $\overline{p_T}$ and |Δy| of the leading jets.

Results

The good performance of the weighting can be seen in figure 4.6. The total number of events is normalized to that of the previous weighted sample for the sake of comparison. There is now a greater yield in the bins that were problematic. The complete set of parameters used for event weighting can be found in appendix B.

4.5 Comparison of Weighting Samples

In figure 4.7 the yield for each of the fully simulated PYTHIA samples are compared. The weighted samples can be seen to have a much larger fraction of their events produced in the large |Δy| region. The two-dimensionally weighted samples also typically have the highest number of events in all the samples, although the difference does fall at higher $\overline{p_{\mathrm{T}}}$.

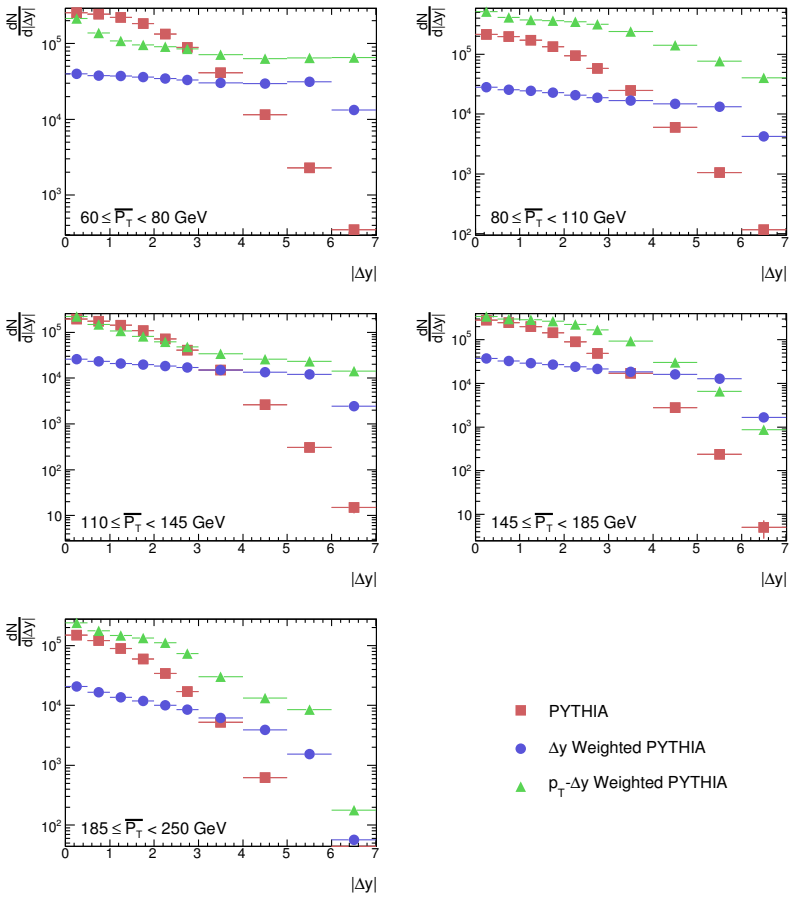


Figure 4.7: Comparison of different PYTHIA events samples and their yields for different regions of $\overline{p_{\mathrm{T}}}$.

Figure 4.8 shows that for all the different samples generated the weighted samples predict consistent cross sections within the statistical uncertainties of the samples. The kinematic suppression of large rapidity intervals can be seen to increase with the higher samples, the highest parton p_{T} slice for example varies over nearly 10 orders of magnitude.

The utility of these weighted samples can be best seen in the improvement of the statis-

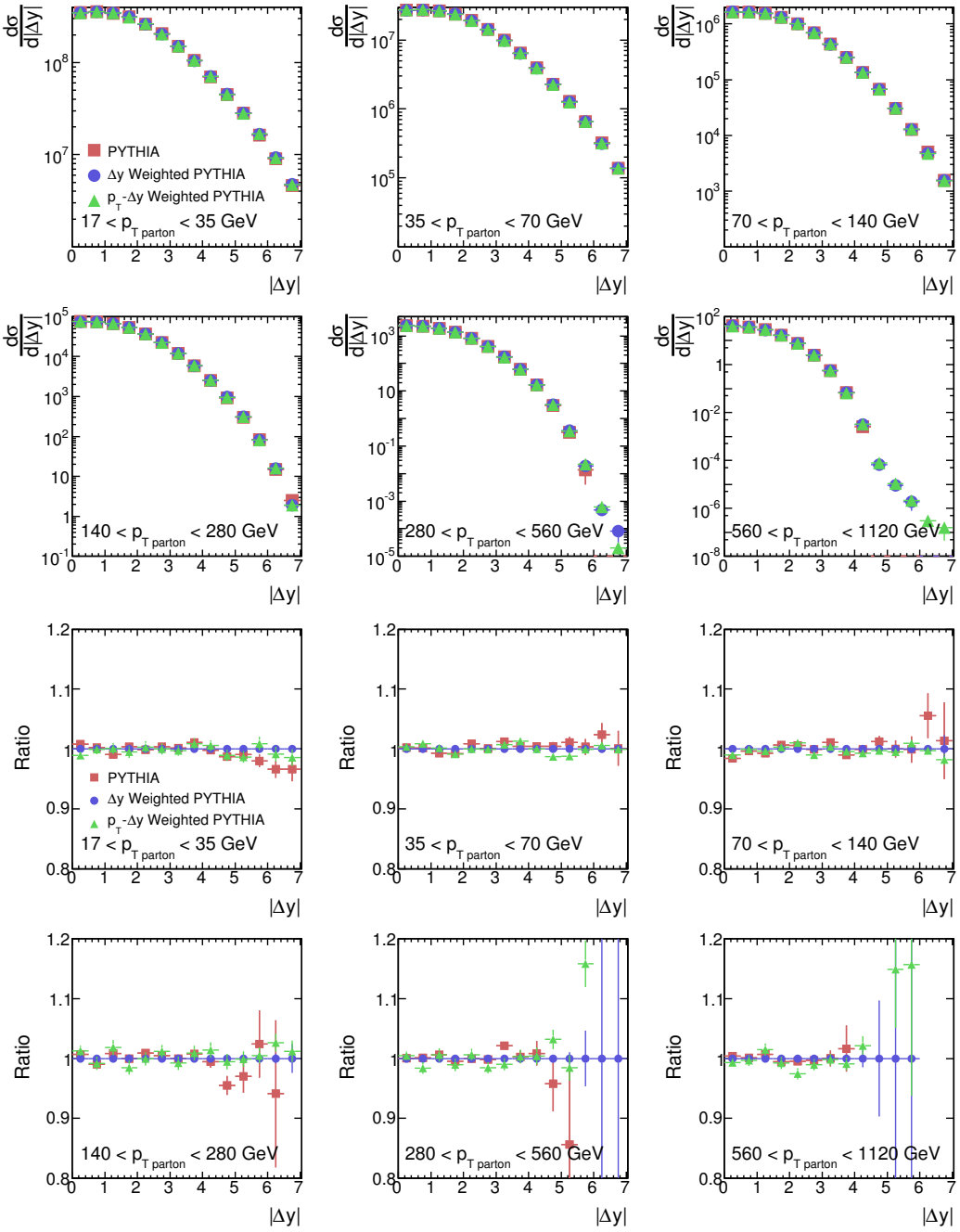


Figure 4.8: Comparison of weighted and unweighted PYTHIA events. The samples are divided into the different parton p_T ranges, with the only requirement that two jets above 12GeV are required. The ratios are shown with respect to the |Δy| weighted events distribution.

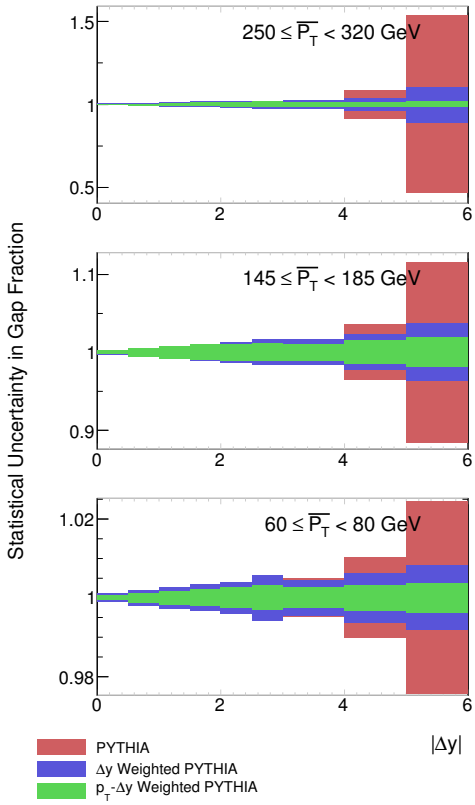


Figure 4.9: Comparison of the statistical uncertainty for the gap fraction as a function of |Δy|.

tical uncertainty in the observables of the dijet with a central jet veto analysis. In figure 4.9 the fractional statistical uncertainty is shown for a few of the gap fraction distributions which are investigated. The growth of the uncertainty with the size of the interval is suppressed in the weighted samples relative to the default PYTHIA samples, where uncertainties are seen to grow to 50%. It is also observed that the p_T-|Δy| weighted sample has a lower statistical uncertainty than the |Δy| weighted sample, where the greatest difference is observed in events with both large |Δy| and $\overline{p_{\mathrm{T}}}$.

Bibliography

[1] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[2] Georges Aad et al. Charged particle multiplicities in pp interactions at $\sqrt{s}$ = 0.9 and 7 TeV in a diffractive limited phase-space measured with the ATLAS detector at the LHC and new PYTHIA6 tune. Technical Report ATLAS-CONF-2010-031, CERN, Geneva, Jul 2010.

[3] A. Sherstnev and R. S. Thorne. Parton Distributions for LO Generators. *Eur. Phys. J.*, C55:553–575, 2008.

[4] S. Gieseke et al. Herwig++ 2.5 Release Note. 2011.

[5] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[6] G. Corcella et al. HERWIG 6.5 release note. 2002.

[7] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637–646, 1996.

[8] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002.

[9] Rosa Maria Duran Delgado, Jeffrey R. Forshaw, Simone Marzani, and Michael H. Seymour. The dijet cross section with a jet veto. *JHEP*, 08:157, 2011.

[10] G. Aad et al. The ATLAS Simulation Infrastructure. *European Physical Journal C*, 70:823–874, December 2010.

[11] S. Agostinelli and et al. G4–a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[12] Rene Brun and Fons Rademakers. ROOT - An Object Orientated Data Analysis Framework. *Nucl. Inst. & Meth. in Phys. Res. A*, 386:81–86, 1997. See also http://root.cern.ch.

[13] David Cournapeau et al. Scikits.learn: machine learning in python. http://scikit-learn.sourceforge.net/stable/index.html, 2011.

Chapter 5

NLO Dijet Production

In this chapter the POWHEG NLO Dijet generator [1, 2, 3] will be described including a discussion of uncertainties related to the choice of PDF, renormalisation scale and factorisation scale. Generation issues encountered whilst producing the predictions and strategies to mitigate their effect are described. Finally the parameters used to generate the predictions are listed and a selection of predictions shown for inclusive distributions.

5.1 POWHEG

POWHEG-BOX provides a new framework to produce NLO events which subsequently can be passed through any of the standard Monte Carlo generators, such as PYTHIA or HERWIG. The dijet process [4] within POWHEG should provide NLO-accurate predictions of dijet production. By passing the POWHEG events through the standard MC generators the effects of parton shower, hadronisation and MPI are taken into account. This is significantly easier than previous methods that were applied to NLOJet++ [5]. With NLOJet++ it had been necessary to derive corrections for the soft physics effects by running Monte Carlo with underlying event and hadronisation switched on and off.

A number of parameters were chosen when generating the events in POWHEG. These include the PDF, scale, weighting and a number of other POWHEG specific parameters.

5.1.1 Parton Distribution Function Choice

Two PDF sets, MSTW2008NLO [6] and CT10 [7], were investigated for the production of the hard scatter with POWHEG-BOX. These NLO PDF sets, produced by seperate collaborations, use some of the most recent hadronic collider data and latest theoretical methods to provide knowledge of the content of interacting hadrons. The CT10 PDF uses more recent data sets including updated inclusive jet data from the Tevatron. In addition to changing the PDF, the full eigenvector error sets for each PDF set were investigated.

The determination of the uncertainties due to PDF choice was carried out though the technique known as PDF re-weighting. In this approach, the effect of using a different PDF is established by applying a weight, w, to each event. The weight is defined as

$$w = \frac{f^{new}(x_1, q, fl_1) f^{new}(x_2, q, fl_2)}{f^{old}(x_1, q, fl_1) f^{old}(x_2, q, fl_2)} \tag{5.1}$$

where f is the parton distribution function evaluated for the given momentum fraction $x_{1,2}$, energy scale q of the interaction and parton flavour $fl_{1,2}$. The "new" and "old" refer to the PDF set used, the old being the default central PDF set used in the initial generation. Re-weighting was necessary because the computational cost of generating POWHEG events for all 53 PDF eigenvectors in the CT10 PDF set and 41 PDF eigenvectors in the MSTW2008 NLO PDF would have been too great.

For each PDF set, the asymmetric uncertainties in an observable F were then calculated from the PDF eigenvectors through the use of the following formulas

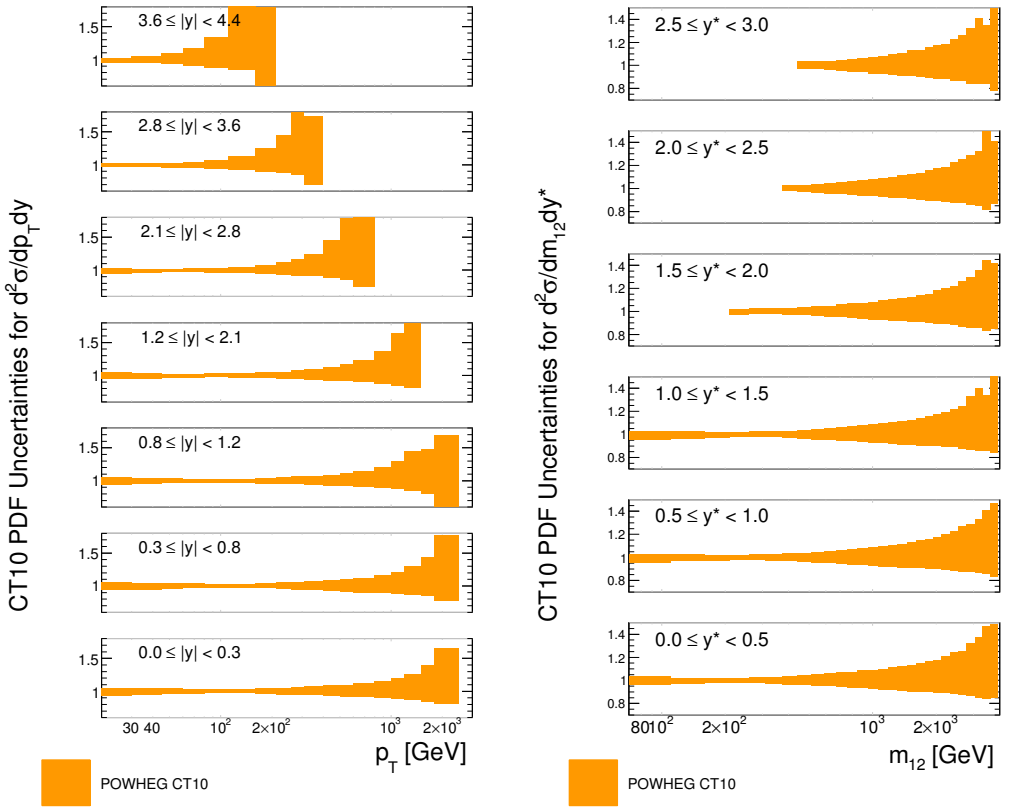
$$(\Delta F)_+ = \sqrt{\sum_{k=1}^{n} \{\max[F(S_k^+) - F(S_0), F(S_k^-) - F(S_0), 0]\}^2}, \tag{5.2}$$

$$(\Delta F)_- = \sqrt{\sum_{k=1}^{n} \{\max[F(S_0) - F(S_k^+), F(S_0) - F(S_k^-), 0]\}^2}, \tag{5.3}$$

where S_0 and $S_k^{\pm}$ are the central and k$-th$ error eigenvector PDF sets respectively. This is the defined way of doing this [6].

Inclusive Jet Observables

In figure 5.1 the CT10 PDF uncertainties for two doubly differential inclusive distributions are shown. CT10 PDF was the default choice for the POWHEG prediction in another ATLAS publication in progress. The uncertainty grows substantially for events containing jets with high transverse momentum and pairs of jets with high dijet mass. The uncertainty is highly asymmetric, where the positive uncertainty is over 50% in the highest p_{T} bins. The dominance of the gluon PDF in the high dijet mass region and high y^* is a source for the large uncertainty in figure 5.1.

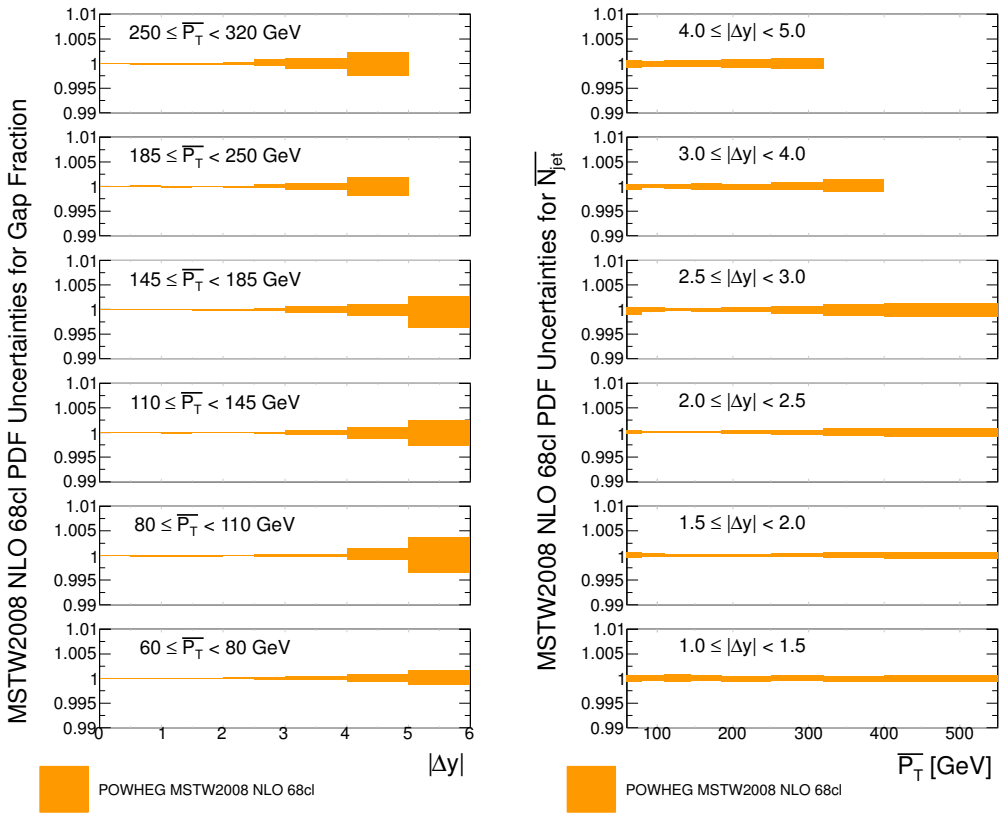


(a) (b)

Figure 5.1: The relative CT10 PDF uncertainties for inclusive single jet cross sections (a) and dijet mass cross sections (b). The POWHEG events are considered at the parton level using the anti-k_t $R = 0.6$ algorithm to make parton jets. The orange band corresponds to the results of equations 5.2-5.3.

Veto Jet Observables

The work carried out for the veto analysis happened before CT10 became the default PDF choice for generating POWHEG events, MSTW 2008 NLO was used instead. The size of the PDF uncertainty for the dijet with central veto jet measurement was found to be small. In figure 5.2 the typical uncertainty for the MSTW 2008 NLO PDF set is far below 1%. The size of the uncertainty does increase with |Δy| but is still considered to be negligible relative to other theoretical uncertainties and could possibly be explained by the poor statistics in these regions. The small uncertainty is not limited to the MSTW 2008 NLO PDF set, both CTEQ66 and NNPDF have also been observed to produce consistent predictions with the same low uncertainty originating from the PDF error sets.



(a) (b)

Figure 5.2: The relative PDF uncertainties for a selection of gap fractions vs |Δy| (a) and average veto jet number vs $\overline{p_T}$ (b) distributions. The POWHEG events are considered at the parton level using the anti-k_t $R = 0.6$ algorithm to make parton jets.The orange band corresponds to the results of equations 5.2-5.3.

5.1.2 Scale Choice

In POWHEG-BOX the dijet process by default uses the p_{T} of the underlying leading order process for the renormalisation and factorisation scales on an event by event basis. It has been suggested that this default choice of scale $\mu = p_T$ can lead to instability in a purely NLO prediction. An alternative renormalisation and factorisation scale choice suggested in [8] was made using

$$\mu = \frac{m_{jj}}{2\cosh(0.7y^*)}, \tag{5.4}$$

where m_{jj} is the mass and y^* is the separation in rapidity of the two partons that make up the underlying leading order system. This choice of scale can be considered to be an interpolation between the dijet mass and the p_{T} as scale choices. In addition to an alternative scale choice, scales were varied to estimate the effect of missing higher orders in the perturbative expansion. Scale variation required additional events where the factorization and renormalisation scales had an additional multiplicative factor applied. The renormalisation and factorization scales (μ_R, μ_F) were varied by factors of (2, 2), (0.5, 0.5), (2, 1), (1, 2), (0.5, 1.0) and (1.0, 0.5). The envelope of these variations was taken to be the total scale uncertainty.

Inclusive Jet Observables

A comparison of the uncertainties from scale variation are shown for the single jet cross sections and dijet mass cross sections in figure 5.3. The PDF uncertainties are shown as well for reference. For both choices of scale there was a similar result for the inclusive jet cross section, where there was an $\approx 10\%$ uncertainty for the entire p_{T} range in each $|y|$ region. The size of the uncertainty for scale variation and PDF are of similar magnitude at low p_{T}. At high p_{T} the PDF uncertainties have become the dominant source of uncertainty.

For the dijet mass distributions in regions of low y^* both scale choices produced similar uncertainties, however as the separation of the leading jets increased the uncertainties

diverged. In the highest separation region considered here, $2.5 < y^* < 3.0$, the scale uncertainty in the negative direction was greater than 50% for the p_{T} scale choice. The new scale choice did not see the same growth of uncertainty with increasing y^*, indicating it may be the preferable choice especially if the uncertainty for the default scale choice continues to grow and possibly becomes unstable.

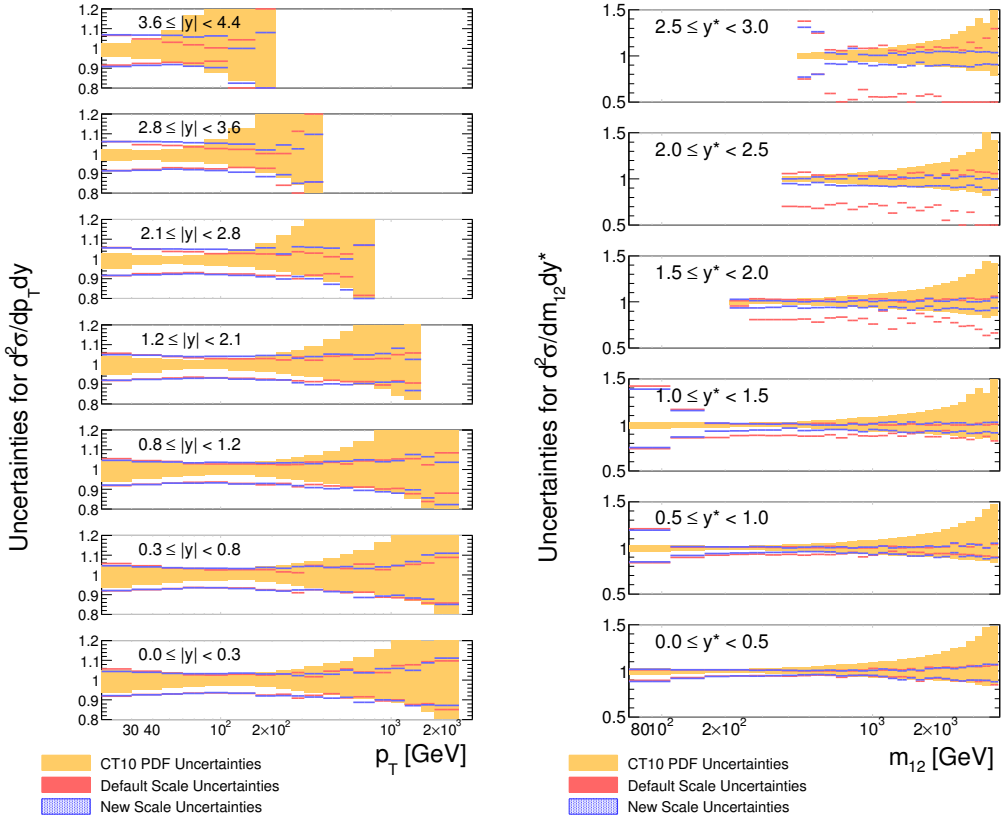


Figure 5.3: The relative PDF uncertainties and scale uncertainties for inclusive single jet cross sections (a) and dijet mass cross sections (b). The POWHEG events are considered at the parton level using the anti-k_t $R = 0.6$ algorithm to make parton jets. The extent of the PDF uncertainties are indicated by the yellow lines when the scale uncertainties are larger.

Veto Jet Observables

The uncertainty from varying the scale by a factor of 2 for the veto observables was found to be small, with the exception of the largest $|\Delta y|$ bin where low event numbers make the uncertainty dominated by the statistics. In figure 5.4 the uncertainty from scale varia-

tion was found to be much smaller than the difference between POWHEG+PYTHIA and POWHEG+HERWIG samples. The small size of the uncertainties from scale variation seemed to indicate that there was only a small effect from missing higher order terms in the perturbative expansion, however [9] found that a large theoretical uncertainty can be obtained through varying the arguments in the resummation of their calculation. Given that fact, it is the difference between the PYTHIA and HERWIG showers that demonstrate the missing higher order terms. It was therefore decided not to display uncertainties from scale variation on any of the subsequent predictions for the veto observables and instead rely upon PYTHIA and HERWIG differences, this strategy was adopted for the ATLAS publication [10].

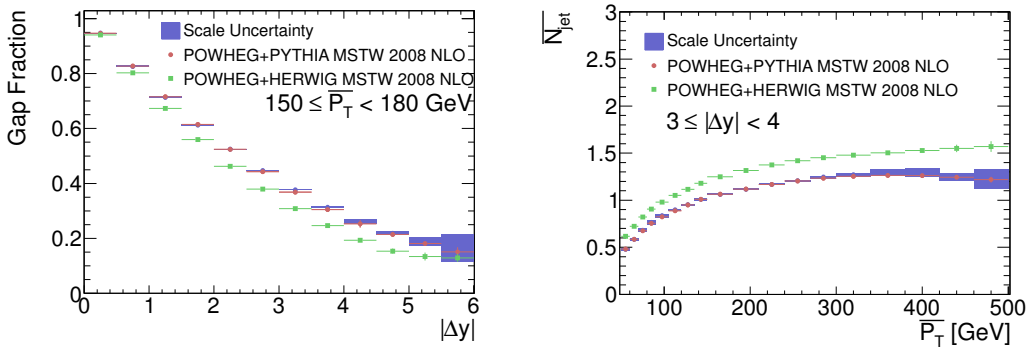


Figure 5.4: Comparison of POWHEG events showered with PYTHIA (AUET1) and HERWIG (AMBT1) and with the scale uncertainties shown for the POWHEG+PYTHIA prediction.

Figure 5.5 shows that the effect of the different scale choice was less than a few percent for a selection of veto observables; with more statistics this effect may reduce further. This does not rule out the possibility that higher |Δy| could be affected, especially if the fixed NLO prediction does actually produce a negative cross section.

5.1.3 Event Weighting

Events in the POWHEG-BOX dijet process were generated with weights. These weights were derived from properties at the underlying leading order level. The default function used to calculate the event weight is

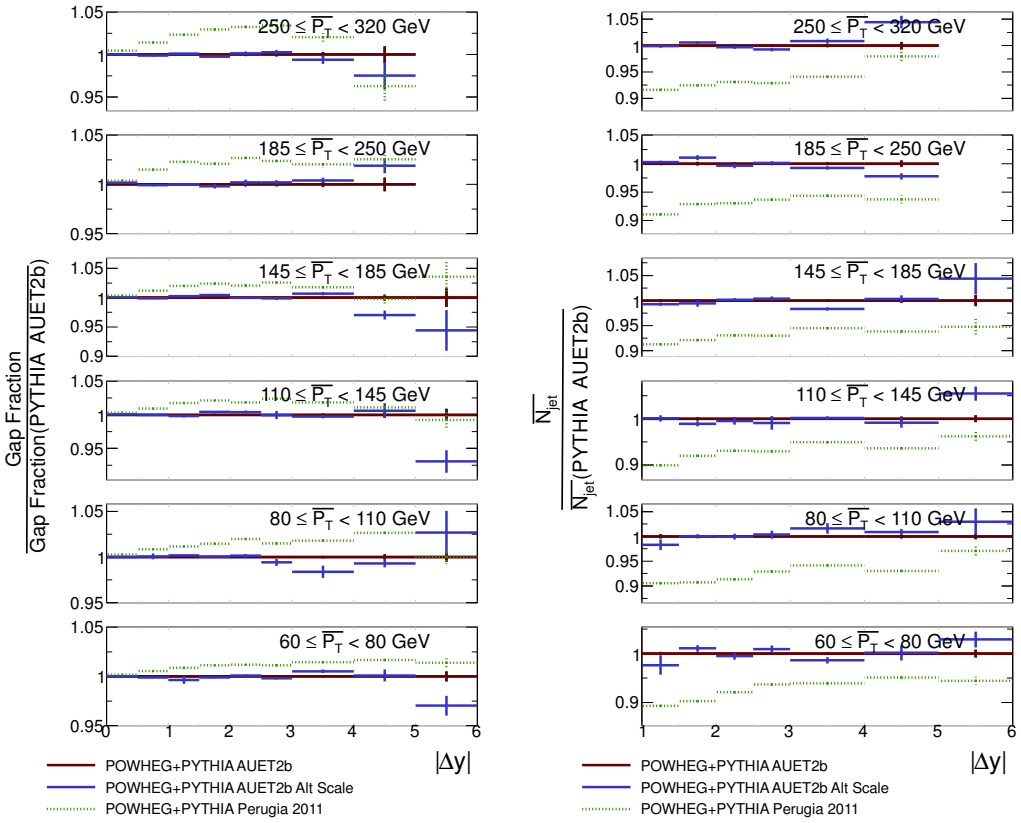


Figure 5.5: The ratio of different PYTHIA showered POWHEG predictions to the default PYTHIA choice. Included is a sample of POWHEG events generated and showered with the alternative scale choice in equation 5.4

$$f(p_T) = \left(\frac{p_T^2}{p_T^2 + p_{T\,supp}^2} \right)^3 , \tag{5.5}$$

where p_{T} is the underlying leading order transverse momentum and $p_{T\,supp}$ is a user defined parameter which controls the level of weighting. Events which have $p_T \gg p_{T\,supp}$ are kept with a weight $\rightarrow 1$, whilst for $p_{T\,supp} \gg p_T$ events will receive a very large weight. Samples have been generated where the outermost exponent has been changed from 3 to 2, reducing the suppression of the very lowest transverse momentum events. This was found to be useful for handling problems discussed in the next section.

The default p_{T} suppression is capable of producing events across a wide range of p_{T} values, however it suffers from the same deficiencies as the standard PYTHIA samples, where large $|\Delta y|$ events are unlikely to occur. It is therefore necessary to generate very large samples

of events, in the hundreds of millions, to cover the regions which this thesis investigates.

5.1.4 POWHEG Generation Issues

A problem observed when generating weighted POWHEG events was that on a small number of occasions the highest p_{T} partons generated were much lower in p_{T} than the hardest jets of the event. For the events produced in POWHEG to be accepted by a showering MC generator they must be put into the correct format, where the hardest radiation parton must be assigned a parent. Consider the possible angles which the emitted parton can be assigned in figure 5.6 depending on the choice of parent. The probability to be assigned the larger θ_1 angle is suppressed but not zero. The choice of the larger angle results in a larger energy scale available for the subsequent parton shower. The increased scale means that much higher p_{T} objects can be generated in the showering MC generator.

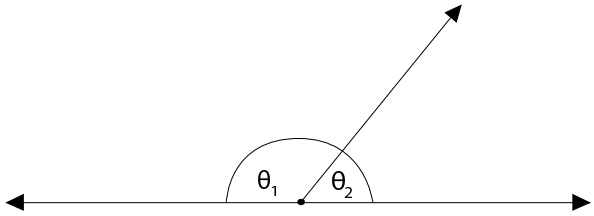


Figure 5.6: Assignment of the hardest radiation generated by POWHEG.

These events are not unphysical but they do pose a problem through the size of their weights. They have low parton p_{T} and so through equation 5.5 they will have a large event weight. These large weighted events with high p_{T} generated by the showering MC enter regions of observables where the typical weight is much lower. An example of a large fluctuation caused by one of these events can be seen in figure 5.7. Although generating more events would reduce the size of these fluctuations, the number needed is too large for the computing resources available. The first method to mitigate the effect of the events was to change the exponent in equation 5.5 from three to two. This reduced the size of weights in the very lowest p_{T} events. The second method involves the filtering of events using parton level quantities. Specifically, events were rejected if they satisfied

$$p_{T\ jet} > \mathrm{C} p_{T\ parton}, \tag{5.6}$$

where C is a free parameter, $p_{T\ jet}$ is the highest transverse momentum jet and $p_{T\ parton}$ is the highest transverse momentum parton. These events are a genuine component of the cross section and so their removal can cause a bias. Figure 5.7 shows an example of how the pathological events exhibit themselves in the veto observables. A large fluctuation was present in the gap fraction for the events without the cut applied. For each of the curves with the cut applied with different choices of threshold, the large fluctuation was removed without causing significant bias. The levels of bias were evaluated by dividing a generated sample into a set of subsamples and calculating the average value in each bin with and without the cut applied.

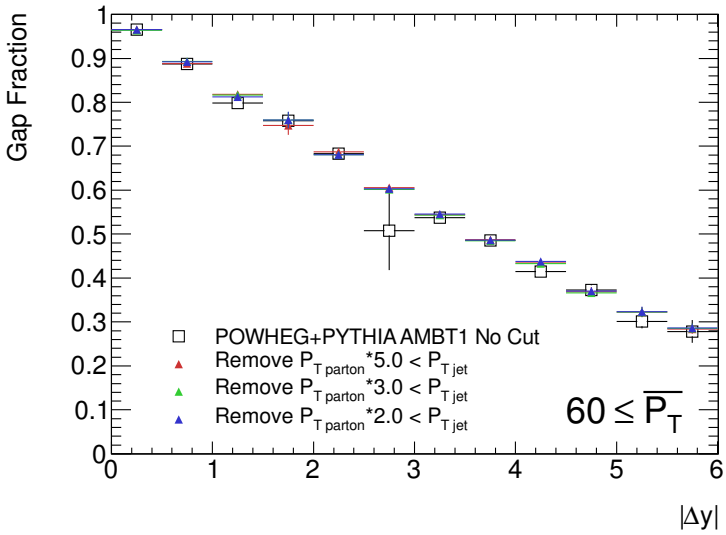


Figure 5.7: Comparison of different cuts to remove events with large interaction scales and low parton p_{T}.

For investigating dijet with a central jet veto the parameter C was chosen to be three; sufficient for the removal of all the large fluctuations seen.

5.2 Generation Parameter Summary

Here are the list of parameters considered for the generation of POWHEG dijet events.

$p_{T\,min} = 5\,\text{GeV}$	The minimum p_{T} of the hard scatter that POWHEG will generate. It is sufficiently low that there is no bias present for jets with $p_T > 20\,\text{GeV}$.
CT10 MSTW2008 NLO	The choices of PDF used in the generation of the hard scatter. Choice seen to have little effect on dijet with veto observables
$p_{T\,supp} = 100\,\text{GeV}$ $p_{T\,supp} = 250\,\text{GeV}$ $p_{T\,supp} = 400\,\text{GeV}$	The suppression factor as defined in equation 5.5.

Showering Monte Carlo Settings

A number of different tunes were used to shower the POWHEG events for this analysis. For PYTHIA the tunes which have been considered are AMBT1, AUET2b and Perugia 2011. For HERWIG+JIMMY AUET1 and AUET2 have been considered.

The authors of POWHEG found that it was necessary to limit the MPI scale to that of the jets in the primary interaction [4]. This is achieved in PYTHIA by setting the parameter MSTP(86)=1. The same requirement is applied for HERWIG+JIMMY, where events with the additional scatters with a p_{T} greater than the hard scatter p_{T} are rejected.

5.3 Inclusive Jet Predictions

In figure 5.8 the range of predictions for POWHEG are shown for the inclusive jet cross section. The ratio plots show the size of the systematic with both the scale variations and

the CT10 PDF errors added in quadrature. Both POWHEG PYTHIA distributions produce predictions within 10% of each other whilst the POWHEG HERWIG prediction starts to relatively increase as the rapidity of the jets becomes more forward. Additionally the POWHEG HERWIG predictions are more strongly peaked at low p_{T} in most regions of rapidity. This good agreement between POWHEG+PYTHIA tunes is also seen in the dijet mass spectrum shown in figure 5.9. The differences observed between the HERWIG and PYTHIA showered events could arise from the non-peturbative elements of the predictions, and therefore impose a large additional systematic uncertainty on the POWHEG predictions above and beyond the scale and PDF uncertainties. This is not unprecedented as in predictions made by NLOJet++ [5] there are also significant uncertainties arising from non-peturbative physics models. This interpretation of the differences observed is not quite complete as there are also significant differences observed in regions where the peturbative physics should be dominant, in this case it would appear that peturbative showers within PYTHIA and HERWIG are playing a more significant role than expected.

The agreement of POWHEG predictions of inclusive jet distributions with data has been seen to be good in [11]. However the complete range of distributions is still yet to be published for the 2010 ATLAS data, including the data for the systems with the greatest rapidity separations where the most significant differences between HERWIG and PYTHIA showered events occurred. A future publication by the ATLAS collaboration will show the POWHEG distributions described in this chapter compared to the improved measurement.

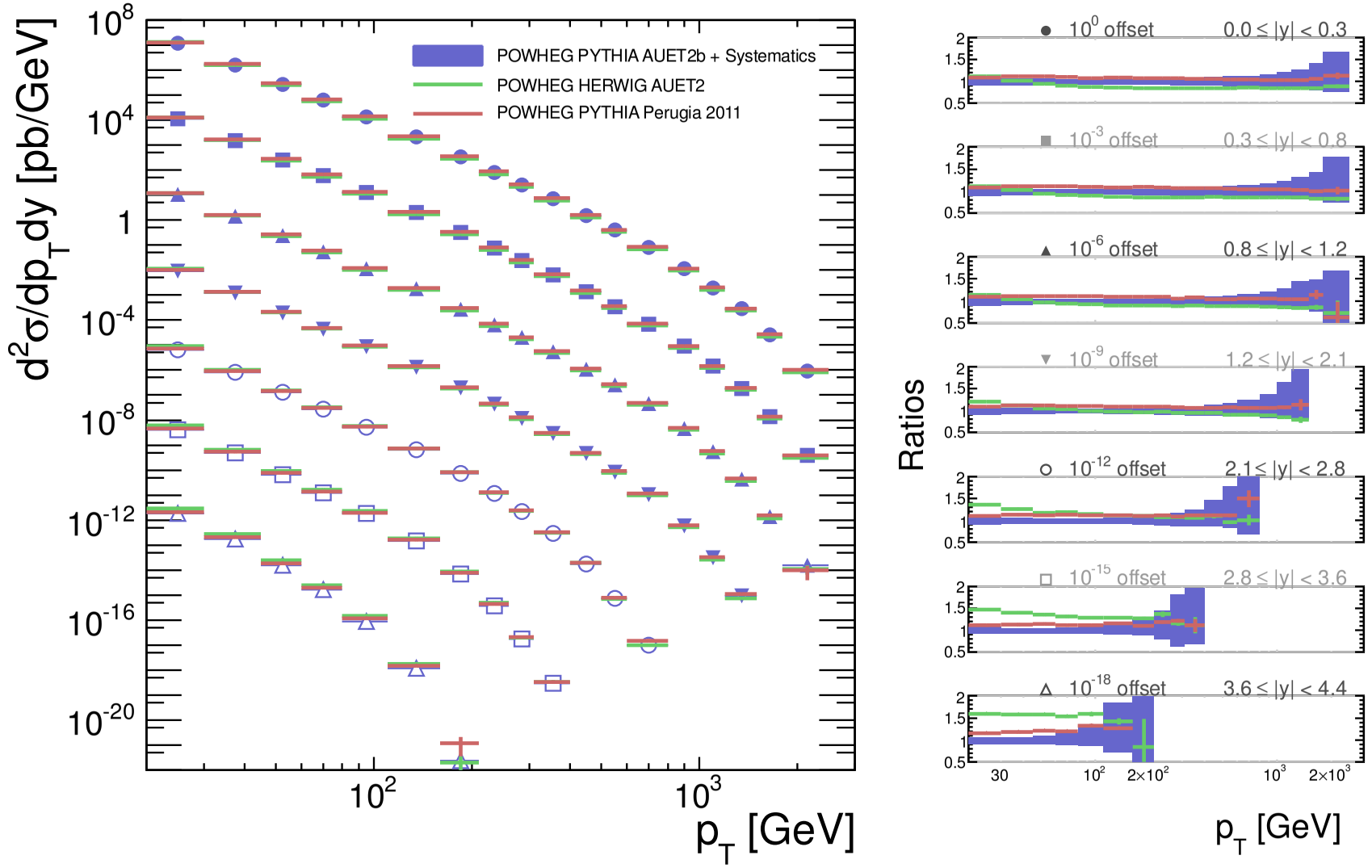


Figure 5.8: The inclusive jet cross section for a number of different rapidity slices. The jet algorithm used is anti-k_t $R = 0.6$. The prediction shown is the POWHEG+PYTHIA with the AUET2b tune.

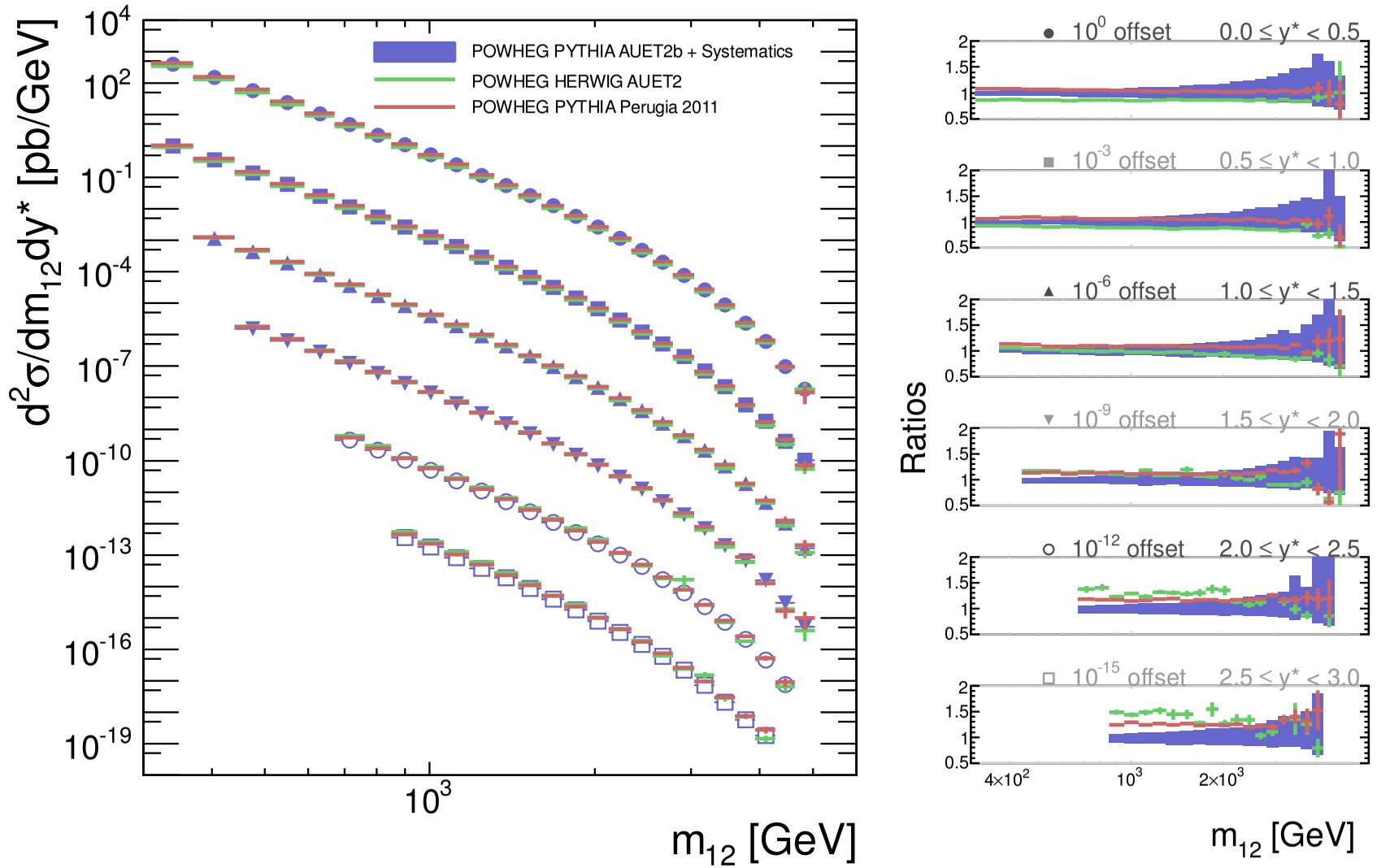


Figure 5.9: The dijet mass jet cross section for a number of different rapidity slices. The jet algorithm used is anti-k_t $R = 0.6$. The prediction shown is the POWHEG+PYTHIA with the AUET2b tune.

5.4 Dijet Veto Predictions

Detailed discussion and comparison of the POWHEG predictions to data for the dijet veto analysis are given in chapters 6 and 7

Bibliography

[1] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[2] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[3] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[4] Simone Alioli, Keith Hamilton, Paolo Nason, Carlo Oleari, and Emanuele Re. Jet pair production in POWHEG. *JHEP*, 04:081, 2011.

[5] Zoltan Nagy. Next-to-leading order calculation of three jet observables in hadron hadron collision. *Phys. Rev.*, D68:094002, 2003.

[6] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *European Physical Journal C*, 63:189–285, September 2009.

[7] M. Guzzi, P. Nadolsky, E. Berger, H.-L. Lai, F. Olness, and C. Yuan. CT10 parton distributions and other developments in the global QCD analysis. *ArXiv e-prints*, January 2011.

[8] Stephen D. Ellis, Zoltan Kunszt, and Davison E. Soper. Two-jet production in hadron collisions at order α^3 in qcd. *Phys. Rev. Lett.*, 69:1496–1499, Sep 1992.

[9] Rosa Maria Duran Delgado, Jeffrey R. Forshaw, Simone Marzani, and Michael H. Seymour. The dijet cross section with a jet veto. *JHEP*, 08:157, 2011.

[10] Georges Aad et al. Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector. *Journal of High Energy Physics*, 2011:1–36, 2011. 10.1007/JHEP09(2011)053.

[11] Measurement of inclusive jet and dijet cross sections in proton-proton collision data at 7 TeV centre-of-mass energy using the ATLAS detector. Technical Report ATLAS-CONF-2011-047, CERN, Geneva, Mar 2011.

Chapter 6

Dijet Veto Measurement

In this chapter the remaining elements of the dijet with a central jet veto analysis are discussed. An outline is given to show how events were selected from data in the 2010 period and what quality requirements were made.

With great detail this chapter also examines the determination of the data corrections that were critical for this analysis. The correction that has been derived translated the measured observable quantities from the detector level to the hadron level definition. The utility of the PYTHIA Monte Carlo samples in determining the corrections was examined by comparing the full detector simulation events with uncorrected data, and were found to be in sufficient agreement. The requirements for the use of the detector correction technique were then shown to be satisfied through the investigation of event migration in bins of the relevant kinematic variables. A set of conservative systematic uncertainties were determined through the consideration of both theoretical modeling uncertainties and measurement limitations of the ATLAS detector. The complete set of corrections are presented for the main analysis selection criteria.

The remaining important sources of systematic uncertainty for the measurement are identified. The chapter ends with a summary of the results which were published in the ATLAS dijet with a central jet veto paper [1].

6.1 Analysis Selection

The default kinematic selection as defined in chapter 2 is used for selections A, B and C. In all cases only anti-k_t jets with a radius parameter $R = 0.6$ were considered.

6.1.1 Data Selection

Good Data Selection

This analysis was dependent upon many detector subsystems of ATLAS and the operation of accelerator. It was therefore necessary that a rigorous selection was applied to the complete data recorded in the 2010 period. ATLAS records information pertaining to the quality of the data on a luminosity block basis, where a luminosity block is a period of data taking of $\approx$ 1 minute. The data quality stored in each luminosity block is derived from comparisons to benchmark distributions for each sub-detector and when agreement was seen, the sub-detector was considered to be operating in a good or "green" state. In 2010 all sub-detectors were flagged as green for over 90% of the data [2]. In this analysis the calorimeters, tracking system, magnet system and triggers were all required to be operating at the green level. In addition to the good operation of the ATLAS detector, further requirements were placed on the accelerator operation. The LHC was required to be operating in stable beam conditions and colliding protons at $\sqrt{s} = 7$TeV. After making all these quality requirements the amount of integrated luminosity fell from the $45\,\text{pb}^{-1}$ delivered to the ATLAS detector [3] to $35.6\,\text{pb}^{-1}$ available for the analysis.

Jet Quality

Data quality requirements were also made for the jet objects by requiring that there were no loose-bad or ugly jets with $p_T > 20\,\text{GeV}$ in each event [4]. The jets defined as bad have been found to be fake or not associated with in-time energy deposition; they may for example originate from cosmic rays or noise in the calorimeter. Ugly jets were associated

with in-time energy deposition but have occurred in a region where the measurement was not considered reliable, such as in transition regions between detectors. The jet requirements have been observed to have a high efficiency in rejecting bad events with only a sub-percent effect on all the observables [1]. These same jet quality requirements were tested in the Monte Carlo samples which had the full detector simulation applied, and were seen to have only a small effect. Figure 6.1 shows in a selection of observables that the effect of the jet quality requirement was typically less than 1%, comparable to the effect observed in data. Although these requirements were not applied to the final MC results because of an incomplete modeling of all the jet quality variables in the simulation at the time.

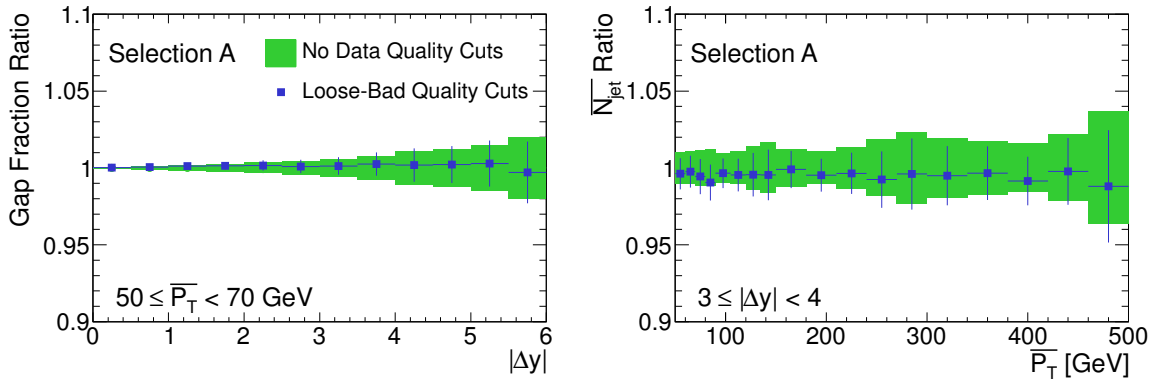


Figure 6.1: Comparison of simulated PYTHIA Monte Carlo events at the reconstruction level with and without the application of jet quality requirements. Events use the selection A criteria.

Trigger Requirements

The dijet events which were investigated in this analysis had $\overline{p_{\mathrm{T}}}$ in the range 50 GeV to 500 GeV. The production rate for the low $\overline{p_{\mathrm{T}}}$ events was very high requiring ATLAS to reject a large fraction of these events using highly prescaled jet triggers. ATLAS used a series of single jet triggers with different energy thresholds to collect jet events, where lower energy thresholds required large prescales to cope with the high event rate by the end of 2010. The trigger strategy that was used in this analysis was the division method [5], where separate trigger items are used to fill different regions of phase space. In this case the $\overline{p_{\mathrm{T}}}$ distribution

was divided into seven regions in which different single jet triggers were used to select events. These regions were selected by finding the minimal jet $\overline{p_{\mathrm{T}}}$ where the jet trigger is more than 99% efficient. The choice of threshold for each division is listed in table 6.1. More details about the trigger strategy can be found in chapter 7.

$\overline{p_{\mathrm{T}}}$ range [GeV]	Recorded Luminosity [pb^{-1}]
50-70	0.0246
70-90	0.110
90-120	0.311
120-150	1.57
150-180	8.33
180-210	8.29
> 210	35.6

Table 6.1: List of average transverse momentum ranges for each trigger division and the luminosity available after trigger prescaling.

Table 6.1 shows the recorded luminosities for each $\overline{p_{\mathrm{T}}}$ division for the 2010 running. The effect of trigger prescales, deadtime and data quality is taken into account in figure 6.2, where the large trigger prescales are apparent at low $\overline{p_{\mathrm{T}}}$, reducing the yield of events substantially.

Vertex Selection

Events are only kept if they have exactly one good vertex. The definition of a good vertex was one that had $N_{tracks} > 4$ associated with it, where each track had $p_{T\,track} > 100\,\mathrm{MeV}$. This requirement was made to remove the effect of in-time pile up and to allow jet reconstruction access to the position of the interaction point as part of the standard jet energy correction in ATLAS. In figure 6.2 the effect of this vertex requirement was quite clearly seen to reduce the sample size by a significant fraction. In the low $\overline{p_{\mathrm{T}}}$ region the fraction of data lost to the cut was approximately 35% and rose to 70% in the highest $\overline{p_{\mathrm{T}}}$ regions.

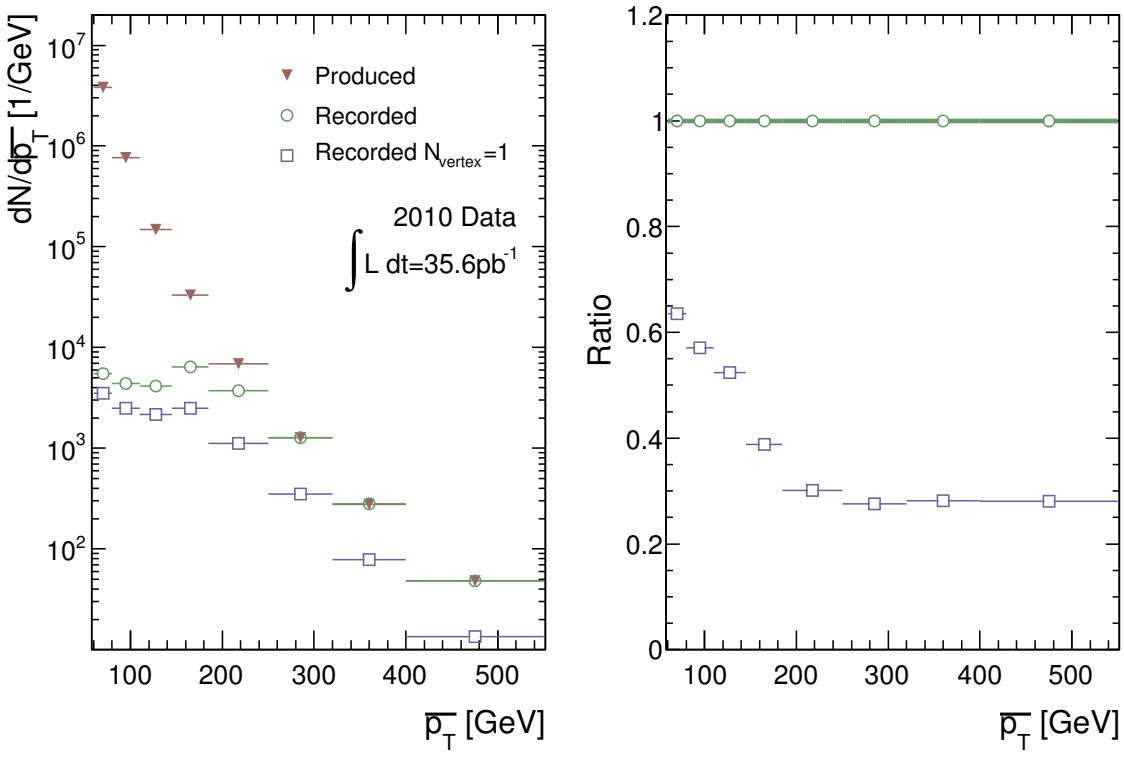


Figure 6.2: The yield of events that was produced by the LHC, collected by ATLAS and passed the data quality requirements. The recorded distributions include the effect of both the trigger selection and the vertex requirement. The ratio plot is given with respect to the recorded event number.

Example event

An example of an event that satisfied the event selection criteria is shown in figure 6.3 using the Atlantis event display system [6]. The leading jets of this event have $\overline{p_T} = 510\,\text{GeV}$ and a separation of over four units of rapidity. The presence of an additional jet between the leading jets with $p_T = 31\,\text{GeV}$ caused the event to fail the veto requirement imposed for both selection A and B, however it did pass selection C.

6.2 Data Correction

Using the correction outlined in the following sections of this chapter the effects of the detector were removed from the data. By correcting the data to the hadron level it may be

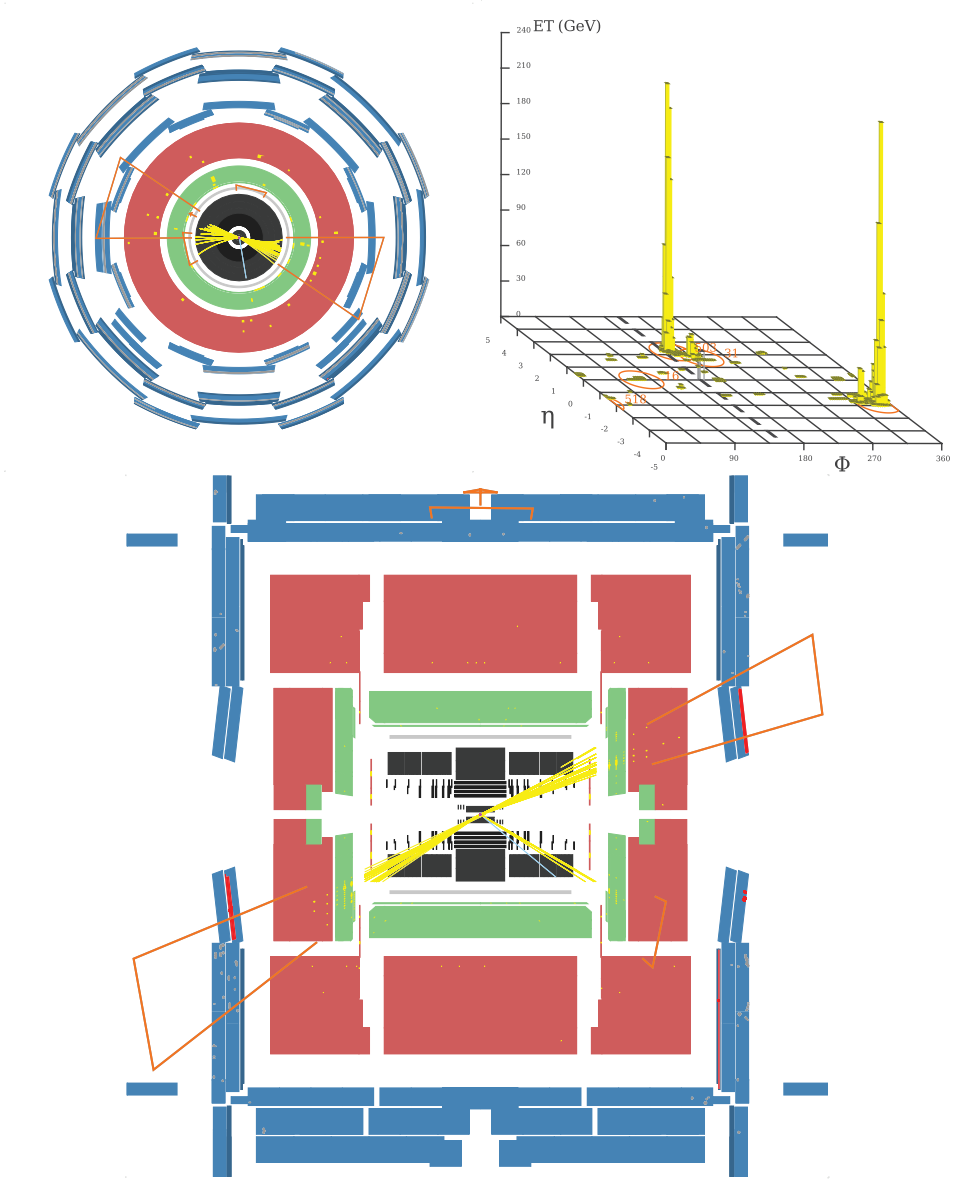


Figure 6.3: An event display for event number 9435121 from run 167607 collected by ATLAS. The top left diagram shows the ATLAS detector in the x-y plane and the bottom figure shows the ATLAS detector in the y-z plane. The top right diagram shows the energy measured in the calorimeter system in the $\eta - \phi$ plane.

compared to many of the standard Monte Carlo dijet predictions and indeed analytical calculations without the need for the full detector simulation. It additionally allows comparisons to measurements made by other experiments. The correction was applied by a simple multiplicative factor, the correction factor, on a bin-by-bin basis for each observable considered and was derived by the comparison of Monte Carlo samples at the truth level and after de-

tector simulation. The correction factor for the i-th bin is defined as

$$C_i = \frac{O_i^{truth}}{O_i^{simul}}, \tag{6.1}$$

where O represents the observable to be corrected, O_i^{simul} is the observable after full ATLAS detector simulation and reconstruction and O_i^{truth} is the observable derived from the truth record. When the reconstructed data is multiplied by this correction factor it should ideally produce an estimate of the real life physics distributions free from detector effects. The requirements for the use of bin-by-bin corrections such as these are that the measurement values must be very close to the true values and that the bin size should be much larger than the resolution to minimize the migrations between bins [7].

Monte Carlo Samples

Any correction technique can suffer from large statistical fluctuations if the MC samples have poor statistics. The result of the poor statistics is that the systematic uncertainty for the data correction would be larger, significantly reducing the precision of the measurement. This problem has been reduced through the generation of large |Δy| weighted PYTHIA event samples, which was discussed in chapter 4. In the derivation of both the correction factors and their systematic uncertainties, the unweighted and |Δy| weighted PYTHIA 6 AMBT1 dijet event samples were combined. The combination was performed using

$$f(x) = \frac{w_1(x)f_1(x) + w_2(x)f_2(x)}{w_1(x) + w_2(x)}, \tag{6.2}$$

where $f_i(x)$ is the observable value in the x-th bin for the i-th PYTHIA sample and $w_i(x)$ is the number of events in the x-th bin for the i-th PYTHIA sample. The independently generated samples combined in this case are the unweighted PYTHIA 6 and |Δy| weighted PYTHIA 6 samples. By combining these samples the number of events available in the large |Δy| region was greatly increased without loss of precision in the small |Δy| region. Comparisons of the event yields in each sample can be seen in chapter 4 in figure 4.7. The

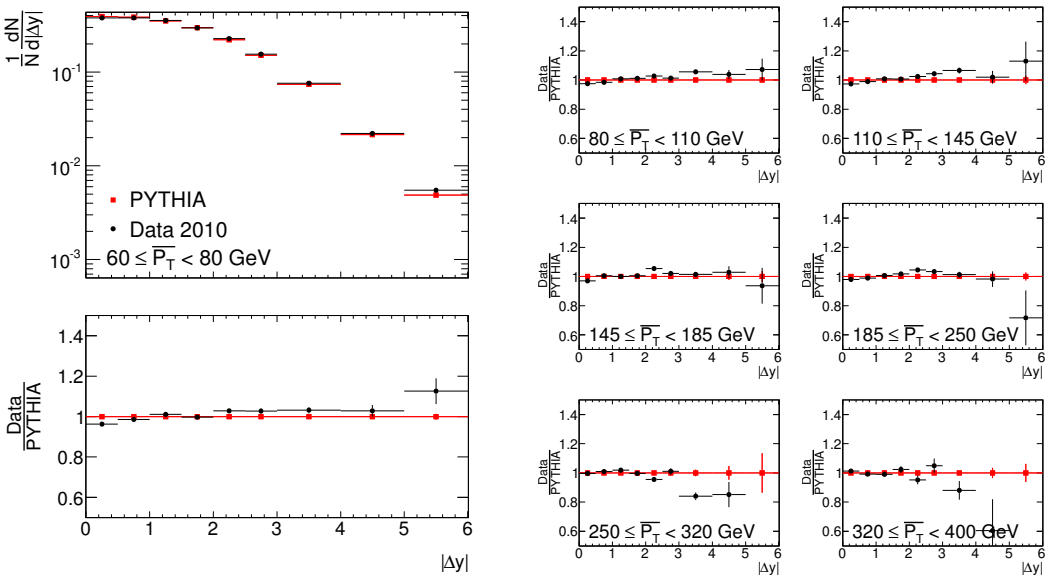
p_T and |Δy| weighted samples had not yet been made when this work was carried out.

6.3 Uncorrected Data Comparisons

The utility of a simulated event sample to derive the correction factors is strongly dependent on the agreement between the MC model and data. A simple test of this involved the comparison of the MC events after simulation with the reconstructed data distributions. As stated earlier in the chapter the events were required to be taken during adequate detector operation with the quality of jets classified as green. The control plots that demonstrate data/MC applicability are

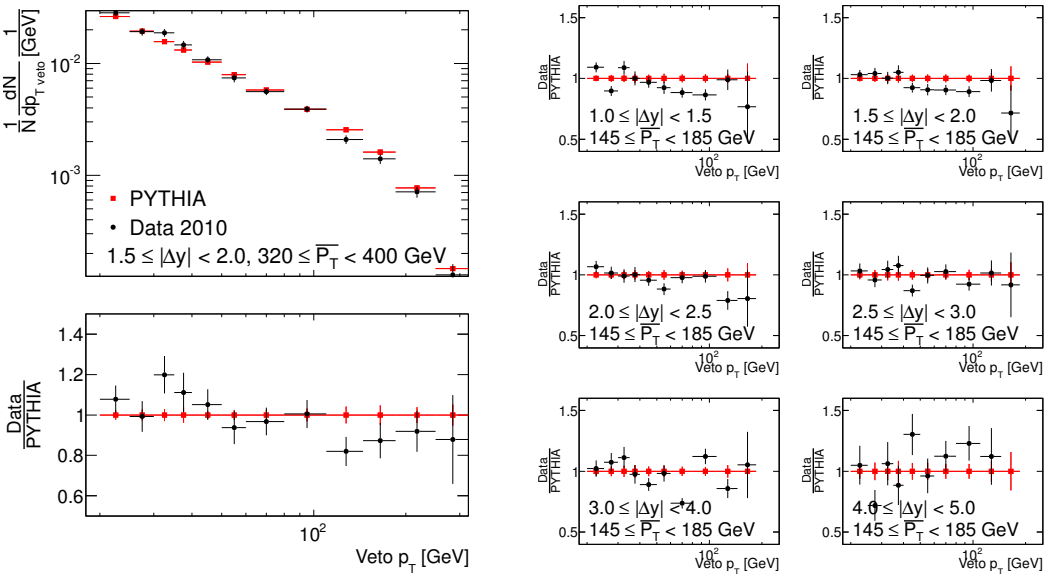
$$\frac{1}{N_{event}}\frac{\mathrm{d}N}{\mathrm{d}|\Delta y|}, \quad \frac{1}{N_{event}}\frac{\mathrm{d}N}{\mathrm{d}p_{T\,veto}}, \quad \frac{1}{N_{event}}\frac{\mathrm{d}N}{\mathrm{d}N_{veto}} \text{ and } \frac{1}{N_{event}}\frac{\mathrm{d}N}{\mathrm{d}|\Delta\phi|} \tag{6.3}$$

where N_{event} is the total number of events in each selection region, |Δy| is the interval between the two highest p_{T} jets, $p_{T\,veto}$ is the transverse momentum of the highest p_{T} jet inside the interval region, N_{veto} is the number of jets meeting the veto requirement between the interval jets and lastly $\Delta\phi$ is the difference in azimuthal angle between the interval defining jets. The following figures 6.4-6.7 show comparisons of uncorrected 2010 data with the combination of unweighted and |Δy| weighted PYTHIA 6 AMBT1 dijet events after detector simulation. The errors shown are the statistical uncertainties only. The theoretical uncertainties for PYTHIA would be expected to be relatively large. There was reasonable agreement seen in the shape for each of the distributions and in each of the subdivisions of the kinematic regions shown. This initially gave confidence that the corrections for detector effects can be calculated for the observables in this analysis using these MC samples.



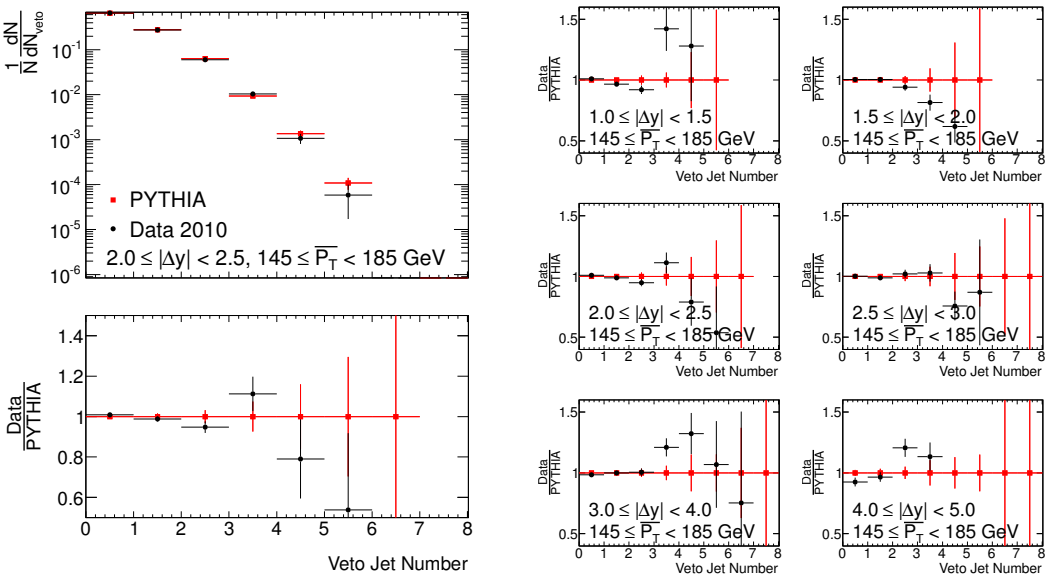
(a) (b)

Figure 6.4: Comparison of the uncorrected data distribution and the PYTHIA 6 AMBT1 event samples with both distributions normalized to unit area. $|\Delta y|$ is defined to be the rapidity interval between the two highest p_T jets with average transverse momentum $\overline{p_T}$. Only the statistical uncertainties are shown. (b) shows the ratio plots for further $\overline{p_T}$ slices.



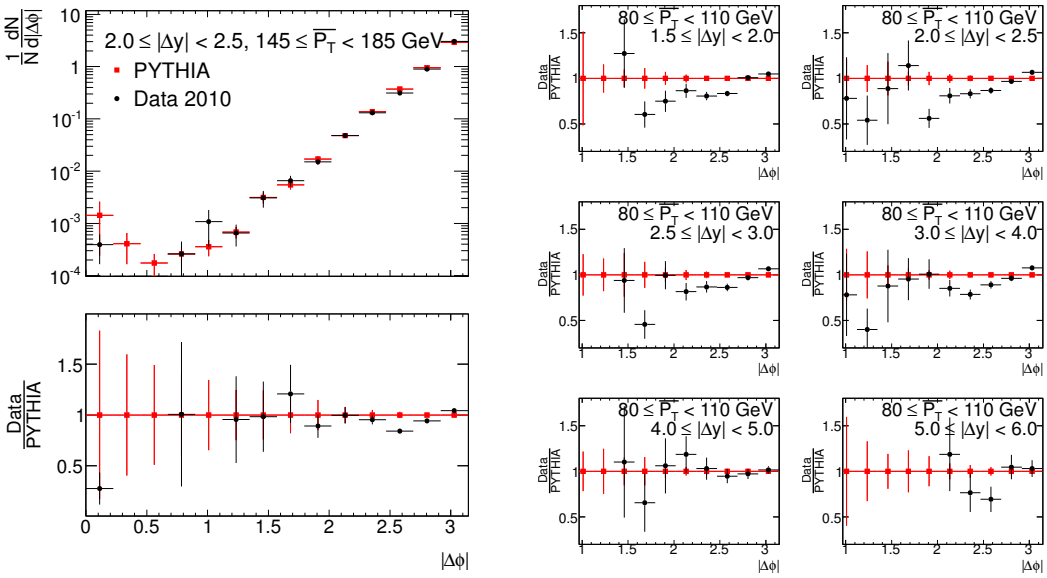
(a) (b)

Figure 6.5: Comparison of the uncorrected data distribution and the PYTHIA 6 AMBT1 event samples with both distributions normalized to unit area. The veto p_T is defined to be the highest p_T between the two highest p_T jets. Only the statistical uncertainties are plotted. (b) shows the ratio plots for further slices.



(a) (b)

Figure 6.6: Comparison of the uncorrected data distribution and the PYTHIA 6 AMBT1 event samples with both distributions normalized to unit area. N_{veto} is defined to be the number of jets above the veto scale between the highest p_{T} jets. Only the statistical uncertainties are plotted. (b) shows the ratio plots for further slices.

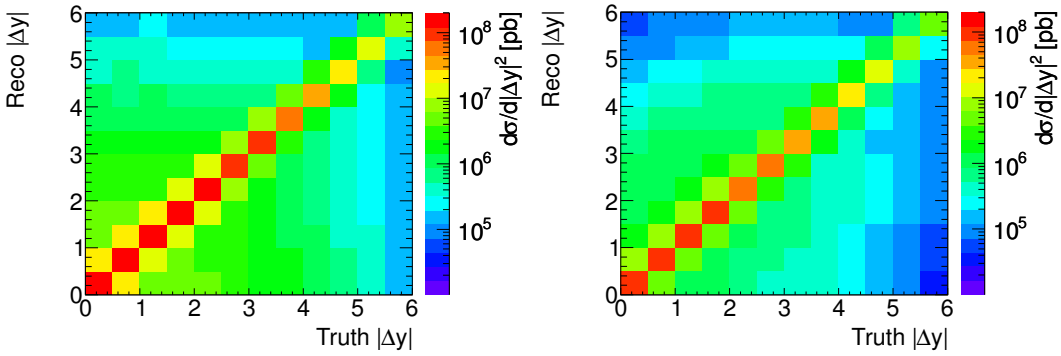


(a) (b)

Figure 6.7: Comparison of the uncorrected data distribution and the PYTHIA 6 event samples with both distributions normalized to unit area. $|\Delta\phi|$ is defined to be azimuthal difference between the two highest p_{T} jets. Only the statistical uncertainties are plotted. (b) shows the ratio plots for further slices.

6.4 Migrations

The bin-by-bin correction technique assumes that there is a low level of migration of events between bins. It was therefore important to confirm that the level of migration seen between the Truth MC and simulated MC was sufficiently small for the method to be applicable [8]. In figure 6.8 the |Δy| migration matrices for both Selection A and B, where $\overline{p_T} > 50\,\text{GeV}$, show a strong correlation between the truth |Δy| and the detector simulated value. The small fraction of events which migrate over a large range of rapidity separation values was caused by misidentifying the interval defining jets. A more quantitative picture of the migration of events is given by the purity variable defined in the next section.



(a) Selection A (b) Selection B

Figure 6.8: The differential cross section for PYTHIA 6 events generated with the AMBT1 tune where the x and y axis are the truth and reconstructed |Δy| variables respectively. There was a single requirement on the dijet system that $\overline{p_T} > 50\,\text{GeV}$. The two selection requirements for the dijet system are shown (a) The highest p_T jets (b) The most forward and backward jets.

Purity

The purity of a bin is defined by

$$P^i = \frac{N^i_{\text{truth+simu}}}{N^i_{\text{simu}}}, \tag{6.4}$$

where $N^i_{\text{truth+simu}}$ is the number of events where both the truth and simulated quantities of an event are in bin i, and N^i_{simu} is the total number of MC simulated events that lie in bin i.

The value of the purity will be high if the migration of events between bins is rare and the efficiency is high.

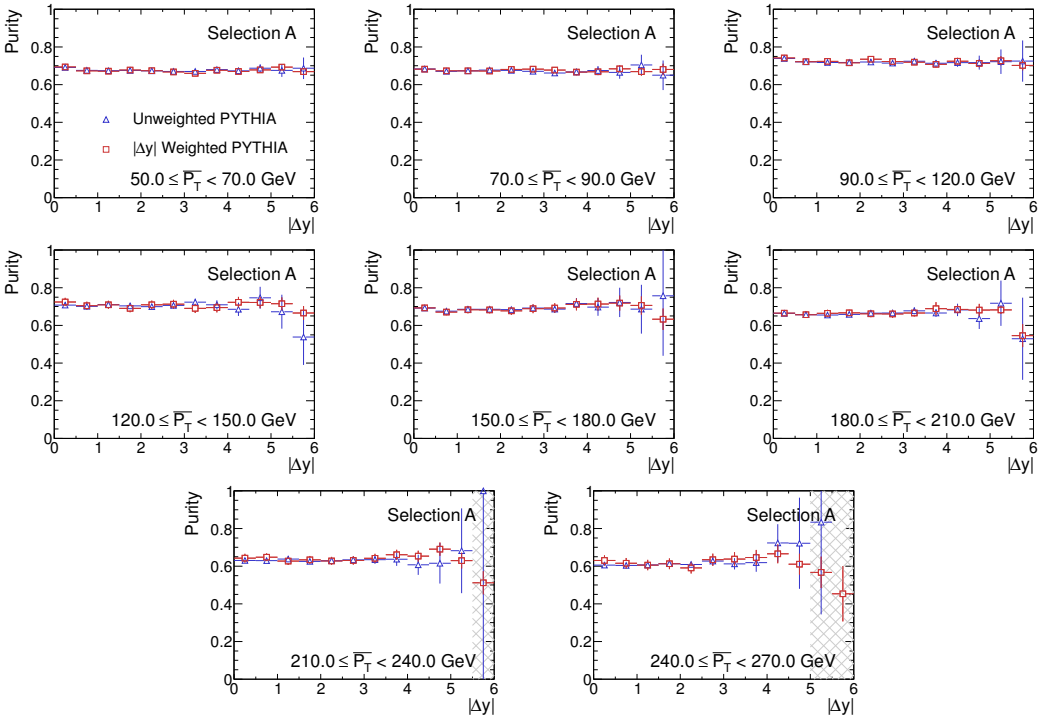


Figure 6.9: Selection A purities for the variable |Δy| where the MC events used are PYTHIA 6 with the AMBT1 tune. The regions of the graphs with a hatched background are not needed to unfold data because these regions lack events in the data.

The figures 6.9-6.11 show the selection A purities in bins of |Δy|, $\overline{p_{\mathrm{T}}}$ and Q_0 in all regions of phase space considered. These variables show a purity greater than 50% in all the bins that were needed for data correction in subsequent chapters. It should be noted that the bins in each figure are essentially two dimensional, so the purity of $> 50\%$ is actually reasonably high. The structure visible in figure 6.10 is due to the way the bin size varies with $\overline{p_{\mathrm{T}}}$. An increase in the bin size leads to an increase in purity. The purity plots for the other selections can be found in appendix section C.1, with similar results compared to selection A.

All the distributions show that the weighted and unweighted PYTHIA 6 samples were in agreement and importantly the |Δy| weighted sample provided lower statistical uncertainties in the large |Δy| region.

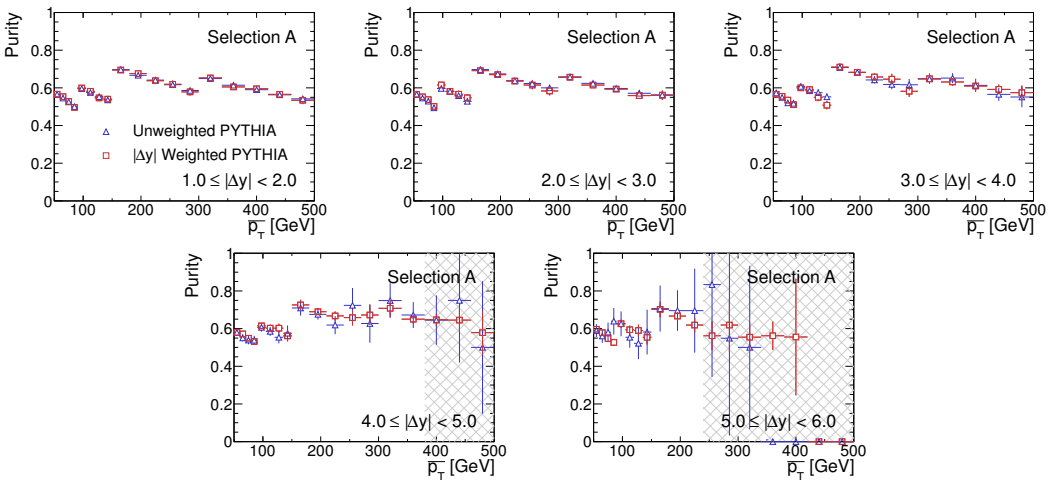


Figure 6.10: Selection A purities for the variable $\overline{p_T}$ where the MC events used are PYTHIA 6 with the AMBT1 tune. The regions of the graphs with a hatched background are not needed to unfold data because these regions lack events in the data.

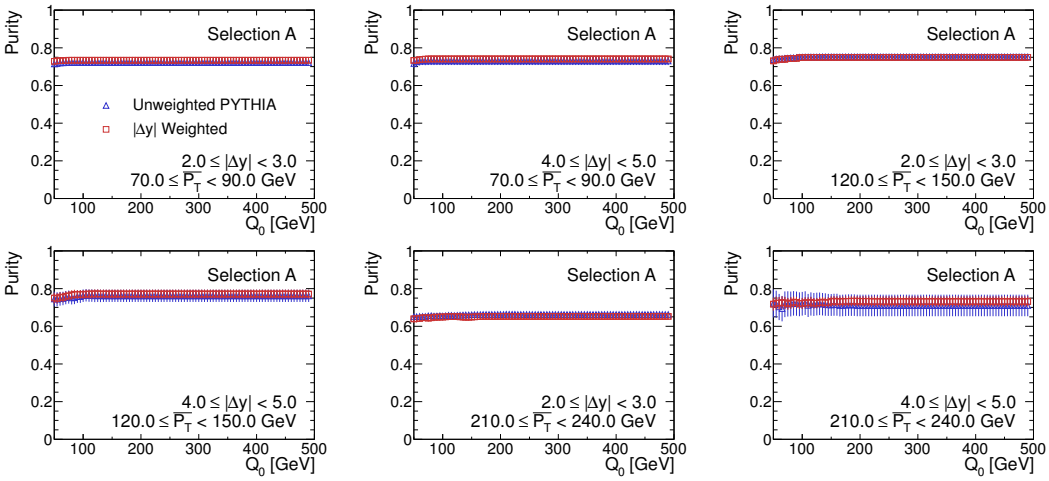


Figure 6.11: Selection A purities for the variable Q_0 where the MC events used are PYTHIA 6 with the AMBT1 tune.

6.5 Systematic Uncertainty Determination

There were a number of possible sources of systematic uncertainty for the data corrections derived in this chapter. These include the uncertainty in the physics modeling, increased jet energy resolutions, uncertainty of the jet efficiency and the uncertainty in the position of the primary vertex. The final total systematic uncertainties for the detector corrections are a quadratic sum of these effects as well as a statistical uncertainty on the MC samples.

6.5.1 Physics Modeling Uncertainty

Modifications applied

The sensitivity of the detector corrections to variations in the physics modeling in MC was tested. The distributions considered were the same as those which were being corrected, so the distributions of $\overline{p_{\mathrm{T}}}$, $|\Delta y|$ and Q_0. The modification of the distributions was achieved by re-weighting the events such that the shape varied within constraints derived using the JES uncertainty. This represents the maximal variation allowed by data as the data and MC agree within the JES uncertainty. The functions used to reweight the events and alter the shape were: -

$|\Delta y|$ **Shape** Event weight multiplied by two different functional forms

$\alpha|\Delta y| + 1$

$\alpha(10 - |\Delta y|) + 1$

where $\alpha \in [-0.05, 0.05]$.

$\overline{p_{\mathrm{T}}}$ **Shape** Event weight multiplied by two different functional forms

$\alpha\overline{p_{\mathrm{T}}} + 1$

$\alpha(500\,\mathrm{GeV} - \overline{p_{\mathrm{T}}}) + 1$

where $\alpha \in [-10^{-3}, 10^{-3}]$.

Q_0 **Shape** Event weight multiplied by two different functional forms

$\alpha Q_0 + 1$

$\alpha(150\,\text{GeV} - Q_0) + 1$

where $\alpha \in [-3 \times 10^{-3}, 3 \times 10^{-3}]$.

The JES uncertainties used to determine the range of modification applied were from an older more conservative estimation of the JES uncertainty. This has resulted in the modifications applied here to also be conservative.

Correction Uncertainties

The uncertainty in the correction factor originating from the modification of the physics modeling is defined to be the fractional difference with respect to the default correction factor

$$\Delta(F) = \frac{F_{\text{new}} - F_{\text{default}}}{F_{\text{default}}}. \tag{6.5}$$

where F is the observable being considered. The correction factor uncertainties after the modifications made to each of the truth distributions are shown in figure 6.12-6.16 for selection A. The equivalent results for the selections B and C can be found in appendix C, it should be noted selection C can be seen to have significantly higher uncertainties than the other two selections.

In selection A the largest contribution to the uncertainty from the physics modeling considerations originated from variation in the third jet p_{T} spectrum. This was seen to be usually less than 1% in the gap fraction distributions with small increases with larger $|\Delta y|$. The physics modifications were more important for the average number of jets above the veto threshold, where at low $|\Delta y|$ the uncertainty is greater than 5%. The other physics modelling sources were almost negligble in all cases. For selection B the uncertainties are of similar magnitude whilst for selection C the uncertainties are large, reaching 10% for the average number of veto jets.

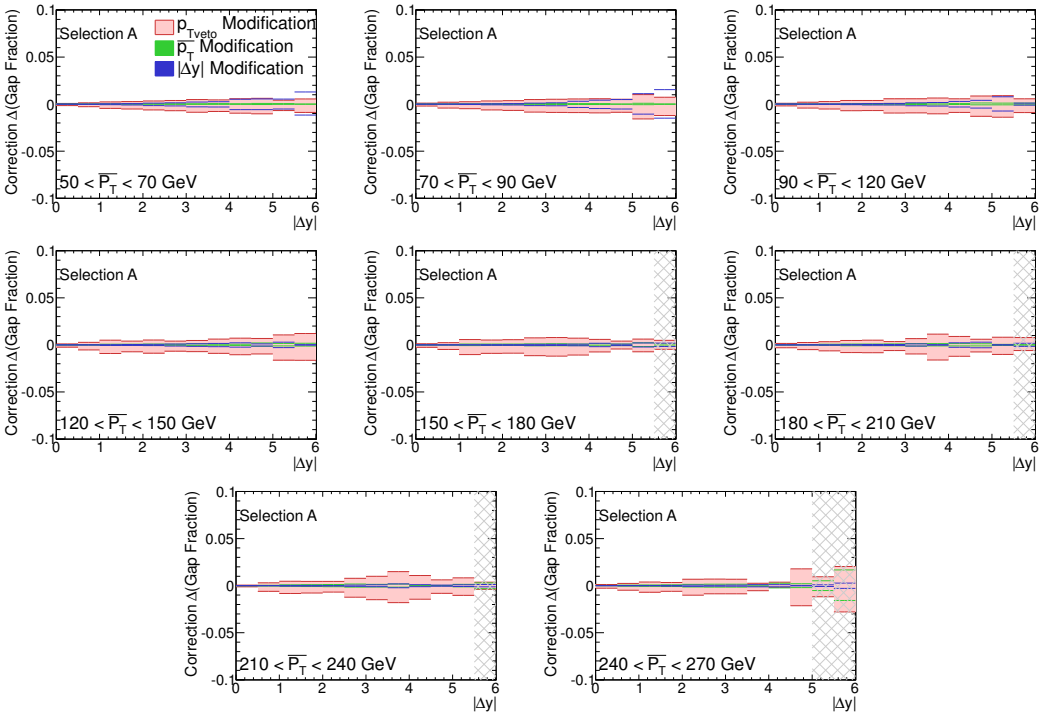


Figure 6.12: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection A is used.

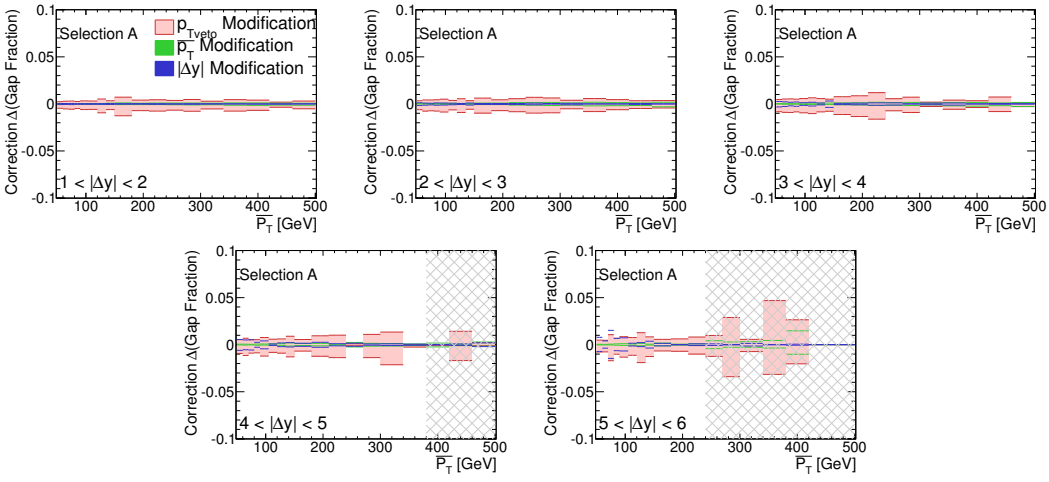


Figure 6.13: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_T}$ distribution. Selection A is used.

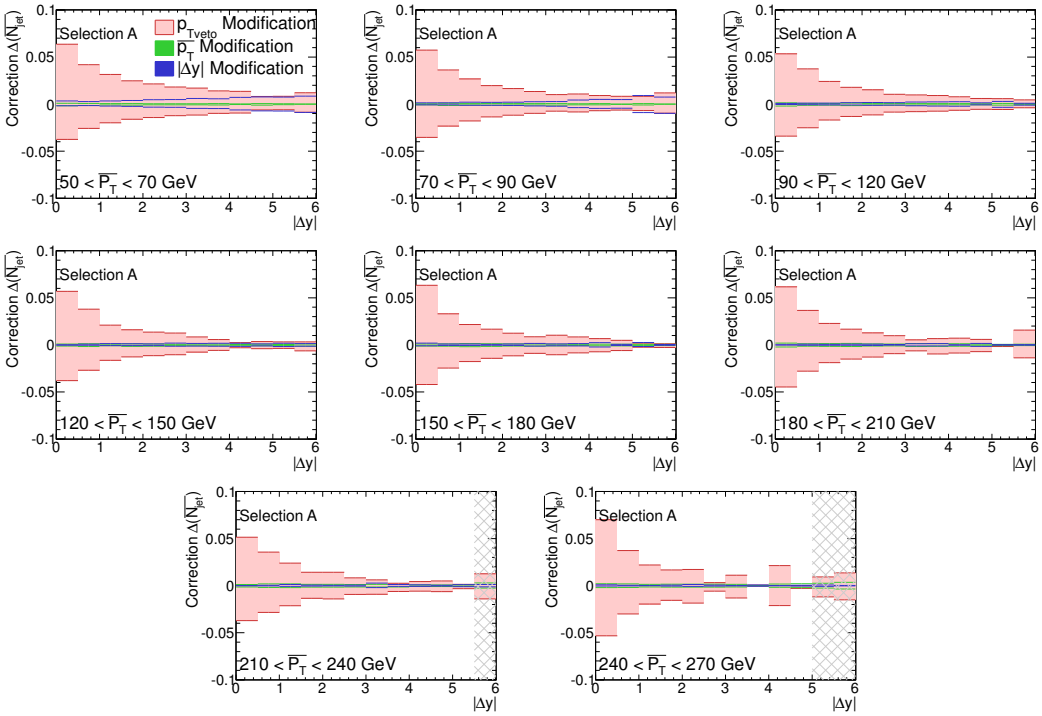


Figure 6.14: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection A is used.

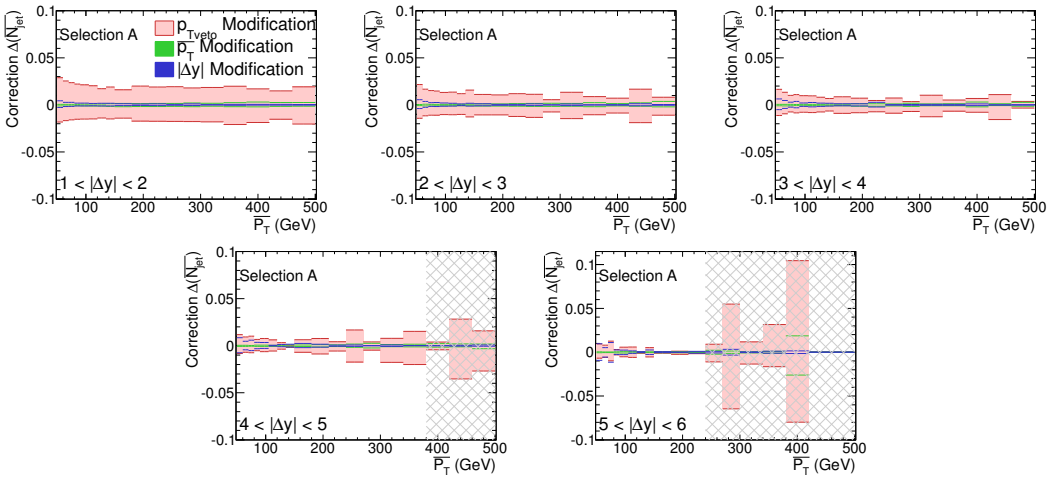


Figure 6.15: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_T}$ distribution. Selection A is used.

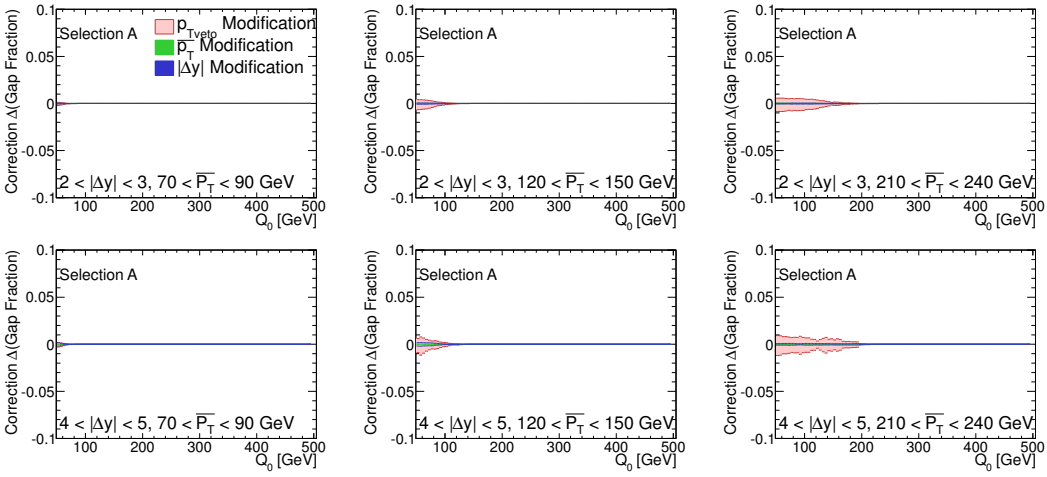


Figure 6.16: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection A is used.

6.5.2 Detector Modeling Uncertainty

Jet Energy Resolution

The jet energy resolution (JER) has been measured by ATLAS [9] and there is a relative uncertainty that varies with the p_{T} and η of the jet, some example values are shown in table 6.2.

p_T [GeV]	$\Delta\sigma(0 < \eta < 0.9)$	$\Delta\sigma(0.9 < \eta < 2)$	$\Delta\sigma(2 < \eta < 4.4)$
15	0.125	0.25	0.35
20	0.10	0.20	0.28
30	0.05	0.10	0.14

Table 6.2: The jet resolution uncertainties for a number of jet energies for different η regions of the detector. Jets with $p_T > 30\,\mathrm{GeV}$ have the same resolutions as the $p_T = 30\,\mathrm{GeV}$ case.

The effect of a larger resolution was estimated by smearing the p_{T} of each jet by an additional amount equal to the uncertainty in the resolution for that particular jet. Given that it was not possible to use this same method to improve the resolution, the uncertainty was symmetrised around the nominal correction value, for example $\Delta(F)_{up} = \Delta(F)_{down}$.

Jet Efficiency Loss

There is an uncertainty in the measurement made for the jet reconstruction efficiency of $\approx 2\%$ for jets with $p_T < 30$ GeV [9]. The jet efficiency in the low p_{T} region was artificially decreased by removing at random 2% of the jets which have $p_T < 30$ GeV. The effect of removing these jets caused a rise in the MC simulated gap fraction. This resulted in the correction factor for the gap fraction to be lower than the nominal value. As in the case for the jet energy resolution the uncertainty was symmetrised, $\Delta(F)_{up} = \Delta(F)_{down}$.

Vertex Position

Over the period in which the data was collected in 2010 the width of the vertex z position distribution varied significantly between runs. This variation is not simulated in the MC

samples. The effect this would have on the correction factor was estimated by re-weighting events to vary the width of the vertex z position. The range of weights applied to events based upon the vertex z position are listed below

Vertex Z Shape Event weight multiplied by four different functional forms

$1.0 + |z|/200.0$

$1.0 - |z|/200.0$

$1.0 + (|z|/200.0)^2$

$1.0 - (|z|/200.0)^2$

where $|z|$ is the absolute position of the vertex in the z direction. The maximum $|z|$ allowed is 200mm.

Correction Uncertainties

The uncertainty in the correction factors are shown in figures 6.17-6.21 for the other sources. Each of the sources have been seen to produce similar uncertainties in the correction factor, usually $\lesssim 2\%$ across the whole phase space and often consistent with zero. The uncertainties in the average number of veto jets are yet again the largest of all the observables. There was also often seen large fluctuations present from the vertex Z position re-weighting. This was understandable when you consider the small number of events produced with large values of $|z|$. There was similar behavior seen for the other selections as before, and they are displayed in appendix C.

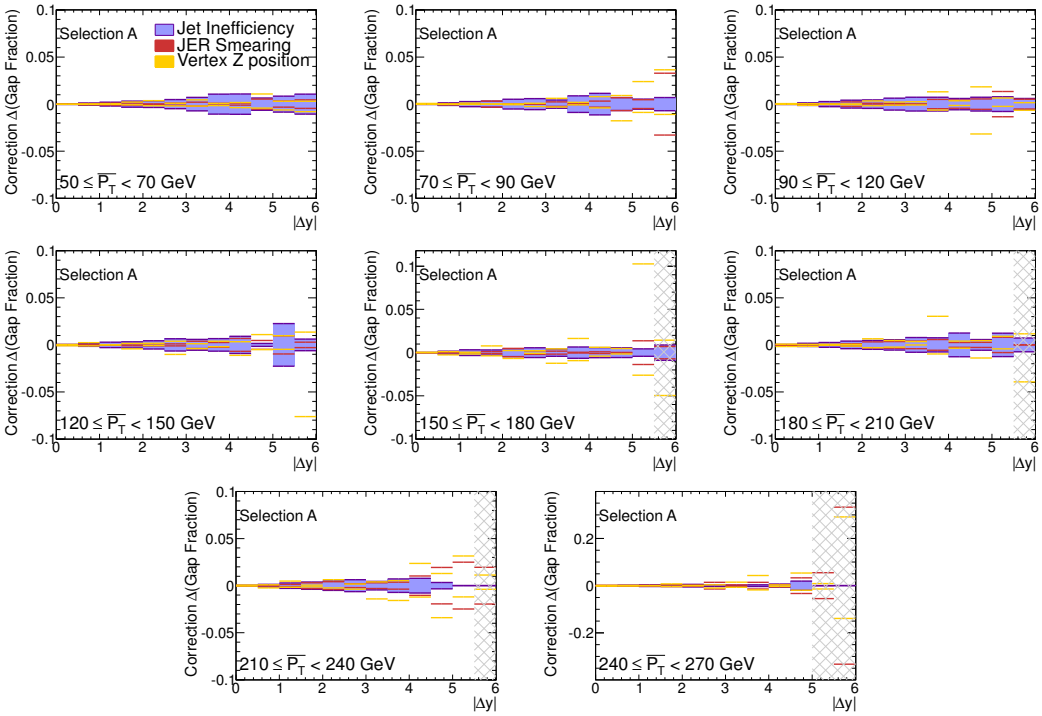


Figure 6.17: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection A is used.

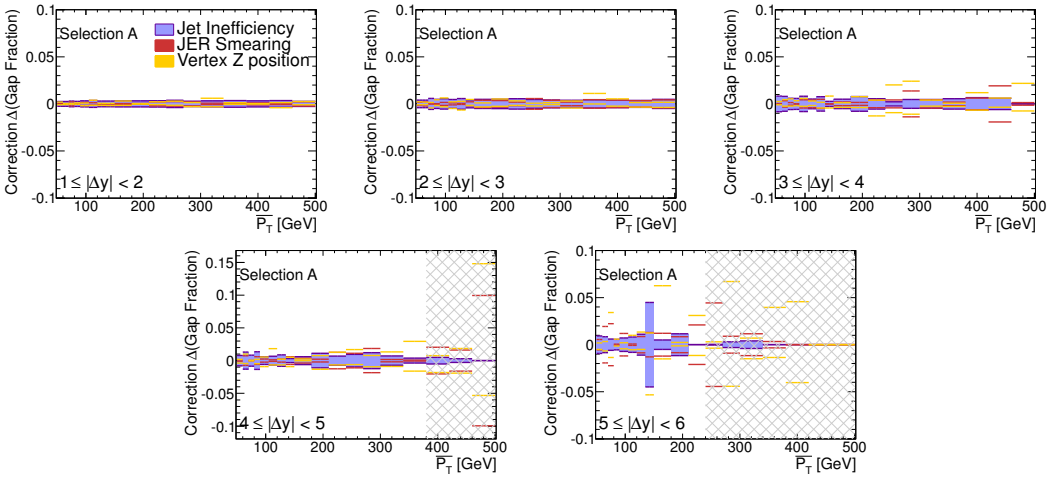


Figure 6.18: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_{\mathrm{T}}}$ distribution. Selection A is used.

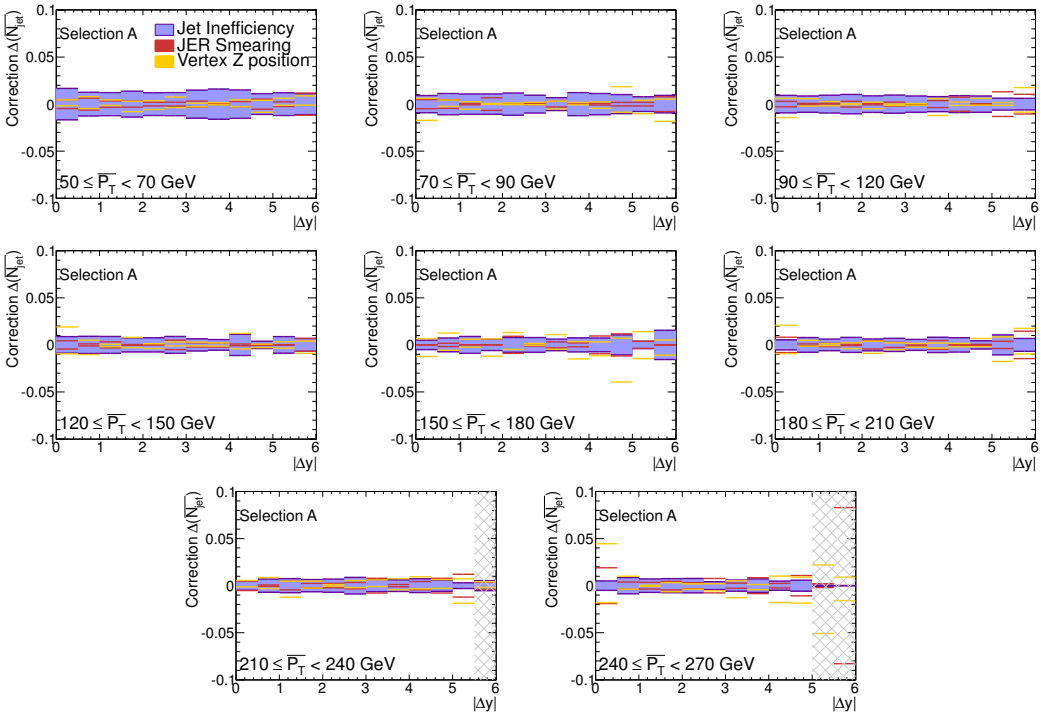


Figure 6.19: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection A is used.

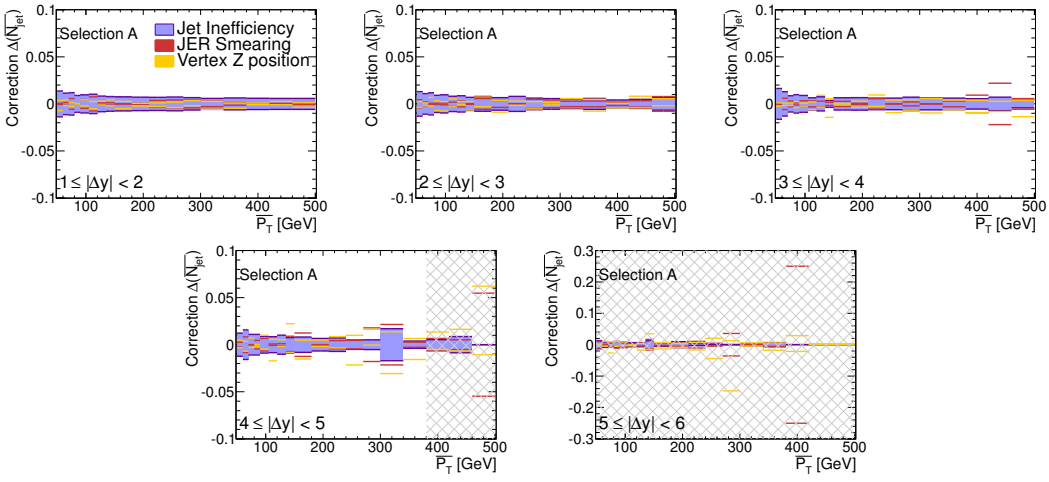


Figure 6.20: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_T}$ distribution. Selection A is used.

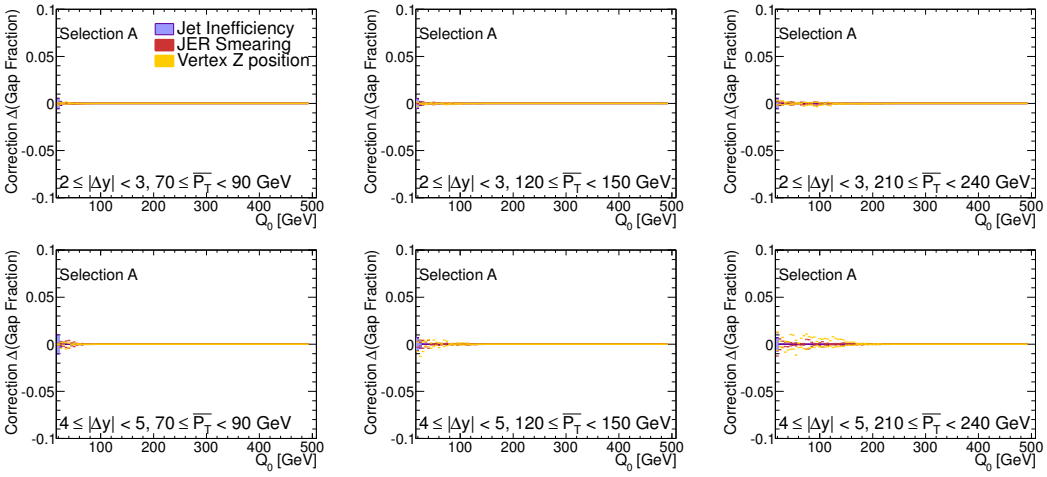


Figure 6.21: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection A is used.

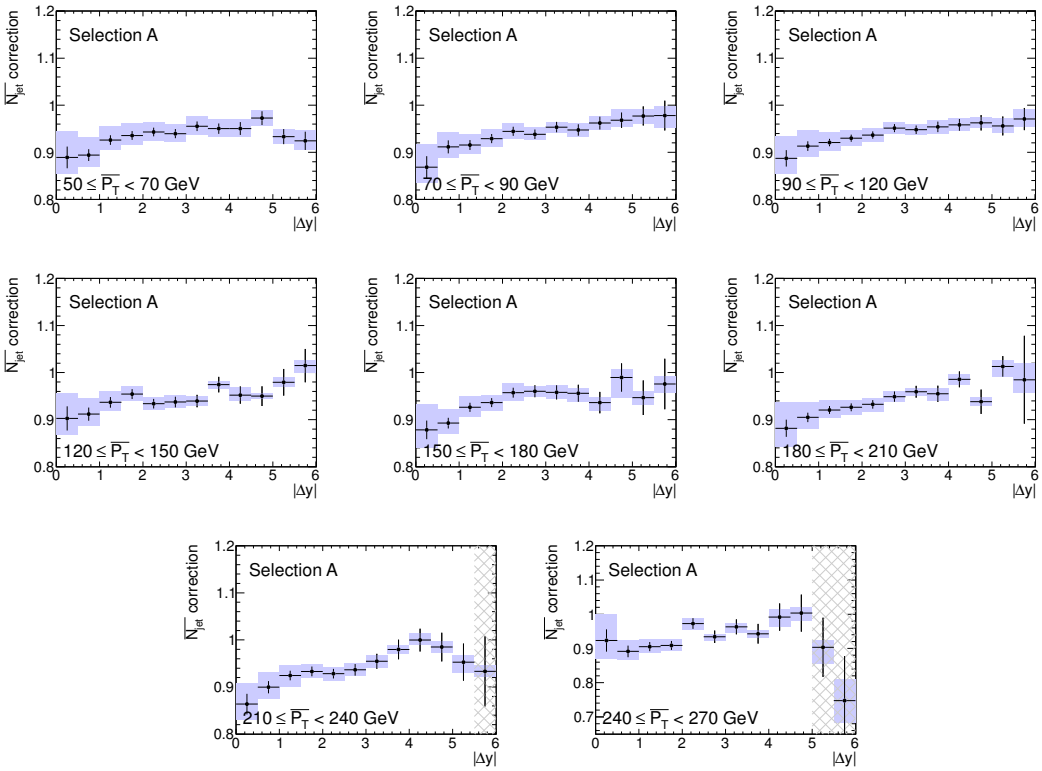


Figure 6.22: Correction factors for $N^{jet}_{average}$ vs |Δy| for selection A.

6.6 Final Correction Factors

The correction factors are shown in figures 6.22-6.25 where the statistical errors are displayed as error bars and the total systematic uncertainty is represented by the blue band. The total systematic uncertainty was evaluated by summing in quadrature the uncertainties obtained from truth modification, JER smearing, jet efficiency loss and the vertex z position re-weighting. All the data correction factors were found to be close to one for the gap fraction distributions and close to 0.9 for the average number of veto jet distributions, giving confidence in the use of the bin-by-bin method. Even with the additional event samples the statistical uncertainties are still a major component of the uncertainty in the data corrections. The correction factors for the other selections can be found in appendix C. Selection C has overall the highest uncertainties.

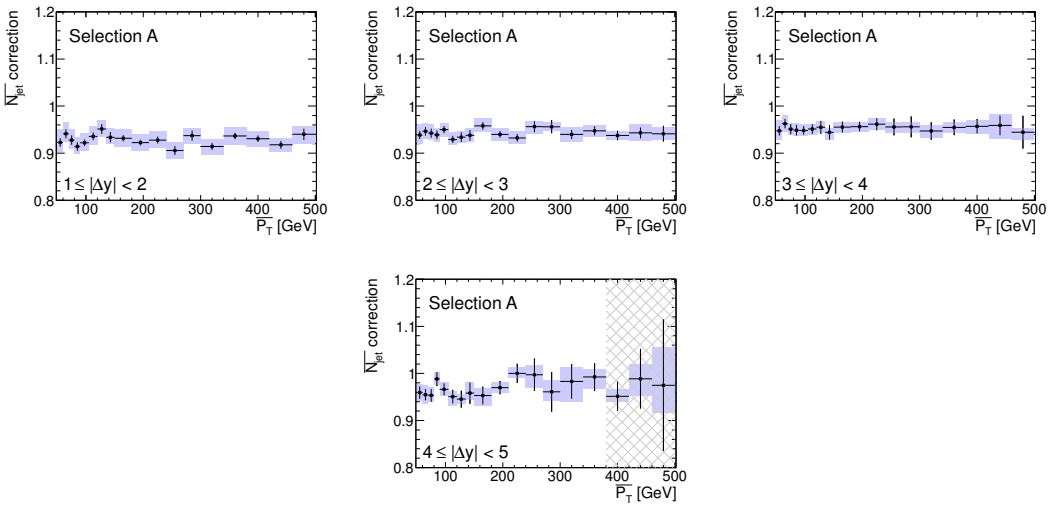


Figure 6.23: Correction factors for $N^{jet}_{average}$ vs $\overline{p_T}$ for selection A.

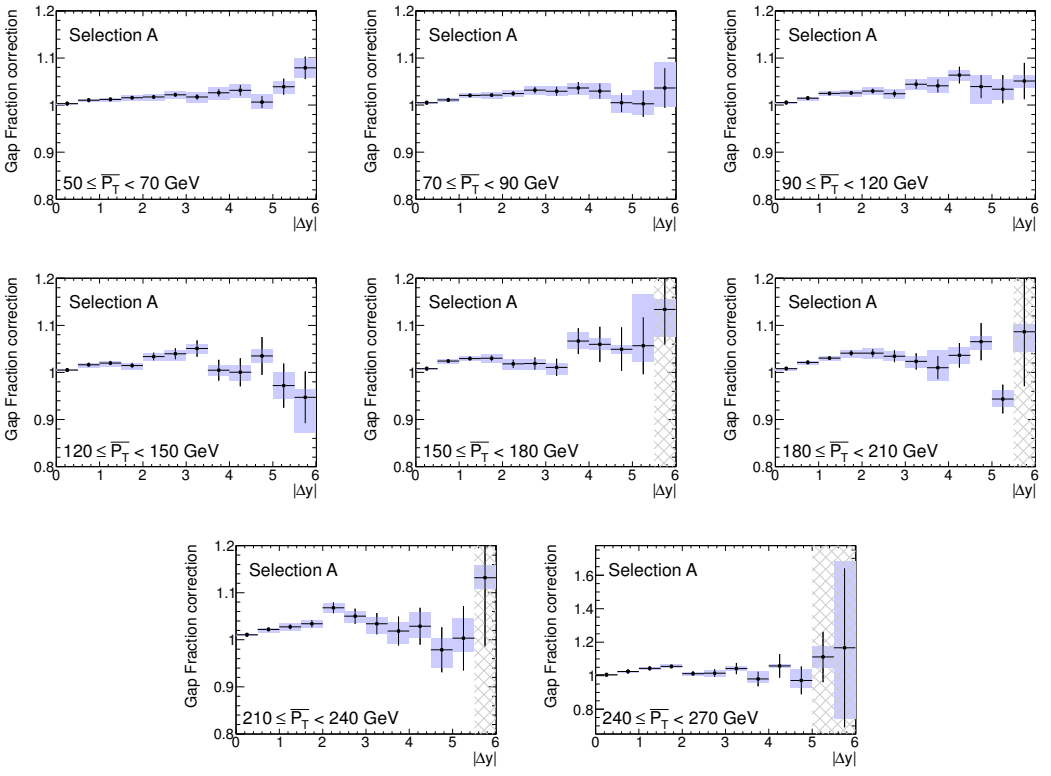


Figure 6.24: Correction factors for gap fraction vs |Δy| for selection A.

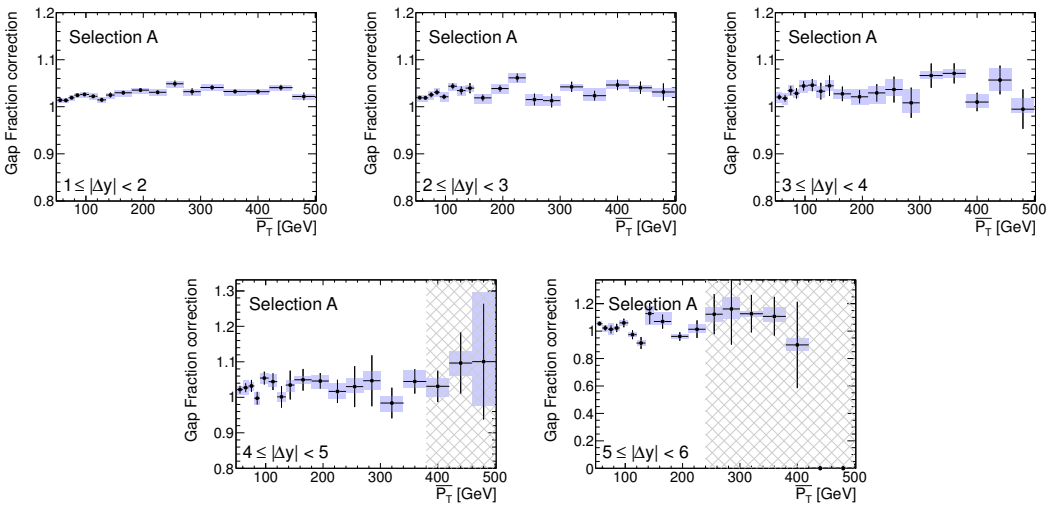


Figure 6.25: Correction factors for gap fraction vs $\overline{p_{\mathrm{T}}}$ for selection A.

6.7 Additional Systematic Uncertainties on Measurement

The systematic uncertainty for the data corrections has already been shown in section 6.2. Additional sources for uncertainty in the measurements were considered. These included the jet energy scale, the trigger and pile-up. These were determined by other members of ATLAS and will be described briefly in the following sections.

6.7.1 Jet Energy Scale Uncertainty

One of the most significant uncertainties in any measurement that uses jets arises from the uncertainty in the jet energy scale (JES). The effect on this analysis was estimated by varying the JES of each jet within the allowed limits of the JES uncertainty. The JES uncertainty was determined with both a correlated JES shift for jets in all η regions of the detector and an uncorrelated JES shift for jets in different η regions. It was always found to be the case that the correlated JES uncertainty was much larger than the uncorrelated JES uncertainty. The final JES uncertainty was therefore taken to be just the correlated case. In figure 6.26, a summing of the systematic uncertainties for the JES and data corrections, labeled unfolding, are shown for a selection of regions. The uncertainty for the data corrections was found to be typically lower than the JES uncertainty in most region of $|\Delta y|$ and $\overline{p_T}$ considered.

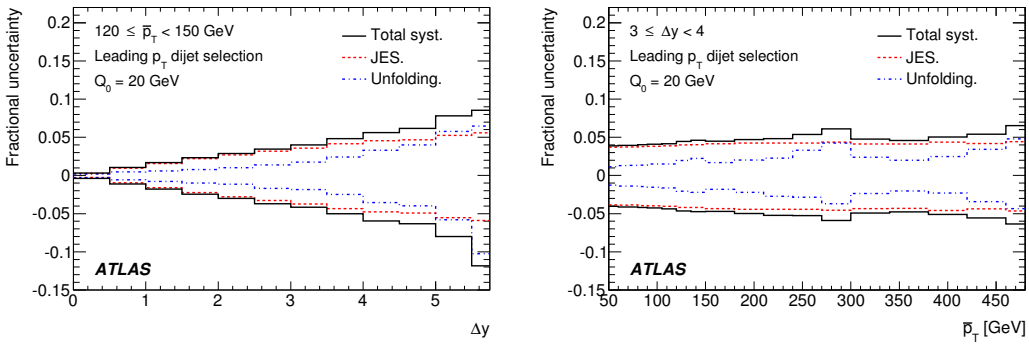


Figure 6.26: The relative uncertainties for the jet energy scale correction and the data corrections. These results are taken from [1]

6.7.2 Trigger and Pile-up Uncertainty

The trigger strategy was tested for possible bias by checking that the efficiency for finding events with an empty interval was the same as the efficiency for a filled interval. The bias present was found to contribute < 0.5% to the systematic uncertainty [1]. The effectiveness of the vertex requirement in removing in-time pile-up was tested by considering how the measured observables varied as a function of the mean number of interaction per bunch crossing, which increased with time over the period of 2010 data taking. The deviation was also seen to be < 0.5% in the measured observables as the number of interactions increased.

6.8 Results

In this section the results from the dijet with a central jet veto paper [1] are discussed, where the data has been corrected and compared to a number of theoretical predictions. These predictions, excepting HEJ, have been discussed in previous chapters.

6.8.1 Comparison to Monte Carlo Samples

Comparisons of the fully corrected data were made to a number of leading order Monte Carlo generators. In figure 6.27 a selection of the regions investigated is shown. The spread of the leading order MC generators was large compared to the experimental uncertainty in the measurement. The ALPGEN+HERWIG/JIMMY predictions show the greatest deviation from the data, where the predicted jet content is too high leading to a gap fraction which is significantly lower than the data.

The spread of the leading order MC generator predictions reflects a large uncertainty in the theory. Hence while the PYTHIA predictions seem to match very well there is no reason to believe one MC model over another one.

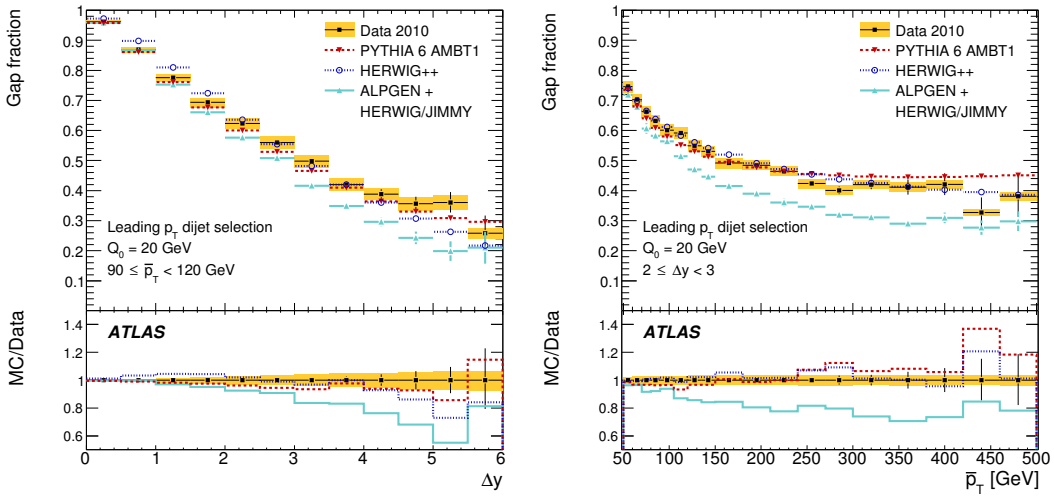


Figure 6.27: Comparisons with data of the PYTHIA, Herwig++ and Alpgen+HERWIG/JIMMY samples generated earlier. These results are taken from [1]

6.8.2 Comparison of data to POWHEG and HEJ

HEJ

The higher energies that the LHC provides have altered the behavior of the typical jet events. There is now more likely to be a larger number of high p_{T} jets in each event. High energy jets (HEJ) [10][11] is a new framework that provides an all-order description of processes containing more than two partons that have a large rapidity separations, inspired from BFKL formalism. There was an expectation that in the limit of $p_T << m_{jj}$ that the HEJ prediction should perform well. The samples that were generated and compared to the measurement did not contain a parton shower, but the effect from soft corrections was estimated to be small. As in the case for the other predictions, the anti-k_t jet algorithm was applied with a radius parameter of $R = 0.6$, clustering the partons to produce infrared and collinear-safe predictions.

Results

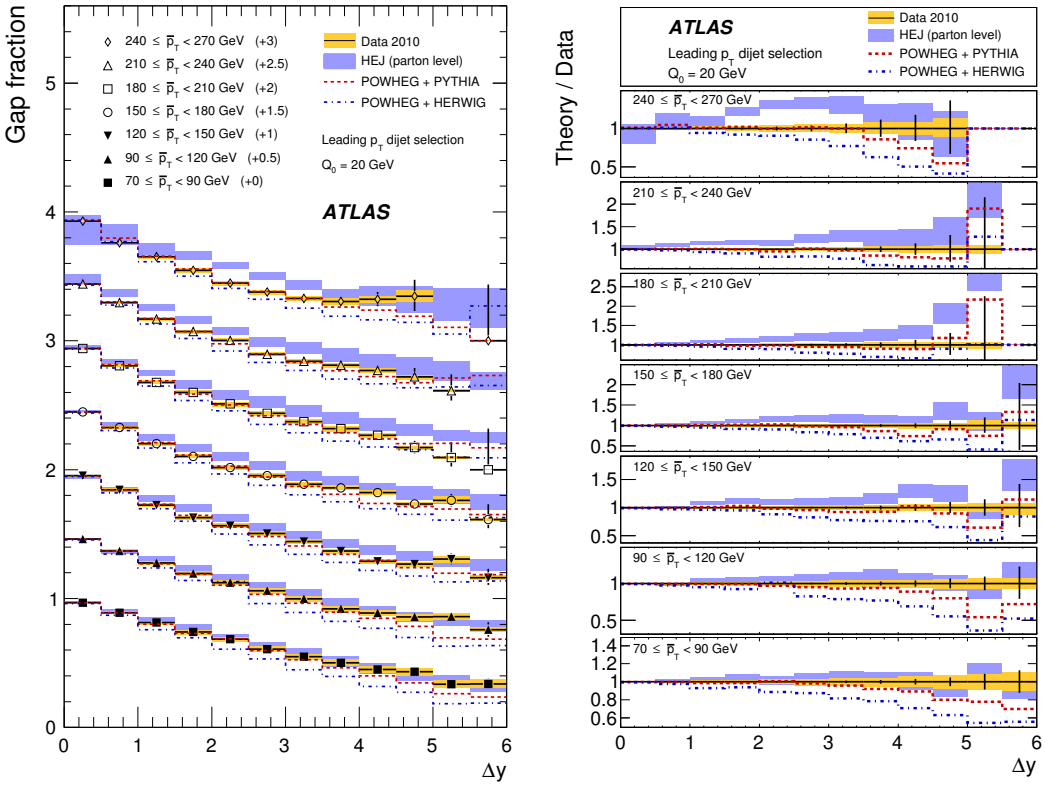
In this section the results are presented for the fully corrected data collected in 2010. The data has been compared to HEJ, POWHEG+PYTHIA and POWHEG+HERWIG/JIMMY.

The POWHEG+PYTHIA shown here uses the MSTW 2008 NLO PDF for the generation of the hardest scatter and the tune used for PYTHIA showering was AMBT1. The POWHEG+HERWIG/JIMMY shown also uses MSTW 2008 NLO PDF for the hardest scatter and the AUET1 tune in HERWIG/JIMMY. The HEJ curves contain both a scale and PDF uncertainty, whilst the POWHEG curves show only the statistical uncertainty. The systematic error on the data values contains the uncertainty from the data correction and the JES uncertainty. Figures 6.28 and 6.29 show the gap fractions vs $|\Delta y|$ and $\overline{p_{\mathrm{T}}}$. Figure 6.30 shows the average number of jets above the veto threshold inside the gap region vs $\overline{p_{\mathrm{T}}}$. For each figure the left-hand side plot shows the observable with each curve offset in the y-axis direction, as indicated by the value in the legend. The right-hand side plot shows the ratio of each theoretical prediction to data.

The POWHEG+PYTHIA predictions for the gap fraction as a function of $|\Delta y|$ for selection A, figure 6.28, were generally seen to be in good agreement although a trend appeared at large $|\Delta y|$ for the prediction to lie below the data. POWHEG+HERWIG/JIMMY has significantly greater disagreement for $|\Delta y| > 3$. The Multi-Regge Kinematic (MRK) limit which HEJ uses is only applicable for the case when the outgoing partons are well separated in rapidity and have similar transverse momenta. Events with high $\overline{p_{\mathrm{T}}}$ combined with a low veto threshold do not satisfy the MRK requirement. This is consistent with HEJ performing best when $70 < \overline{p_T} < 90\,\mathrm{GeV}$ with a growing disagreement for higher $\overline{p_{\mathrm{T}}}$.

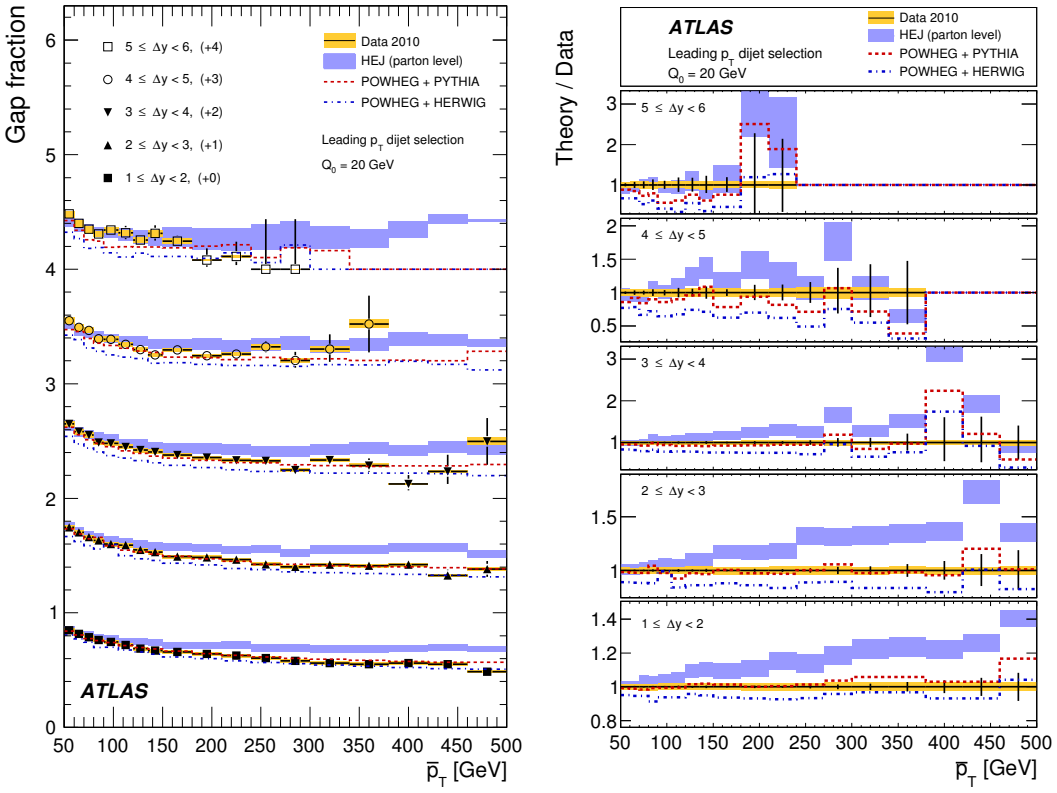
The mean number of jets observed in the gap as a function of $\overline{p_{\mathrm{T}}}$, figure 6.30, shows similar behavior. As before POWHEG+PYTHIA provides the best description of the data. POWHEG+HERWIG has more than 20% extra jets above the veto threshold in the gap region over large ranges of $\overline{p_{\mathrm{T}}}$, these differences are seen to worsen at low $\overline{p_{\mathrm{T}}}$. HEJ predicts too few jets above the threshold over almost all of the $\overline{p_{\mathrm{T}}}$ range considered. An explanation for this may be related to the lack of parton shower in the HEJ predictions, and so cannot properly reproduce the multiplicity of jets measured.

The comparisons to measurements for the other selections B and C are available in the dijet with a central jet veto paper [1] and in additional material [12]. In selection B it was



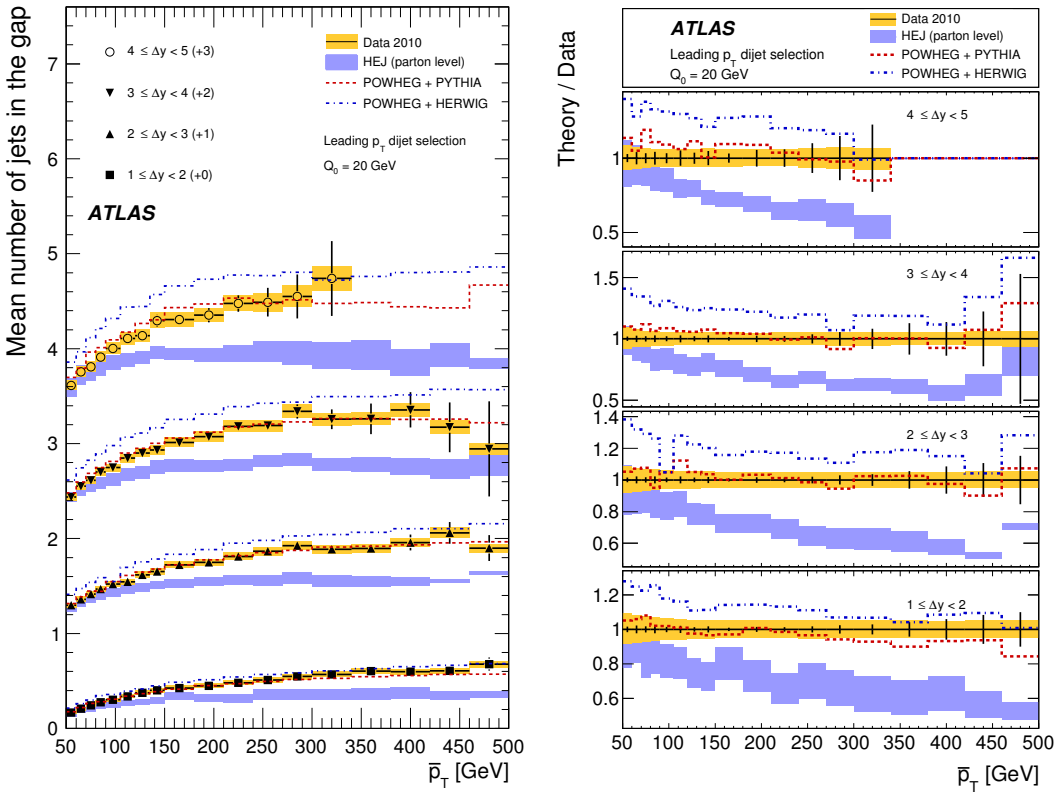
(a) (b)

Figure 6.28: Gap fraction vs |Δy| for the corrected 2010 data compared to the predictions from POWHEG and HEJ in selection A. The ratios of the predictions to data are shown in (b). The error bars on the data points are the statistical uncertainty and the yellow bar is the systematic uncertainty. The systematic uncertainty for HEJ is represented by the blue bar. These results are taken from [1].



(a) (b)

Figure 6.29: Gap fraction vs $\overline{p_{\mathrm{T}}}$ for the corrected 2010 data compared to the predictions from POWHEG and HEJ in selection A. The ratios of the predictions to data are shown in (b). The error bars on the data points are the statistical uncertainty and the yellow bar is the systematic uncertainty. The systematic uncertainty for HEJ is represented by the blue bar. These results are taken from [1].



(a) (b)

Figure 6.30: Average veto jet number vs $\overline{p_{\mathrm{T}}}$ for the corrected 2010 data compared to the predictions from POWHEG and HEJ in selection A. The ratios of the predictions to data are shown in (b). The error bars on the data points are the statistical uncertainty and the yellow bar is the systematic uncertainty. The systematic uncertainty for HEJ is represented by the blue bar. These results are taken from [1].

seen that the HEJ predictions are more consistent with the data for the gap fraction vs Δy, however there were still seen to be major discrepancies with the average number of veto jets in the gap. In selection C the MC shower dependence was seen to become much smaller, where the POWHEG+PYTHIA and POWHEG+HERWIG predictions produced identical results and were both consistent with the data.

6.8.3 Public Code

A public version of code was written that can produce the dijet with a central jet veto distributions. It is available in the Rivet framework [13], which is a toolkit for the validation of Monte Carlo event generators. Running the analysis code `ATLAS_2011_S9126244` will produce all the predictions for the currently publicly available ATLAS results [1], which can all be obtained from either Rivet or the HEPDATA archive (`http://hepdata.cedar.ac.uk/view/p8047`).

Bibliography

[1] Georges Aad et al. Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 2011:1–36, 2011. 10.1007/JHEP09(2011)053.

[2] A Corso-Radu et al. Data Quality Monitoring Framework for ATLAS Experiment: Performance Achieved with Colliding Beams at LHC. Technical Report ATL-DAQ-PROC-2011-007, CERN, Geneva, Jan 2011.

[3] Georges Aad et al. Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector at the LHC. *Eur.Phys.J.*, C71:1630, 2011.

[4] M Baak, M Petteni, and N Makovec. Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV. Technical Report ATLAS-COM-CONF-2010-038, CERN, Geneva, May 2010.

[5] Victor Lendermann et al. Combining Triggers in HEP Data Analysis. *Nucl. Instrum. Meth.*, A604:707–718, 2009.

[6] N P Konstantinidis et al. The Atlantis event visualisation program for the ATLAS

experiment. *Computing in High Energy Physics and Nuclear Physics*, page pp.361, 2004.

[7] Glen Cowan. A survey of unfolding methods for particle physics. *Proc. Conf. on Advanced Statistical Techniques in Particle Physics*, page 248, 2002.

[8] Volker Blobel. Unfolding methods in high energy physics experiments. *Proceedings of the 1984 CERN School of Computing*, 1984.

[9] Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $\sqrt{s}$ = 7 TeV. Technical Report ATLAS-CONF-2010-054, CERN, Geneva, Jul 2010.

[10] Jeppe R. Andersen and Jennifer M. Smillie. High Energy Description of Processes with Multiple Hard Jets. *Nucl. Phys. Proc. Suppl.*, 205-206:205–210, 2010.

[11] Jeppe R. Andersen and Jennifer M. Smillie. Multiple Jets at the LHC with High Energy Jets. *JHEP*, 1106:010, 2011.

[12] Georges Aad et al. Additional material for measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector, https://atlas.web.cern.ch/ATLAS/GROUPS/PHYSICS/PAPERS/STDM-2011-03, August 2011.

[13] A. Buckley et al. Rivet user manual. *ArXiv e-prints*, March 2010.

Chapter 7

Measurement Extension

In this chapter the use of the anti-k_t jet algorithm with radius parameter $R = 0.4$ was explored for the dijet with a central jet veto measurement. The use of an alternative jet radius allowed a test of the sensitivity of the veto observables to multiple interactions, underlying event and peturbative radiation. The contributions to the jet energy from each of these sources should vary as a function of the jet radius. This required work to evaluate the trigger strategy and the jet energy scale uncertainty for the new radius. In addition the vertex requirement was loosened and the impact of additional pile-up vertices was tested. Finally the new measurements were compared to the latest POWHEG predictions.

7.1 Analysis Changes

A number of changes with regards to chapter 6 have been made in the consideration of these extensions to the dijet with a central jet veto analyses. The minimum veto scale was been raised from 20GeV to 30GeV, reducing the effect which pile-up could have on the veto observable and in general reducing the total systematic uncertainty as well. The increased veto scale did come at the cost of reduced sensitivity to the large angle soft gluon radiation, although this is offset by the greater access to high $\overline{p_{\mathrm{T}}}$ events. As the veto scale has been increased so were the minimum p_{T} values for the leading jets, both $p_{T\,1} > 30\,\mathrm{GeV}$ and $p_{T\,2} >$

30 GeV were required for the first two jets. The only selection considered in this chapter is type A, where the leading two jets define the interval region. In addition to the previously used PYTHIA samples a new sample of Monte Carlo events became available. This two dimensionally weighted sample, p_{T} and $|\Delta y|$, yielded a higher fraction of large interval events The new sample of events results in an improved determination of the unfolding and jet energy scale systematic uncertainties, which had been previously statistically limited in the high $|\Delta y|$ region.

7.1.1 Distribution Selection

To alleviate some issues seen in chapter 6 with regards to the purity distributions, where purity levels reached a nadir of 60%, a manual re-optimisation of the binning was made. The bin edges for the $|\Delta y|$ variable were changed to [0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0] and the bin edges for the $\overline{p_{\mathrm{T}}}$ variable were changed to [60.0, 80.0, 110.0, 145.0, 185.0, 250.0, 320.0, 400.0, 550.0] GeV. An increased bin size was observed to result in increased purity values. In figures 7.1 and 7.2 the purity is shown to be approximately 80% and flat as a function of $\overline{p_{\mathrm{T}}}$. The increase in the purity gave an improved confidence in the data correction procedure. Larger bins not only improved the purity observed but also increased the number of events in each bin reducing the statistical uncertainty. The downside of an increased bin size was that observable features were limited to a larger size. The figures also show that there is no significant change in purity when anti-k_t $R = 0.4$ jets are considered.

7.2 Trigger

In chapter 6 the choice of radius used for the anti-k_t jet algorithm was restricted to $R = 0.6$. As part of the work to understand the veto measurement in the presence of pile up another radius, $R = 0.4$, is now also considered. The triggering procedure used to make the previous measurement using the original jet radius needed to be evaluated for $R = 0.4$. Figure 7.3

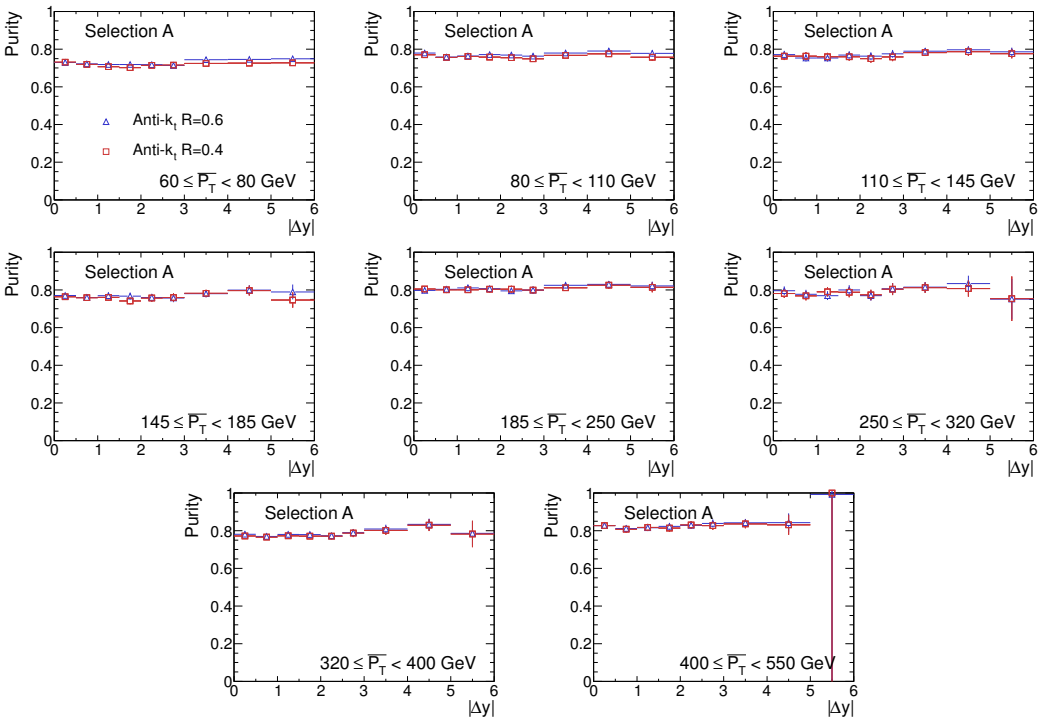


Figure 7.1: Selection A purities for the variable |Δy| where the MC events used were PYTHIA 6 with the AMBT1 tune.

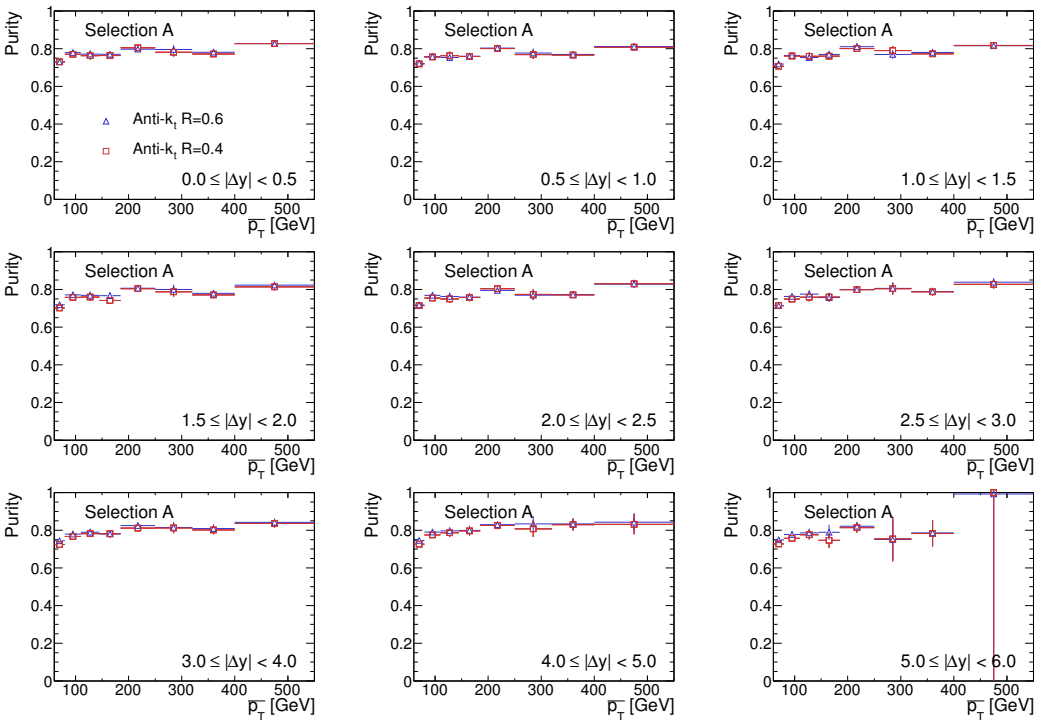


Figure 7.2: Selection A purities for the variable $\overline{p_T}$ where the MC events used were PYTHIA 6 with the AMBT1 tune.

shows the trigger efficiency, the fraction of events as a function of $\overline{p_{\mathrm{T}}}$ passing the three lowest threshold jet triggers given that the event has passed the L1_MBTS_1 trigger. The L1_MBTS_1 trigger used requires that there be at least one hit in the minimum bias trigger scintillator, its efficiency as shown in [1] is high even for very low multiplicity events, thus allowing the inference of the absolute trigger efficiency. Data recorded in an early period of running is necessary because the luminosity conditions were relatively low, which allowed the use of the L1_MBTS_1 trigger with a low prescale value. It is observed when the events have $\overline{p_{\mathrm{T}}} > 60.0$GeV the three jet triggers considered L1_J5, L1_J10 and L1_J15 are all above 99% efficiency. The turn-on points for anti-k_t $R = 0.4$ and anti-k_t $R = 0.6$ jets are consistent.

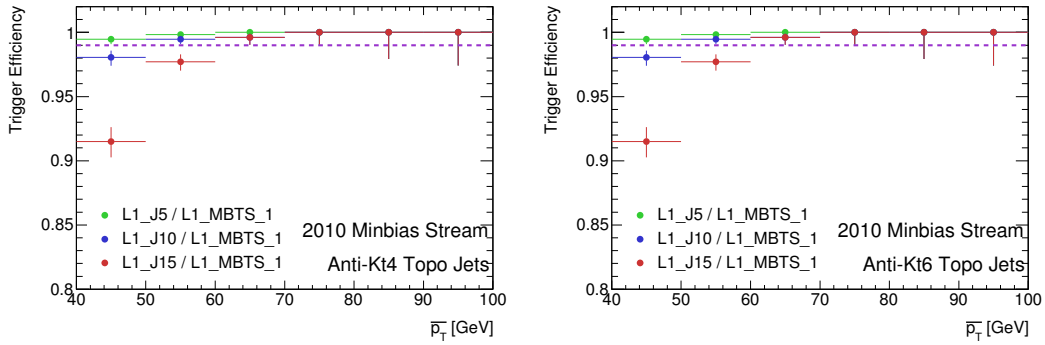


Figure 7.3: Trigger efficiency for the three lowest threshold L1 jet triggers L1_J5, L1_J10 and L1_J15. The L1_MBTS_1 trigger is used as the reference trigger. The dashed line indicates the 99% efficiency point. Standard jet good run list requirements. No ugly or loose bad jets with $p_T > 20$GeV. The vertex requirement of one good vertex is required

There is a relatively small amount of data available in the Minbias stream with high p_{T} jets and insufficient events to obtain the trigger efficiency curves for the higher threshold jet triggers. A solution is to use the technique commonly referred to as bootstrapping. This is where a jet trigger is used as a reference trigger to evaluate a higher threshold trigger's efficiency, assuming that the absolute trigger efficiency is available for the new reference trigger. This allows the use of a larger sample of events originating from the jet stream in the evaluation of the jet trigger efficiency for trigger thresholds above L1_J15. In figure 7.4 the plateau regions for the other thresholds can be observed, where again there can be seen to be

a great deal of similarity between the two choices of jet radius.

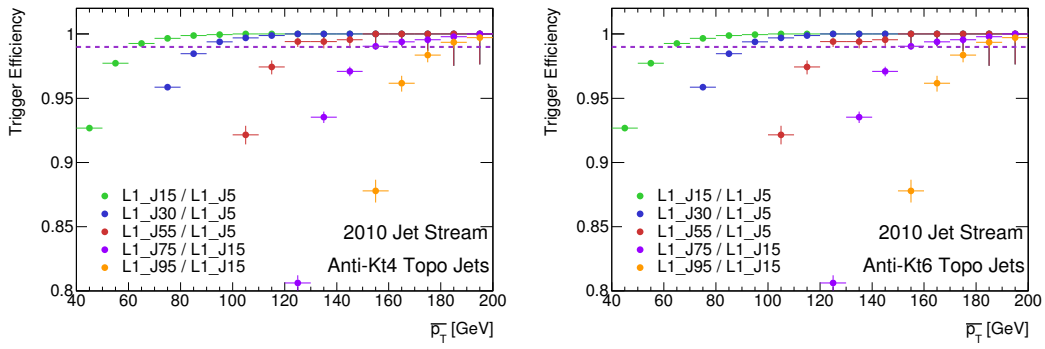


Figure 7.4: Trigger efficiency for the higher threshold L1 jet triggers. Softer jet triggers are considered as the reference triggers. The dashed line indicates the 99% efficiency point. Standard jet good run list requirements. No ugly or loose bad jets with $p_T > 20$GeV. The vertex requirement of one good vertex is required

Given that the behavior of the triggers are observed to be the same as in the anti-k_t $R = 0.6$ jet case, the trigger division strategy for anti-k_t $R = 0.4$ can use the same triggers and $\overline{p_T}$ divisions for event selection. The choice of trigger divisions are then taken to be the same as those listed in table 6.1. The size of bias for the new radius was also observed to be negligible as in the anti-k_t $R = 0.6$ case.

7.3 Vertex Cut Loosening

Figure 7.5 shows that the average number of vertices in 2010 data rises with $\overline{p_T}$. This is because the triggers in which low $\overline{p_T}$ events are collected had larger prescales applied as the luminosity increased, such that a larger fraction of events are recorded during the early running with low pile-up conditions. There is also notably very little dependence on the $|\Delta y|$ between the leading jets or the choice of jet algorithm radius.

The minimal effect of the pile-up present in 2010 data can be easily observed by comparing the gap fraction and average veto jet number observables with and without the $N_{vertex} = 1$ criterion applied. In figure 7.6 the distributions are seen to be consistent within the statistical uncertainties. Some of the largest deviations occur when there has been a large fluctuation in

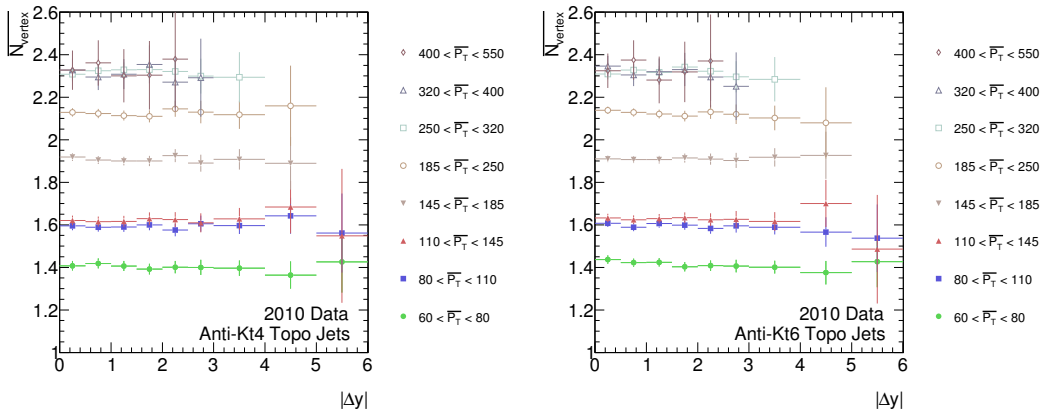
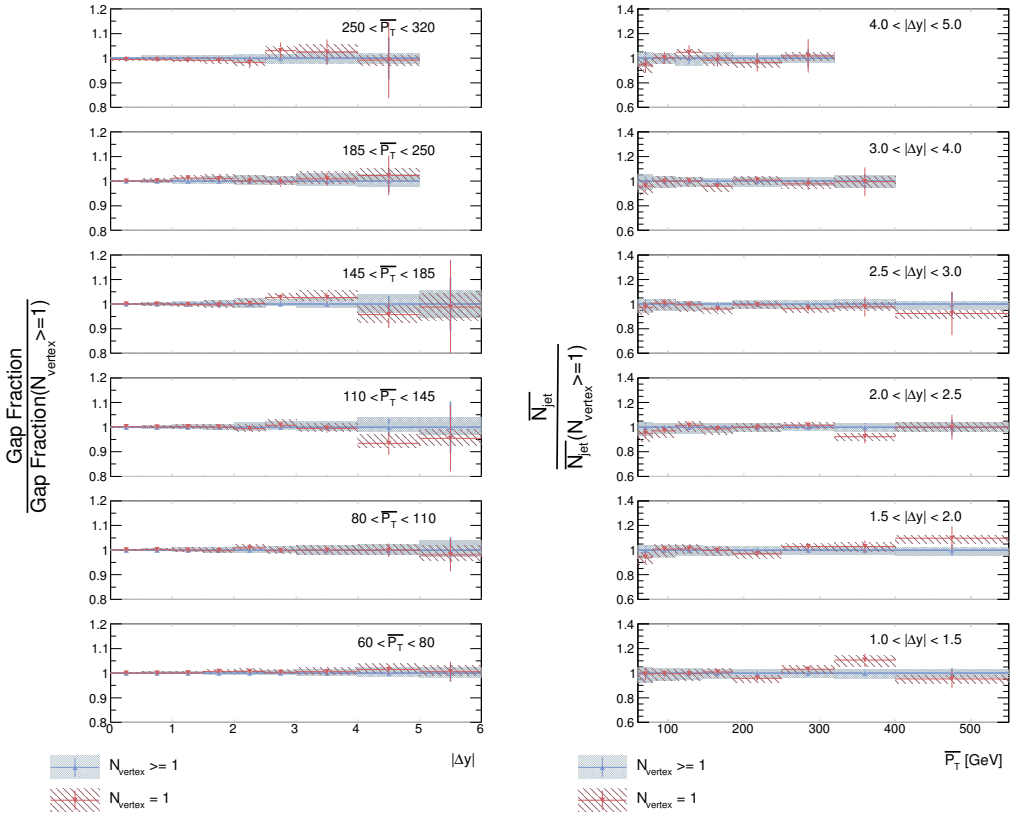


Figure 7.5: The average number of good vertices in events recorded during 2010. The events are split into several different $\overline{p_{\mathrm{T}}}$ regions as indicated in the legend.



(a) (b)

Figure 7.6: (a) Ratios of the gap fraction as a function of |Δy| with and without additional vertices allowed. (b) Ratios of the average veto jet number as a function of $\overline{p_{\mathrm{T}}}$ with and without additional vertices. The complete ATLAS 2010 data is shown for the anti-k_t $R = 0.4$ jets. The filled regions and error bars represent the systematic and statistical uncertainties respectively.

the $N_{vertex} = 1$ distribution; the additional statistics available after loosening this requirement can be clearly seen in the reduction of the statistical uncertainties. The other observables and anti-k_t $R = 0.6$ jets show similar levels of agreement.

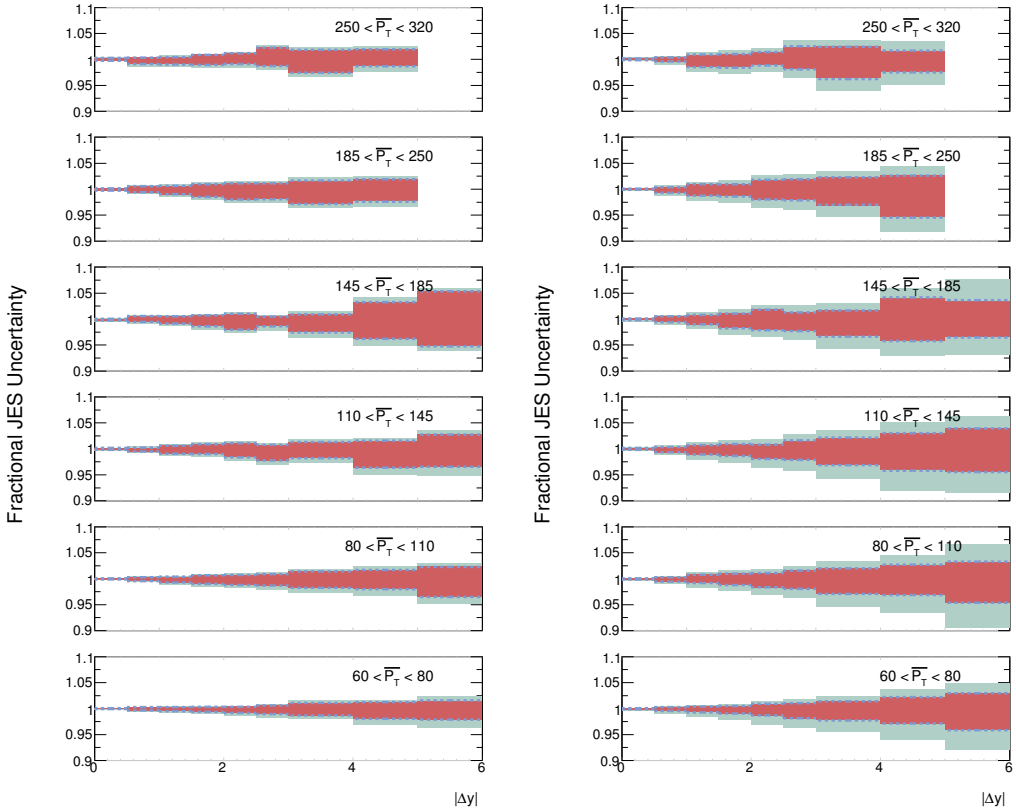
7.4 Jet Energy Scale

One of the largest sources of systematic uncertainty for the measurement of dijets with a central jet veto was the jet energy scale. The jet energy scale uncertainty was evaluated using the simulated Monte Carlo event samples, with the addition of the p_{T}-$|\Delta y|$ weighted PYTHIA sample. As was performed in chapter 6 the uncertainty for the veto observables was evaluated by considering a shift in energy up and down of each jet across the entire detector equal to the jet energy scale uncertainty.

7.4.1 JES Systematic Uncertainty

Figures 7.7, 7.8, 7.9 and 7.10 show the effect that pile-up inclusion has on the relative jet energy scale uncertainty for each of the veto observables. A feature typical of these plots is that the uncertainty derived for anti-k_t $R = 0.4$ jets is almost universally smaller than for anti-k_t $R = 0.6$ jets. Whilst there is a difference in the magnitude of the jet energy scale uncertainty, the dependence as a function of $|\Delta y|$ and $\overline{p_{\mathrm{T}}}$ is the same for both radii considered. Considering firstly the case of events with only one vertex the behaviour is seen to be consistent between the two jet radii. In figure 7.7 the jet energy scale uncertainty for the gap fraction grows with $|\Delta y|$, approaching 2.5% uncertainty at the largest separations considered and is true for each $\overline{p_{\mathrm{T}}}$ region. On the other hand the gap fraction vs $\overline{p_{\mathrm{T}}}$ in figure 7.8 has a constant behavior. The average number of jets above the veto scale, in figures 7.9 and 7.10, have a systematic uncertainty that is flat as a function of both $|\Delta y|$ and $\overline{p_{\mathrm{T}}}$.

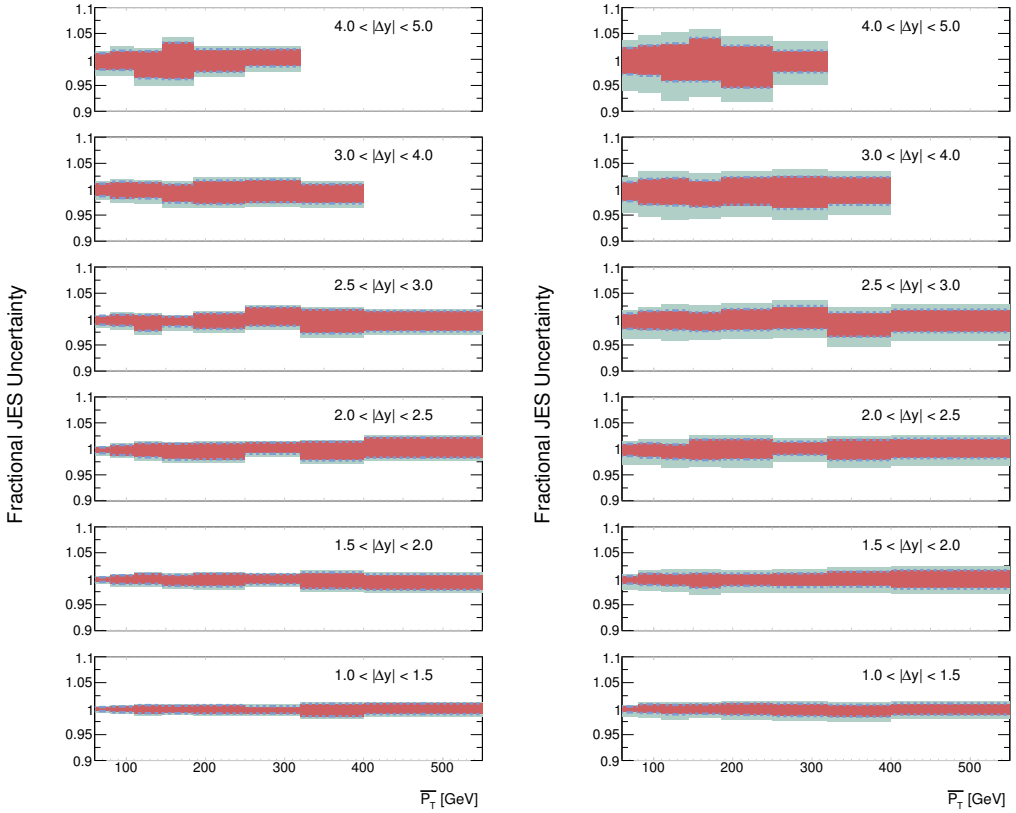
Given that we have removed the constraint on vertex number previously imposed, the jet energy scale uncertainty needed to be determined again. The previous method used Monte



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.7: The relative jet energy scale systematic uncertainty for the gap fraction as a function of $|\Delta y|$. The red region represents the uncertainty for data containing 1 vertex events only. Dashed blue line represents the uncertainty for 2010 data without the 1 vertex restriction and is always found at border of red. The pale green region is for the case where the average number of vertices is extrapolated to 7.

Carlo events to infer the systematic uncertainty. These events however do not have a vertex number distribution consistent with the data, as they had not been simulated with pile-up conditions. The samples can still be used but there was a need to evaluate the average number of vertices in each bin for each observable from data. Combining this with the knowledge of the jet energy scale uncertainty for each vertex number from Monte Carlo, an extrapolation of the jet energy scale uncertainty for the true number of vertices in each bin can be made. The inclusion of events with more than one vertex was seen to have minimal effect on the final jet energy scale uncertainty, where the dashed blue line in figures 7.7 to 7.10 are almost entirely consistent with the one vertex case. This is not too surprising as the pile-up correction applied

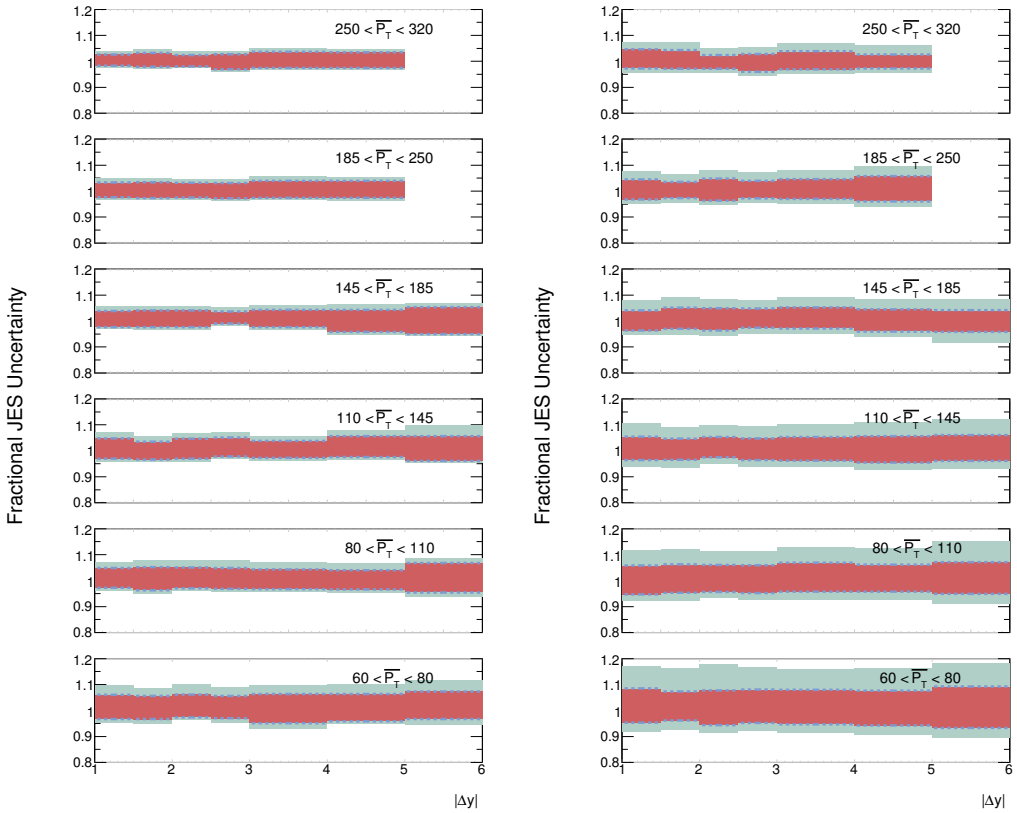


(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.8: The relative jet energy scale systematic uncertainty for the gap fraction as a function of $\overline{p_{\mathrm{T}}}$. The red region represents the uncertainty for data containing 1 vertex events only. Dashed blue line represents the uncertainty for 2010 data without the 1 vertex restriction and is always found at border of red. The pale green region is for the case where the average number of vertices is extrapolated to 7.

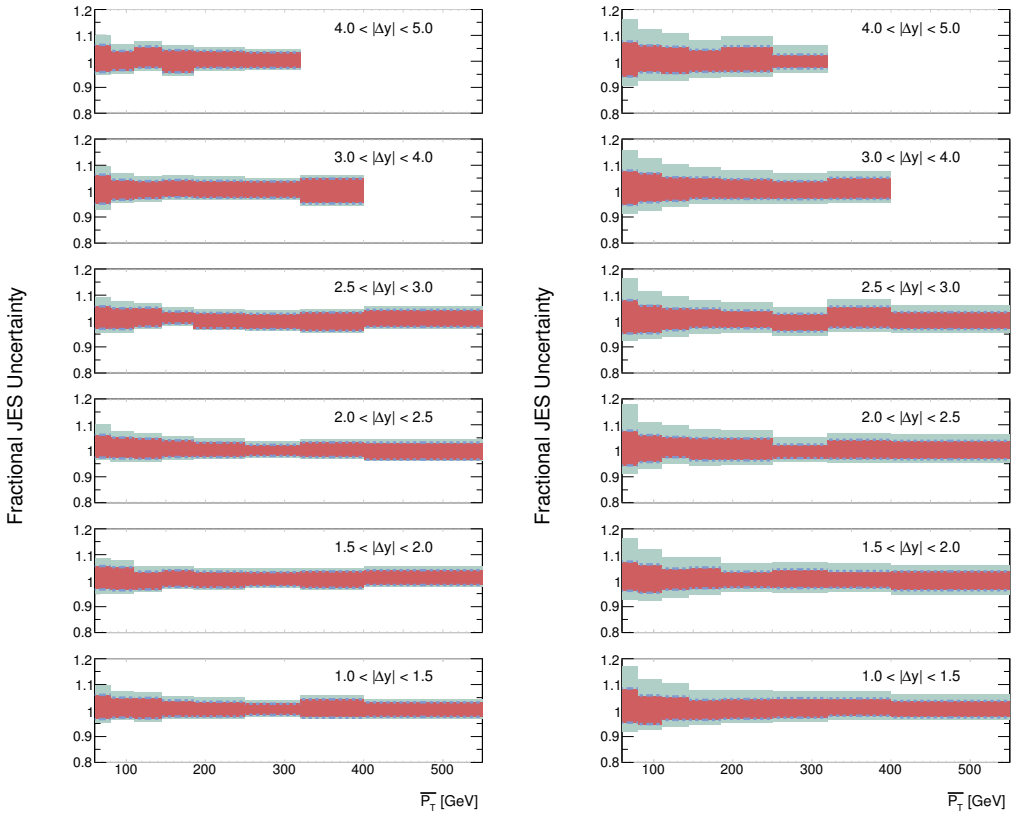
is so small.

The most interesting case is when the effect of the jet energy scale uncertainty is extrapolated to the case where $\overline{N_{vertex}} = 7$, shown as the pale green region in the relative uncertainty figures. There is now a significant rise of uncertainty beyond the zero pile-up case. In figure 7.7 the uncertainty rises above 5% in the anti-k_t $R = 0.6$ jet case and in figure 7.9 and 7.10 there is a significant rise in the low $\overline{p_{\mathrm{T}}}$ region. These effects are seen to a much smaller extent in the anti-k_t $R = 0.4$ jet case, as expected from previous studies on the radius dependence of pile-up in jets shapes.



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.9: The relative jet energy scale systematic uncertainty for the average number of veto jets as a function of $|\Delta y|$. The red region represents the uncertainty for data containing 1 vertex events only. Dashed blue line represents the uncertainty for 2010 data without the 1 vertex restriction and is always found at border of red. The pale green region is for the case where the average number of vertices is extrapolated to 7.



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.10: The relative jet energy scale systematic uncertainty for the average number of veto jets as a function of $\overline{p_{\mathrm{T}}}$. The red region represents the uncertainty for data containing 1 vertex events only. Dashed blue line represents the uncertainty for 2010 data without the 1 vertex restriction and is always found at border of red. The pale green region is for the case where the average number of vertices is extrapolated to 7.

7.5 Jet Radius Comparison

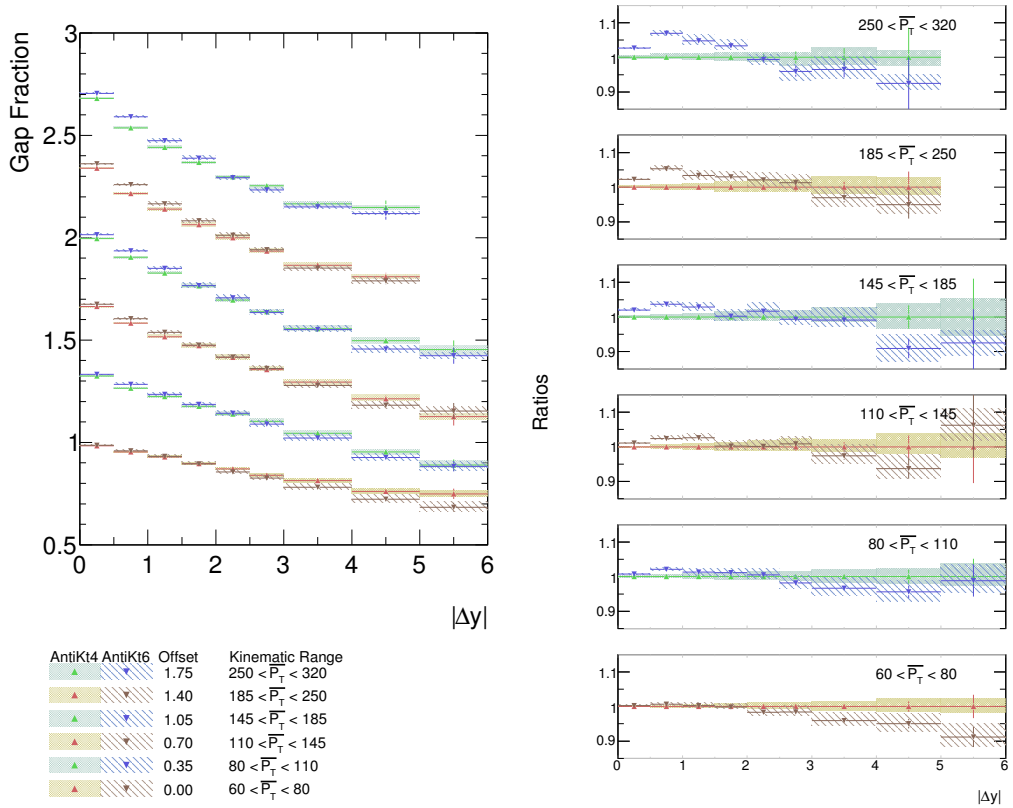


Figure 7.11: The gap fraction as a function of |Δy| is given for both anti-k_t $R = 0.4$ (filled areas) and anti-k_t $R = 0.6$ (diagonal line shaded) jets using ATLAS 2010 data. The filled regions and error bars denote the systematic uncertainty and statistical uncertainty respectively. Each gap fraction curve has been offset by the amount specified in the legend. The plots on the right are the ratios with respect to the anti-k_t $R = 0.4$ data.

With the effect of pile-up in the 2010 data seen to be minimal for each of the jet radii considered and with the systematic uncertainty determined, comparisons to data can be now made. Figures 7.11, 7.12, 7.13 ,7.14 and 7.15 show comparisons of the observables calculated using anti-k_t $R = 0.4$ and anti-k_t $R = 0.6$ jets. Whilst in general there is good agreement between the different choice in jet radii, some regions do show significant differences. In figure 7.11 there is a trend for the gap fraction to be lower at large separations of leading jets for $R = 0.6$.

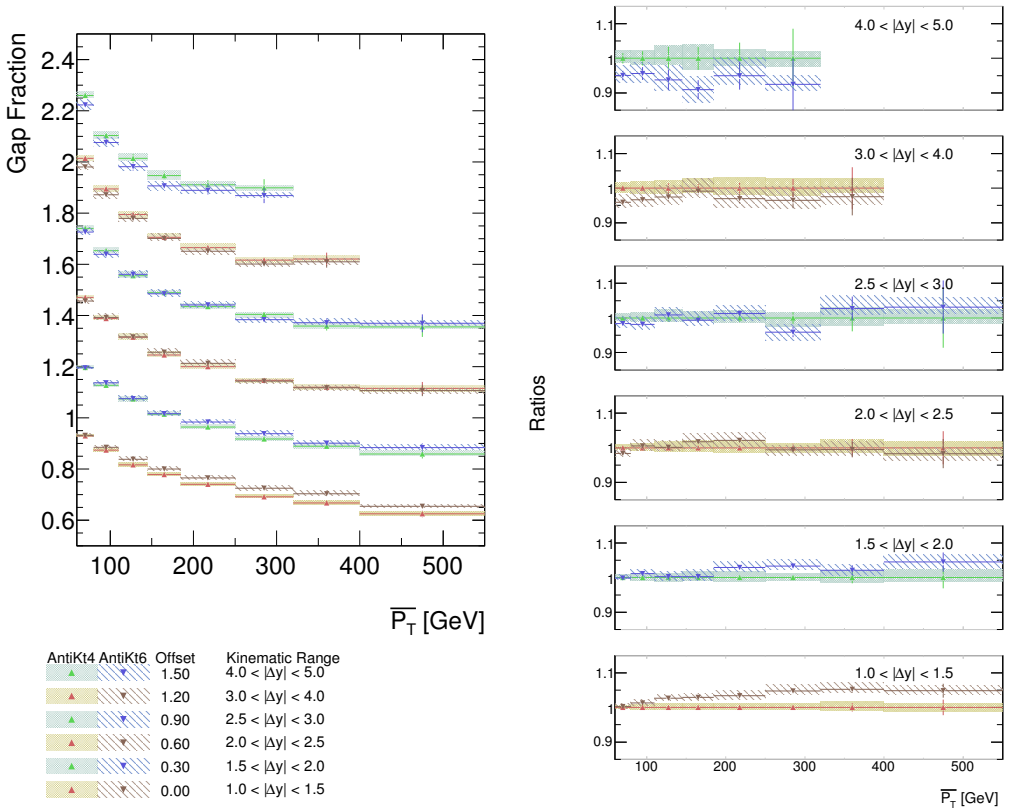


Figure 7.12: The gap fraction as a function of $\overline{p_T}$ is given for both anti-k_t $R = 0.4$ (filled areas) and anti-k_t $R = 0.6$ (diagonal line shaded) jets using ATLAS 2010 data. The filled regions and error bars denote the systematic uncertainty and statistical uncertainty respectively. Each gap fraction curve has been offset by the amount specified in the legend. The plots on the right are the ratios with respect to the anti-k_t $R = 0.4$ data.

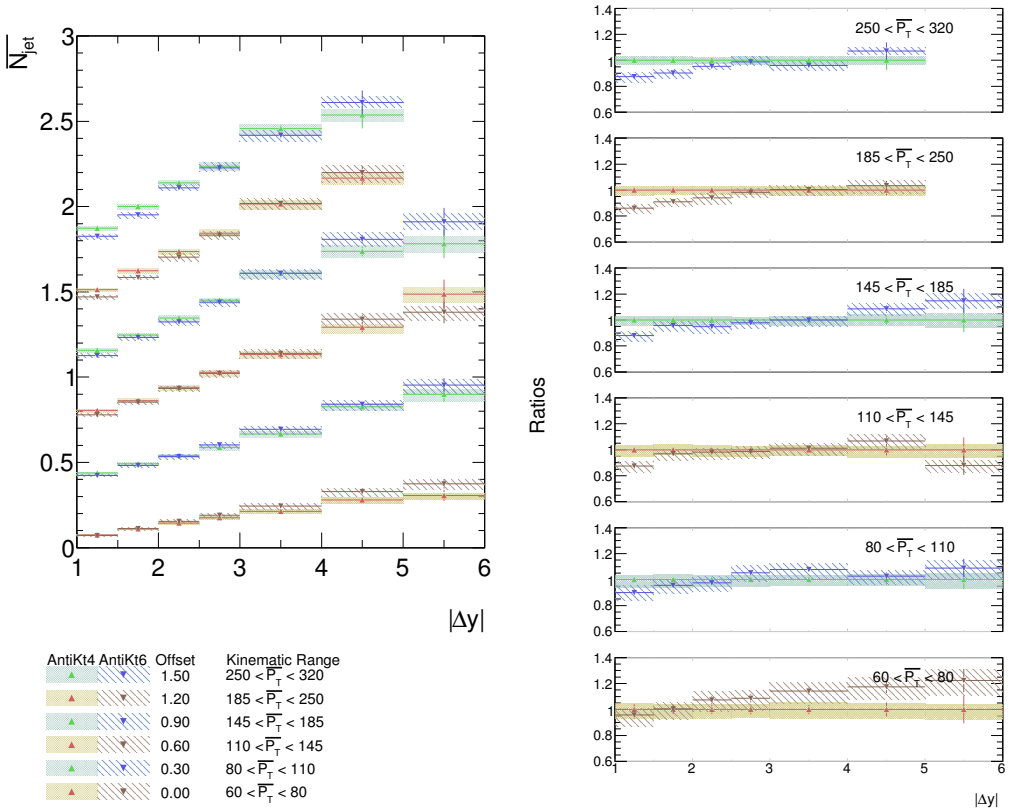


Figure 7.13: The average veto jet number as a function of |Δy| is given for both anti-k_t R = 0.4 (filled areas) and anti-k_t R = 0.6 (diagonal line shaded) jets using ATLAS 2010 data. The filled regions and error bars denote the systematic uncertainty and statistical uncertainty respectively. Each gap fraction curve has been offset by the amount specified in the legend. The plots on the right are the ratios with respect to the anti-k_t R = 0.4 data.

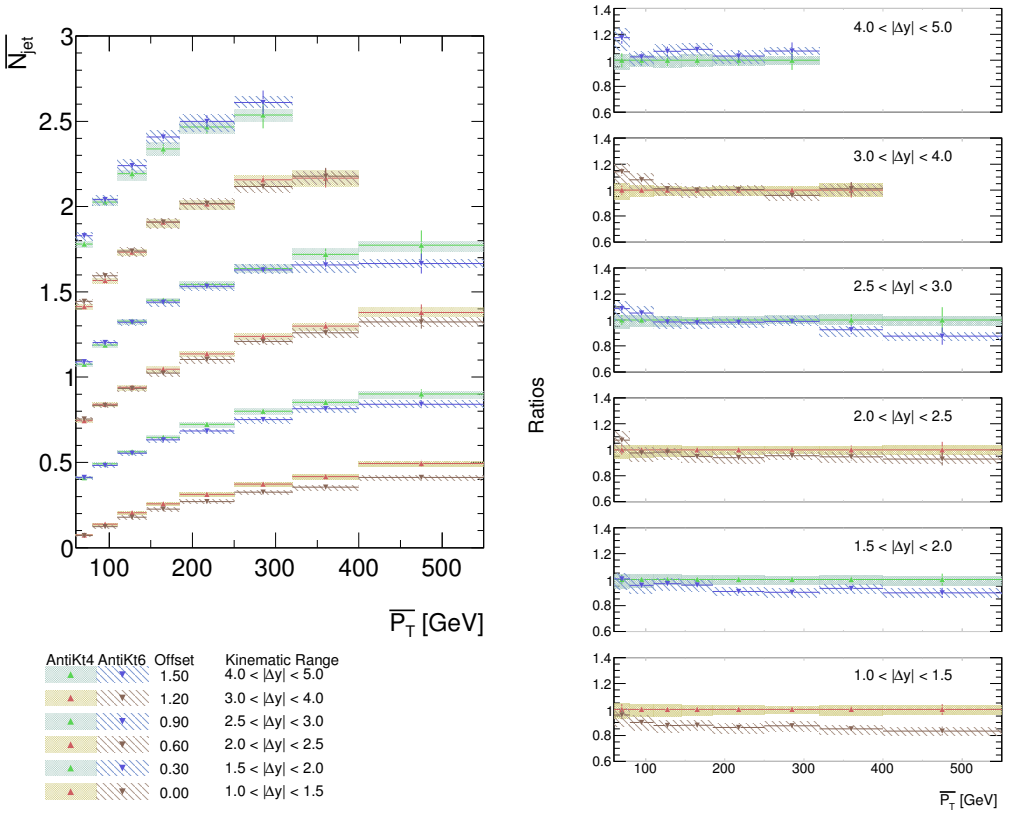


Figure 7.14: The average veto jet number as a function of $\overline{p_T}$ is given for both anti-k_t $R = 0.4$ (filled areas) and anti-k_t $R = 0.6$ (diagonal line shaded) jets using ATLAS 2010 data. The filled regions and error bars denote the systematic uncertainty and statistical uncertainty respectively. Each gap fraction curve has been offset by the amount specified in the legend. The plots on the right are the ratios with respect to the anti-k_t $R = 0.4$ data.

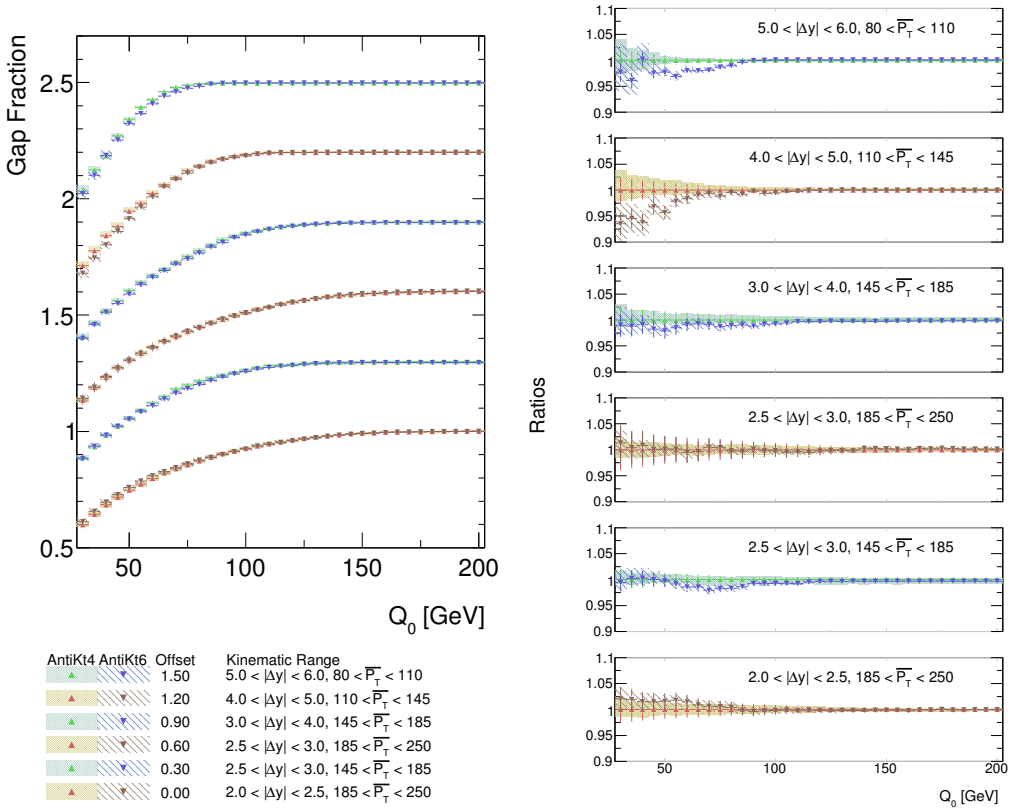
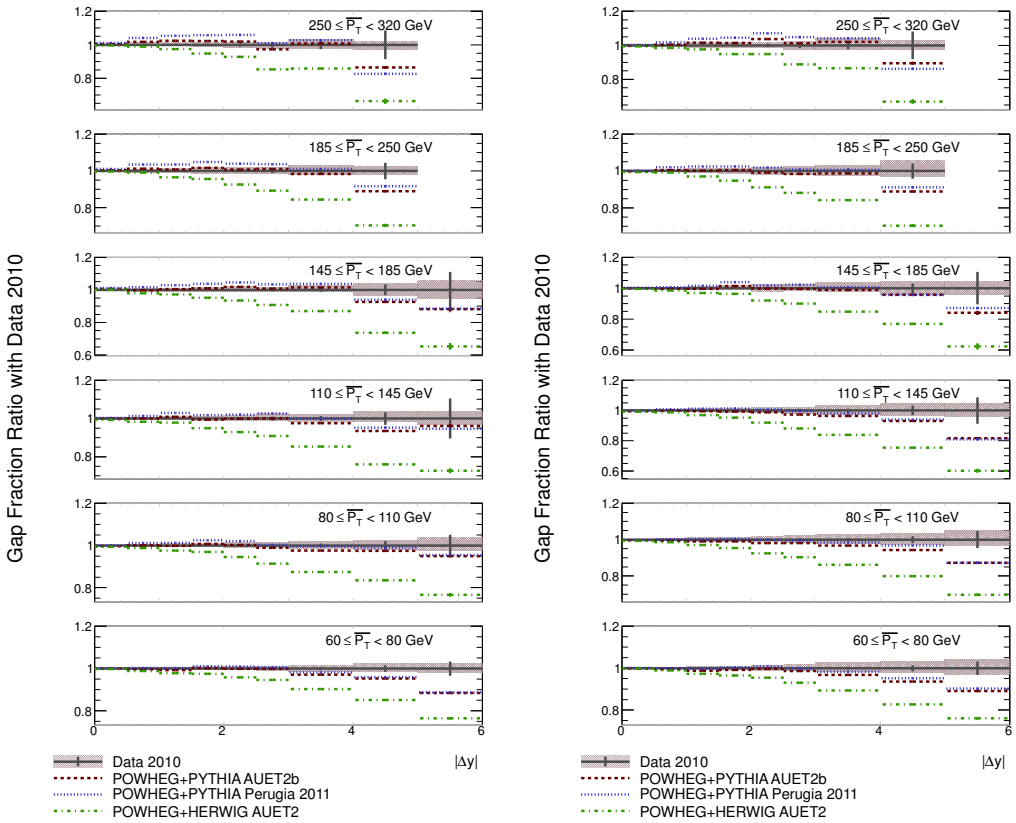


Figure 7.15: The gap fraction as a function of Q_0 is given for both anti-k_t $R = 0.4$ (filled areas) and anti-k_t $R = 0.6$ (diagonal line shaded) jets using ATLAS 2010 data. The filled regions and error bars denote the systematic uncertainty and statistical uncertainty respectively. Each gap fraction curve has been offset by the amount specified in the legend. The plots on the right are the ratios with respect to the anti-k_t $R = 0.4$ data.

7.6 Final Theory Comparisons

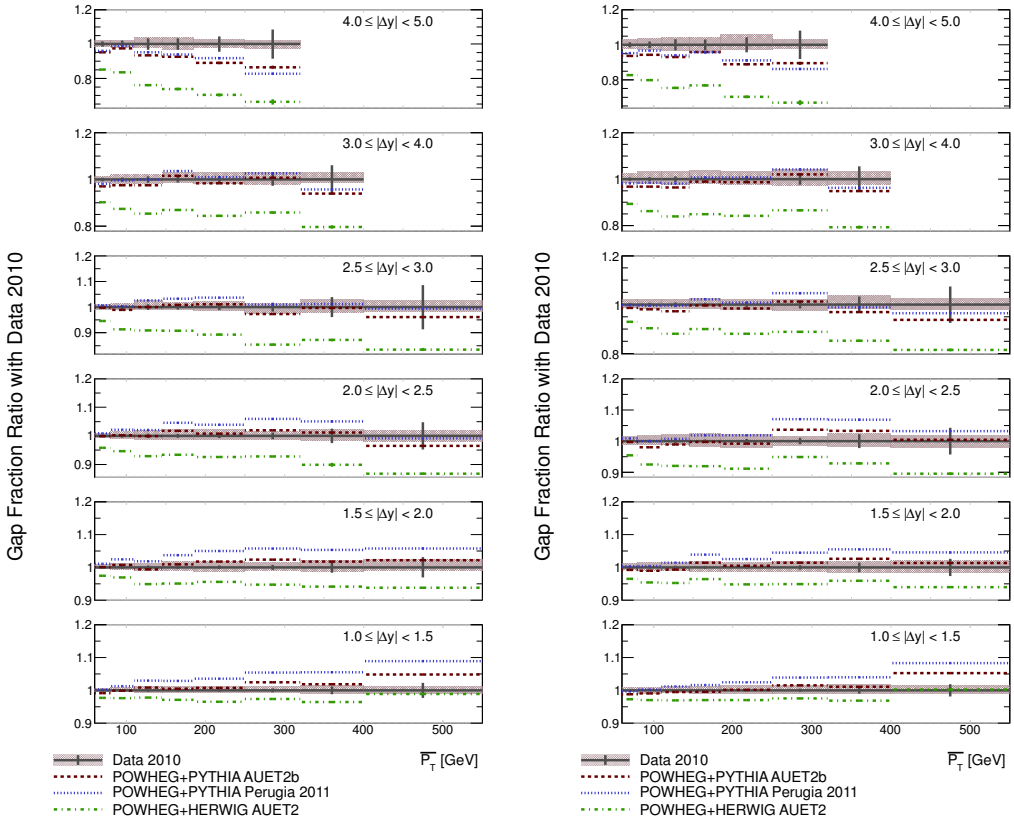


(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.16: The ratio of gap fraction as a function of |Δy|. Comparison is made to the 2010 ATLAS data with the statistical and systematic uncertainties shown separately. The POWHEG predictions are given for three different tunes.

The final results make comparisons of each of the observables against three theoretical predictions. POWHEG predictions are used where the chosen PDF is CT10. Of the three tunes considered, two use FORTRAN PYTHIA and one HERWIG+JIMMY. The PYTHIA tunes chosen were the author tune Perugia 2011 and ATLAS tune AUET2b. The HERWIG+JIMMY tune is restricted to AUET2. The ATLAS tunes were optimized for the ATLAS underlying event measurements made at 7TeV. The Perugia 2011 tune also makes use of the recent 7TeV measurements.

In the gap fraction vs |Δy| ratio plots in figure 7.16 the two PYTHIA tunes are relatively consistent. All the tunes have a trend to lie below the data as the interval size tends to large

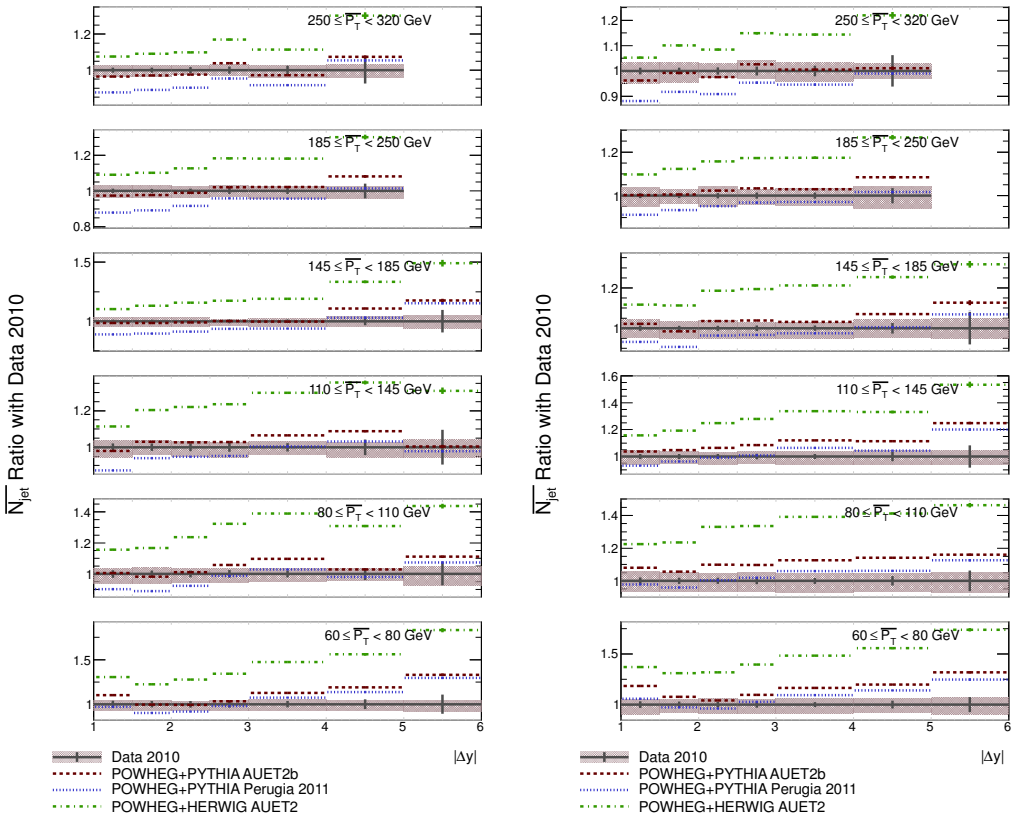


(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.17: The ratio of gap fraction as a function of $\overline{p_T}$. Comparison is made to the 2010 ATLAS data with the statistical and systematic uncertainties shown separately. The POWHEG predictions are given for three different tunes.

values, with the HERWIG+JIMMY tune showing the largest discrepancy with data. There is also a tendency for the PYTHIA tunes to have higher gap fractions than data at small |Δy| and large $\overline{p_T}$. As has been seen previously the behavior of $R = 0.4$ is consistent with $R = 0.6$. In figure 7.17, the gap fraction vs $\overline{p_T}$, it is now possible to see that there is a difference between the two PYTHIA tunes in the low |Δy| region. The Perugia 2011 tune is seen to have a gap fraction around 5% higher.

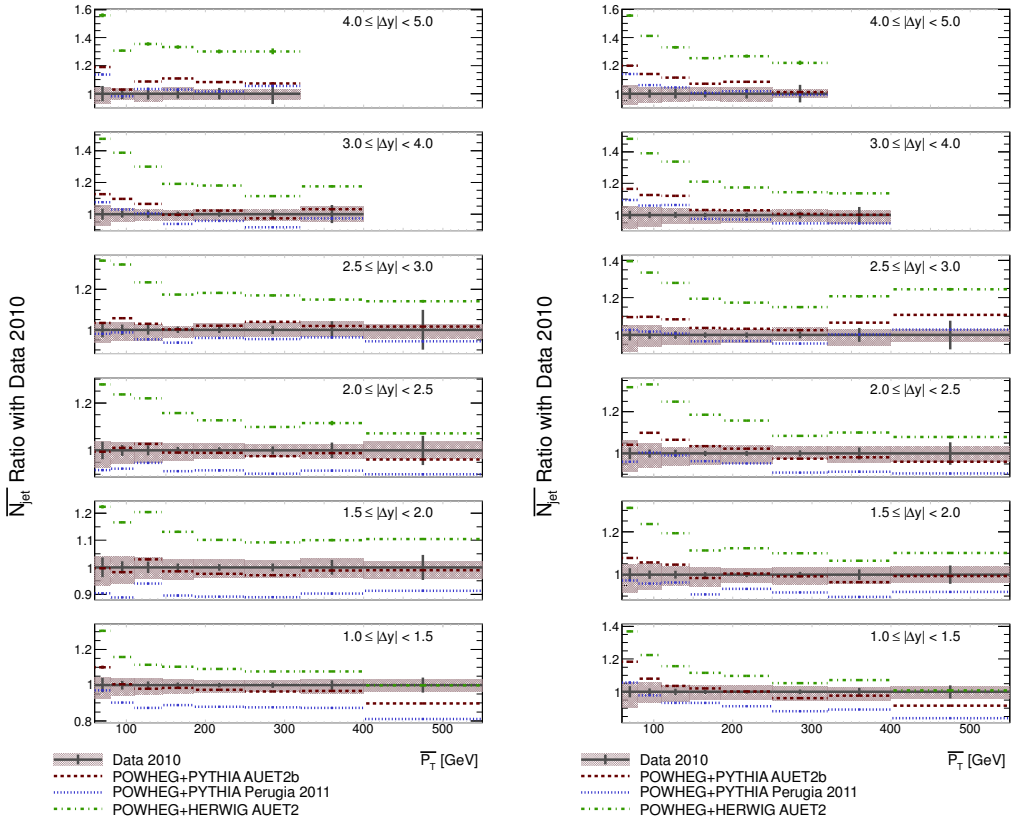
The behavior in the average veto jet number plots as a function of |Δy|, figure 7.18, and $\overline{p_T}$, figure 7.19 show a consistent behavior where the HERWIG+JIMMY tune predicts too many jets above the veto scale in the interval region. The collective behavior of the different tunes shows good agreement of the POWHEG predictions with data, with the exception of the high |Δy| regime.



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

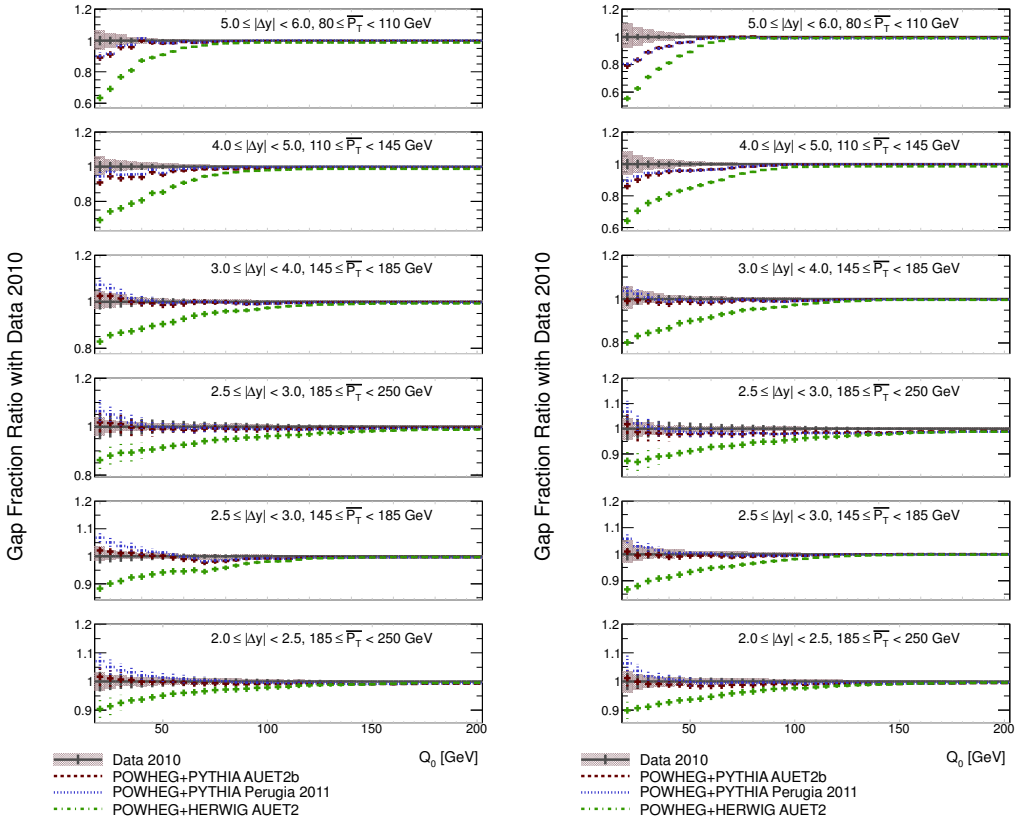
Figure 7.18: The ratio of the average veto jet number as a function of $|\Delta y|$. Comparison is made to the 2010 ATLAS data with the statistical and systematic uncertainties shown separately. The POWHEG predictions are given for three different tunes.

Finally in the gap fraction vs Q_0, figure 7.20, there is still a trend for the HERWIG+JIMMY tune to lie beneath the data even at high Q_0 values. This indicates that the differences are probably not simply due to the model of the underlying event, but rather that the parton shower differences appear to be playing a role in the disagreement. There is also an interesting difference observed in comparing the two ratio plots for the kinematic region $5.0 < |\Delta y| < 6.0$ and $80 < \overline{p_T} < 110$, where the $R = 0.6$ jets gap fraction lies below data for all the tunes whilst $R = 0.4$ jets show good agreement for the PYTHIA tunes.



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.19: The ratio of the average veto jet number as a function of $\overline{p_{\mathrm{T}}}$. Comparison is made to the 2010 ATLAS data with the statistical and systematic uncertainties shown separately. The POWHEG predictions are given for three different tunes.



(a) anti-k_t $R = 0.4$ (b) anti-k_t $R = 0.6$

Figure 7.20: The ratio of gap fraction as a function of Q_0. Comparison is made to the 2010 ATLAS data with the statistical and systematic uncertainties shown separately. The POWHEG predictions are given for three different tunes.

Bibliography

[1] Performance of the minimum bias trigger in p-p collisions at $\sqrt{s} = 7$ TeV. Technical Report ATLAS-CONF-2010-068, CERN, Geneva, Jul 2010.

Chapter 8

Summary and Conclusions

The work presented in this thesis documents a major contribution to two of the early ATLAS jet measurements:'Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ using the ATLAS detector' [1] and 'Measurement of inclusive jet and dijet cross sections in proton-proton collision data at 7 TeV centre-of-mass energy using the ATLAS detector' [2]. An extension to the inclusive jet and dijet cross section analysis is currently in collaboration review.

In this thesis next-to-leading order QCD predictions for a number of analyses have been made using the POWHEG-BOX dijet process. In doing so, a number of interesting features have arisen. The choice of scale has been seen to have a dramatic impact on the stability of the predictions for the inclusive dijet mass, while having almost no effect on predictions related to the dijet with central jet veto process. The use of scale variation was also found to be unable to properly account for the missing higher order terms in the prediction. It was instead necessary to rely upon the variation amongst the different shower Monte Carlo generators to give an estimate of the uncertainty.

Data corrections have been derived for the dijet with central jet veto analysis, with conservative estimates for the systematic uncertainty. In each of the analysis selections considered the size of the corrections were small. A large reduction in the systematic uncertainty was achieved through the development of new weighting strategies for Monte Carlo events. This

allowed the extension of the measurements to large values of $|\Delta y|$ with low total systematics. These samples have also been found to be useful in other analyses in ATLAS.

The theoretical predictions found to be in best agreement with the dijet with central jet veto measurement were the POWHEG+PYTHIA predictions using either of the ATLAS MC tunes, AMBT1 or AUET2b. Other theoretical predictions showed varying levels of agreement, most notably HEJ performed well in the limit of large $|\Delta y|$ and small $\overline{p_T}$. An extension to the analysis found that the measurement could be improved through the use of anti-k_t jets of radius $R = 0.4$, where not only were the systematic uncertainties found to be reduced but the sensitivity to the presence of pile-up was reduced. This allowed the loosening of the vertex requirement and subsequently increased the number of events by a factor of between two and three.

Overall there was observed to be a large spread of MC predictions relative to a small uncertainty in the measurement. This has already led to a response from the theoretical community [3], where it is stated

> "The message is clear: the accuracy of the ATLAS data already demands better theoretical calculations."

Bibliography

[1] Georges Aad et al. Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 2011:1–36, 2011. 10.1007/JHEP09(2011)053.

[2] Measurement of inclusive jet and dijet cross sections in proton-proton collision data at 7 TeV centre-of-mass energy using the ATLAS detector. Technical Report ATLAS-CONF-2011-047, CERN, Geneva, Mar 2011.

[3] Rosa Maria Duran Delgado, Jeffrey R. Forshaw, Simone Marzani, and Michael H. Seymour. The dijet cross section with a jet veto. *JHEP*, 08:157, 2011.

Appendix A

Software

In this appendix a short description of some of the software development work carried out so that data may be analysed for the research described in this thesis. The process of distilling the ATLAS data is outlined and some of the benefits of the method chosen are detailed. A short comparison is made to one of the standard ATLAS methods of data analysis.

A.1 Ntuple making

During LHC operation the ATLAS detector is able to collect an enormous amount of data, which originates from both the high collision rate and fine granularity of the detector for many physics objects. From the initial signal to the final analysis ready objects, the data passes through a number of different formats. It starts as a byte-stream from the detectors themselves and is processed into a format known as a Raw Data Object (RDO). The RDO objects need to have the offline algorithms applied to produce calibrated physics objects. The secondary stage is the Event Summary Data (ESD), which provides a sufficient amount of information to re-run reconstruction algorithms. The final official format is the Analysis Object Data (AOD), which has reduced sufficiently the size of the files to be practical for analyses. While the volume of data is reduced tremendously it is still too copious in the AOD form for it to be analysed locally. It is instead the preferred mechanism to access

the data through the world wide computing network commonly referred to as the Grid. As with the case of moving from RDO to AOD, further refinement of the object selection can be made to reduce the size of the dataset necessary to run over. These derived datasets are typically referred to as ntuples. Where the derivation can involve removing entire collections of objects, a reduction of the information stored in each object or filtering out entire events that do not match required criteria. To test out new ideas for ntuple creation for analysis the Python programming language [1] and standard library was used for its ease of use and its growing popularity in experimental particle physics.

A.1.1 Pickling

The key element upon which this ntuple scheme is dependent upon is the process of pickling in the Python programming language. Pickling is Pythons standard mechanism for persisting an objects state such that it may be saved to and retrieved from disk. The reason why this is so attractive from the point of view of analysis is that the objects which are saved are arbitrary Python objects. With very little difficulty a programs state can be frozen and re-used at a later time, in a completely different computing environment.

There are some limitations to what pickle objects can store. They are not able to store file objects, network connections and alike that are dependent on transient behavior. However these are uncommon elements to find in a particle physics ntuple. The function definitions of objects are also not stored, only the class names are stored. This requires the class definition to be available when unpickling and gives the flexibility of changing or adding function behavior.

Utilizing pickle objects on their own to store physics objects is not viable due to the fact that Python loads the whole of the pickle into memory at once. This has performance implications when the amount of data exceeds even just a few hundred megabytes in size, a typical reality of ntuple based analysis today.

A.1.2 ZODB

Alternatively an object database can be used, maintaining the flexibility and ease of storing python objects via the pickle mechanism but with the practicality of handling large amounts of data. The object database for ZOPE (ZODB) [2] was chosen. It allows arbitrary python objects to be stored in a database. The database keeps tracks of Persistent objects and will update the database if objects are changed in memory. Through a transaction mechanism it is also able to create a history of the file, allowing the possibility to revert the database to an earlier version. The main advantage over simply using pickle objects in storing data is that the database only loads objects into memory when necessary. This reduces the memory usage when considering a large number of events.

Code listing A.1 shows the relatively small amount of work necessary to prepare a database for reading and writing.

```
from ZODB import DB
from ZODB.FileStorage import FileStorage
from ZODB.PersistentMapping import PersistentMapping
from ZODB.PersistentList import PersistentList
import zc.zlibstorage

storage = zc.zlibstorage.ZlibStorage(FileStorage(''myDatabase.fs''))
db = DB(storage)
connection=db.open()
root=connection.root()

#Commit changes
import transaction
transaction.commit()
```

Listing A.1: Example code to open a database for reading and writing

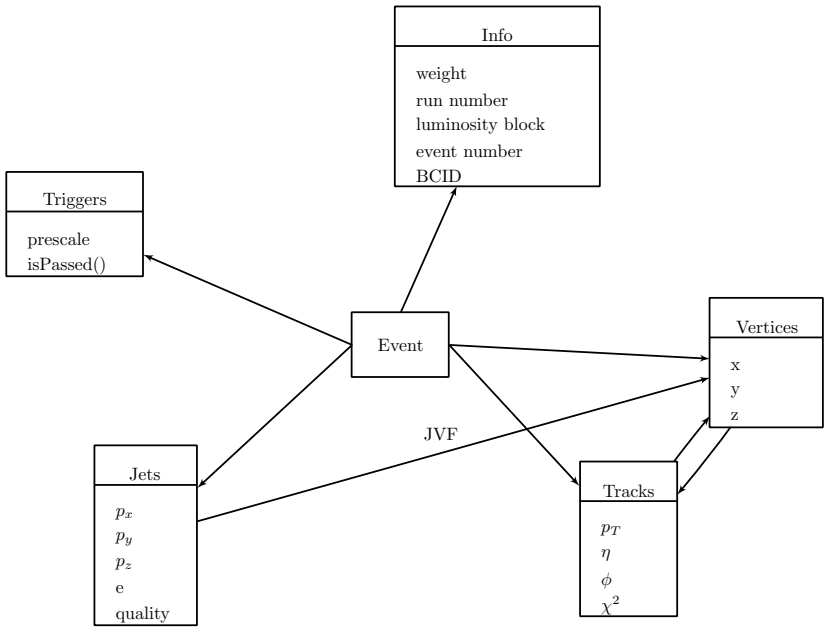


Figure A.1: Example contents of a standard ntuple with some of the more important contents and relationships indicated.

A.1.3 A Generic Event

As stated earlier an advantage of using the ZODB to store event objects is that relationships are very easily maintained with little effort. In figure A.1 you can see a selection of the important objects that make up an event object. You should note that of the basic objects being stored jets, vertices and tracks that there are references to each other. For example jets have a quantity known as the jet vertex fraction, which describes the fraction of overlapping tracks from each vertex. This relationship is useful to maintain as it allows some simple corrections for the effect of pile-up, which has been seen to be very important for typical hadron collider environments.

The event objects are stored within a container known as a B+-tree [3], which has a number of useful features. Access and insertion times have the behavior of $O(\log_b n)$, where b is the branching factor. The 'Info' object contained in the event is used as the key in the BTree, where the ordering of events inside the tree is decided by the event number, luminosity block and run number. With the implementation of B+-tree in ZODB it is possible to specify ranges of events by these properties and retrieve only those from the database.

Performing a range query, such that k-events of a given period are selected, is $O(\log_b n + k)$.

A.2 Skimming and Slimming

While exposing physics objects for use in Python based analysis is of primary import. It is also useful to consider a further reduction in dataset volume, such that analysis can be run more often and quite possibly locally. Once the objects are contained within ZODB format it becomes very simple to modify them. Thus events, physics objects and individual object properties can be removed or altered to reduce the overall disk size. This can all be achieved by the use of the *del* python statement. The code in listing A.2 shows how simple it is to apply the *del* statement to objects at various hierarchies within the ntuple.

```
#Remove first event
del root['Events'].values()[0]

#Remove jets from an event
myEvent = root['Events'].values()[0]
del myEvent.jets

#Remove quality information from only the first jet
myEvent = root['Events'].values()[1]
del myEvent.jets[0].quality

#Commit changes
import transaction
transaction.commit()
```

Listing A.2: Example code to remove objects from events in a ZODB database

If the *del* statement is applied to references it will not automatically remove the object from the database, it is only when the number of references to the object falls to zero that it will no longer be saved. This behavior is very similar to how the memory management works

within Python.

A.3 Performance

In addition to testing the flexibility of various analysis methods it is also illustrative to compare the speed in which certain tasks complete. The tools tested here are ZODB and the AOD format. In figure A.2 the size of the input file grows as the number of events increases. The AOD files size is 5GB for 5000 events. Typically the size of the compressed ZODB file is half the size of the normal ZODB file. The impact of the slimming and skimming is quite visible in the comparison between AOD and ZODB file size, where the size of the AOD files severely restricts the total number of events which can be run over.

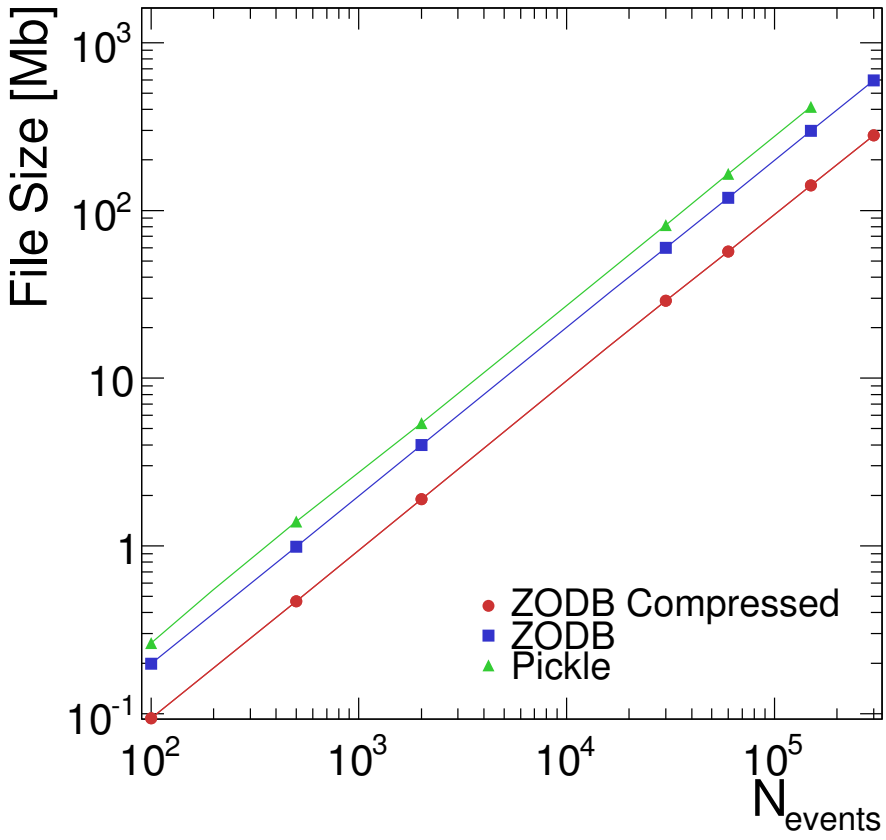


Figure A.2: Comparison of Ntuple size in a a number of different forms.

In figure A.3, slowest by a significant margin is the AOD file. The reason for this is the

long period of Athena software initialization, this can be seen in the significant pedestal seen in figure A.3. Given that AOD file size is so large, it is likely that many simple analyses would be bound by initialization time.

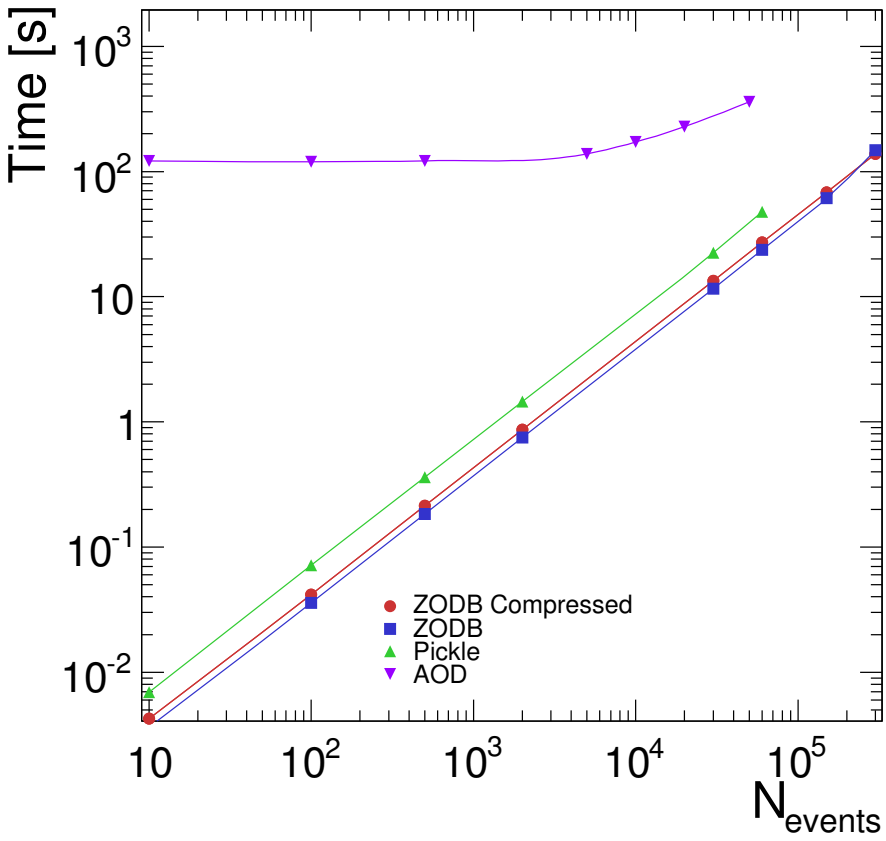


Figure A.3: Comparison of the time to analyse different ntuples.

Although python is in some cases many magnitudes slower in general when compared to other ntuple strategies available, the flexibility provided more than makes up for this. In the future the discrepancy between Python and C++ should decrease. With the advent of such Python implementations as PyPy the performance should be eventually become comparable to that of C++. The JiT compilation seen in PyPy is able to analyse code and build a more efficient program. Perhaps in some cases the performance exceeds C++, mainly due to some of the more arduous compilation tools that C++ must use.

Bibliography

[1] Guido van Rossum. Python tutorial, May 1995. CWI Report CS-R9526.

[2] Zodb - a native object database for python. www.zodb.org.

[3] Douglas Comer. The ubiquitous b-tree. *ACM Computing Surveys*, 11:121–137, 1979.

Appendix B

Monte Carlo Weighting Parameters

The parametrisation of the p_T-$|\Delta y|$ distributions used in weighting events for MC event generation are shown in the following tables. Each table contains all the parameters necessary to describe the distribution for each PYTHIA jet sample, where each PYTHIA sample covers a different range of parton transverse momentum. Within each table the models are sorted by the size of the weight parameter w. For completeness all of the samples generated are included in this section, however the tables highlighted with red rows were not used in the rest of this thesis.

w	μ_{p_T} [GeV]	$\mu_{\|\Delta y\|}$	σ_{p_T} [GeV]	$\sigma_{\|\Delta y\|}$	ρ
0.29	14.91	1.23	1.44	0.81	$-4.48 \cdot 10^{-2}$
0.25	17.74	1.37	2.43	0.86	-0.21
0.21	15.67	3.81	1.95	1.38	$4.12 \cdot 10^{-2}$
0.18	20.28	2.18	3.09	1.22	-0.19
$6.96 \cdot 10^{-2}$	24.21	2.83	4.62	1.68	-0.27

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $8 < p_T < 17$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.3	21.61	1.26	3.87	0.82	$1.7 \cdot 10^{-2}$
0.24	19.99	2.83	3.37	1.6	0.15
0.23	25.53	1.59	5.17	1.05	$4.47 \cdot 10^{-2}$
0.17	30.33	1.83	5.64	1.21	$-9.83 \cdot 10^{-2}$
$6.23 \cdot 10^{-2}$	36.64	2.72	8.46	1.68	-0.3

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $17 < p_T < 35$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.3	41.19	1.19	6.84	0.77	$-2.38 \cdot 10^{-2}$
0.25	46.83	1.3	9.84	0.85	$-5.2 \cdot 10^{-3}$
0.22	39.41	2.77	7.04	1.41	$8.37 \cdot 10^{-2}$
0.17	56.42	1.72	10.53	1.1	-0.11
$6.24 \cdot 10^{-2}$	67.6	2.32	15.34	1.43	-0.23

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $35 < p_T < 70$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.29	81.1	1.09	11.01	0.71	$-3.17 \cdot 10^{-2}$
0.25	89.62	1.16	18.22	0.76	$-1.79 \cdot 10^{-3}$
0.22	78.43	2.46	11.9	1.24	$5.49 \cdot 10^{-2}$
0.18	107.87	1.5	19.28	0.97	$-8.9 \cdot 10^{-2}$
$6.42 \cdot 10^{-2}$	127.22	1.97	27.71	1.24	-0.2

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $70 < p_T < 140$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.28	159.28	1.03	16.86	0.68	$-2.95 \cdot 10^{-2}$
0.26	171.87	1.01	32.15	0.67	$-1.25 \cdot 10^{-2}$
0.2	155.65	2.18	19.8	1.1	$4.13 \cdot 10^{-2}$
0.19	201.75	1.38	36.82	0.91	$-1.43 \cdot 10^{-2}$
$7.05 \cdot 10^{-2}$	240.84	1.37	47.29	0.93	$-7.28 \cdot 10^{-2}$

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $140 < p_T < 280$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.28	311.23	1	25.03	0.66	$-1.7 \cdot 10^{-2}$
0.27	338.99	0.87	50.49	0.58	$2.94 \cdot 10^{-2}$
0.2	374.4	1.21	76.65	0.81	$5.88 \cdot 10^{-2}$
0.18	308.46	1.85	32.57	0.98	$4.39 \cdot 10^{-2}$
$6.56 \cdot 10^{-2}$	458.79	0.93	78.89	0.64	$3.98 \cdot 10^{-3}$

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $280 < p_T < 560$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.36	602.53	1.13	33.86	0.75	$-6.97 \cdot 10^{-3}$
0.28	650.62	0.84	62.72	0.58	0.15
0.19	731.82	0.89	96.2	0.62	0.17
0.11	598.24	0.91	106.45	0.63	$-8.05 \cdot 10^{-2}$
$6.28 \cdot 10^{-2}$	876.31	0.64	130.17	0.45	0.14

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $560 < p_T < 1120$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.38	1,173.97	0.82	40.7	0.54	$-1.64 \cdot 10^{-2}$
0.3	1,235.66	0.69	77.86	0.47	0.17
0.22	1,337.2	0.6	115.55	0.42	0.2
$6.73 \cdot 10^{-2}$	1,496.69	0.41	192.12	0.29	0.14
$4.14 \cdot 10^{-2}$	1,102.68	0.72	219.68	0.51	−0.11

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying $1120 < p_T < 2240$ GeV.

w	μ_{p_T} [GeV]	$\mu_{\lvert\Delta y\rvert}$	σ_{p_T} [GeV]	$\sigma_{\lvert\Delta y\rvert}$	ρ
0.37	2,279.93	0.43	27.16	0.29	$-2.16 \cdot 10^{-3}$
0.29	2,326.14	0.37	51.46	0.26	0.16
0.23	2,389.56	0.33	77.74	0.23	0.21
$9.38 \cdot 10^{-2}$	2,478.34	0.22	120.06	0.16	0.13
$1.24 \cdot 10^{-2}$	2,052.39	0.45	306.07	0.35	−0.21

Parameters for the mixture model composed of a set of five bivariate distributions. Derived from the standard PYTHIA AMBT1 dijet events with the leading parton satisfying 2240 GeV $< p_T$.

Appendix C

Detector Correction Additional Material

Here the distributions are presented for the alternative selections.

C.1 Purities

The purities for selection B as shown in figures C.1-C.3 have the same shape as the versions for selection A but overall are seen to have a slightly lower purity. These also represent the purities for selection C.

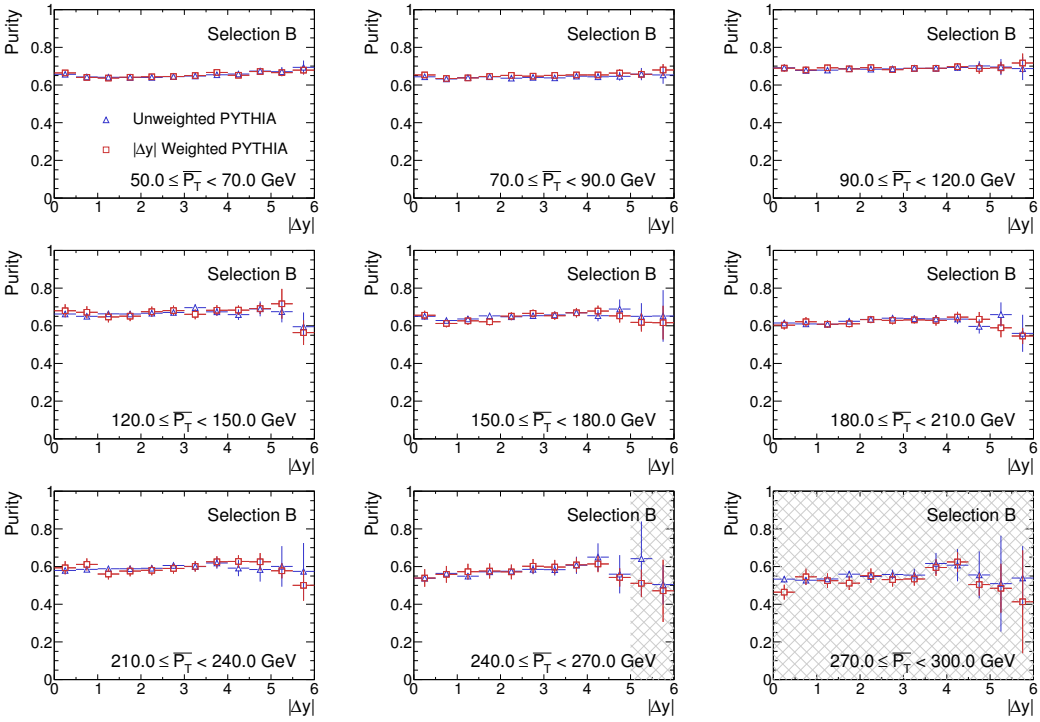


Figure C.1: Purities selection B for the variable |Δy| where the monte carlo events used are PYTHIA with the AMBT1 tune. The regions of the graphs with a filled background are not needed to unfold data.

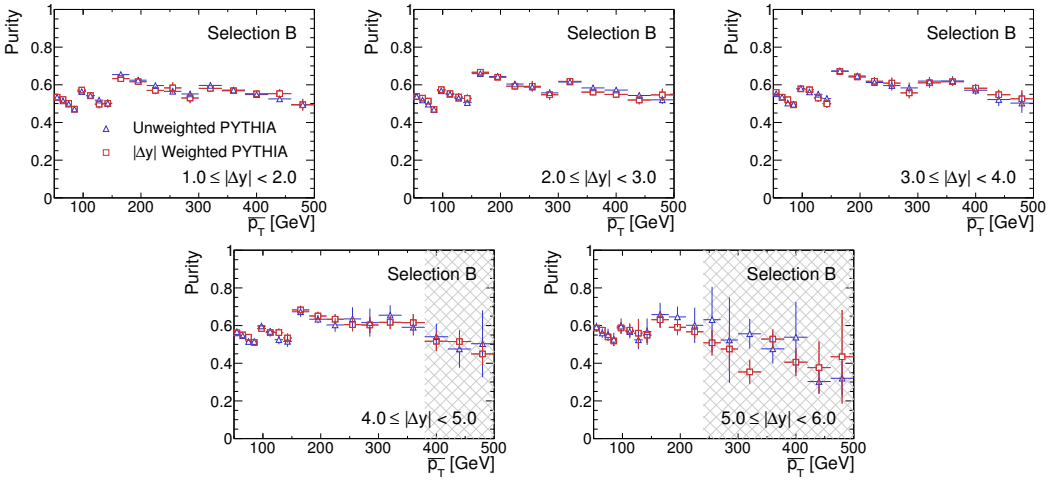


Figure C.2: Purities selection B for the variable $\overline{p_{\mathrm{T}}}$ where the monte carlo events used are PYTHIA with the AMBT1 tune. The regions of the graphs with a filled background are not needed to unfold data.

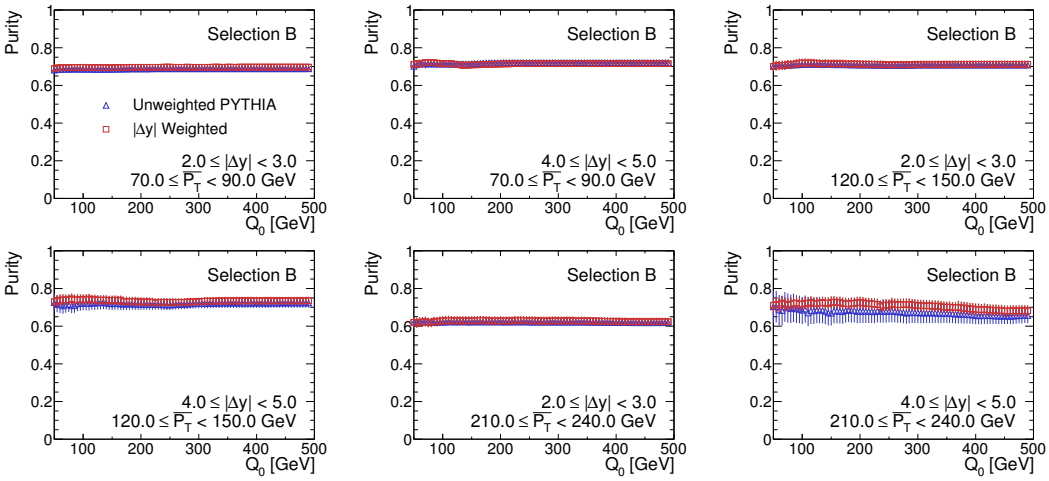


Figure C.3: Purities selection B for the variable Q_0 where the monte carlo events used are PYTHIA with the AMBT1 tune.The regions of the graphs with a filled background are not needed to unfold data.

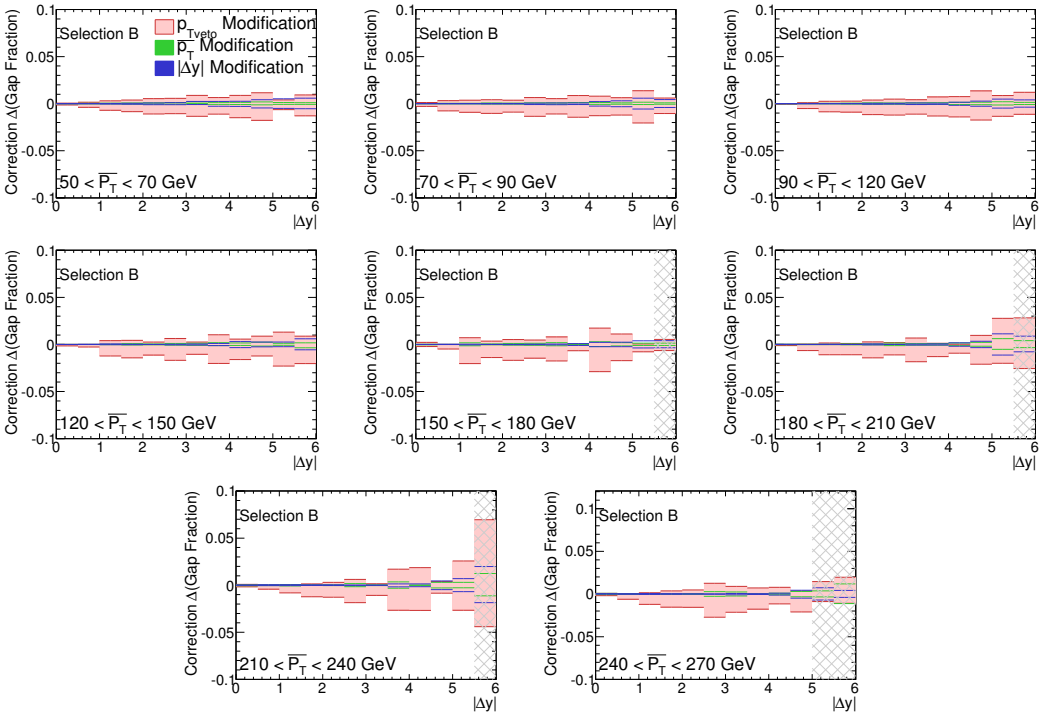


Figure C.4: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection B is used.

C.1.1 Varying Truth Distributions

Selection B

The results of applying the truth modifications for the derivation of the correction factors is shown in figures C.4-C.8 for selection B.

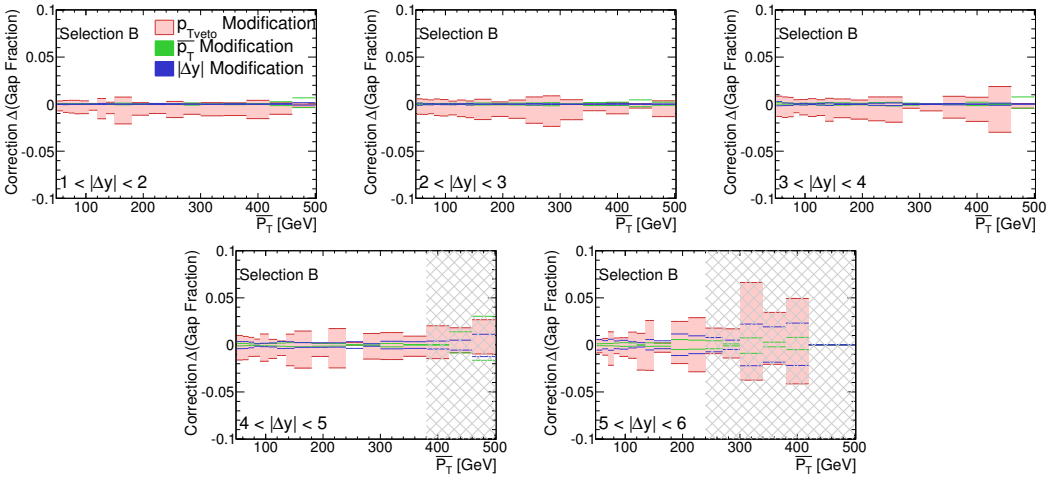


Figure C.5: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_T}$ distribution. Selection B is used.

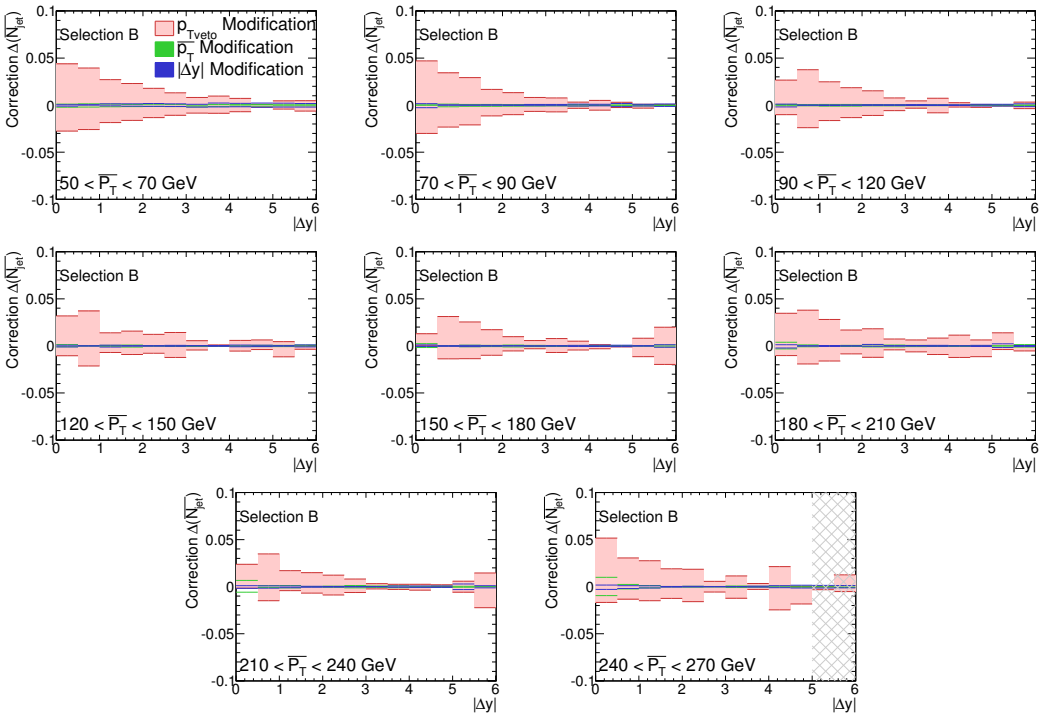


Figure C.6: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection B is used.

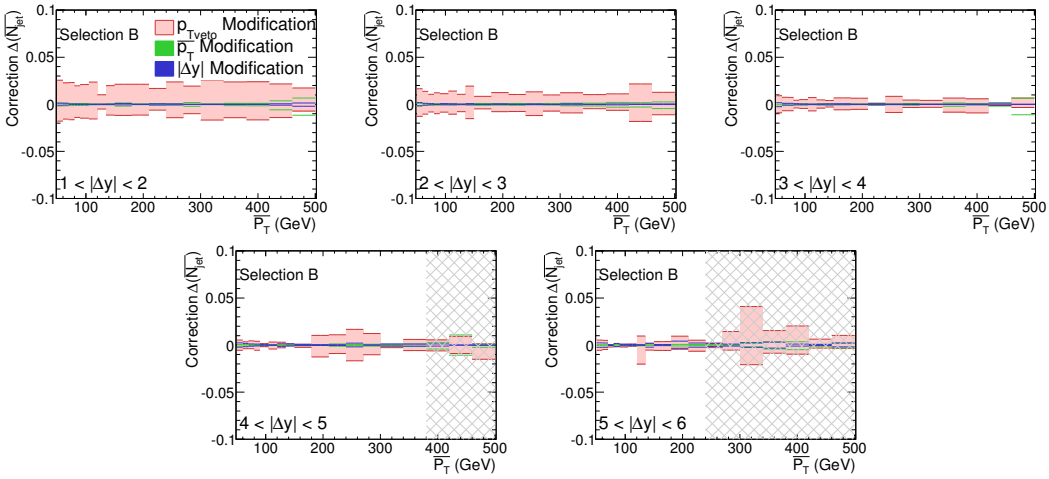


Figure C.7: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_T}$ distribution. Selection B is used.

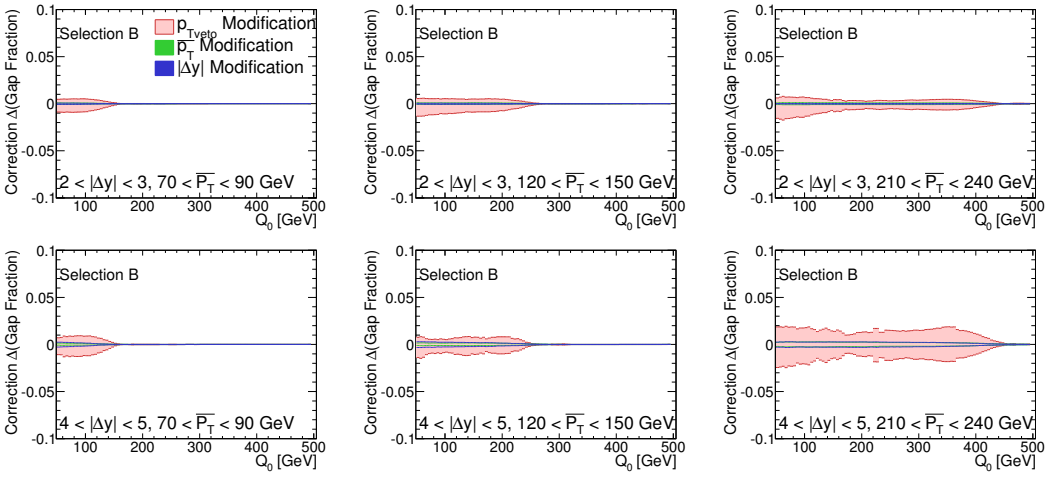


Figure C.8: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection B is used.

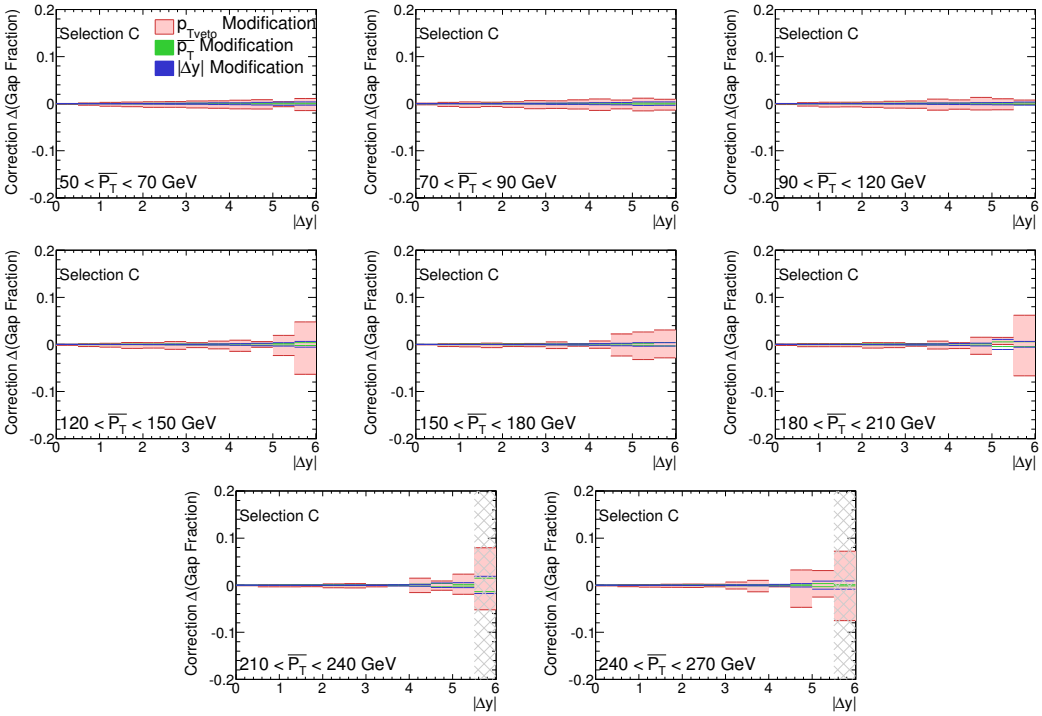


Figure C.9: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection C is used.

Selection C

The results of applying the truth modifications for the derivation of the correction factors is shown in figures C.9-C.13 for selection C.

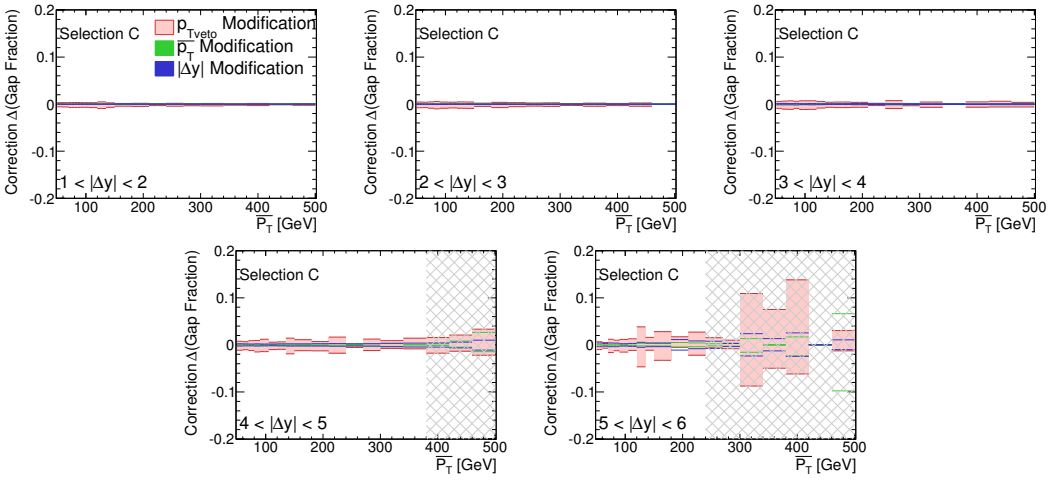


Figure C.10: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_T}$ distribution. Selection C is used.

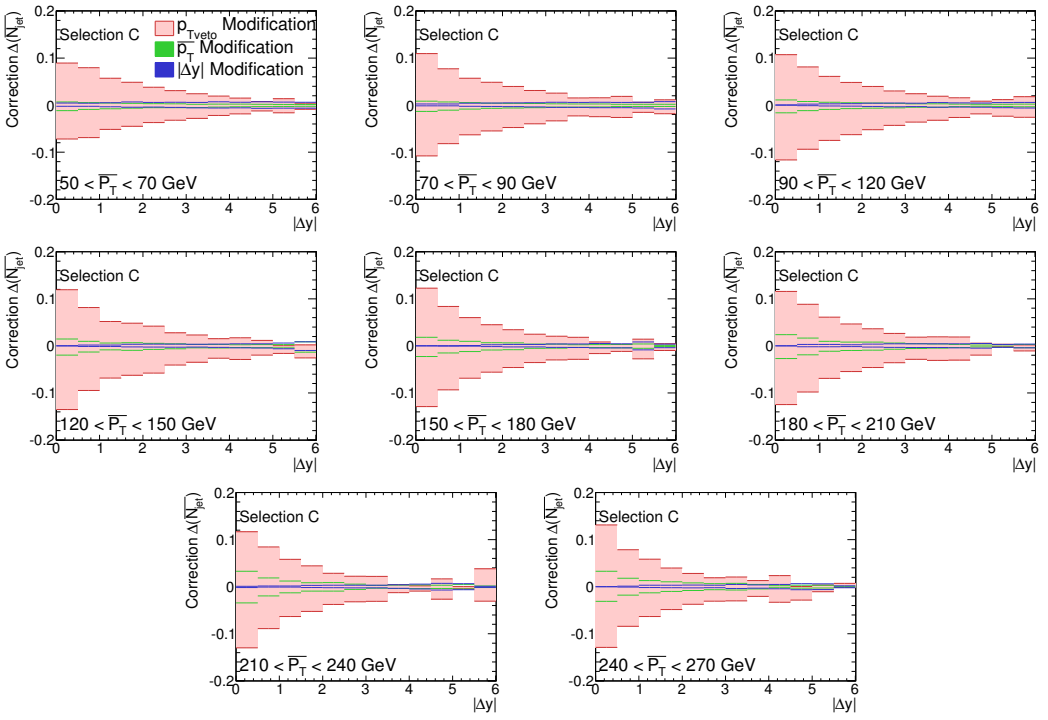


Figure C.11: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection C is used.

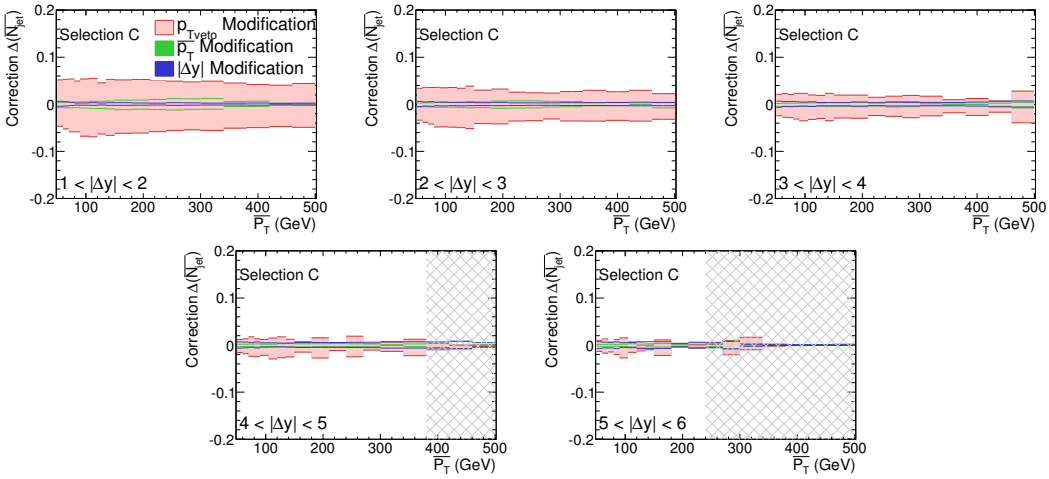


Figure C.12: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_T}$ distribution. Selection C is used.

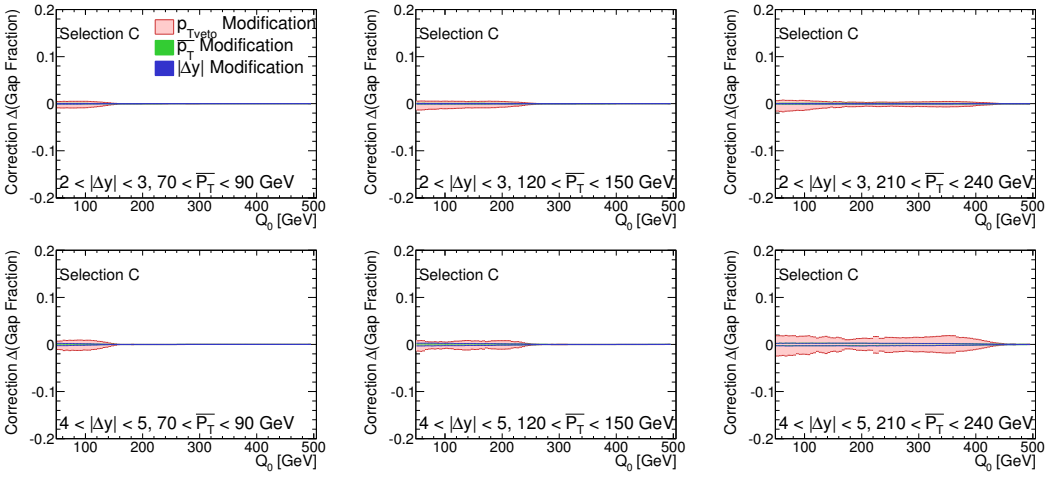


Figure C.13: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection C is used.

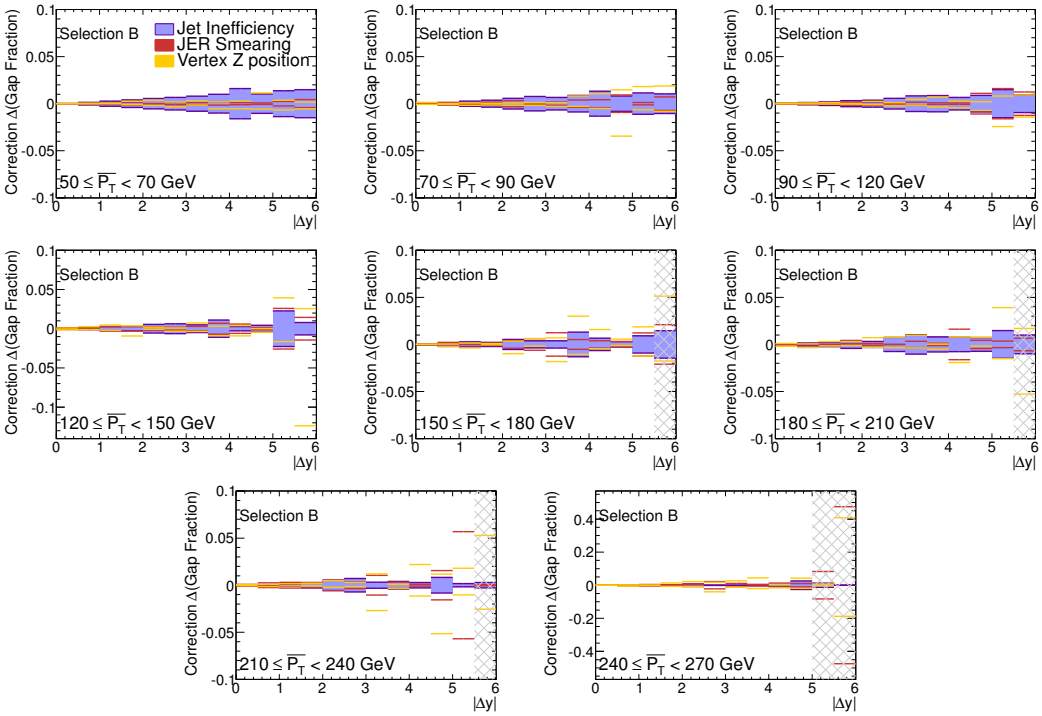


Figure C.14: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection B is used.

C.1.2 Other Systematic Sources

Selection B

The correction uncertainties derived from jet energy resolution, additional jet inefficiency and vertex z position variation are shown in figures C.14-C.18 for selection B.

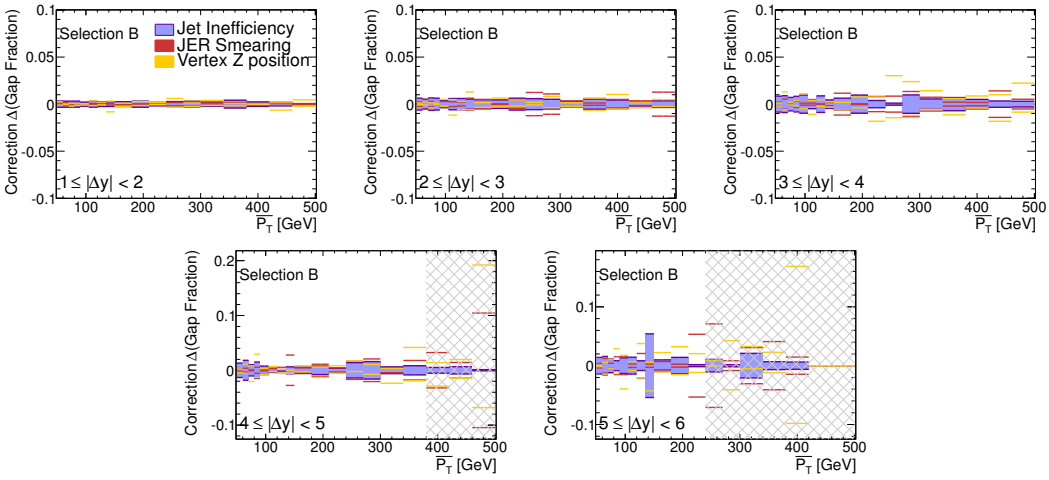


Figure C.15: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_{\mathrm{T}}}$ distribution. Selection B is used.

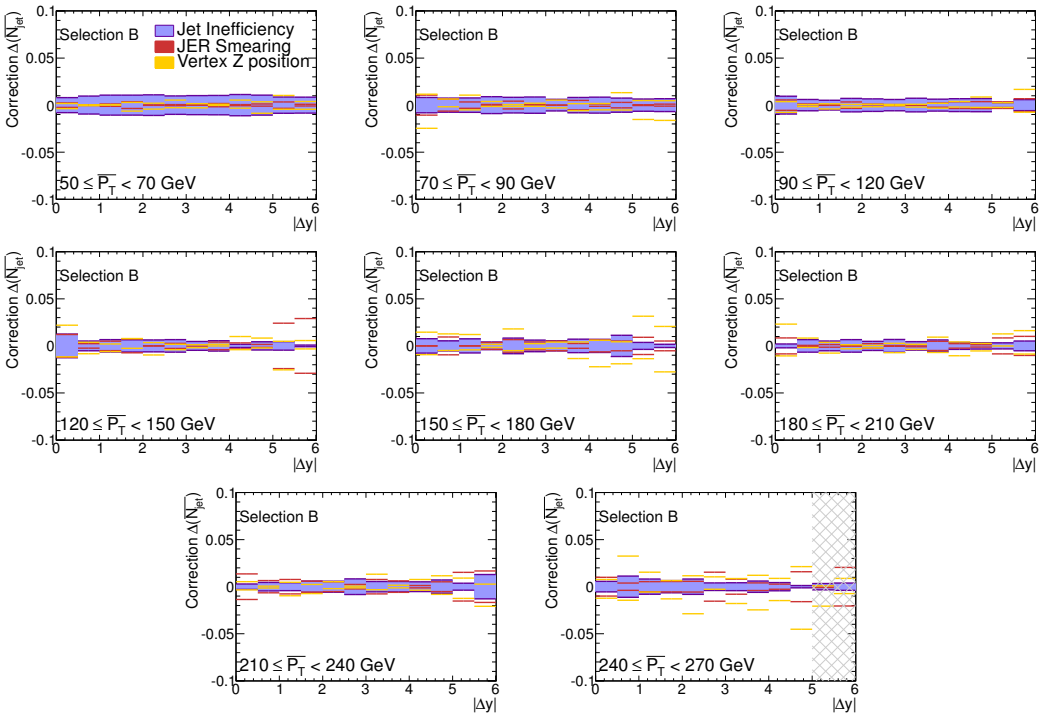


Figure C.16: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection B is used.

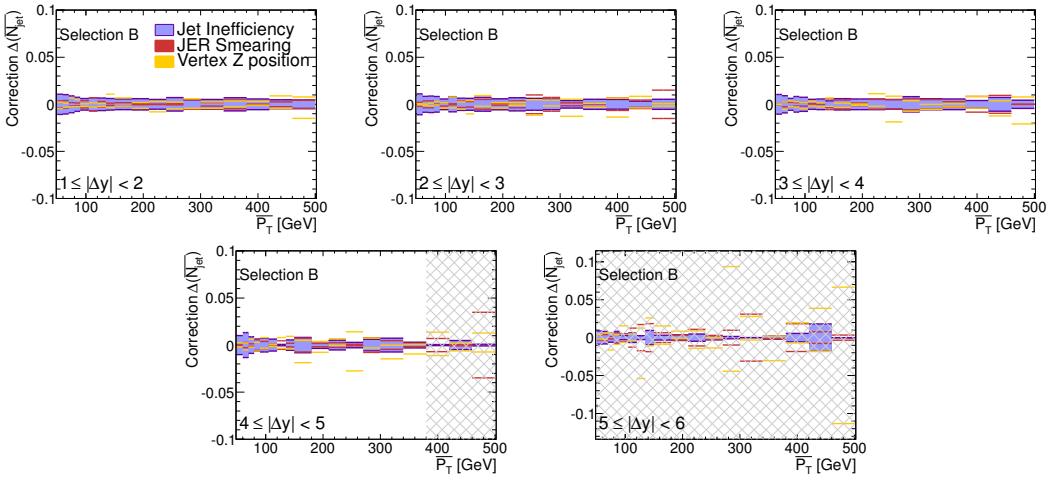


Figure C.17: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_{\mathrm{T}}}$ distribution. Selection B is used.

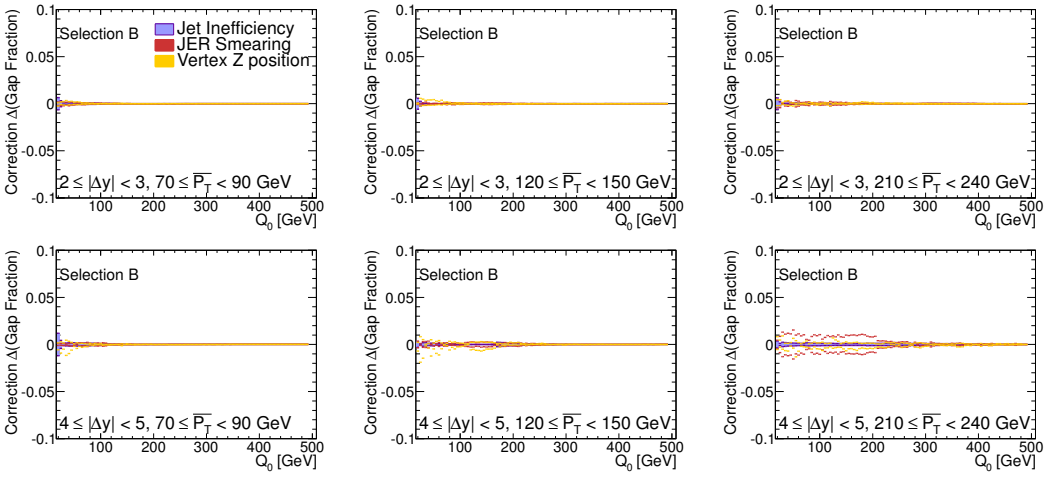


Figure C.18: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection B is used.

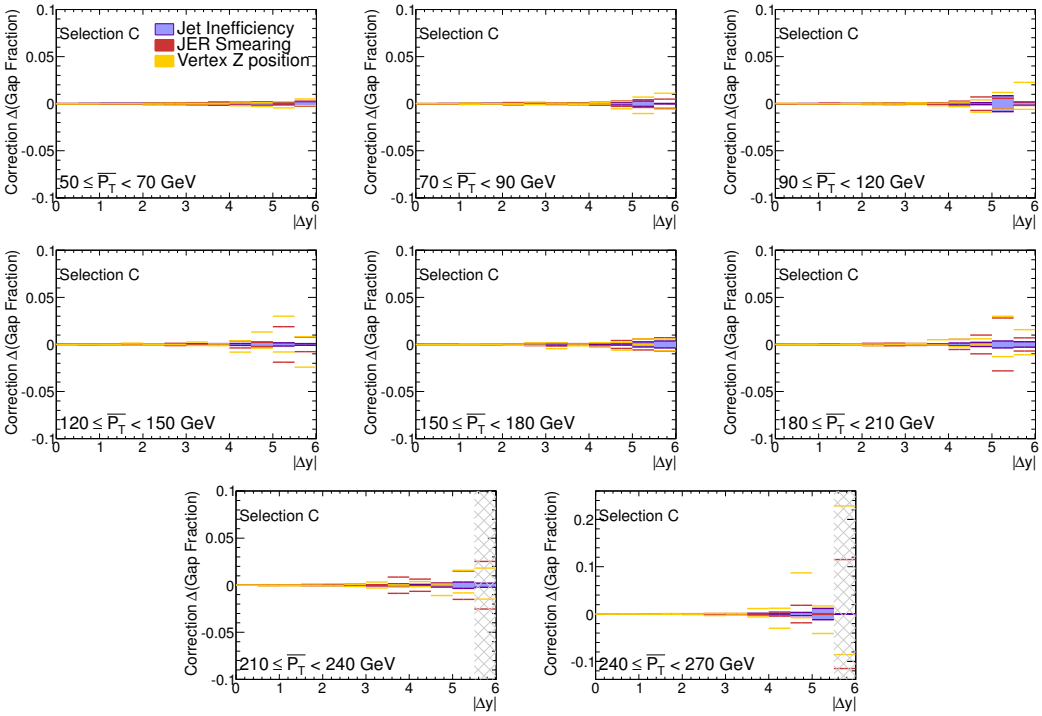


Figure C.19: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs |Δy| distribution. Selection C is used.

Selection C

The correction uncertainties derived from jet energy resolution, additional jet inefficiency and vertex z position variation are shown in figures C.19-C.23 for selection C.

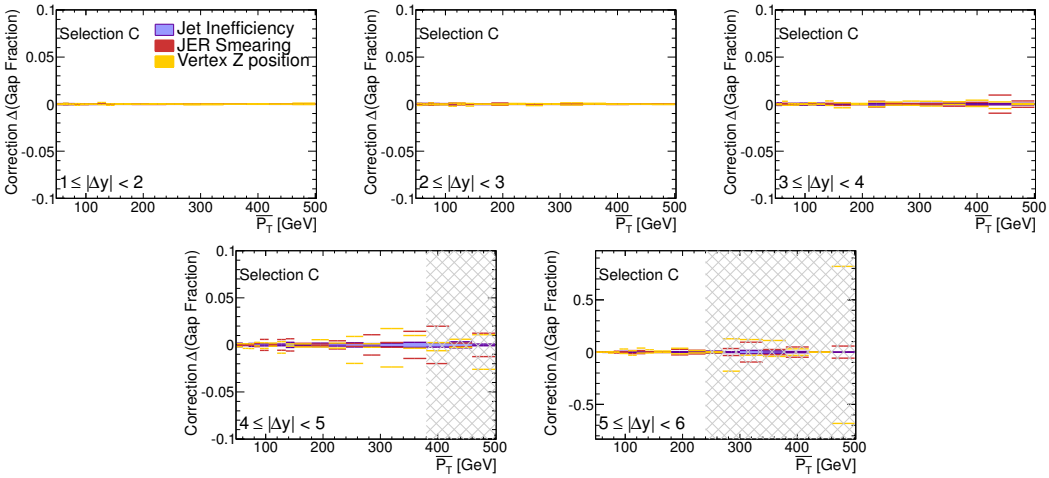


Figure C.20: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs $\overline{p_{\mathrm{T}}}$ distribution. Selection C is used.

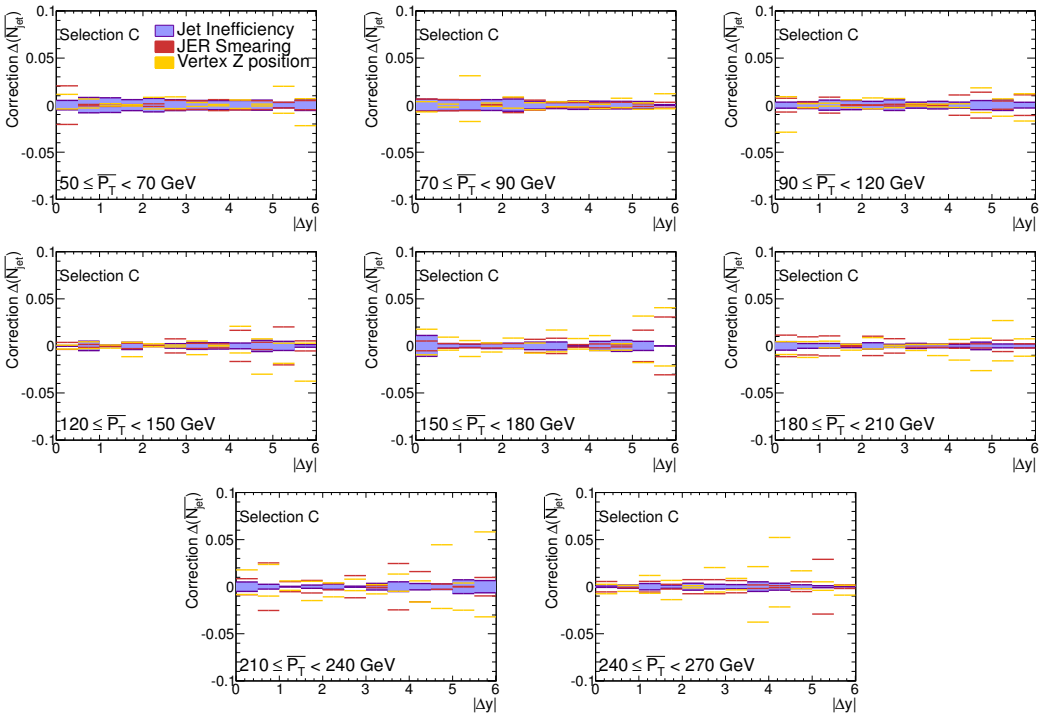


Figure C.21: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs |Δy| distribution. Selection C is used.

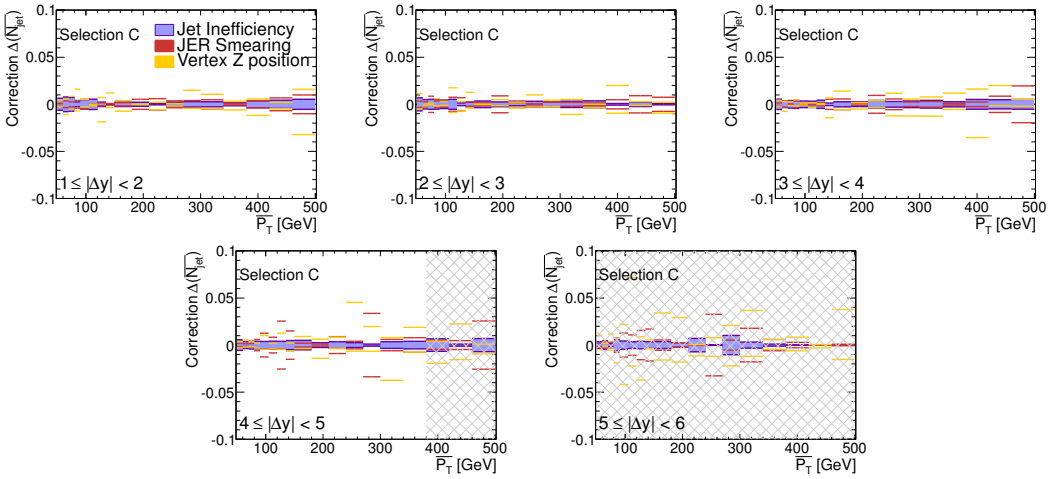


Figure C.22: Uncertainties from modification of the truth distributions for the correction factors in the average veto jet number vs $\overline{p_{\mathrm{T}}}$ distribution. Selection C is used.

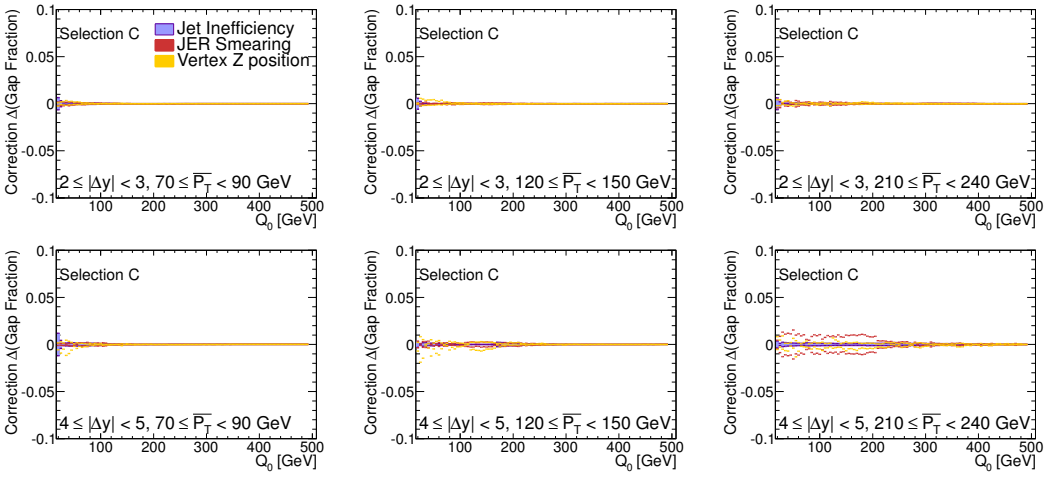


Figure C.23: Uncertainties from modification of the truth distributions for the correction factors in the gap fraction vs Q_0 distribution. Selection C is used.

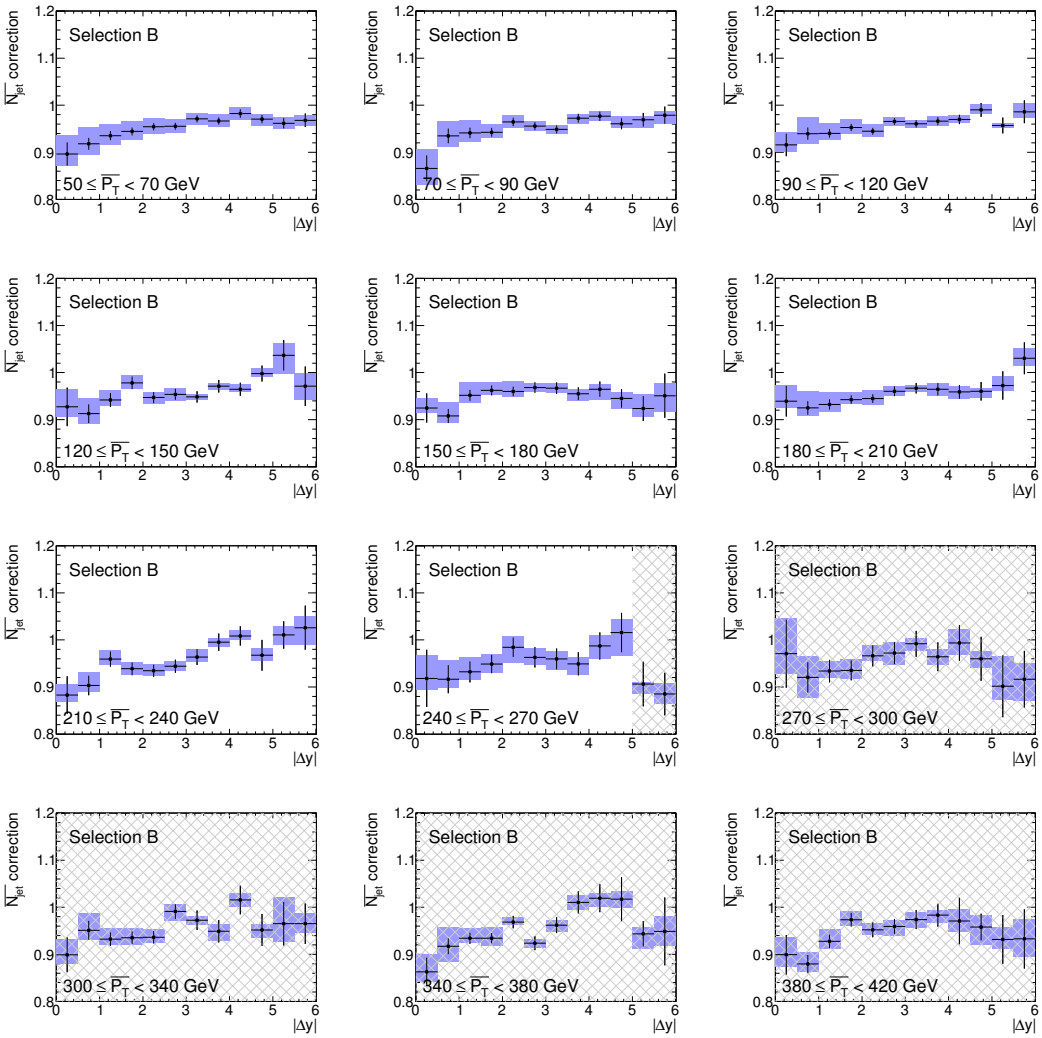


Figure C.24: Correction factors for $N^{jet}_{average}$ vs |Δy| for selection B.

C.1.3 Final Unfolding Factors

Selection B

The final unfolding factors and uncertainties for selection B is shown in figures C.24-C.27.

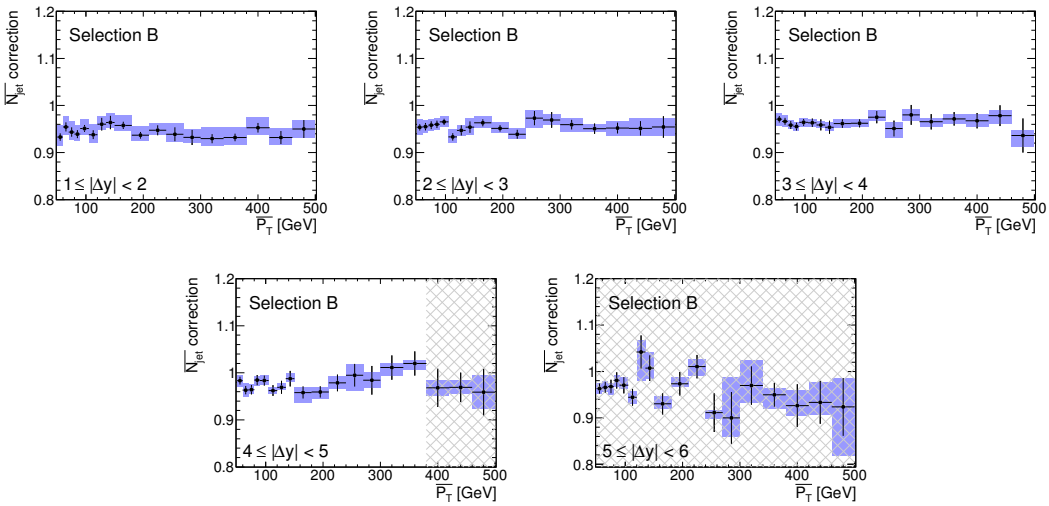


Figure C.25: Correction factors for $N^{jet}_{average}$ vs $\overline{p_{\mathrm{T}}}$ for selection B.

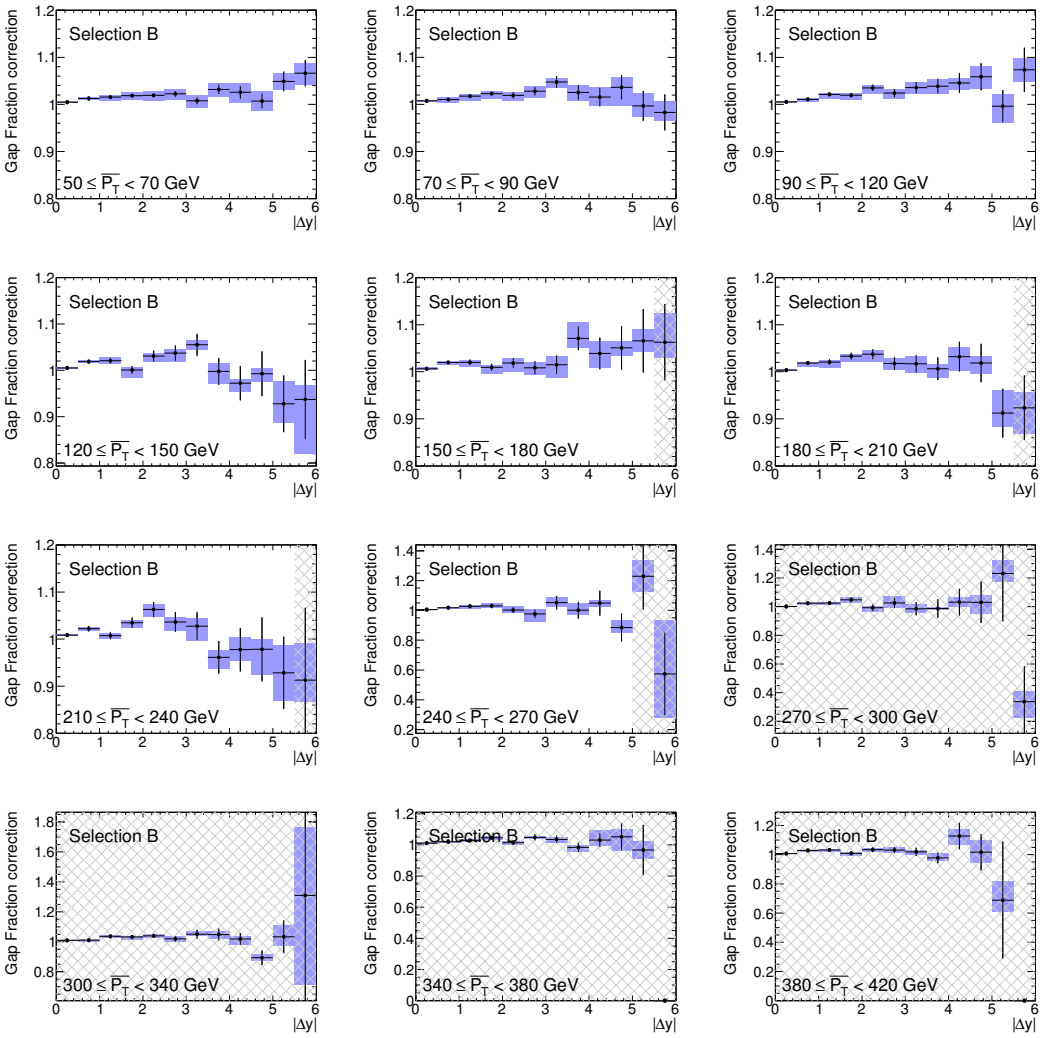


Figure C.26: Correction factors for gap fraction vs |Δy| for selection B.

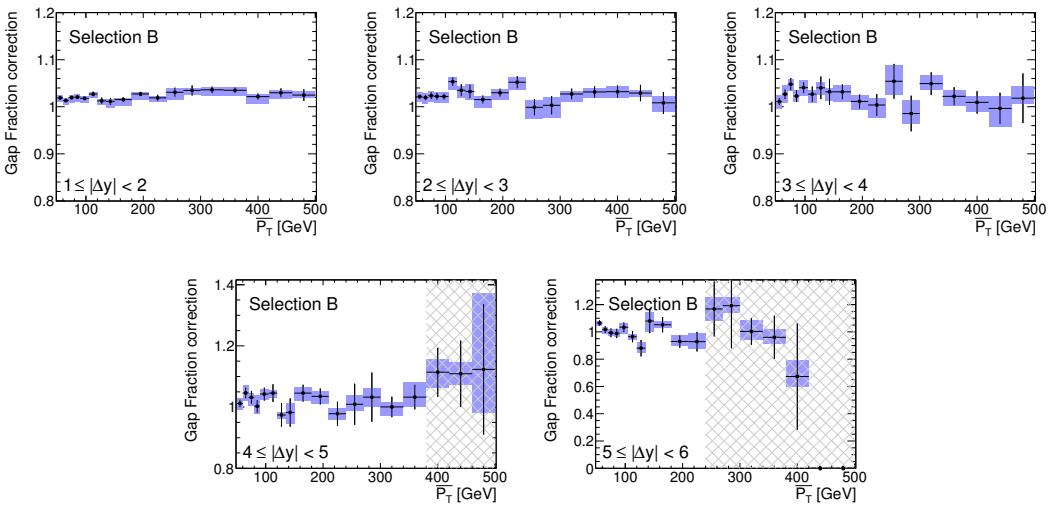


Figure C.27: Correction factors for gap fraction vs $\overline{p_T}$ for selection B.

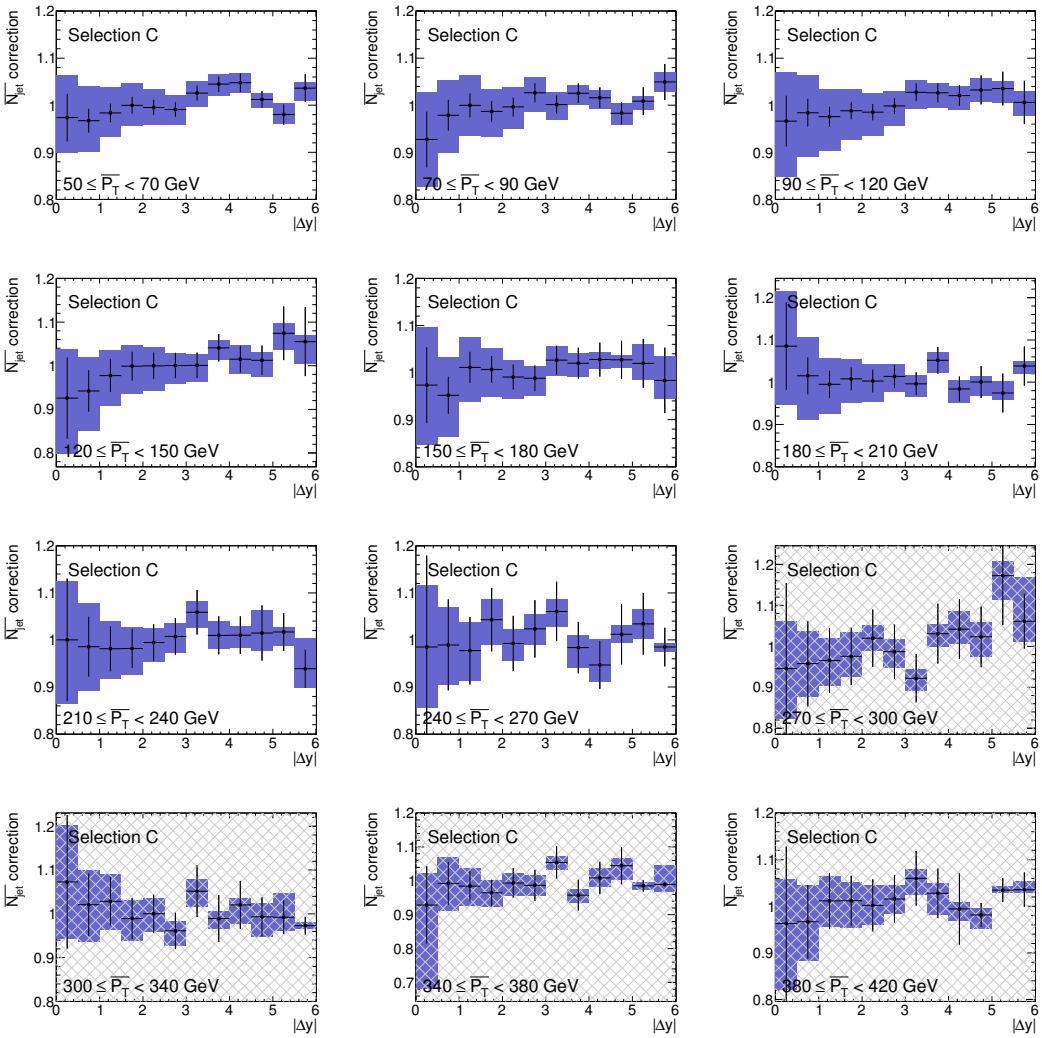


Figure C.28: Correction factors for $N^{jet}_{average}$ vs |Δy| for selection C and $Q_0 = \bar{p_T}$.

Selection C

The final unfolding factors and uncertainties for selection C is shown in figures C.28-C.31.

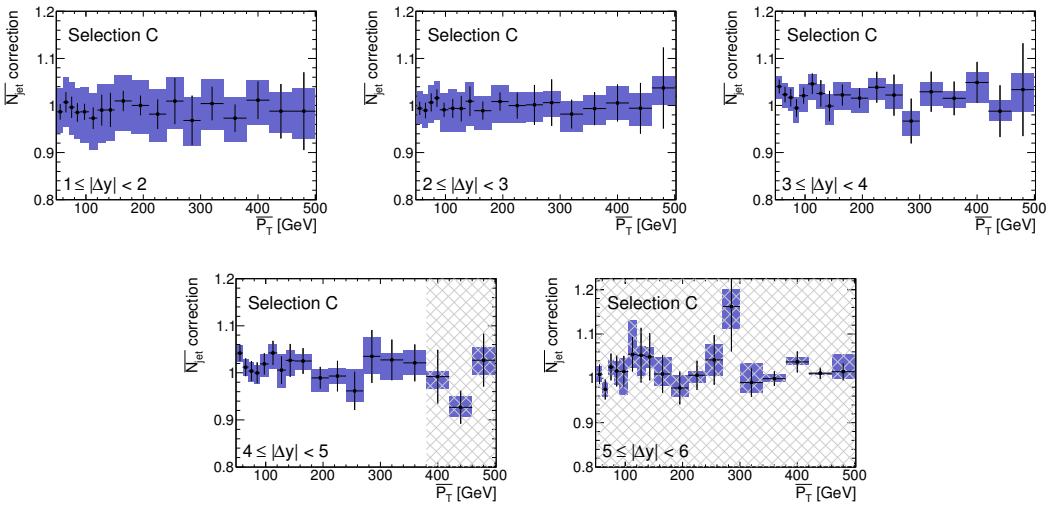


Figure C.29: Correction factors for $N^{jet}_{average}$ vs $\overline{p_{\mathrm{T}}}$ for selection C and $Q_0 = \bar{p_T}$.

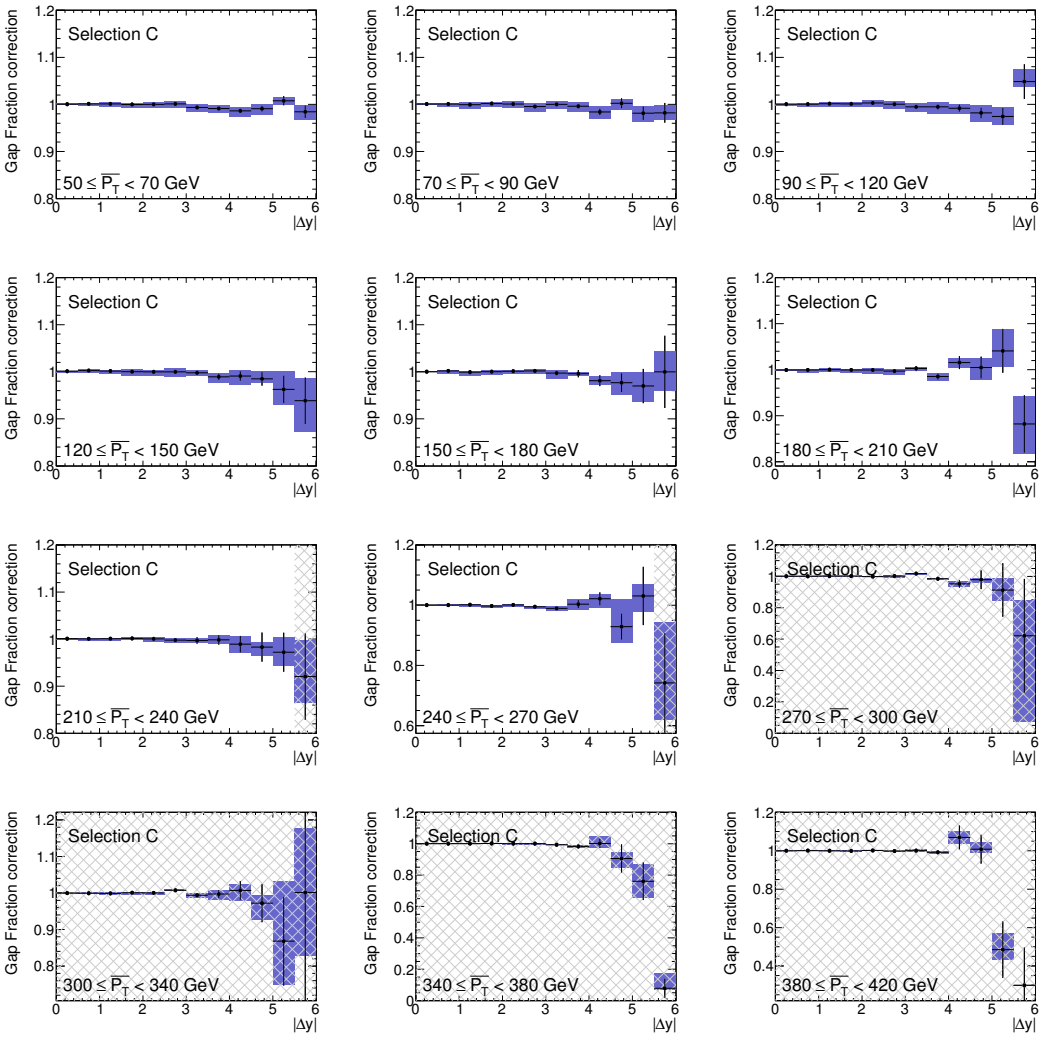


Figure C.30: Correction factors for gap fraction vs |Δy| for selection C and $Q_0 = \bar{p_T}$.

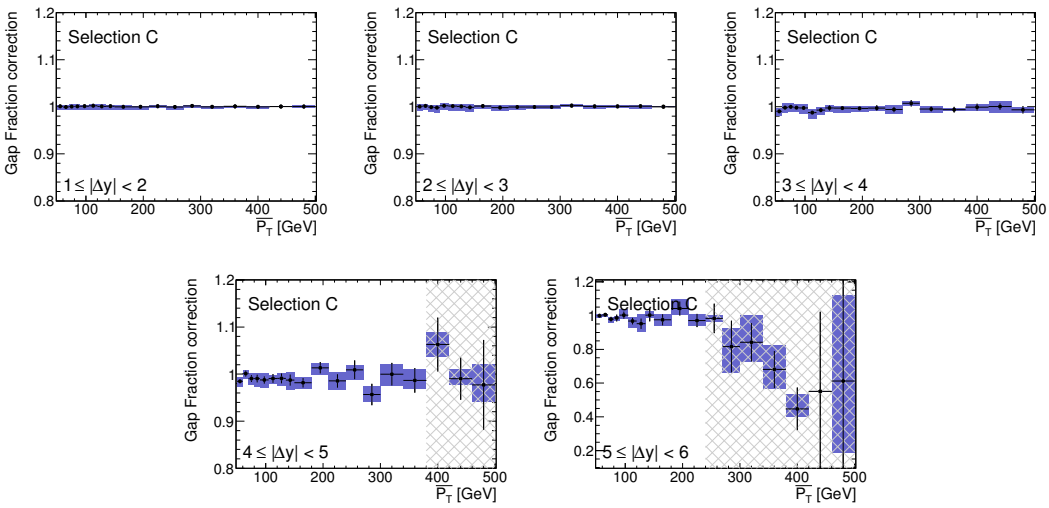


Figure C.31: Correction factors for gap fraction vs $\overline{p_{\mathrm{T}}}$ for selection C and $Q_0 = \bar{p_T}$.