A Study Of Longitudinal Hadronic Shower Leakage and the Development of a Correction for its Associated Effects at $\sqrt{s} = 8$ TeV with the ATLAS Detector

Shaun Gupta

Keble College, University of Oxford

Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Trinity Term, 2015

A Study Of Longitudinal Hadronic Shower Leakage and the Development of a Correction for its Associated Effects at $\sqrt{s} = 8$ TeV with the ATLAS Detector

Shaun Gupta
Keble College, University of Oxford

Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

September 2015

Abstract

In the high energy environment of the Large Hadron Collider, there is a finite probability for the longitudinal tail of the hadronic shower represented by a jet to leak out of the calorimeter, commonly referred to as longitudinal hadronic shower leakage, or jet "punch-through". This thesis prescribes a method for identifying such "punch-through" jets via the use of muon activity found behind a jet in the ATLAS muon spectrometer, finding an occurance rate of up to 18% in the worst affected regions. "Punch-through" jets were found to degrade the measured jet energy scale by up to 30%, and jet energy resolution by a factor of 3. A correction to remove these effects was developed in Monte Carlo and validated in data, with associated systematic uncertainties derived. The correction was found to negate the degredation of the measured jet energy scale, improving the jet energy resolution by up to 10% in the worst affected regions, and up to 1.6% overall. The correction was integrated into the final 2012 ATLAS jet energy calibration scheme as the fifth step of the Global Sequential corrections. The prescription developed in this thesis to derive the correction is currently being used by ATLAS in Run II of the Large Hadron Collider.

Acknowledgements

I have received help from many people during the course of my DPhil that have made this thesis possible. First and foremost, I would like to thank my supervisor, Çiğdem İşsever, who provided me with valuable feedback, support and guidance throughout my DPhil, without which this thesis would not have been possible.

From the ATLAS JetEtMiss group, Caterina Doglioni gave me much help when starting to work on jet punch-through and also whilst performing studies for the Dijet Mass Resonances analysis. Steven Schramm provided helpful insight during meetings when developing the jet punch-through correction. Integration of the correction into the Global Sequential framework was made possible through long Skype chats with Santiago Batista, whom together with Reina Camacho Toro and Joe Taenzer formed a great team that I had the pleasure of working with when producing the Global Sequential CONF note.

My time spent at Ferney-Voltaire on LTA at CERN wouldn't have been the same without the LTA and Citadines crew. In particular, thanks goes to Carl, Rich and Gary for all of the beers and kebabs shared.

Within the Oxford ATLAS group, I thank Todd Huffman for helping me to find a position on ATLAS when I initially was not sure about which project to work on. For the office banter, after-work drinks and ever interesting lunchtime discussions, I give a big thanks to Craig, Jim, Mireia, Kate, Lucy and David, and also to Nazim, Scott, Neven, Krish, Amy and Donal. Finally, for helping make bureaucracy easy I would like to thank Sue Geddes and Kim Proudfoot, and the Oxford IT support for providing speedy help when using the Oxford cluster.

During my time at Keble College, particularly in my first and third year, I was lucky enough to make some great friends with whom many bops and evenings at Meeples were enjoyed. For this, I would like to thank Abe, Liam, Sam, Ellie, Robbie, Frank, Dan, Dave M., Dave B., as well as the rest of the Keble MCR.

For sparking my interest in Particle Physics with inspiring lectures, and giving me the opportunity to work on a summer project as part of the High Energy Physics department at the University of Toronto, I thank Robert S. Orr.

My family and close friends from home and my time at UCL must also receive my thanks for supporting me prior to and during my DPhil.

I also thank STFC for providing me with financial support for the duration of my degree.

Lastly, a huge amount of thanks must go to Rachel for her constant support and encouragement during the difficult times, and also for putting up with my crazy sleeping schedule whilst writing this thesis - I couldn't have done this without her.

Preface

The work in this thesis was done as part of the ATLAS collaboration's jet and missing transverse momentum group, and as a result was dependent on the work of thousands of other scientists that helped to design, construct and run the detector. The author of this thesis was the sole analyser for all studies, producing all figures aside from the top-left and top-right plots in Figure 2.7, in the Appendices and the following chapters:

Chapter 4. The author performed studies that, coupled with the studies in Appendices A and B (also performed by the author), led to $N_{segments}$ being used as the primary indicator for jet punch-through and motivated the need for a jet punch-through correction.

Chapter 5. The author developed a procedure to calibrate jets using $N_{segments}$ to remove jet punch-through effects. Initial code to derive and test the jet punch-through correction was implemented mainly from scratch, using a modified version of the jet response fitting tool. This was later incorporated by the author into the main framework used by ATLAS to derive the Global Sequential corrections.

Chapter 6. The author proposed the dijet balance method used to validate the correction and derive its associated systematic uncertainties, implementing and running the code.

Chapter 7. The author performed all performance studies in this chapter, modifying the derivation code to test the impact of the correction on jets, and modifying the dijet analysis code to test the effect of the correction on the dijet mass resonances search.

The work in this thesis led to the jet punch-through correction being incorporated into the final 2012 ATLAS jet energy scale calibration as the fifth step in the Global Sequential calibrations, as described in Chapter 8. The performance of these final ATLAS 2012 Global Sequential corrections, together with the procedure used to validate them in data (also performed by the author as detailed in Appendix E), is published in Reference [1].

The procedure developed and implemented in this thesis to derive the jet punch-through correction is currently being used by ATLAS in Run II, with improvements planned in the future as will be mentioned in the Conclusions.

In addition to the work in this thesis, the author also performed studies for the ATLAS Dijet Mass Resonances search, contributing to initial 14 TeV sensitivity studies documented in Ref. [2], and fit stability tests for the 8 TeVanalysis published in Ref. [3].

Contents

Introduction

The LHC completed Run I in 2012, with ATLAS collecting 5.46fb^{-1}and 22.8fb^{-1}of data at 7 and 8 TeV respectively. During this period, the Standard Model was probed to scales never reached before, and the existence of the Higgs boson finally confirmed by ATLAS and CMS in July 2012 [4, 5]. Measurements of hadronic final states were integral to both this, and to searches for New Physics where they appear as both backgrounds and signatures for new particles.

Hadronic final states appear in the detector as hadronic showers. Hadronic jets are a fundamental ingredient in clustering these hadronic showers into objects from which the kinematic properties of the hadronic final states can be measured. They are also indispensable theoretical tools, allowing theoretical predictions to be compared with experimental results. Accurately measuring the energy of hadronic jets is therefore integral to the success of New Physics searches and precision measurements of the Standard Model.

One key factor affecting the accuracy of energy measurements of hadronic jets is how well the hadronic shower represented by the jet is contained within the hadronic calorimeter of the detector. Longitudinal leakage of a jet's hadronic shower outside of the hadronic calorimeter, commonly referred to as 'jet punch-through', can degrade the accuracy of the measured energy of the hadronic jet. Leakage can also induce non-Poissonian fluctuations in the measured energy, degrading the jet energy *resolution*, an important quantity in determining the quality of a calorimeter, and the quality of the overall energy measurement. The initial ATLAS 2012 jet energy scale was found to be sensitive to these effects.

This thesis describes the development, derivation, validation and performance of a new

correction, the 'jet punch-through correction', designed to remove the effects of longitudinal hadronic shower leakage from the ATLAS jet energy scale. It begins with an overview of the Standard Model, the theoretical constraints that give rise to hadronic showers and necessitate the use of jets, and the basic calorimetry concepts integral in measuring the energy of hadronic showers, in Chapter 1. A summary of the Large Hadron Collider and ATLAS experiment is given in Chapter 2. This is followed by an explanation of jet reconstruction within ATLAS in Chapter 3. An explanation of how punch-through jets are identified, and the subsequent impact of jet punch-through on the initial ATLAS 2012 jet energy scale is explored in Chapter 4, which motivates the need for the correction derived in this thesis. The procedure used to derive this correction is described in Chapter 5, with the data validation of the correction performed in Chapter 6, from which systematic uncertainties on the correction were derived. The performance of this correction in terms of its impact on the ATLAS jet energy scale and a search for New Physics is assessed in Chapter 7. Lastly, the final ATLAS 2012 jet energy scale calibration that incorporates this correction is briefly outlined in Chapter 8.

The jet punch-through correction was derived for four anti-k_t jet collections in total: EM+JES+GS and LCW+JES+GS scale jets with R = 0.6 and R = 0.4. In the main chapters of this thesis, plots corresponding to studies done using jets at the EM+JES+GS-scale with R = 0.6 are shown. Studies using jets at the LCW+JES+GS-scale with R = 0.6 can be found in Appendix D.

Chapter 1

Theoretical Background

The Standard Model (SM) of particle physics summarises our current understanding of the fundemental interactions of the *strong*, *weak* and *electromagnetic* forces. It catalogues the gauge bosons that mediate these interactions, along with the known leptons and quarks that form the fundemental building blocks of all matter. It is the culmination of decades of experimental and theoretical progress in quantum physics, built upon the foundations of **Q**uantum **F**ield **T**heory (**QFT**).

Whilst it has had unparalleled success in predicting physical quantities to unprecendented accuracies, it is only an effective theory, describing the world down to the scale of ~ 1 TeV. In order to develop a more completely theory, it is necessary to search for new physics at the high energy frontier, currently being probed by the Large Hadron Collider. Accurately measuring the kinematics of hadronic final states is critical to the success of such searches, and also for precision measurements of the Standard Model.

This chapter gives the theoretical motivation for the use of jets to measure the kinematics of hadronic final states, and the need for methods that ensure the measured jet energy is unaffected by longitudinal hadronic shower leakage, motivating the correction described in this thesis. The chapter begins by giving an overview of the Standard Model in Section 1.1. This is followed in Section 1.2 by an outline of the theory of the strong force - Quantum Chromodynamics - and the theoretical constrains that necessitate the use of jets to link theoretical

predictions of strong processes with experimental measurements. Processes important in making energy measurements using calorimeters, as well as the problems associated with longitudinal hadronic shower leakage, are described in Section 1.3. Section 1.4 summarises the Monte Carlo simulations used in this thesis.

1.1 The Standard Model of particle physics

The **S**tandard **M**odel is a relativistic, non-abelian[1], locally gauge invariant Quantum Field Theory [6–9]. It's local gauge invariance under the symmetry group SU(3)⊕SU(2)⊕U(1) gives rise to conserved currents under Noether's theorem [10], mediated by massless gauge bosons. SU(3), represented by **Q**uantum **C**hromo**d**ynamics (**QCD**), couples to the colour charge, mediated by eight gluons with different permutations of colour. $SU(2)_L \oplus U(1)$ represents the **E**lectro-**W**eak (**EW**) force that gives rise to the weak hypercharge, mediated by the $W^{\pm}$, W^0 and B^0 bosons. Gravity is not described by the SM.

The symmetry of the $SU(2)_L \oplus U(1)$ group is broken (known as electro-weak symmetry breaking) using the Higgs mechanism [11, 12] via spontaneous symmetry breaking, resulting in the formation of two new gauge bosons through the mixing of W^0 and B^0- the photon (γ) and Z boson[2]. The photon couples to electric charge, mediating the electromagnetic force described by U(1) and **Q**uantum **E**lectro**d**ynamics (**QED**). The $W^{\pm}$ and Z bosons become massive through EW symmetry breaking and couple to weak isospin, mediating the weak force described by SU(2). Another consequence of electo-weak symmetry breaking is the formation of a condensate manifesting as a scalar field, the Higgs field, mediated by the massive and neutral Higgs boson that couples to mass. The existence of the Higgs boson was confirmed by ATLAS and CMS on 4^{th} July 2012 [4, 5].

The interactions of the strong, weak and electromagnetic forces manifest as 24 half-integer spin fermions, each with a respective anti-fermion that has inverted quantum numbers. These

[1] The transformations of the weak (SU(2)) and strong (SU(3)) symmetry groups do not commute, resulting in self-interactions of the gauge bosons.

[2] This mixing is defined by the *Weinberg* or *weak mixing* angle, θ_W.

fermions combined with the gauge bosons (integer spin) and the scalar Higgs boson (spin 0) make up the known particles of the SM, as summarised in Figure 1.1. The fermions are split into three generations of six quarks and six leptons based upon their interactions and mass, with each successive generation being increasingly massive. Quarks carry non-integer electric charge and colour, interacting via the strong and electromagnetic forces. Isolated quarks are never observed due to asymptotic freedom, but instead form colourless particles called hadrons, as will be explained in Section 1.2.1. Leptons are further split into three with unit charge and three neutral, with all carrying weak isospin and no colour charge, thus being invisible to the strong force.

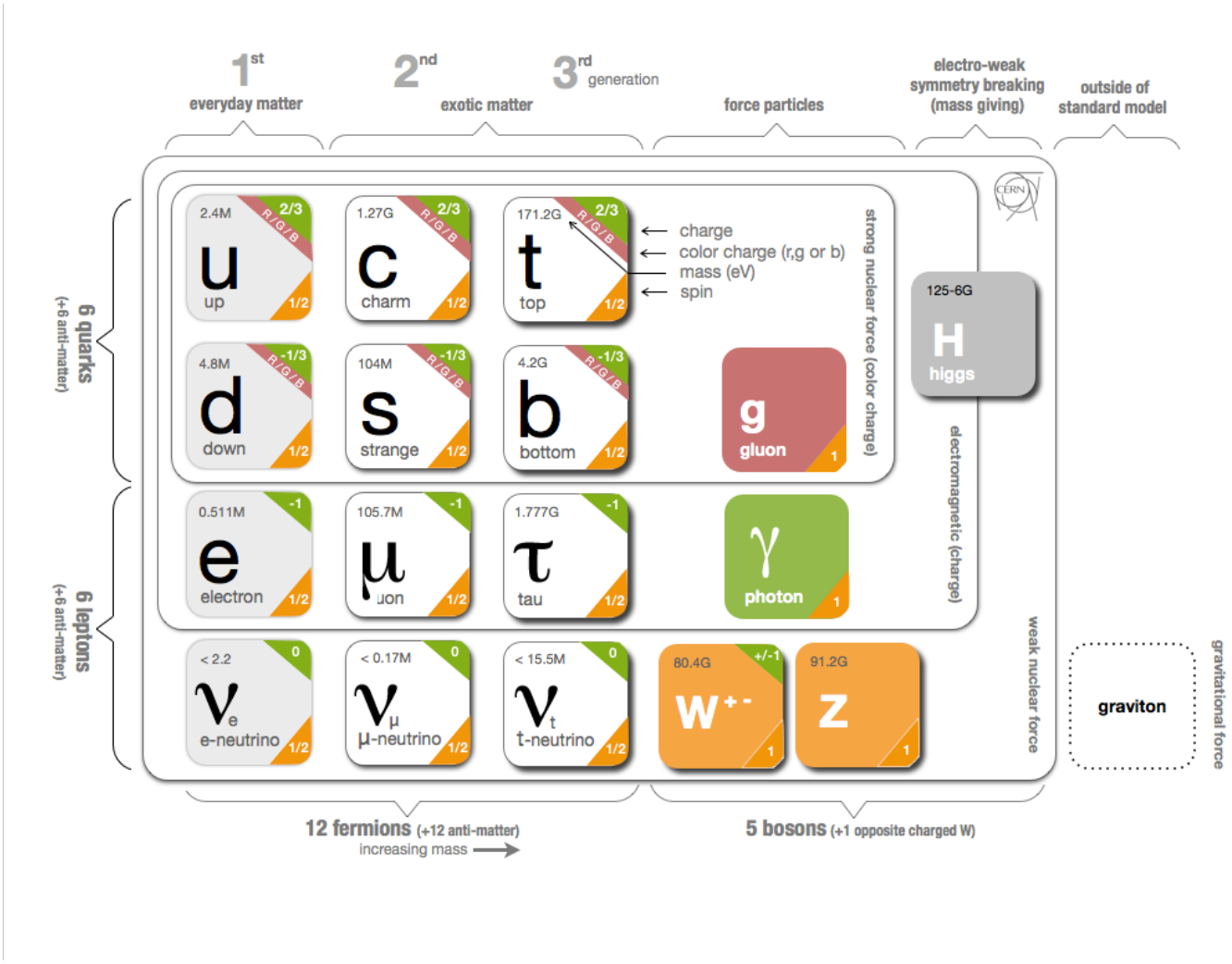


Figure 1.1: The particle content of the SM [13].

1.2 Quantum Chromodynamics

The theory of strong interactions began in the 1930's with the development of isospin by Heisenberg, used to explain the similarities between the proton and neutron (nucleons) [14]. Yukawa then introduced the exchange of *mesons* as a method of generating a nuclear force capable of binding nuclei together [15]. Flavour symmetries that extended isospin to SU(3) led to the development of the *Eightfold Way* by Gell-Mann and Ne'eman [16, 17], resulting in predictions of the Ω^- and colour, both later experimentally verified [18]. Gell-Mann [19] and Zweig [20] observed this framework could be explained by introducing an underlying point-like structure of hadrons - *quarks*. Bjorken [21] postulated that this structure should result in the electron proton (ep) scattering cross-section in the limit of infinite momentum transfer (Q^2), also known as the *deep inelastic limit*, scaling as:

$$x = \frac{Q^2}{2p \cdot q} \tag{1.1}$$

where $Q^2 = -q^2 = -(p - p')^2$ is the difference between the initial (p) and final (p') scattering particle momentum. This was confirmed by an MIT-SLAC collaboration [22–24], resulting in Feynman calling these point-like constituents *partons* [25]. This led to Bjorken's x being identified as the fraction of the longitudinal hadron momentum carried by a given parton.

Further deep inelastic scattering experimental results [22–24] showed the probes scattered from point-like spin-1/2 constituents, leading to the introduction of *colour* as an additional quantum number to satisfy Fermi-Dirac statistics [26, 27]. From the cross-section for electronpositron annihilation into hadrons, it was inferred that three colours existed - *red*, *blue* and *green*. The spin-1 gluon and structure of the proton were then inferred kinematically from momentum distributions of the partons, suggesting the proton is made from 3 valence quarks (up, up, down) mixed with a 'sea' of lower energy quark/anti-quark pairs and gluons. The quark model eventually led to the development of Quantum Chromodynamics.

Two key features of QCD are *confinement* and *asymptotic freedom*. Confinement results

in quarks and gluons being bound inside colour neutral states such as nucleons, however at sufficiently high energies (small distance scales) they act as free - asymptotic freedom. These features prevent the experimental observation of free quarks and gluons. Instead, quarks and gluons are measured by the detector as collimated showers of hadrons clustered into *jets*.

1.2.1 Singularities, asymptotic freedom and confinement

Perturbative expansions of calculations involving loops, such as the loop in Figure 1.2, result in singularities called *ultra-violet divergences*.

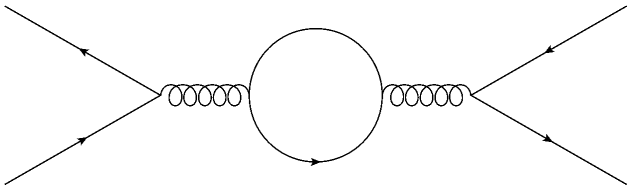


Figure 1.2: Fermionic loop correction to S-channel annihilation [28].

In order to remove these non-physical infinities to give physically meaningful results, a renormalisation procedure is necessary. This is performed through the subtraction of the infinities through counter-terms embedded in *bare* parameters that are unmeasurable. These subtractions are introduced via corrections to the 'renormalised parameter' at a scale known as the renormalisation scale μ_R. As μ_R is introduced by hand, any physical variable calculated must be independent of μ_R when summing over all orders of perturbation theory, however the calculation still depends on μ_R when truncating the calculation at a given order. This results in the renormalisation correction being dependant on μ_R and Q^2. An important renormalised parameter is the strong coupling constant, which can be calculated to **L**eading **O**rder (**LO**) via [29]:

$$\alpha_{\mathrm{S}}(Q^2) = \frac{\alpha_{\mathrm{S}}(\mu_R^2)}{1 + \alpha_{\mathrm{S}}(\mu_R^2)\beta_0 \ln(Q^2/\mu_R^2)} \tag{1.2}$$

The renormalised coupling constant is thus a function of the scale of the process, and *runs* with energy. This can be seen from the experimental measurements of α_{S} as a function

of Q, shown in Figure 1.3. Making the substitution $\ln(\Lambda^2_{QCD}) = \ln(\mu_R^2) - \frac{1}{\beta_0 \alpha_S(\mu_R^2)}$, $\alpha_S(Q^2)$ can also be expressed as [29]:

$$\alpha_S(Q^2) = \frac{1}{\beta_0 \ln(Q^2/\Lambda^2_{QCD})} \tag{1.3}$$

where Λ^2_{QCD} can be interpreted as the non-perturbative limit, at which point perturbative methods break down due to $\alpha_s \to \infty$ as $Q^2 \to \Lambda^2_{QCD}$. This phenomenon of the strength of the strong force increasing with distance[3] leads to *confinement* of quarks within colour neutral states. Conversely, as $Q^2 \to \infty$, $\alpha_s \to 0$, resulting in quarks behaving as free particles at high Q^2 (small distance scales) - *asymptotic freedom*.

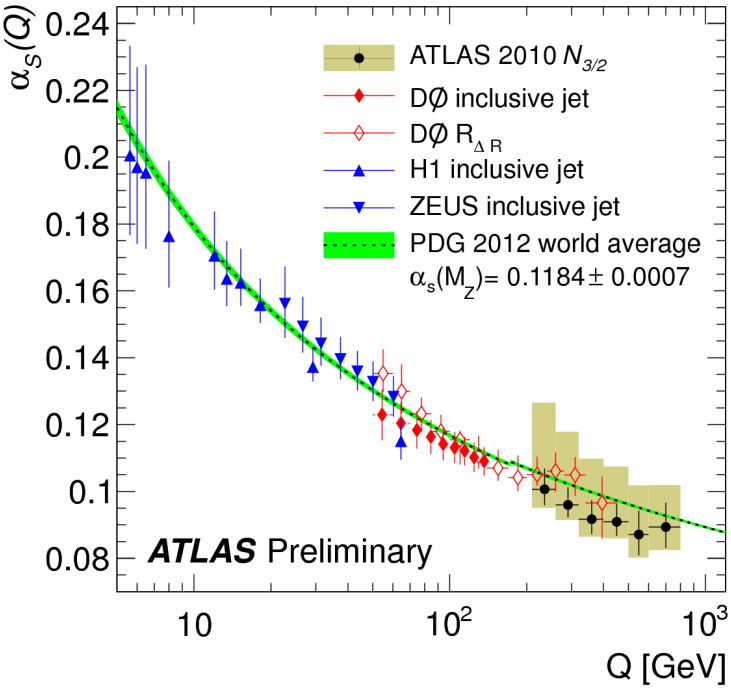


Figure 1.3: α_S as a function of Q [30].

1.2.2 Hadronization

Experimentally, asymptotic freedom results in high energy probes being necessary to eject quarks and gluons from the nucleon, while confinement prevents quarks and gluons from

[3] $\lambda \sim \frac{1}{\sqrt{Q^2}}$, where λ is the wavelength of the probing particle.

being observed directly. Instead, they appear in the detector as collimated sprays of particles. This process can be summarised for hadronic collisions (such as the pp collisions at the Large Hadron Collider) as follows [31]:

1. Two hadrons collide at high Q^2, where each hadron can be viewed as a 'bag of partons'.

2. A collision between one parton from each hadron (*incoming partons*) results in the hard scatter of interest, giving the *outgoing partons*.

3. Bremstrahlung results in emissions from the incoming (***I**nitial **S**tate **R**adiation* or ***ISR***) and outgoing (***F**inal **S**tate **R**adiation* or ***FSR***) partons.

4. Additional partons within the hadrons may collide within a single hadron-hadron collision - *multiple interactions*.

5. As the distance between the outgoing partons increases ($Q^2 \to \Lambda^2_{QCD}$), $\alpha_s \to \infty$. This results in a non-perturbative process, whereby the colour force can be seen as a string connecting two particles with opposite colour charges. The colour potential in the string increases as the separation between the two colour charges increases, eventually resulting in the string breaking and creating two new colour charges at its endpoints. When two opposite colour charges are close together they will recombine into hadrons. This process is called *hadronization*, and results in collimated sprays of particles.

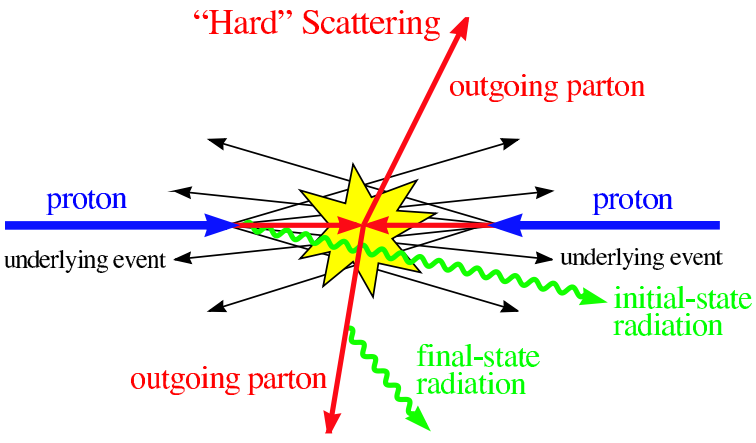


Figure 1.4: Pictorial representation of $2 \to 2$ body hard scattering event [32].

These steps are represented schematically in Figure 1.4.

The collision remnants from the hard scatter not identified as interesting (ISR, FSR, multiple interactions) are referred to as the *underlying event*.

1.2.3 Measuring inclusive QCD processes: jets

The emission of infinitely soft particles or particles at small angles in QCD processes can cause *infra-red* and *collinear* divergences respectively. These can cause infinities during calculations, that result in a divergent inclusive QCD cross-section. The **K**inoshita-**L**ee-**N**auenberg (**KLN**) theorem [33, 34] states these divergences cancel when using inclusive observables insensitive to processes involving a variable number of final-state partons. Such inclusive observables are called *infra-red safe*, and have kinematics that reflect the kinematics of the final state partons. This requirement to use such infra-red safe observables can be likened to the finite spatial resolution of the detector preventing individual final state partons from being resolved.

In 1977, Sternman and Weinburg [35] introduced the concept of *jets* - collimated groups of particles generated by hadronization - as inclusive objects with a rigorous definition that could be used to connect calculations from perturbative QCD to experimental measurements of hadronic final states. Their definition was based upon the assumption that two partons with sufficiently high momentum will produce back-to-back jets. The jet cross-section was then defined as the number of events containing two cones of opening angle δ, that contain a fraction $(1-\epsilon)$ of the total energy of the event, where ϵ is the fraction of energy in the event not contained within the cone. Provided δ and ϵ are not too small, the result is perturbatively well behaved. Today, many different jet definitions are available. Chapter 3 contains more information on these definitions and jet reconstruction in the ATLAS experiment.

1.3 Energy measurement using calorimeters

In modern particle physics experiments, the energy of final-state particles is commonly measured through total absorption in a block of matter called a *calorimeter*. Calorimeters also provide additional information for particle ID, event selection, triggering and precision measurements. The measurement process of a calorimeter is *destructive*, meaning particles are no longer available for inspection after measurement (with the exception of muons).

They are popular due to a number of useful features. Their energy resolution[4], σ, improves with increasing energy, making them ideal for use in detecting the high energy products of collisions in high energy physics experiments. They are relatively compact due to the amount of material necessary to contain a particle shower $\propto \log(E)$, as will be discussed more in Section 1.3.3, making the material cost manageable. Finally, they are sensitive to both charged and neutral particles.

Calorimeters make use of a range of different technologies, including scintillation, Čerenkov radiation, ionization, and cryogenic phenomena. There are two main types of calorimeter: *homogeneous* and *sampling* calorimeters. Homogeneous calorimeters are made only out of material sensitive to particles (active material) such as scintillating crystals. Sampling calorimeters are made from alternating layers of active materials (scintillators, semiconductors, noble liquids) and high density passive materials that act as absorbers (Fe, Cu, Pb, etc.). Sampling calorimeters are used by ATLAS because they require less material and are thus cheaper, however their energy resolution is poorer than homogeneous calorimeters.

Different particles react in different ways with the calorimeter, depositing energy via a variety of processes. Electrons and photons form *electromagnetic showers* in the calorimeter, whilst hadronic final states form *hadronic showers*. The different interactions through which they lose and deposit energy in the calorimeter are important considerations in the design of the calorimeter, especially in terms of the amount of material necessary to fully contain the showers. Shower containment is vital in providing accurate energy measurements.

[4] σ and E are the width and mean of the measured energy distribution respectively.

1.3.1 Electromagnetic showers

When interacting with the calorimeter, electrons and positrons lose energy via *bremsstrahlung* and *ionisation*, while photons lose energy via *electron-positron pair production*, and the *photoelectric* and *Compton* effects. At low energy, ionisation and the photoelectric/Compton effects dominate. At high energy however, such as the environment at the Large Hadron Collider, bremsstrahlung radiation and electron-positron pair production dominate, resulting in the production of *electromagnetic showers*.

A convenient distance measure for electromagnetic showers is the radiation length (X_0), defined to be the mean length of material required to reduce the energy of a photon or electron by $1/e$. The longitudinal depth of the shower maximum in units of X_0, t_{max}, is then given by [36]:

$$t_{max} = \frac{\ln(E_0/E_c)}{\ln 2} \tag{1.4}$$

where E_0 is the initial electron or photon energy, and E_c is the critical energy at which energy loss rates due to bremsstrahlung and ionisation are equal. Here, $E_c = 610\ \text{MeV}/(Z + 1.24)$ [37], and is only dependant on the atomic number of the material Z. Therefore the longitudinal depth of the shower maximum, and thus the amount of material required to contain 95% of the shower, increases logarithmically with energy of the electron or photon.

1.3.2 Hadronic showers

The presence of strong hadronic interactions with matter makes hadronic showers more complicated than electromagnetic ones. The strong interaction results in the following processes (summarised in Figure 1.5):

- The production of hadronic shower particles through hadronization and the decay of these hadrons. Neutral pions (π^0) account for $\sim \frac{1}{3}$ of the produced particles, which subsequently decay into two photons ($\pi^0 \to \gamma\gamma$), producing electromagnetic showers.

These electromagnetic showers develop in the same way as those initiated by high energy electrons or photons.

- Nuclear reactions resulting in the release of neutrons and protons. The nuclear binding energy released by these interactions is not normally absorbed by the detector, resulting in *invisible energy*. These processes can also result in the production of neutrinos (and muons) that leave the detector unobserved, contributing to *escaped energy*.

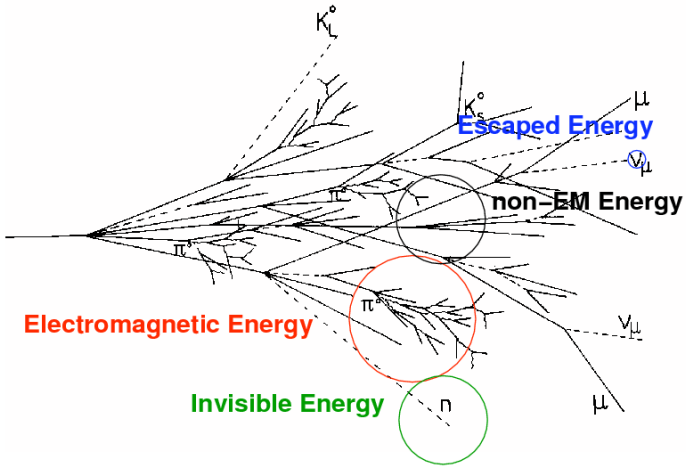


Figure 1.5: Schematic of hadronic shower [38].

As a result of these additional interactions, the energy *response* of a hadron, defined to be the ratio between the detected (reconstructed) energy and the true energy of the incident particle, is smaller than the energy response of an electron of the same nominal energy. Two approaches are generally taken to improve the response of hadrons:

- *Compensation*: Calorimeters that use specialised active layers to capture invisible energy are called *compensating calorimeters*. Examples of this are calorimeters that use high-Z materials such as uranium (^{238}U) combined with scintilator plastic to ensure electrons and hadrons have the same response. When using uranium, nuclear reactions will also produce neutrons, which can further produce additional neutrons and γ-rays by inducing fission of other nuclei, enhancing the amplitude of the hadronic shower.

- *Non-compensation*: Calorimeters that have a different response to electrons and hadrons are called *non-compensating calorimeters*. Instead of using specialised active layers to measure invisible energy, non-compensation is accounted for through the use of additional calibrations which restore the hadron response to the electron response. ATLAS uses a non-compensating calorimeter.

The ratio of the electron response to the pion response (e/π) is one method of assessing the level of non-compensation, where an $e/\pi > 1$ suggests non-compensation. The hadron response (and thus e/π) in non-compensating calorimeters will be lower for showers with a lower electromagnetic component. The fraction of electromagnetic energy in a hadronic shower, f_{em}, is shown in Figure 1.6 (left) to increase as a function of the initial pion energy, resulting from higher energy pions having more energy available for hadron (π^0) production. Thus e/π for a non-compensating calorimeter will increase with initial pion energy.

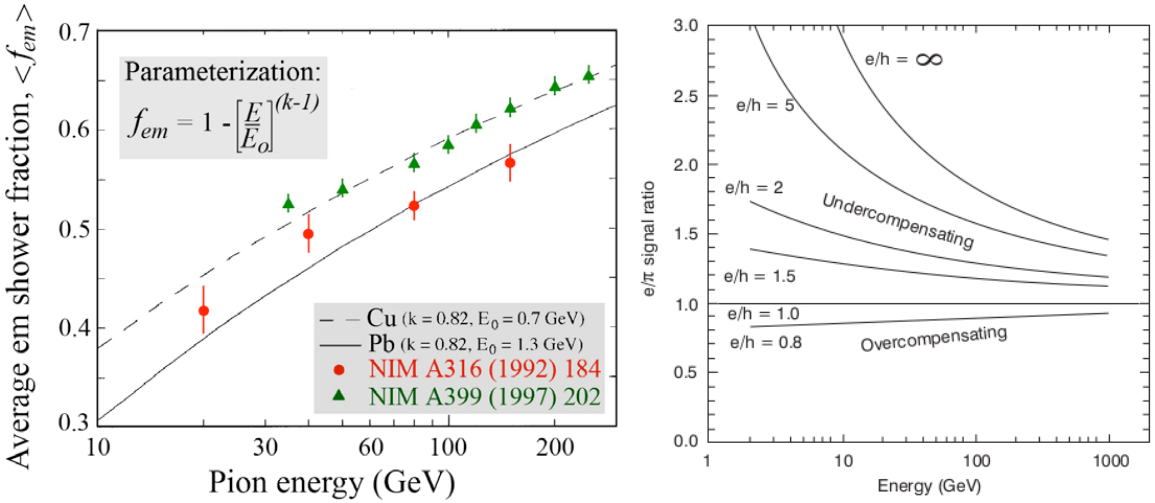


Figure 1.6: Fraction of electromagnetic energy in hadronic shower (left) [39] and e/π for various e/h values (right) [40] as a function of initiating pion energy.

An energy independent way of assessing non-compensation is to use the ratio of the calorimeter response to the electromagnetic and non-electromagnetic components of a hadronic shower, e/h [41]. One parameterisation for extracting e/h from measurements of e/π is [42]:

$$\frac{e}{\pi} = \frac{e/h}{1 + (e/h - 1) \cdot k \cdot \ln E} \tag{1.5}$$

Here, $k = \frac{E_e/E}{ln(E)}$, where E is the total hadronic shower energy and E_e is the contribution from the electromagnetic part of the shower. Calorimeters with $e/h > 1$, $e/h = 1$ and $e/h < 1$ are called *undercompensating*, *compensating* and *over-compensating* respectively. Non-compensating calorimeters are undercompensating. Figure 1.6 (right) shows e/π as a function of initial pion energy for different e/h values, which confirms e/π for non-compensating calorimeters increases with initial pion energy. ATLAS has $e/h = 1.362 \pm 0.006$ [43].

Hadronic shower profiles are governed by the *nuclear interaction length* (λ_{I}), defined to be the average distance hadrons travel before inducing a nuclear interaction [44]:

$$\lambda_{\mathrm{I}} \approx 35 g/cm^2 A^{1/3} \tag{1.6}$$

where A is the nuclear mass number. λ_{I} is generally much larger than X_0 e.g. For copper, $X_0 = 1.4cm$, while $\lambda_{\mathrm{I}} = 15cm$ [40]. As a result, hadronic calorimeters are generally much deeper than electromagnetic calorimeters.

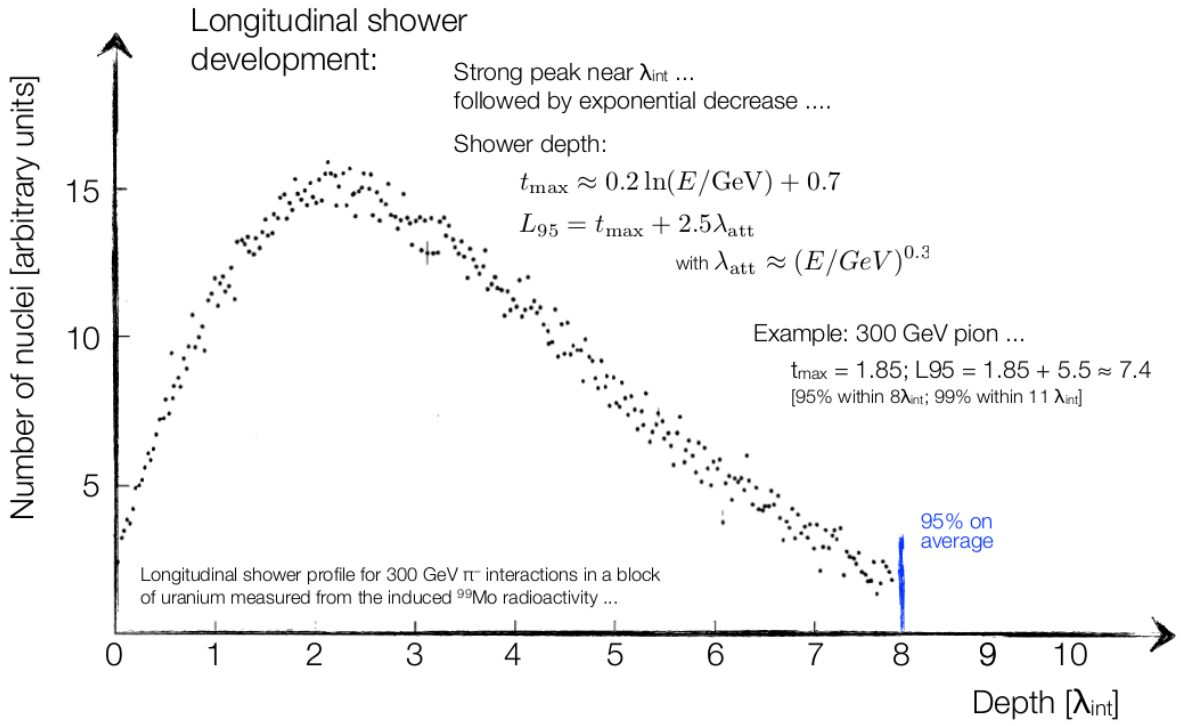


Figure 1.7: The longitudinal shower profile for 300 GeV pion interactions in a block of uranium, measured from the induced ^{99}Mo radioactivity [45]

The longitudinal shower profile for 300 GeV pion interactions in a block of uranium, measured from the induced ^{99}Mo radioactivity, is shown in Figure 1.7. The figure shows how the number of nuclei in the shower changes with the depth of the shower measured in λ_{I}. The shower maximum, t_{max} is approximated to $0.2 \ln E[GeV]+0.7 = 1.85\ \lambda_{\mathrm{I}}$ in this case, with 95% of the shower contained at a depth of L_{95}, where $L_{95} = t_{max} + 2.5(E[GeV])^{0.3} \approx 7.4\ \lambda_{\mathrm{I}}$ for this shower. Thus for a 300 GeV pion, uranium with a depth of 8 λ_{I} is required to achieve 95% containment. An alternative form for the amount of material required to contain 95% of the hadronic shower is [46]:

$$L(95\%) = (9.4\ \ln E[\text{ GeV}] + 39)[cm\ Fe] \tag{1.7}$$

Whilst these definitions do not give identical results, they both suggest that the length of material needed to contain a hadronic shower grows with the logarithm of the energy of the particle that initiated the shower, similar to what is seen for electromagnetic showers. For a 300 GeV pion in iron, Equation 1.7 gives ~ 92.62 cm, while ~ 124 cm is obtained from the other definition when taking λ_{I} for iron to be 16.77 cm [47].

1.3.3 Issues associated with longitudinal hadronic shower leakage

Longitudinal hadronic shower leakage refers to the leakage that occurs when the longitudinal tail of the hadronic shower is not contained within the hadronic calorimeter. Longitudinal leakage, for the purposes of this thesis, does not include *escaped energy* - energy lost due to neutrinos and muons produced during the hadronic shower escaping the calorimeter undetected. As seen from Equation 1.7 the amount of material required to contain a hadronic shower increases with $\ln E$. Longitudinal hadronic shower leakage is thus more likely to occur when E is larger, or when the shower passes through a layer of the detector with less material. In addition, the probability of longitudinal leakage is higher when the shower begins later in the calorimeter.

When leakage occurs, the energy of the leaked shower is not measured, lowering the

hadronic response. The process is stochastic, resulting in non-Poissonian fluctuations of the measured hadronic energy. As well as affecting the mean of the measurement, this can have a negative impact on the *resolution* of the calorimeter, where the resolution is a measure of the fluctuation of the measured energy around the mean measured energy. The resolution is a commonly used measure of the quality of a detector - the better the resolution, the better the detector. For non-compensating hadronic calorimeters, the resolution, $\frac{\sigma_E}{E}$, can be expressed via [45]:

$$\frac{\sigma_E}{E} = \frac{A}{\sqrt{E}} \oplus B \oplus \frac{C}{E} \tag{1.8}$$

$\frac{A}{\sqrt{E}}$ is affected by fluctuations in the invisible energy content of the shower, a result of fluctuations in the electromagnetic shower fraction, nuclear excitations/fission/binding energy, and heavily ionising particles. B is a constant term affected by inhomogeneities and longitudinal shower leakage. $\frac{C}{E}$ is affected by electronic noise and sampling fluctuations. $\oplus$ denotes addition in quadrature. The energy resolution therefore gets worse as a result of longitudinal hadronic shower leakage, a result of the increasing contribution of B.

These effects were confirmed in the study in Reference [48] that assessed the effect of leakage on the energy response and energy resolution of the ATLAS hadronic tile calorimeter. The study was performed using 100 GeV pion test beams fired at a prototype calorimeter configuration composed of five sector modules with total thickness 9.4 λ_{I}. Each module (visible in Figure 1.8) consists of 57 'periods', with each period consisting of four layers. The first and third layers are formed by large trapezoidal steel plates, whilst the second and fourth layers consist of alternating steel plates and scintillator tiles [48]. The modules were mounted onto a scanning table (Figure 1.8), allowing for high precision movements along any direction [49]. The test beam was fired at incidence angle $\Theta = 10°$, with impact point Z along the axis parallel to the face of the calorimeter varied from -36 to 20 cm.

In order to detect for leakage, plastic scintillator detectors called 'muon walls' were placed behind and at the right side of the calorimeter modules, resulting a setup similar to that

shown in Figure 1.9. An event was tagged as having longitudinal or lateral leakage if at least one hit was detected in the 'muon walls' behind or to the side of the calorimeter respectively.

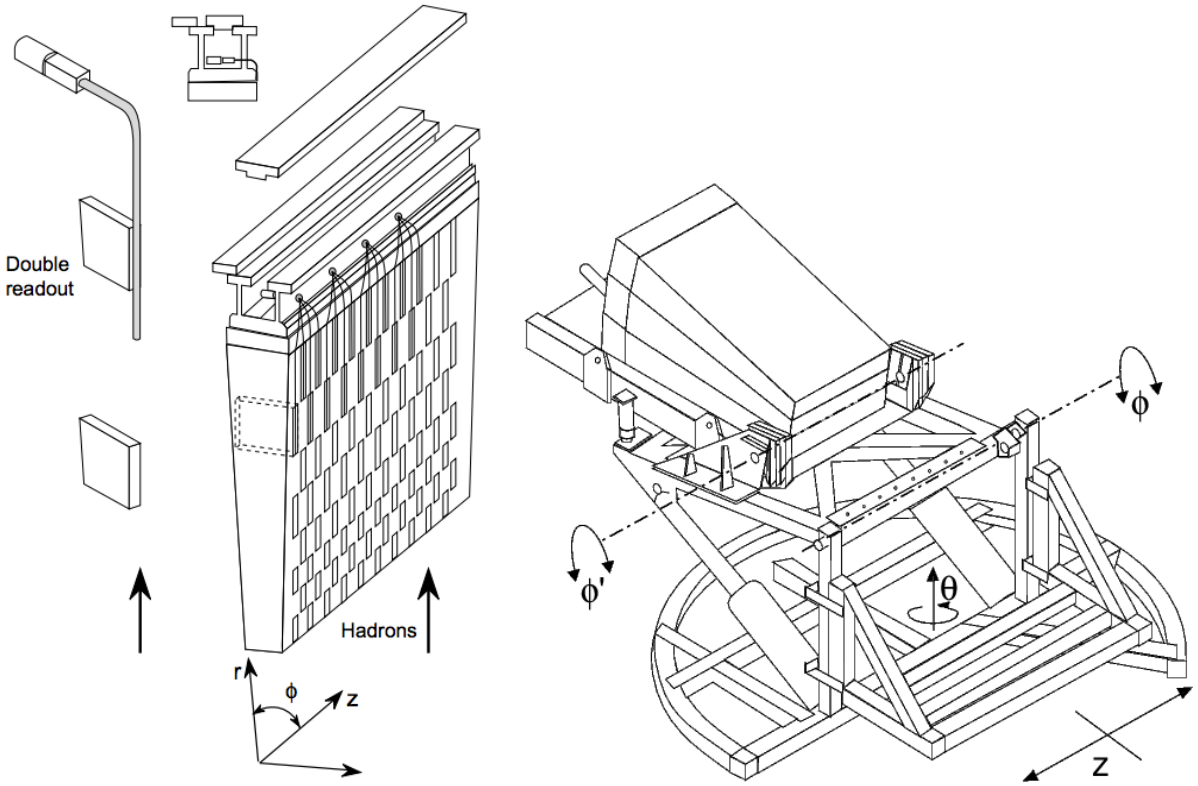


Figure 1.8: Prototype tile calorimeter module (left) [48] mounted onto the scanning table (right) [49].

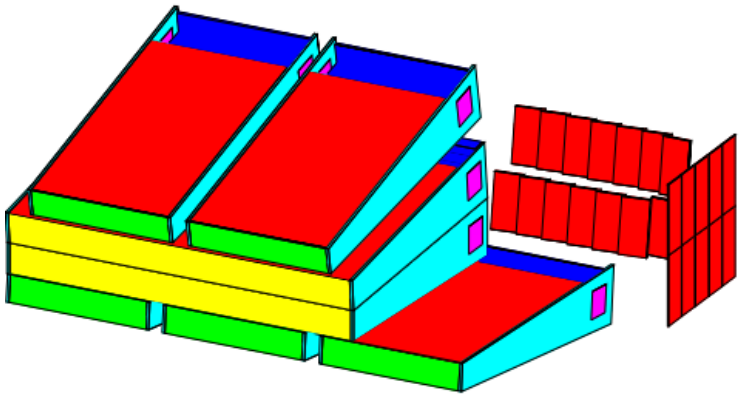


Figure 1.9: Example arrangement of 'muon walls'. [50].

Figure 1.10 taken from this study shows the mean energy (left) and energy resolution (right) as a function of Z for hadronic showers experiencing various types of leakage. The

energy resolution is characterised by looking at the RMS of the energy distribution divided by the mean energy, $RMS(E)/E$. The 'no leakage' case is represented by black squares, while 'longitudinal leakage' is represented by open squares. The results clearly show longitudinal leakage reduces the measured mean energy from 100 GeV to 90-92 GeV(an 8-10% reduction). It also drastically worsens the energy resolution, from ~ 0.1 to between 0.15 and 0.3.

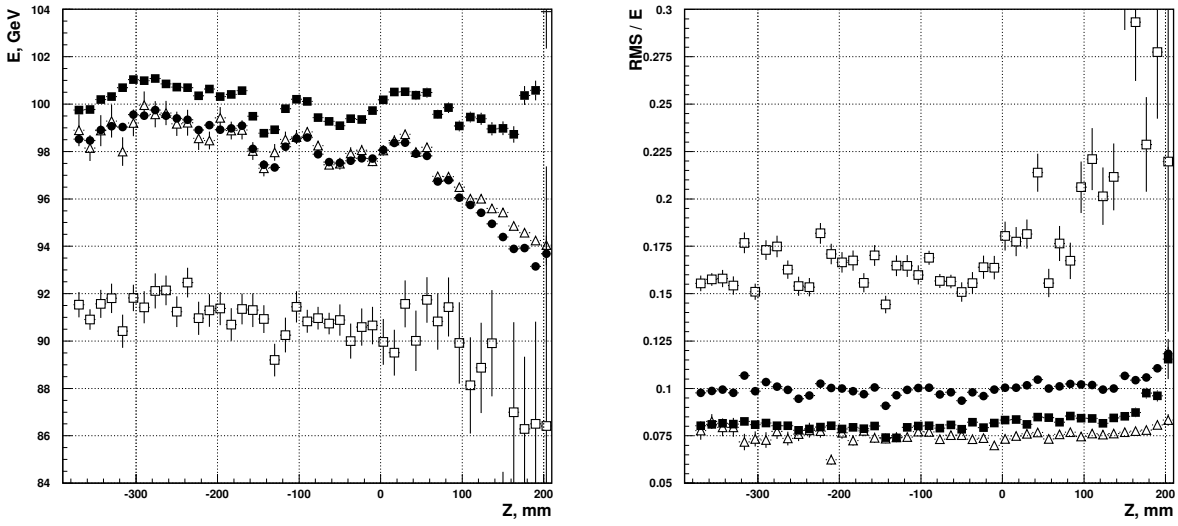


Figure 1.10: The mean energy (left) and energy resolution characterised as the RMS of the measured energy distribution divided by the mean energy, $RMS(E)/E$, (right) as a function of Z for different types of leakage: (a) black squares - no leakage, (b) open squares - longitudinal leakage, (c) open triangles - lateral leakage, (d) black circle - all events [48].

The degradation of the resolution due to longitudinal leakage can be parameterised as [48, 51] using data from the CITF collaboration [52]:

$$\frac{(\sigma_l - \sigma_0)}{\sigma_l} = 0.9 \cdot \sqrt{\frac{\langle E_0 \rangle - \langle E_l \rangle}{\langle E_0 \rangle}} \tag{1.9}$$

where σ_l and E_l are the resolution and mean energy respectively for events with longitudinal hadronic shower leakage, while σ_0 and E_0 are the resolution and mean energy respectively for events without longitudinal hadronic shower leakage. The resolution degradation from Equation 1.9 was found to be within reasonable agreement with the observed degradation in Figure 1.10 [48].

Longitudinal hadronic shower leakage is thus an important consideration in calorimeter design, and also in the subsequent calibrations involved in measuring the energy of the hadronic shower from the energy depositions the calorimeter. The correction described in this thesis aims to remove the affect of longitudinal hadronic shower leakage on the energy response and the energy resolution for hadronic showers measured using jets by ATLAS.

1.4 Monte Carlo simulation of QCD processes

Monte Carlo (MC) event generators are used to simulate the energy and direction of particles produced in pp collisions. Simulations are used to assess the true energy of jets that interact with the detector, allowing their energy response and resolution to be easily measured (see Section 3.4). MC simulations are therefore extremely useful in studying the affect of longitudinal hadronic shower leakage on the measured jet energy, and are used in this thesis to derive a correction to remove these effects. An overview of the MC event generators used in ATLAS can be found in Ref. [53].

The baseline simulation samples used in this thesis are dijet events produced using Pythia8 [54] (version 8.160) with the CT10 PDF set [55] and the underlying-event tune AU2 CT10 [56]. Hadronic showers are simulated with the QGSP BERT model [57]. Multiple parton interactions are simulated separately using Pythia8 as minimum bias events using the A2M tune [56] and the MSTW2008LO PDF [58]. The simulated multiple parton interaction detector signals are then overlaid on-to the hard-scatter detector signals [1].

In this thesis, this Pythia8 sample is referred to as the nominal Pythia8 Monte Carlo sample. The full sample contains approximately 38.9 million simulated jet events. The samples are generated in eight transverse momentum bins (X slices), generated at $\sqrt{s} = 8$ TeV. Each slice is generated with a flat jet transverse momentum spectrum. In order to form the final event sample with the correct shape in the jet transverse momentum, the i-th event

is given the weight w_i:

$$w_i = \frac{w_{MC} \cdot \epsilon_{filter,X} \cdot \sigma_X \cdot \mathcal{L}}{N_{events,X}} \tag{1.10}$$

where w_{MC} is an event-by-event weight designed to restore the correct shape in the transverse momentum spectrum, $\epsilon_{filter,X}$ is the Monte Carlo filter efficiency designed to account for events filtered out during generation of slice X, σ_X is the cross section in nb at which events in slice X were generated, $\mathcal{L}$ is the total luminosity of data in nb^{-1} the Monte Carlo sample is being compared to, and $N_{events,X}$ is the total number of events generated in sample X. These values (excluding w_{MC} which is generated on an event-by-event basis) along with the transverse momentum range of each sample are displayed in Table 1.1.

X Slice	Truth Transverse Momentum Range [GeV]	$\epsilon_{filter,X}$	σ_X	$N_{events,X}$
0	0-20	9.8557×10^{-1}	7.2850×10^{7}	5977995
1	20-80	1.2909×10^{-4}	7.2850×10^{7}	5991878
2	80-200	3.9901×10^{-3}	2.6359×10^{4}	5999034
3	200-500	1.2220×10^{-3}	5.4418×10^{2}	5997214
4	500-1000	7.0839×10^{-4}	6.4453	5977255
5	1000-1500	2.1516×10^{-3}	3.9739×10^{-2}	2991083
6	1500-2000	4.6773×10^{-3}	4.1610×10^{-4}	2988652
7	2000+	1.4595×10^{-2}	4.0636×10^{-5}	2961957

Table 1.1: Parameters for Equation 1.10 for each generated Monte Carlo slice. Values are taken from **A**tlas **M**etadata **I**nterface (**AMI**) [59].

Chapter 2

The LHC and the ATLAS Experiment

The **L**arge **H**adron **C**ollider (**LHC**), situated at CERN in Geneva, is a high centre-of-mass energy proton-proton (pp) collider, designed to push the boundaries of our understanding of the fundemental nature of the universe at the Terascale. The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) experiment is located at interaction point 1 on the LHC [60]. It is a general purpose detector, designed to search for a broad range of new phenomena, discover the Higgs boson, and also provide precision measurements of the standard model processes at the energy frontier of particle physics. At the time of writing, the LHC had just begun 'Run II', with ATLAS recording data from LHC collisions at a centre-of-mass energy of 13 TeV. 'Run I' refers to the data taking period from 2010-2012.

This chapter will briefly describe the LHC in Section 2.1, followed by an overview of the ATLAS detector and its sub-systems in Section 2.2.

2.1 The Large Hadron Collider

The LHC is the worlds largest and most energetic particle accelerator. Situated on the French-Swiss border in a 26.7 km tunnel at a mean depth of 100 meters, it is the result of over 20 years of design, development, construction and testing. The LHC accelerates two beams of protons (or lead ions) circulating in opposite directions to high energies ($\mathcal{O}$(TeV)), bringing

them to collision within detectors situated at four interaction points, depicted in Figure 2.1. LHCb is a specialised b-physics experiment [61], ALICE is a heavy-ion experiment utilising data from heavy (lead) ion collisions [62], and ATLAS [60] and CMS [63] are 'general purpose' detectors designed to study a multitude of different phenomena. The design centre-of-mass collision energy of the LHC is 14 TeV, although for Runs I and II the operational centre-of-mass energy was lower, as will be discussed in Section 2.1.1.

Hydrogen molecules from a standard gas bottle provide the protons that are injected into the CERN accelerator complex displayed in Figure 2.1. To reach the operational centre-of-mass energy, the protons are accelerated through a chain of accelerators: LINAC 2 (50 MeV) $\rightarrow$ **P**roton **S**ynchrotron (**PS**) Booster (1.4 GeV) $\rightarrow$ PS (28 GeV) $\rightarrow$ **S**uper **P**roton **S**ynchrotron, **SPS** (450 GeV) $\rightarrow$ LHC ($\mathcal{O}$(TeV)).

The LHC accelerator contains a total of 9593 magnets, including 1232 dipole magnets to bend the beam, 392 quadropole magnets to focus the beam, and 8 radio-frequency cavities per beam to provide acceleration. For Run I, the main dipoles generate unprecedented 8.4 T magnetic fields, achieved by using superfluid liquid helium to cool the magnets to $< 1.9\ K$.

The number of events, N_i, produced by a process with cross section σ_i is given by:

$$N_i = \sigma_i \int \mathcal{L} dt \tag{2.1}$$

where $\mathcal{L}$ is the instantaneous luminosity. Thus, in order for the LHC experiments to be sensitive to rare processes with low cross sections, the instantaneous luminosity delivered by the collider must be optimised to be as high as possible. The instantaneous luminosity is dependant on the LHC beam parameters:

$$\mathcal{L} = \frac{N_1 N_2 n_b f_r}{4\pi \Sigma_x \Sigma_y} F \tag{2.2}$$

whereby the LHC proton beams are arranged into *'trains'* of *'bunches'* of protons, such that, N_1 and N_2 represent the number of particles per bunch in each beam, n_b is the number of

bunches per beam, f_r is the rotation frequency, F is a reduction factor to account for the crossing angle at the interaction point, and Σ measures the beam size in the x and y plane. The LHC is designed to hold 2808 bunches, each containing up to 1.15×10^{11} protons, with a separation of 25 ns between each bunch. It has a design luminosity of $\mathcal{L} = 10^{34} cm^2 s^{-1}$, achieved with beam parameters set to $\Sigma_{x,y} = 16.7 \mu m$ and $f_r = 11.25$ kHz [64].

Measurements of the luminosity can be made by expressing the luminosity as:

$$\mathcal{L} = \frac{\mu n_b f_r}{\sigma_{inel}} = \frac{\mu^{vis} n_b f_r}{\sigma_{vis}} \tag{2.3}$$

where μ is the number of interactions per bunch crossing and σ_{inel} is the inelastic pp cross section. ATLAS uses event counting to measure μ/σ_{inel} in terms of *'visible'* quantities μ^{vis}/σ_{vis} that take into account inefficiencies in the detector.

μ^{vis} is obtained by counting the number of activated readout channels per bunch crossing via multiple methods, including the LUCID and **B**eam **C**onditions **M**onitor (**BCM**) detectors, timing information from the liquid-argon calorimeter, and hit counts in the muon chambers (see Ref. [65] for a full list). Using multiple methods helps to reduce the uncertainty on μ^{vis}. Combining Equations 2.3 and 2.2 allows σ_{inel} to be measured in terms of $\Sigma_{x,y}$, obtained through **v**an **d**er **M**eer (**vdM**) scans. In a vdM scan, the event rate is monitored while the beams are scanned across each other, first in the x direction and then in the y direction, from which $\Sigma_{x,y}$ can be extracted. The uncertainty in the vdM measurement dominates the uncertainty on the measured luminosity.

Increasing the number of protons per bunch to increase the luminosity also increases the average number of interactions per bunch crossing, μ, known as *in-time pile-up*. Increasing the luminosity by increasing the number of bunches per beam can cause the bunch spacing to be less than the detector latency, resulting in *out-of-time pile-up*. Reducing $\Sigma_{x,y}$ will increase both types of pile-up. Pile-up results in additional uninteresting events that can obscure interesting physics events and degrade detector performance, making it an important consideration when optimising beam parameters to increase luminosity.

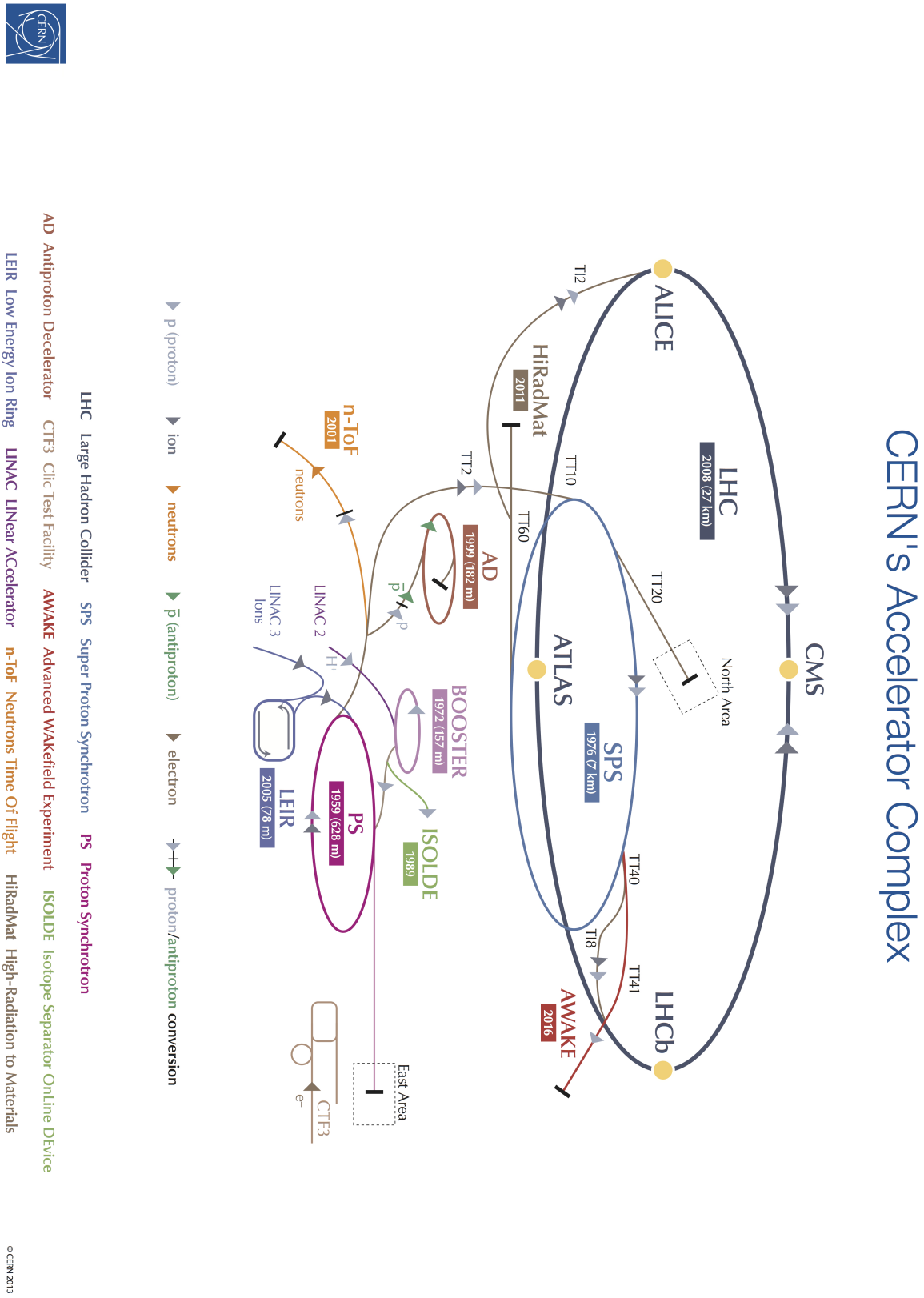


Figure 2.1: The CERN accelerator complex [66]. Yellow circles indicate interaction points.

2.1.1 LHC operating conditions during data collection

The data taking period from 2010-2012 is collectively known as 'Run I', with the LHC having just begun collecting data for 'Run II' at the time of writing. Only data from Run I is used in this thesis. A comparison of the Run I and design operating conditions can be found in Table 2.1 [67].

	Run I			Design
	2010	2011	2012	
Centre-of-mass energy [TeV]	7	7	8	14
Bunch spacing [ns]	150	50	50	25
Peak luminosity [10^{33} cm^{-2} s^{-1}]	0.2	3.6	7.7	10
Delivered luminosity [fb^{-1}]	0.047	5.46	22.8	-
Luminosity uncertainty	3.5%	1.8%	2.8%	-

Table 2.1: LHC operating conditions. For Run II, expected conditions are listed.

Figure 2.2 summarises the measured integrated luminosity and mean number of interactions per bunch crossing in Run I.

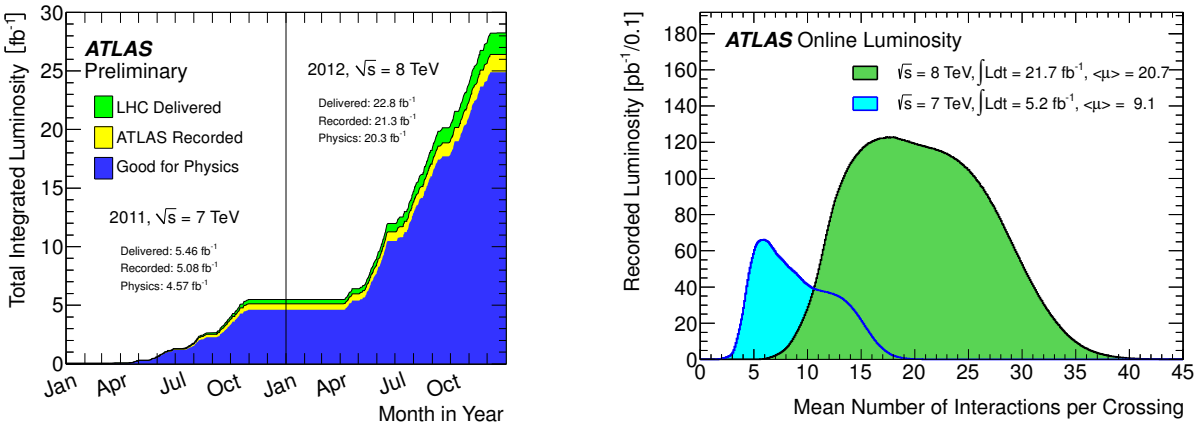


Figure 2.2: Integrated luminosity recorded and pile-up conditions during Run I [68].

2.2 The ATLAS detector

The ATLAS detector is a general-purpose particle detector designed, built and operated by a collaboration of over 2900 physicists from 172 different institutions and universities located

throughout the world. Figure 2.3 displays the layout of the detector.

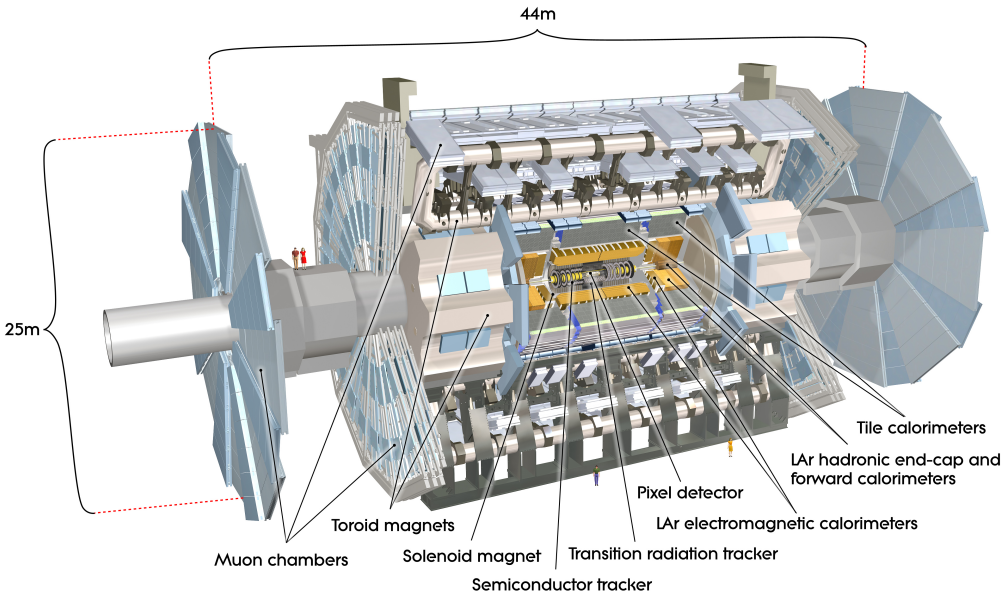


Figure 2.3: Schematic of ATLAS detector [69].

The detector is hermetic, aiming to cover the full 4π steradian range, exhibiting both forward-backward and cylindrical symmetry. It is composed of concentric sub-detector layers that encompass the collision point at their centre.

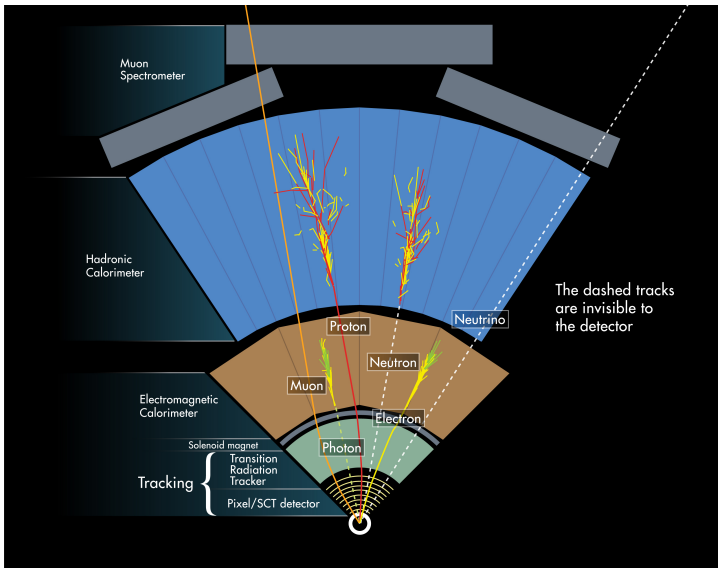


Figure 2.4: Schematic diagram of particles passing through ATLAS detector [70].

Closest to the interaction point is the inner detector, containing the ATLAS tracking system designed to measure the trajectories of charged particles. This is surrounded by the

liquid-argon electromagnetic calorimeter designed to measure the energy of electromagnetic showers, followed by the tile hadronic calorimeter used to measure the energy of hadronic showers. The muon spectrometer forms the outer-most detector layer, used to measure muons. Momenta measurements of charged particles from tracks are made by bending their trajectories using two magnet systems: a 2 T solenoid surrounding the inner detector and a larger 2-8 T toroidal magnet lying outside the calorimeter. A schematic diagram displaying the traversal of various particles through these detector layers is shown in Figure 2.4.

2.2.1 The ATLAS coordinate system

ATLAS uses a right-handed coordinate system with the collision point at its centre. The z-axis runs parallel to the anti-clockwise beam direction, the x-axis points towards the centre of the LHC and the y-axis points upwards. The azimuthal angle, ϕ, is measured in the $x-y$ plane, while the polar coordinate, θ, is measured relative to the z-axis. In most analysis, defining a Lorentz-invariant coordinate axis is useful. As most particles measured will have $E \gg m$, it is convenient to use the pseudorapidity, defined as:

$$\eta = -\ln\tan(\frac{\theta}{2}) \tag{2.4}$$

which is the ultra-relativistic approximation of the rapidity, $y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$, where E is the particle energy and p_z is the z-component of the particle p_{T}. The metric for $\eta - \phi$ space, ΔR, is given by:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.5}$$

2.2.2 The Inner Detector

The ATLAS **I**nner **D**etector (**ID**), displayed in Figure 2.5, is designed to provide high-resolution tracking information achieved through fine detector granularity.

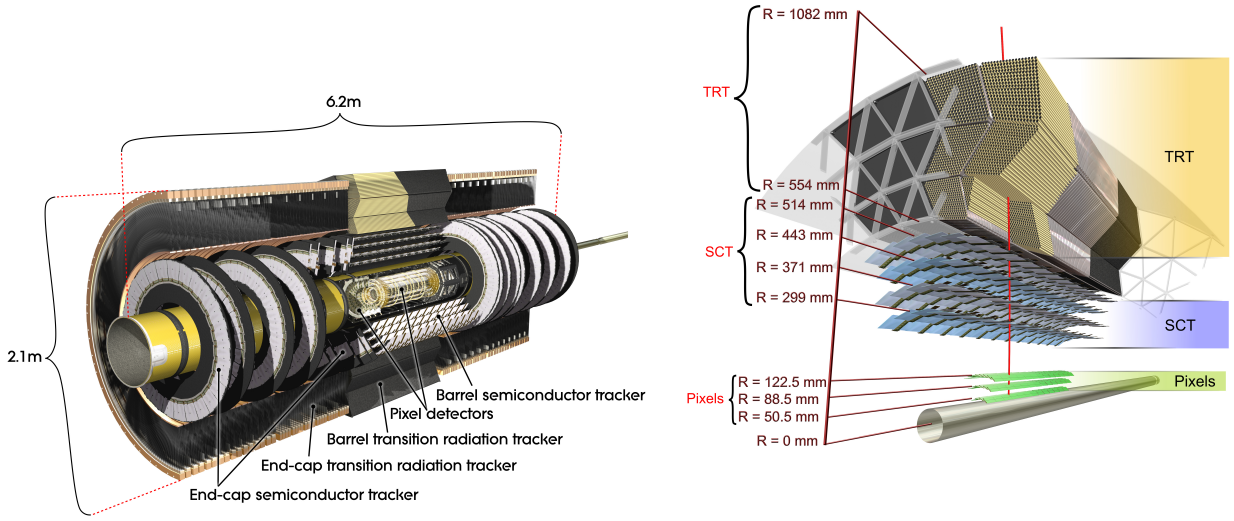


Figure 2.5: Diagram displaying ATLAS Inner Detector [71].

The solenoid magnetic field enables the track momenta to be measured from the curvature of the track, allowing the primary vertex to be identified as the vertex with the highest sum of transverse momentum (p_{T}) of tracks pointing towards it, normalised by the number of tracks. High-resolution tracking also allows the transverse (d_0) and longitudinal (z_0) impact parameters to be precisely determined as the distance between the track and the beamline at perigee[1] and the z-coordinate of the track at perigee respectively. The ID covers the region $\eta < 2.5$, consisting of three concentric sub-detectors:

1. **The Pixel detector** is constructed of 1744 pixel modules installed in three layers closest to the beam pipe. Each module has 47,232 pixels, with 90% (10%) of pixels having area $50\mu m - 400\mu m$ ($50\mu m - 600\mu m$) and thickness $250\mu m$. This high pixel density provides an intrinsic accuracy of $10\mu m \times 115\mu m$ in the $R - \phi$ ($R - R$) plane for the barrel (end-cap) region [72].

2. **The Semiconductor tracker (SCT)** consists of 15,912 silicon strip sensors surrounding the Pixel detector, each with 770 strips. Pairs of sensors are aligned in modules with a small stereo angle of 40 mrad, allowing the measurement of a two-dimensional hit coordinate. There are four concentric barrels and nine end-cap disks, ensuring all tracks with $|\eta| < 2.5$ will pass through a minimum of four SCT layers. These features

[1]The perigee point is the point along the track which is closest to the beam-line.

allow for an intrinsic accuracy of $17\mu m \times 590\mu m$ in $R-\phi$ $(R-R)$ space for the barrel (end-cap).

3. **The Transition radiation tracker (TRT)** consists of 370,000 drift chambers, known as straws, surrounding the SCT, that simultaneously act as a particle tracker and a transition radiation detector to aid in electron identification. The $4mm$ diameter straws are arranged parallel to the beam-line in the barrel and radially in the end-cap. The TRT covers the region $|\eta| < 2.0$, only providing $R-\phi$ position information, with an intrinsic accuracy of $130\mu m$ per straw.

2.2.3 The ATLAS calorimeter system

The ATLAS calorimeter system, displayed in Figure 2.6, is used to measure the energy and position of final state particles through energy depositions of showers in the calorimeter. The calorimeter system employs a sampling technique, whereby layers of active and passive material are sandwiched together to alternate absorption and measurement of energy, allowing for a more compact design with almost complete containment of the particle showers, collectively covering $|\eta| < 4.9$.

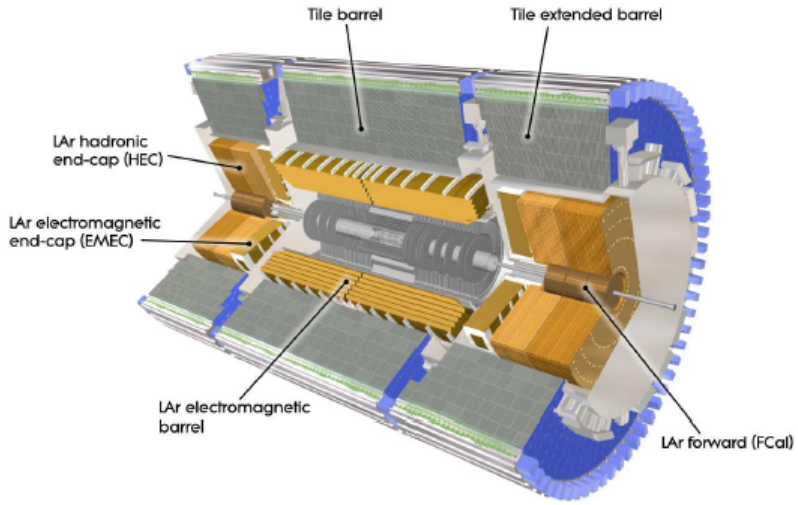


Figure 2.6: Diagram displaying ATLAS calorimeter system [73].

In order to account for the different properties of electromagnetic and hadronic showers, the ATLAS calorimeter is composed of two subsystems - the **E**lectro**M**agnetic (**EM**) and

Hadronic **Cal**orimeters (**ECAL** and **HCal** respectively). As mentioned in Section 1.3.2, the ATLAS calorimeters are non-compensating, with the correct hadronic scale restored via the *Jet Energy Scale calibration*, described in Section 3.5.

In order to ensure the calorimeters provide good containment of electromagnetic and hadronic showers, with minimal probability for longitudinal shower leakage into the muon systems, the ECAL has thickness $\sim 22X_0$ ($\sim 24X_0$) in the barrel (end-cap), while the HCal has thickness $\sim 9.7\lambda_{\mathrm{I}}$ ($\sim 11\lambda_{\mathrm{I}}$) in the barrel (end-cap). A distribution of the calorimeter material as a function of η can be seen in Figure 2.7. It will be shown in Section 4.1 that the ATLAS calorimeters still display a probability for hadronic longitudinal shower leakage, which is highly dependant on this distribution of detector material.

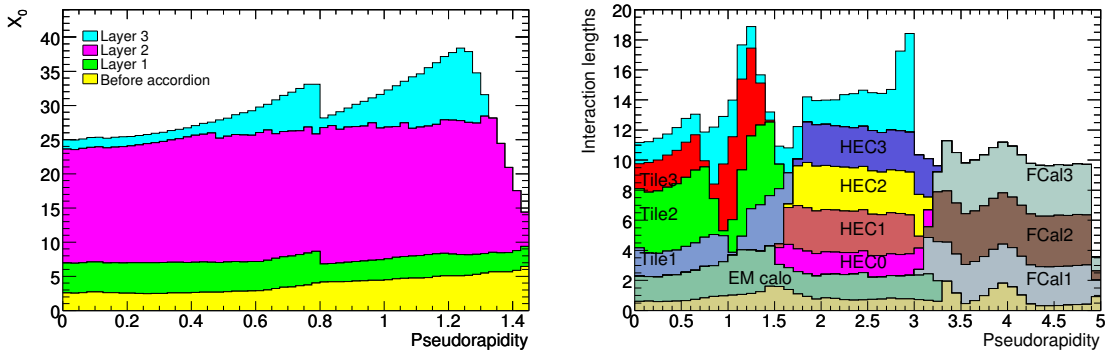


Figure 2.7: Distribution of EM (combined) calorimeter material in units of X_0 (λ_I) vs η_{det} is shown on the left (right). Numbers indicate different detector layers. Turquoise in the right plot indicates muon spectrometer [74].

Electromagnetic calorimeters

The ATLAS electromagnetic calorimeters, commonly known as the *LAr* calorimeters, use liquid-argon as the active material and lead as the absorber that induces showering within the calorimeter. The LAr barrel consists of two identical cylindrical components collectively covering $|\eta| < 1.475$, while the end-caps are separated into two coaxial wheels covering $1.375 \leq |\eta| < 2.5$ and $2.5 \leq |\eta| < 3.2$. The region between the forward and barrel components ($1.375 \leq |\eta| < 1.52$), is the poorly instrumented *'crack'* region, often excluded when reconstructing electrons.

The LAr calorimeter barrel, displayed in Figure 2.8, has an accordion structure designed to avoid azimuthal cracks, provide full ϕ symmetry and allow for fast readout. It is segmented into three layers, each with decreasing granularity. The first layer, used to reconstruct the η position of the shower and provide particle identification, has a granularity of $\Delta\eta \times \Delta\phi = 0.0031 \times 0.098 = 4.69mm \times 147.3mm$. The second (third) layer has a granularity of $\Delta\eta \times \Delta\phi = 0.0245 \times 0.0245 = 37.5mm \times 36.8mm$ ($\Delta\eta \times \Delta\phi = 0.0245 \times 0.05 = 37.5mm \times 75.1mm$), collecting the bulk (tail) of the electromagnetic shower.

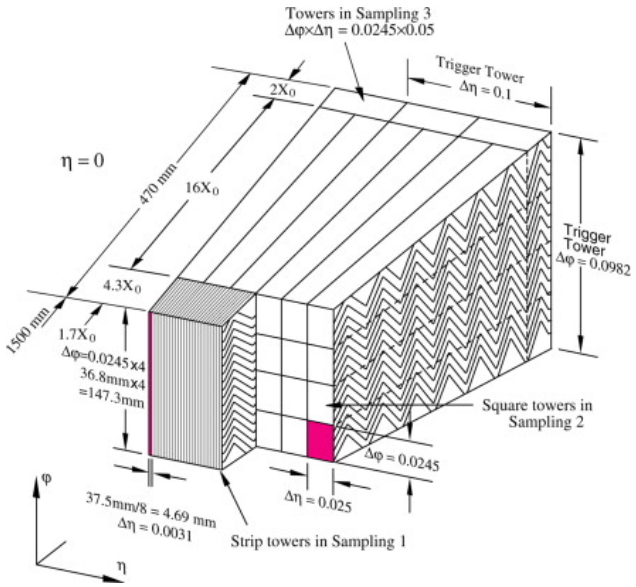


Figure 2.8: Segmentation of electromagnetic barrel calorimeter [73].

Prior to the calorimeters, in the region $|\eta| < 1.8$, particles traverse an 11 (5) mm layer of liquid-argon forming the barrel (end-cap) presampler. Energy from the presampler is readout separately from the rest of the LAr calorimeter, and is weighted to recover energy losses due to particles traversing inactive material upstream of the LAr calorimeter.

Hadronic calorimeters

The ATLAS hadronic barrel calorimeter, commonly known as the *Tile* calorimeter, is split into the barrel and extended barrels covering the regions $0 \leq |\eta| < 0.8$ and $0.8 \leq |\eta| < 1.7$ respectively (displayed in Figure 2.9). There is a gap between the extended tile barrel and tile barrel calorimeter of ~ 600 mm in the region $0.8 < \eta < 1.0$, which is instrumented by the **I**ntermediate **T**ile **C**alorimeter (**ITC**). The ITC consists of a calorimeter in the region

$0.8 < |\eta| < 1.0$ (cells D4 and C10 in Figure 2.9) and scintillators between $1.0 < |\eta| < 1.6$ (E1-E4 in Figure 2.9).

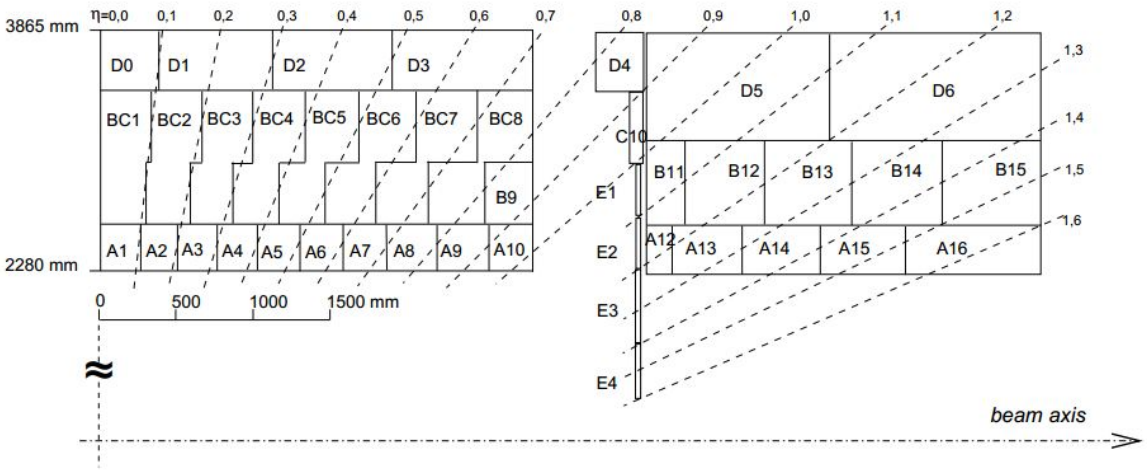


Figure 2.9: Segmentation of tile calorimeter [73].

The Tile calorimeter consists of three layers, with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two layers and $\Delta\eta \times \Delta\phi = 0.1 \times 0.2$ in the final layer. It employs plastic scintillator as its active material and steel for the absorber. The plastic tiles are staggered in depth and connect to photomultiplier tubes via optical fibres at both sides. Particles depositing energy in the scintillators produce scintillation light with intensity proportional to the deposited energy. The light is collected using wavelength-shifting-fibres and funnelled into the photomultiplier tubes to be readout.

Hadronic calorimetry is extended to the forward region via the **H**adronic **E**ndcap **C**alorimeter (**HEC**), which covers $1.5 \leq |\eta| < 3.2$. It is important to note that the transition from the barrel to the HEC is poorly instrumented, and likely to have an increased probability for longitudinal hadronic shower leakage. In order to deal with the larger flux of higher energy particles in the forward region, the HEC employs liquid-argon as its active material and copper as the absorber. The HEC is split into two layers, with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ in the second.

Forward calorimeter (FCal)

Calorimetry is extended to the region $3.1 \leq |\eta| < 4.9$ by the forward calorimeter, commonly known as the *FCal*. It is positioned 1.2 m from the front face of the ECAL in order to reduce the neutron flux, and has a depth of $\sim 10\lambda_{\mathrm{I}}$. As with the HEC, liquid-argon is used as the active material to deal with the high particle fluxes. The FCal is split into three layers, the first of which employs copper as its absorber, while the last two use tungsten.

2.2.4 The ATLAS muon spectrometer

The ATLAS muon spectrometer (displayed in Figure 2.10) is designed to measure the momentum of muons in the region $|\eta| < 2.7$, and provide trigger information in the region $|\eta| < 2.4$. Information from the muon spectrometer is used in this thesis to study the impact of longitudinal hadronic shower leakage.

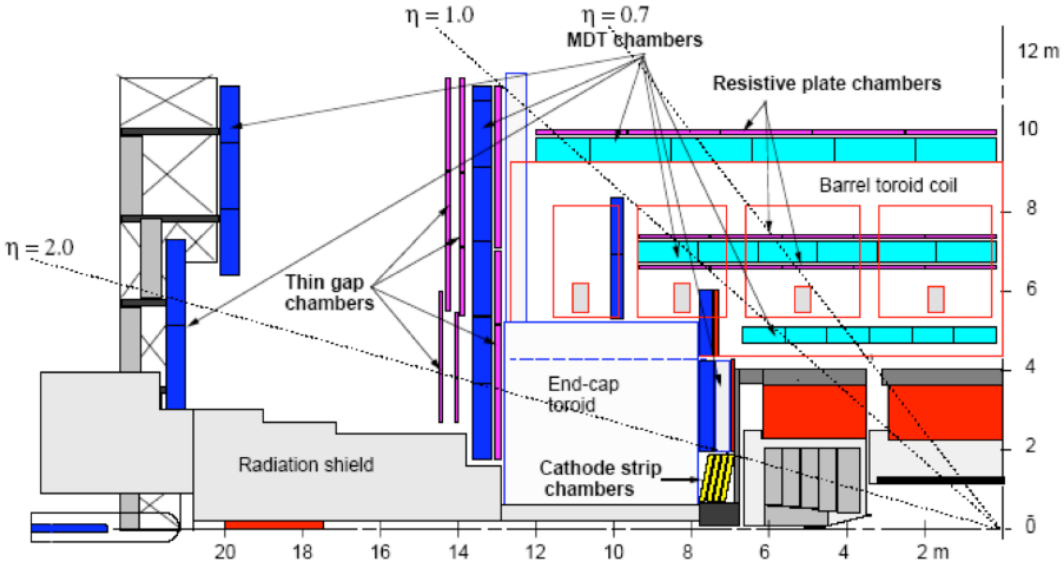


Figure 2.10: Schematic diagram of muon spectrometer [73].

Momentum measurements are made by deflecting muons with the help of large air-core toroid magnets, three in the barrel at 0.5 T and two in the end-cap at 1 T. The muon system consists of four types of tracking chambers installed in three concentric cylindrical shells around the beam axis in the barrel regions ($|\eta| < 1.0$) and four large wheels perpendicular

to the z axis in the end-cap region ($1.0 \leq |\eta| < 2.7$) [73]:

- **Cathode Strip Chambers (CSC)** complement MDTs by providing additional precision tracking in the forward region ($2.1 < |\eta| < 2.7$), where higher particle fluxes require finer granularity. They are multi-wire proportional chambers providing both R and ϕ coordinate information, with a resolution of $\sim 60\mu m$.

- **Resistive Plate Chambers (RPC)** provide bunch-crossing identification and well defined p_{T} thresholds for triggering, as well as additional tracking information, in the $|\eta| < 1.05$ region. They are gaseous parallel electrode-plate detectors operated in avalanche mode. Each chamber has two layers, providing η and ϕ coordinate measurements. Coincidences between tracks in consecutive layers are used to reject background.

- **Thin Gap Chambers (TGC)** provide trigger capability and determination of the ϕ coordinate to complement MDTs in the $1.05 \leq |\eta| < 2.4$ region. They are multi-wire proportional chambers which operate in 'quasi-saturated' mode.

- **Monitored Drift Tubes (MDT)**, shown in Figure 2.11, provide precision tracking in the region $|\eta| < 2.7$ via proportional drift tubes filled with Ar-CO_2. The chambers form three layers in the barrel and four in the end-cap, as visible in Figure 2.10. The tubes are 3 cm in diameter with an average resolution of $\sim 80\mu m$. Each chamber contains two multi-layers of three (four) drift-tubes in the middle/outer (inner-most) layers. The MDTs are essential in reconstructing information relating to muon activity behind jets that are leaking into the muon spectrometer as a result of longitudinal hadronic shower leakage, as will be explained in Section 4.1.

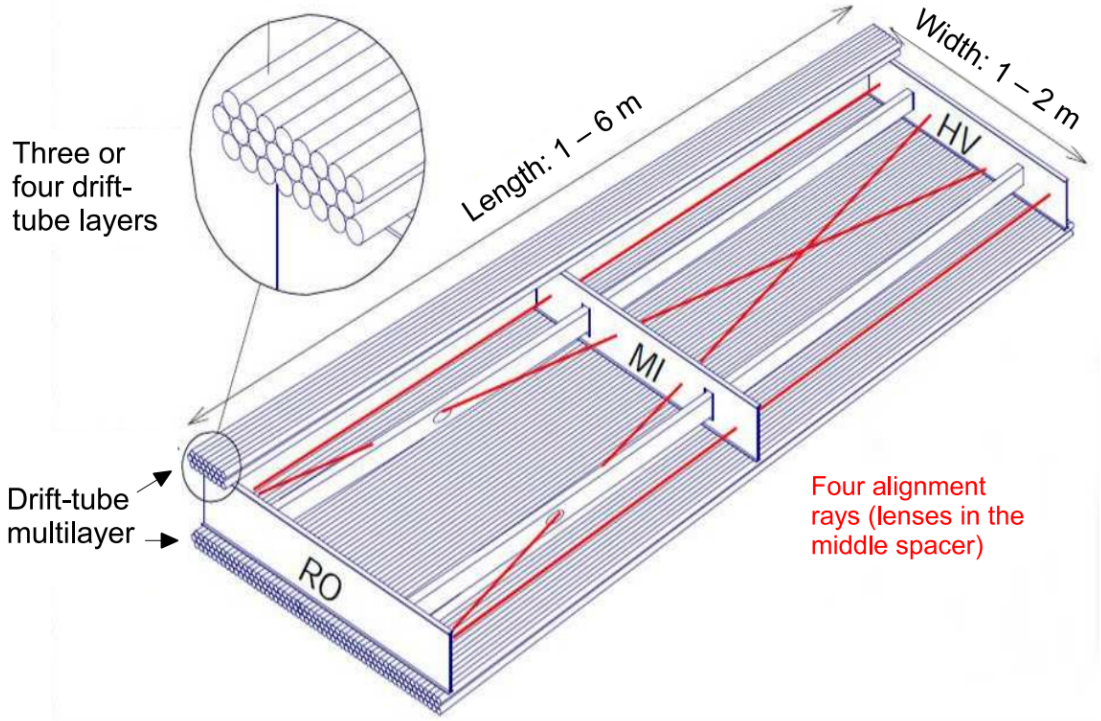


Figure 2.11: Schematic diagram of muon Monitored Drift Chambers [73].

2.2.5 Trigger and Data Acquisition

Technological limitations mean it is only possible to record data at a rate of 200Hz, however the LHC delivers data at a rate of up to 1 GHz (closer to 20 MHz during 2012 [75]), with the majority of these events being uninteresting. In order to select interesting events for analysis offline, ATLAS employs a *trigger system*. The trigger system is designed to reject background without biasing selection.

The ATLAS trigger system consists of three levels: **L**evel **1** (**L1**), **L**evel **2** (**L2**) and the **E**vent **F**ilter (**EF**). The L1 trigger is a hardware base trigger installed onto the detector, while the L2 and EF are software based triggers collectively referred to as the **H**igher **L**evel **T**rigger (**HLT**), implemented on an off-detector computer farm. Triggers are organised into *trigger configurations* (or *trigger menus*), composed of *trigger chains* containing a sequence of triggers from L1 to the EF. Trigger chains may be *prescaled* in order to reduce the event rate for processes with high cross-sections, with prescales adjusted based on the current LHC luminosity and physics program[2]. An overview of the trigger system is shown in Figure 2.12.

[2]A trigger with prescale p will randomly select events with a probability of $1/p$, reducing the event rate.

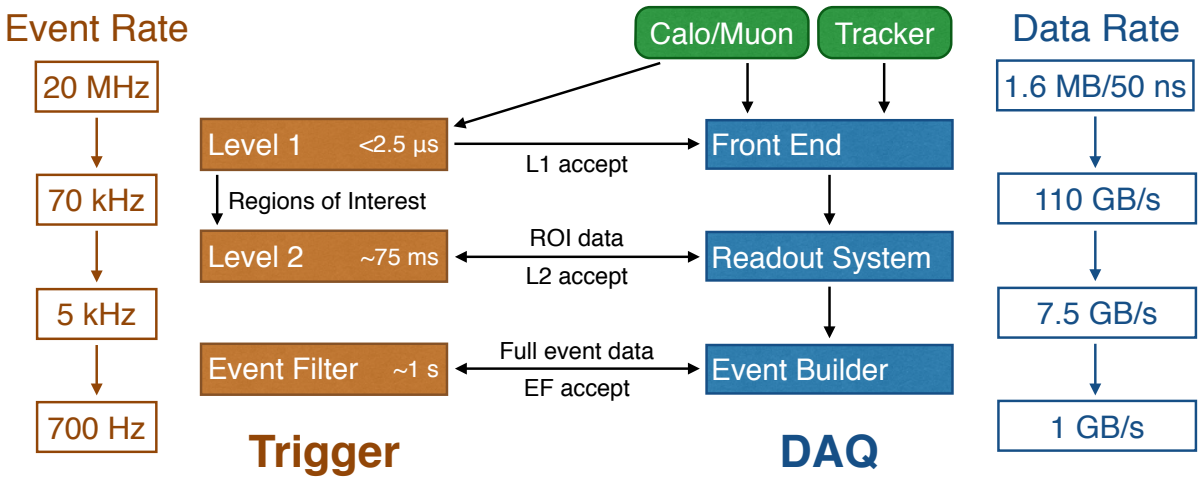


Figure 2.12: Overview of ATLAS trigger and Data Acquisition (DAQ) system [76].

The L1 trigger uses reduced granularity information from a range of detector subsystems to search for signatures from high-p_{T} muons (using the RPCs and TGCs), as well as electrons/photons, jets[3], τ-leptons decaying into hadrons, large missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$)[4] and large transverse momentum ($\sum \mathrm{E}_T$) (using the calorimeter sub-systems). The acceptance rate of the L1 trigger is 75 kHz, with a latency of 25 μs [60].

The L2 trigger is seeded by coordinates of possible trigger objects identified by the L1 trigger, known as **R**egions-**O**f-**I**nterest (**ROI**). Information on the energy and types of signatures (roughly 2% of the total event information) is used to limit the data readout to an event rate below 3.5 kHz, with an average latency of 40 ms. Finally the EF uses reconstructed physics objects containing information close to the fully reconstructed offline physics objects to make a decision, reducing the event rate to 200 Hz with a latency of ~ 4 s. The full trigger system reduces the data retention rate from 64 TB/s to roughly 1 GB/s.

For studies in this thesis, events are selected using single jet triggers. An overview of the ATLAS 2012 jet trigger can be found in Figure 2.13. At L1, decisions are made using the raw calorimeter energy in calorimeter towers (see Section 3.3) of size $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. A sliding window algorithm is used to select local maxima, and if the maxima has E_T above threshold the event is retained, with the ROI passed to L2. In 2012 an additional L1.5 trigger,

[3] See Chapter 3 for an overview of jets and their reconstruction.
[4] $E_{\mathrm{T}}^{\mathrm{miss}}$ is constructed as the vector sum of the momenta of all objects originating from the primary vertex, assuming the initial momentum in the transverse plane is zero.

also known as the **L2 F**ast **S**can (**L2FS**) trigger, was implemented for certain trigger chains. The L1.5 trigger reconstructs anti-k_t jets using all of the L1 trigger towers as inputs, as opposed to just the trigger towers falling inside the ROI, and passes these anti-k_t jets to L2.

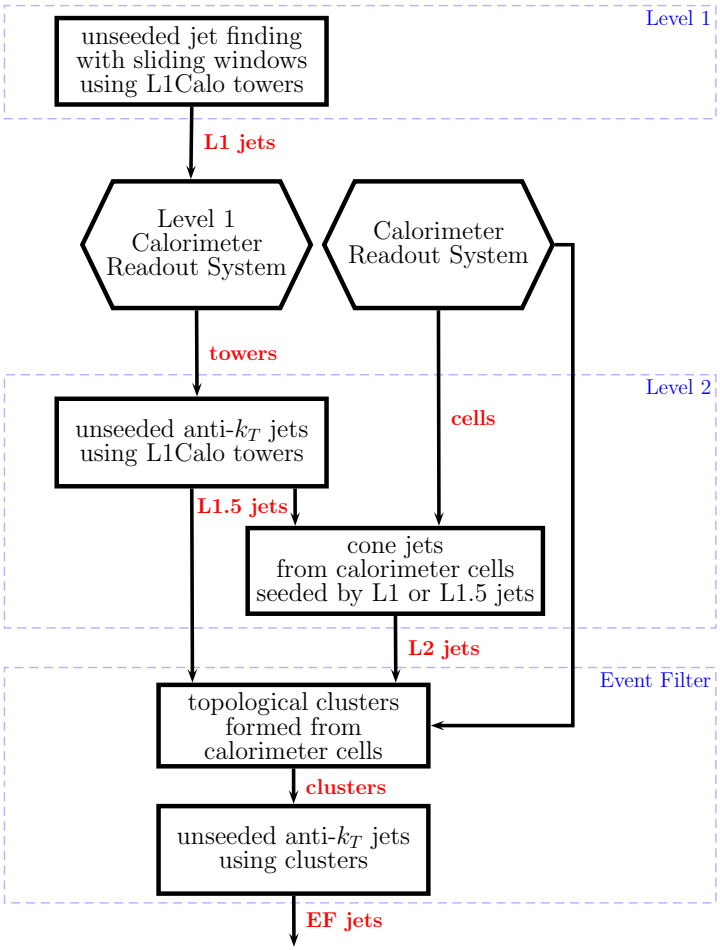


Figure 2.13: Overview of ATLAS 2012 jet trigger [77].

At L2, anti-k_t and cone jets are reconstructed using the L1 or L1.5 trigger inputs as seeds, the energy of which are used to filter events passed to the EF. Decisions at the EF are then performed using fully reconstructed jets. The jet trigger system is split into two independent subsystems - the central jet triggers ($|\eta| < 3.2$) and the forward jet triggers ($3.2 \leq |\eta| < 4.9$). Please see Chapter 3 for a full description of jets and their reconstruction.

Jet triggers are characterised by their 'turn-on' curves which represent the trigger's efficiency (ϵ) as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, whereby the efficiency is defined to be N_{pass}/N_{probe}, where N_{pass} is the number of events that pass the trigger and N_{probe} is the number of events that pass some fully efficient probe trigger. For analyses, triggers are only used in their 'plat-

eau' region, defined to be the region > 99% efficiency, determined from fits to the efficiency curves. Figure 2.14 displays some example trigger efficiency curves as a function of $p_{\mathrm{T}}^{\mathrm{avg}}$ ($p_{\mathrm{T}}^{\mathrm{avg}} = \frac{p_{\mathrm{T}}^{lead}+p_{\mathrm{T}}^{subl}}{2}$, where p_{T}^{lead} and p_{T}^{sub} are the p_{T} of the leading and subleading jet in the event respectively.).

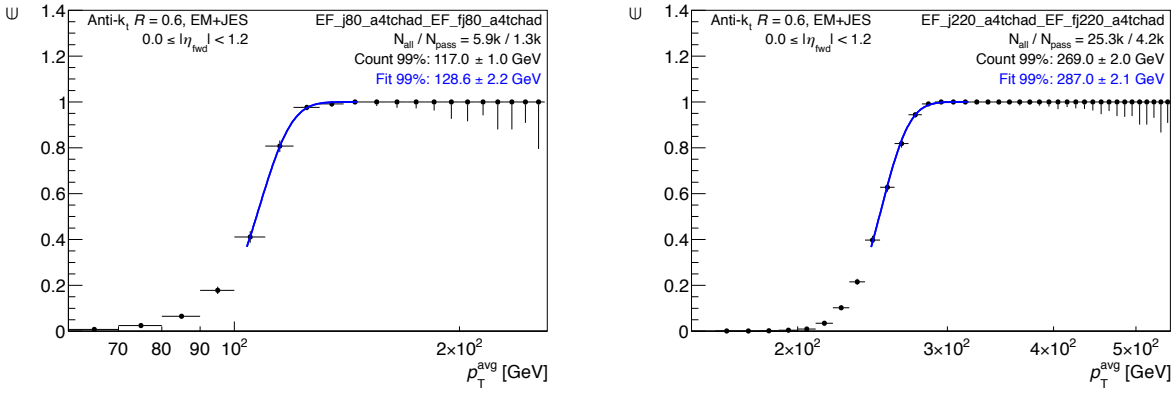


Figure 2.14: Example trigger efficiency curves when requiring the central `j` `OR` the forward `fj` trigger has passed, for a low (left) and high (right) $p_{\mathrm{T}}^{\mathrm{jet}}$ trigger [78].

2.2.6 Overall detector performance

Table 2.2 summarises the overall performance of the ATLAS detector.

Detector subsystem	Resolution	$\lvert\eta\rvert$ coverage	
		Measurement	L1 trigger
Tracker	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 0.03\% \times p_{\mathrm{T}} \oplus 1.5\%$	< 2.5	-
EM calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$	< 3.2	2.5
Hadronic calorimeter	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	< 4.9	4.9
Muon spectrometer	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 4\%$ at $p_{\mathrm{T}} = 100$ GeV	< 2.7	2.4

Table 2.2: The measured performance in the central region of each of the ATLAS sub-detectors [79–82]. Energies and momenta are given in GeV. The quoted MS performance is independent of the ID. The contribution of noise to σ_E/E (a term proportional to $1/E$) is neglected as it depends upon the pile-up environment. Full table taken from [76].

2.3 Simulation of the ATLAS detector

GEANT4[83] combined with the ATLAS simulation framework [84] is used to simulate the ATLAS detector. Depositions in the detector are converted into detector signals and reconstructed with the same format as the ATLAS detector read-out and online software. Simulation is tuned using test-beam data and in-situ measurements. In addition to the full GEANT4 simulation, an ATLAS fast simulation procedure, known as Atlfast-II (AFII), designed to reduce the computational overhead required to produce MC samples, is also implemented. AFII makes use of look-up tables to simulate correlations between variables in the detector. Please see Reference [85] for more information. For all studies in this thesis, only full simulation samples were used, as in 2012 AFII did not simulate the correlation between energy deposits in the calorimeter and hits behind jets in the muon spectrometer.

2.4 Data used in this thesis

Data recorded by ATLAS during 2012 at $\sqrt{s} = 8$ TeV is used in this thesis. Data is divided into *data periods*, *runs* and *luminosity blocks*. Only data deemed reliable for physics analysis were used. In addition, data events are required to have at least one collision vertex with 2 or more tracks, with $p_{\mathrm{T}}^{\mathrm{track}} > 150$ MeV. The longitudinal position of the vertex is required to be within 10 cm of the detector centre. In the presence of multiple pp collisions, the collision vertex with the largest scalar sum of p_{T}^2 of its associated tracks is labelled the primary vertex.

Chapter 3

Jet Reconstruction

It was mentioned in Section 1.3.2 that QCD confinement results in final state quarks and gluons produced from proton-proton scattering manifesting as groups of collimated particles via hadronic showers. In detectors, hadronic showers are clustered into physically meaningful objects called jets. In order for meaningful measurements to be made using jets that can be compared to theoretical predictions, the concept of a jet must be well defined. A schematic representation of a jet is displayed in Figure 3.1. This Chapter details the definition, reconstruction and the initial ATLAS 2012 calibration scheme for jets, prior to the development of the jet punch-through correction in this thesis.

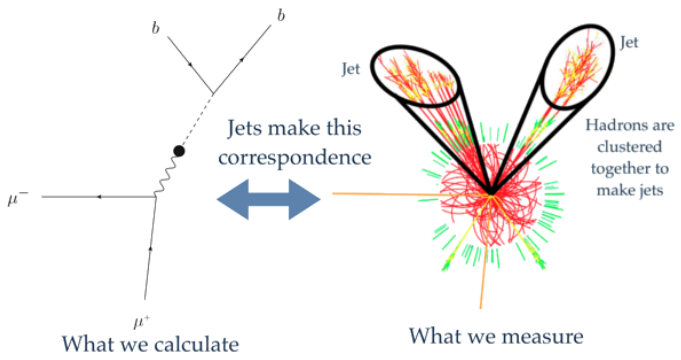


Figure 3.1: Schematic of hadrons from a hard collision clustered into a jet object [86].

The specifications that a jet definition must meet are outlined in Section 3.1. Jet definitions used by ATLAS are given in Section 3.2.1. ATLAS inputs to these jet definitions

are listed in Section 3.3. A method to evaluate the calorimeter jet energy response through comparisons with simulated particles is described in Section 3.4. The initial ATLAS 2012 jet energy scale calibration scheme, designed to obtain jets that are correctly reconstructed at the hadronic scale, whose sensitivity to longitudinal hadronic shower leakage is studied in this thesis, is explained in Section 3.5. Section 3.6 then describes the jet cleaning cuts used in this thesis to ensure only real jets unaffected by detector malfunctions are studied. Techniques used to identify the flavour of the jet initiating parton are discussed in Section 3.7.

3.1 Specifications for a robust jet definition

According to the 2007 Les Houches accord [87], a jet definition must comprise of:

- A **jet algorithm** that clusters energy deposits or particles into a single jet object.
- A **recombination scheme** that defines how the four-momenta of the jet is obtained from the sum of its constituents.
- A full specification of the **jet algorithm parameters**.

A robust jet definition must meet certain theoretical and experimental requirements, discussed in detail in [88–91]. A brief overview of these requirements can be found below.

For a jet algorithm to be *theoretically well behaved*, it must be insensitive to QCD scale effects, giving finite perturbative results at all orders, allowing for meaningful comparisons of data with the most recent theoretical predictions. This is achieved by ensuring the algorithm is insensitive to QCD singularities (described in Section 1.2.3) resulting from:

- *Infra-red* emissions, manifesting via the radiation of very low p_{T} (soft) particles. Jet algorithms insensitive to infra-red emissions are said to be *infra-red safe*. See Figure 3.2 for a schematic example of an infra-red unsafe algorithm.

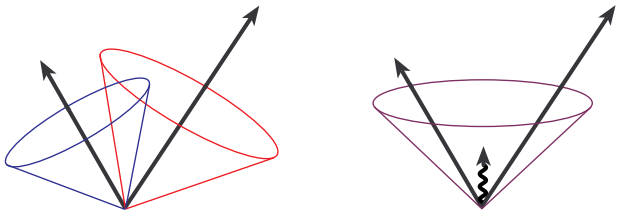


Figure 3.2: Example of an infra-red unsafe jet algorithm producing two jets without (left) and one jet with (right) an infra-red emission present [88].

- *Collinear* emissions, manifesting as particles radiated at a very small angle with respect to the original parton. Jet algorithms insensitive to collinear emissions are said to be *collinear safe*. See Figure 3.3 for a schematic example of a collinear unsafe jet algorithm.

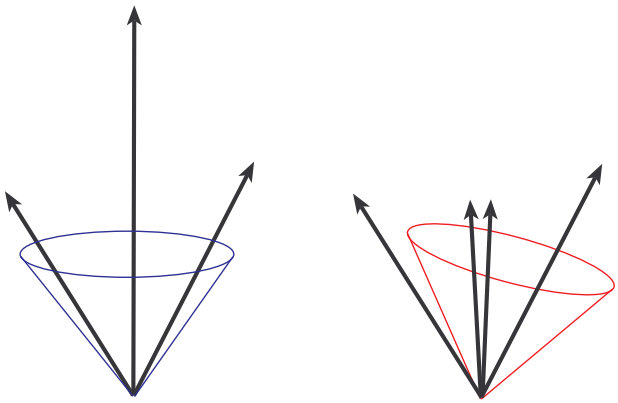


Figure 3.3: Example of a collinear unsafe jet algorithm, whereby the jet axis is shifted in the presence of a collinear emission (right) when compared to the case without one (left) [88].

A jet algorithm must also be *order independent*, reproducing the same physical results irrespective to whether the inputs are partons, particles or calorimeter objects. In addition, it is desirable for the jet algorithm to be *invariant under longitudinal boosts* along the beam axis, aiding comparisons between the CM and lab frame.

Finally, experimental considerations such as the implementation and computational costs, ease of calibration, as well as resilience towards effects such as calorimeter noise and pile-up, must be considered. `FastJet`[92, 93] is used to employ the ATLAS jet reconstruction software in the `ATHENA` framework.

3.2 Jet definition in ATLAS

As mentioned in Section 3.1, a jet definition comprises of a *jet algorithm* and a *recombination scheme*. This section will go into detail on the jet finding algorithms and recombination schemes used by ATLAS.

3.2.1 Jet algorithms used in ATLAS

Jet algorithms help to reduce the complexity of the final state, combining a large number of final-state particles into a few jets. There are two broad classes of jet finding algorithms [94] employed within ATLAS: *cone algorithms* and *successive recombination algorithms*. Each have their own theoretical and experimental strengths that dictate their area of application.

Cone algorithms

Jets are defined as dominant directions of energy flow. For the ATLAS jet trigger software, a simple seeded cone algorithm - the *ATLAS Cone* algorithm - is employed at L2 for fast jet reconstruction (see Section 2.2.5). The *ATLAS Cone* algorithm can be defined as a *seeded iterative* cone algorithm [95], whereby the algorithm starts at some *seed* object that sets some initial direction, around which a cone of radius R in $\eta - \phi$ space is drawn and the four-momenta of all particles in this cone summed. The resultant sum is used as the new seed direction, and the process is iterated until the direction of the resultant cone is stable (does not change significantly between successive iterations). The seed is chosen to be the highest p_{T} input and must be above the 1 GeV threshold.

The resultant jets may overlap, sharing constituents. In order to rectify this, a *split-merge* step is performed in which two jets are merged if more than a fraction (f) of the p_{T} of softer jet is overlapped with the harder jet, otherwise the shared constituents are given to the closer jet. For ATLAS, $f = 0.5$.

The issue with this algorithm is that while being collinear safe, it not infra-red safe.

However, the low computational time associated with this algorithm make it useful for jet finding when making trigger decisions, where fast computational times are essential.

Sequential recombination algorithms

Jets are formed by defining a distance between pairs of input objects (d_{ij}) and performing successive recombinations of the pair of closest input objects, only stopping when all resulting objects are farther apart than some threshold distance, set by the distance between the object and the Beam (d_{iB}[1]). Common d_{ij} and d_{iB} parametrisations are shown in Equations 3.1 and 3.2 respectively:

$$d_{ij} = min(\mathrm{p}_{T,i}^{2\rho}, \mathrm{p}_{T,j}^{2\rho})\frac{\Delta R_{ij}}{\mathrm{R}^2}, \tag{3.1}$$

$$d_{iB} = \mathrm{p}_{T,i}^{2\rho}, \tag{3.2}$$

where i and j are indices of the objects under consideration, p_{T} is the transverse momentum of the object, ΔR_{ij} is the spatial distance between the objects in $\eta - \phi$ space, ρ is a parameter that governs the relative power between the energy and geometric scales, and R is a pre-defined distance parameter - larger R values result in wider jets.

Three commonly used values of ρ are -1, 0 and $+1$, which give the anti-k_t [96], Cambridge/Aachen [97] (C/A) and k_t [98] algorithms respectively. The differences in these three algorithms are down to the momentum weighting and constituent ordering. The anti-k_t and k_t algorithms are p_{T} ordered, whereby the anti-k_t algorithm clusters begins with high p_{T} constituents, while the k_t algorithm does the opposite, beginning with low p_{T} constituents (effectively reversing the QCD parton shower). This is contrasting to the C/A algorithm that is angular-ordered, relying only on distance weighting.

Figure 3.4 shows an example of the anti-k_t, k_t and C/A jet algorithms with $R = 1$,

[1] B represents *Beam*

clustering random soft 'ghost' particles into jets in a simulated event. The anti-k_t algorithm gives circular jets, whereas the C/A and k_t algorithms give somewhat irregular shaped jets.

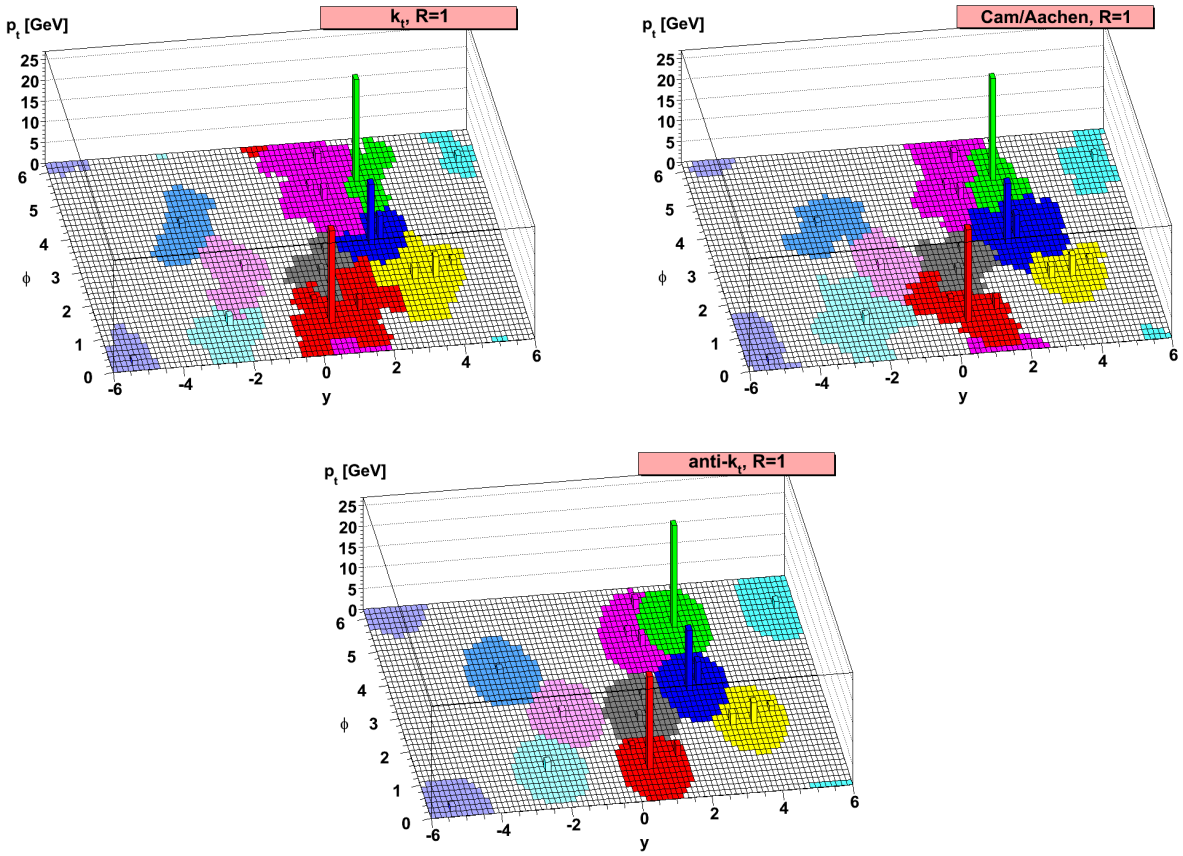


Figure 3.4: A sample parton-level event, together with many random soft "ghosts", clustered with three different jet algorithms - k_t (top-left), C/A (top-right), and anti-k_t (bottom). The distance parameters (R) are set to 1. Illustrated are the "active" catchment areas of the resulting hard jets [96].

The sequential recombination algorithms are all infra-red and collinear safe, and require no split/merge step. Anti-k_t jets are preferred for ATLAS analyses as they demonstrated the best overall performance in the detailed studies performed in [99]. In addition, their circular shape allows for a well defined jet area useful in pile-up subtraction techniques (see Section 3.5.2). One downside of the anti-k_t jet algorithm is the sensitivity to calorimeter noise, however this is compensated for through the use of noise suppressed inputs, as detailed in Section 3.3. Anti-k_t jets with R = 0.4 and R = 0.6 are the focus of the calibration studies in this thesis, and were the most widely used by ATLAS analyses during Run I.

3.2.2 Recombination scheme

ATLAS uses the *four-vector recombination scheme* to obtain the final jet four-momenta from the sum of its components four-momenta, producing a massive jet. This is the most widespread recombination scheme [95] as recommended by [88].

In the *four-vector recombination scheme*, the jet four-momenta is calculated as the sum of the four-momenta of its constituents:

$$P_{jet} = \sum_{i=1}^{N_{constit}} P_i \tag{3.3}$$

where $N_{constit}$ is the total number of constituents that form the final jet, P_i is the four-momenta of the i-th constituent, and P_{jet} is the final jet four-momenta.

3.3 ATLAS inputs to jet finding algorithms

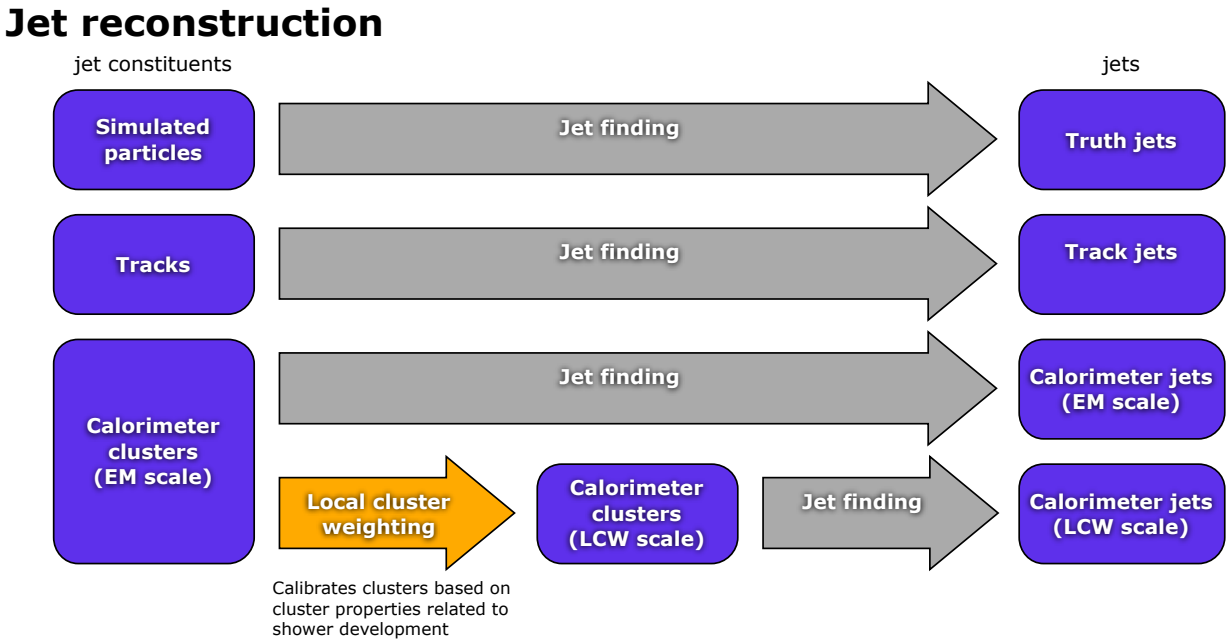


Figure 3.5: Flow-chart showing how jet finding is performed on different inputs to form different types of jets [101].

Different types of jets can be reconstructed by running jet finding algorithms over different

physical input objects. In ATLAS, commonly used input objects include *tracks*, *stable particles*[2] generated by Monte Carlo event generators, and calorimeter based inputs, from which jet finding produces *track jets*, *truth jets* and *calorimeter jets* respectively. Figure 3.5 shows a schematic of this. The most widely used calorimeter input objects for jet finding in ATLAS are topological clusters, *topoclusters* [102–104], from which *topojets* are produced. An alternative calorimeter input exists in the form of *towers* [102], groups of calorimeter cells bounded by a fixed two-dimensional grid in $\eta - \phi$ space, however these are used less frequently due to their increased sensitivity to noise. Therefore, only topojets formed from topoclusters are used in this thesis.

Topoclusters are groups of calorimeter cells designed to follow the shower development, taking advantage of the fine segmentation of the fine ATLAS hadronic calorimeter. They follow a '420' signal-to-noise ratio (t [3]) pattern, designed to find efficiently low energy clusters without being overwhelmed by noise. *Seed cells* are selected to have a signal-to-noise ratio of $t_{seed} > 4$, to which neighbouring cells with a signal-to-noise ratio of $t_{neighbour} > 2$ are iteratively added, forming a *proto-cluster*. Next, all cells neighbouring this proto-cluster are added ($t_{cell} > 0$). Local maxima cells within these proto-clusters are then identified. Proto-clusters containing more than one local maxima are split, using the local maxima as seeds for a new iteration of topoclustering. Final topoclusters are formed from the resultant proto-clusters, whereby the topocluster four-vectors are constructed with energy equal to the sum of the energy of the constituent cells, zero mass and direction as that of the unit vector pointing from the centre of the ATLAS co-ordinate system to the energy-weighted topocluster barycentre [105].

Two topocluster collections are formed, one at the EM-scale, and one at the **L**ocal **C**luster **W**eighting (**LCW**) scale. EM-scale topoclusters are obtained by clustering cells calibrated at the EM-scale as described in Reference [106]. LCW-scale topoclusters are then built by taking the EM-scale topoclusters and applying a local signal cluster weight to the calorimeter

[2] A typical hadron-hadron definition of a *stable particle* is one with lifetime τ satisfying $c\tau \geq 10\ mm$ [100]

[3] $t = \frac{E}{RMS(E)}$, where E_{cell} is the energy in the cell and $RMS(\mathrm{E}_{cell})$ is the RMS of the energy distribution measured from randomly triggered events.

cells within these clusters, designed to correctly reconstruct the response of the calorimeter to hadrons. The weight aims to improve the resolution compared with the EM-scale, and is assigned based on whether a cluster is identified as electromagnetic or hadronic in nature, determined from the energy density and shower shape. For hadronic clusters, the weight is a correction aimed at accounting for the non-compensating nature of the ATLAS detector. For both electromagnetic and hadronic clusters, the weight compensates for energy falling outside clustered cells, estimated from how isolated a cluster is, and energy falling outside the active areas of the calorimeter [4], determined from the position and energy deposited in each layer of the calorimeter. The weight is derived from neutral and charged pion interactions simulated using `Geant4` [105, 107].

Studying the calibration and performance of ATLAS topojets punching through the calorimeter is the focus of this thesis. Truth jets are invaluable reference objects in these studies, used to form a *jet energy response* (see Section 3.4). Truth jets used in this thesis do not include stable muons and neutrinos for consistency with the ATLAS jet calibration (Section 3.5).

3.4 Association of truth jets to calorimeter jets

Truth jets in Monte Carlo can be matched to calorimeter jets to provide information on a jet's constituent particles. For studies in this thesis, this is done via one of two procedures:

1. **Ghost matching:** This involves artificially spraying the calorimeter with particles of infinitesimally small p_T (*ghost particles*), reclustering the particle and calorimeter jets, and then counting the number of overlapping particles. If $> 50\%$ of the particles overlap between a given truth and calorimeter jet, these jets are said to be matched.

2. **Geometrical matching**: Calorimeter jets are matched to truth jets within a cone of $\Delta R < 0.3$. Matching is performed in order of decreasing jet p_{T}, discarding jets

[4] This correction is expected to have some sensitivity to jet punch-through. It will be shown in Appendix D that LCW jets after the full jet calibration scheme still display energy losses due to jet punch-through.

that have already been matched; ambiguities are resolved by choosing the jet with the highest p_{T} as the match.

Matching truth jets with calorimeter jets allows the reconstructed level jet properties to be compared with the particle level jet properties, allowing the *jet energy response* ($\mathcal{R}$) and jet p_{T} response ($\mathcal{R}_{pT}$) to be defined as:

$$\mathcal{R} = \frac{\mathrm{E}_{calo}}{\mathrm{E}_{truth}}, \ \mathcal{R}_{pT} = \frac{p_{\mathrm{T}}^{calo}}{p_{\mathrm{T}}^{truth}}, \tag{3.4}$$

where E_{calo} (p_{T}^{calo}) and E_{truth} (p_{T}^{truth}) are the energy (p_{T}) of the calorimeter jet and its associated truth jet[5] respectively. The jet response provides a measure of the calorimeter response to jets, and is used as a basis for the jet energy calibration, described in Section 3.5. The detector effects mentioned in Section 3.5 can cause the measured jet energy, and thus jet response, to fluctuate, resulting in a Gaussian distribution, as shown in Figure 3.6. The magnitude of these fluctuations can be used as a measure of the *jet energy resolution* (*jet* p_{T} *resolution*), parametrised as the standard deviation of a Gaussian fit to the jet response distribution, $\sigma_{\mathcal{R}}$ ($\sigma_{\mathcal{R}_{pT}}$). The fitting procedure is described in detail in Section 5.4. The resolution is commonly assessed via the fractional resolution, $\frac{\sigma_{\mathcal{R}}}{\mathcal{R}}$ ($\frac{\sigma_{\mathcal{R}_{pT}}}{\mathcal{R}_{pT}}$). The final ATLAS 2012 JER measurement is described in Section 8.3.

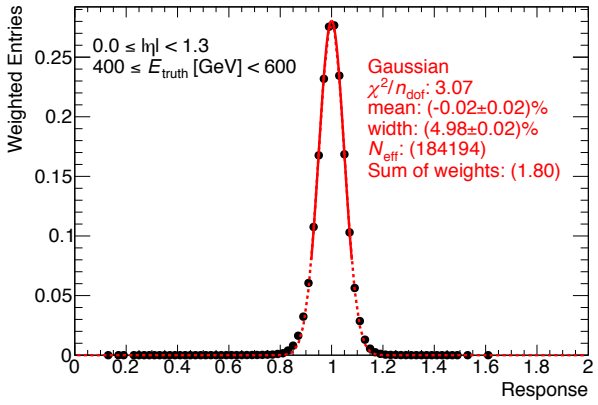


Figure 3.6: An example jet energy response distribution for jets in the region $0 \leq |\eta_{det}| < 1.3$ with $400 \leq E$ [GeV] < 600. Red line is a Gaussian fit to the distribution. Mean and width are quoted as the percentage difference from a response of 1.

[5] In this thesis, E_{truth} will often be referred to as the 'truth energy' of a given calorimeter jet.

It is important to note that this jet response is not sensitive to escaped energy due to muons and neutrinos, a consequence of using truth jets without muons and neutrinos.

3.5 Initial ATLAS 2012 jet energy scale calibration

As mentioned in Section 2.2.3, the ATLAS hadronic calorimeter is non-compensating, having a lower response to hadrons when compared with electrons due to the additional non-ionising hadronic interaction present in hadronic showers, not found in electromagnetic showers (see Section 1.3). This section will detail the derivation of a set of corrections that, in addition to restoring the correct hadronic energy scale - the **J**et **E**nergy **S**cale (**JES**) - from the electromagnetic scale, also aim to remove additional detector effects, including:

- Additional energy deposits due to in-time and out-of-time pile-up.
- Energy losses due to inactive detector regions (dead material).
- Signal losses in calorimeter clustering and jet reconstruction.
- Energy deposits of particles not included in the jet reconstruction.
- Differences in energy reconstruction due to jet flavour (gluon or quark initiated jets).
- Energy losses due to longitudinal hadronic shower leakage, affecting jets in the tail of the jet response distribution. Please note the initial 2012 calibration described in this section was originally still sensitive to these effects, as will be shown in Chapter 4. The development of a correction to remove these effects is the focus of this thesis.

The initial ATLAS jet calibration scheme in 2012 was a multi-stage procedure, involving Monte Carlo corrections derived using the baseline Pythia8 samples (described in Section 1.4). The calibration can be applied to jets clustered from topoclusters at both the EM and LCW scales. The calibration scheme is outlined in Figure 8.1.

First, the jet direction is corrected to point back to the primary vertex via the 'Origin Correction', as explained in Section 3.5.1. Next, the effect of pile-up is removed through the jet area based and residual pile-up corrections, described in Section 3.5.2. The jet energy is then restored to the correct hadronic scale using the 'Absolute EtaJES' correction, as outlined in Section 3.5.3. The **G**lobal **S**equential (**GS**) corrections are then applied, a sequence of four corrections that aim to reduce flavour dependencies of the jet energy scale and improve the jet energy resolution, detailed in Section 3.5.4.

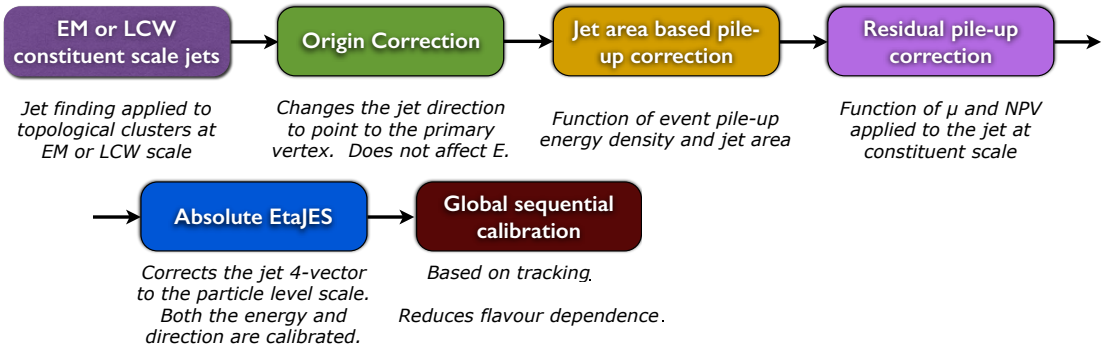


Figure 3.7: Initial calibration chain for ATLAS 2012 jets [108].

In this thesis, EM and LCW scale jets calibrated with up to and including (excluding) the GS corrections are labelled EM+JES+GS (EM+JES) and LCW+JES+GS (LCW+JES) jets respectively.

3.5.1 Origin correction

As mentioned in Section 3.3, topoclusters, and thus topojets, are reconstructed to have four-vectors pointing towards the geometrical centre of the ATLAS detector. The jet four-vector is corrected such that the direction of the jet points towards the hard-scatter primary vertex (defined in Section 2.2.2). The origin correction achieves this by recalculating the kinematic variables for each topocluster using a vector constructed with the primary vertex to the topocluster centroid as its direction. The jet four-vector is then re-defined as the vector sum of the modified topocluster four-vectors. The correction dramatically improves the jet η resolution whilst leaving the jet energy unaffected.

For the remainder of this thesis, the origin corrected η is referred to as η_{origin} or η, while the η coordinate of the jet vector pointing towards the geometrical centre of the ATLAS detector is referred to as η_{det}.

3.5.2 Pile-up corrections

Due to the calorimeter's long integration time, objects reconstructed from calorimeter depositions are especially susceptible to in-time and out-of-time pile-up (described in Section 2.1.1). Pile-up can degrade the jet energy resolution, in addition to shifting the jet energy scale. Such pile-up effects must be corrected for.

The pile-up correction in the ATLAS 2012 jet calibration scheme consists of two parts, outlined in detail in [109, 110]. The first is an area based subtraction technique based on the assumption that pile-up energy depositions can be treated as a uniform, diffuse background, adding signal to the jets. This diffuse background is characterised by measuring the pile-up energy density in the calorimeter in the $\eta - \phi$ plane, ρ[6] , and the measured jet area in this plane, A[7].

The second part in the pile-up correction is a residual η-dependent correction designed to remove any lingering dependence of the jet p_{T} on in-time and out-of-time pile-up. The dependence on in-time pile-up is characterised by the number of reconstructed primary vertices, N_{PV}, through the factor $\alpha\,(\mathrm{N}_{PV} - 1)$. Out-of-time pile-up is accounted for via $\beta\langle\mu\rangle$, where $\langle\mu\rangle$ is the average number of interactions per bunch crossing. The full pile-up correction is given by:

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}} - \rho A - \alpha\,(\mathrm{N}_{PV} - 1) - \beta\langle\mu\rangle, \tag{3.5}$$

where $p_{\mathrm{T}}^{\mathrm{corr}}$ is the pile-up corrected p_{T} of the jet [110].

[6] ρ is defined to be the median of the jet energy density, $\rho_{jet} = \frac{p_{\mathrm{T}}^{\mathrm{jet}}}{A^{jet}}$, where $p_{\mathrm{T}}^{\mathrm{jet}}$ and A^{jet} are the p_{T} and Voronoi measured jet area [111] respectively of an anti-k_t jet with R = 0.4 reconstructed in the central detector region ($|\eta_{det}| < 2$).

[7] A is calculated using an active areas method in which ghost particles, are uniformly added to an event and jets in the event reclustered. The fraction of ghosts clustered in each jet gives an indication of A.

3.5.3 Jet energy scale correction

The jet energy scale correction aims to correct the calorimeter jet four-vector back to the particle level scale, restoring the correct hadronic energy scale. It uses the jet energy response as defined in Equation 3.4 as a basis, and produces jets at the EM+JES and LCW+JES scales from EM and LCW jets respectively.

Only isolated truth and calorimeter jets are used for the calibration. An isolated calorimeter jet is defined to be one with no other calorimeter jet with uncalibrated $p_{\mathrm{T}}^{\mathrm{jet}} > 7$ GeV within a cone of $\Delta R < 1.5$ R, where R is the jet radius. Similarly, an isolated truth jet must have no other truth jets with $p_{\mathrm{T}}^{truth} > 7$ GeV within a cone of $\Delta R < 2.5$ R.

In order to evaluate the jet energy response as defined in Equation 3.4, calorimeter jets are matched to truth jets geometrically as described in Section 3.4. For EM jets, the JES correction uses the jet energy response at the EM-scale, $\mathcal{R}_{EM}$, formed from the EM-scale calorimeter energy of the jet, E_{calo}^{EM}, as its basis. The average EM-scale response, $\langle \mathcal{R}_{EM} \rangle$, is measured from the mean of a Gaussian fit to the $\mathcal{R}_{EM}$ distribution in bins of E_{truth} and η_{det}. For each $(\mathrm{E}_{truth}, \eta_{det})$-bin, the average EM-scale jet energy, $\langle E_{calo}^{EM} \rangle$, is also measured from the mean of the E_{calo}^{EM} distribution. The calibration function, $\mathcal{F}_{calib,k}\left(\mathrm{E}_{calo}^{EM}\right)$, is obtained from a fit to the mean EM-scale response as a function of the average EM-scale jet energy, $\langle \mathcal{R}_{EM}\left(E_{calo}^{EM}\right) \rangle$, in each η_{det} bin k. For LCW jets, the jet energy response at the LCW-scale, $\mathcal{R}_{LCW}$, is used instead, formed from the LCW-scale calorimeter energy of the jet, E_{calo}^{LCW}. The correction function, $\mathcal{F}_{calib,k}\left(\mathrm{E}_{calo}^{\mathrm{LCW}}\right)$, is then formed in the same way as for the EM calibration. The fitting function used is parametrised as [105]:

$$\mathcal{F}_{calib,k}\left(\mathrm{E}_{calo}^{EM \text{ or LCW}}\right) = \sum_{i=0}^{N_{max}} \alpha_i \left(\ln \mathrm{E}_{calo}^{EM \text{ or LCW}}\right)^i \tag{3.6}$$

where α_i are free parameters, and N_{max} is chosen between 1 and 6 depending on the goodness of fit. The final jet energy scale correction, relating the measured calorimeter jet to the true energy, in each η_{det}-bin, is then defined as $1/\mathcal{F}_{calib}\left(\mathrm{E}_{calo}^{EM \text{ or LCW}}\right)$ in Equation 3.7, where

$\mathrm{E}^{\mathrm{EM+JES}\ or\ \mathrm{LCW+JES}}$ is the jet energy at the EM+JES-scale or the LCW+JES-scale [105].

$$\mathrm{E}^{\mathrm{EM+JES}\ or\ \mathrm{LCW+JES}} = \frac{\mathrm{E}_{calo}^{EM\ or\ \mathrm{LCW}}}{\mathcal{F}_{calib}\left(\mathrm{E}_{calo}^{EM\ or\ \mathrm{LCW}}\right)|_{\eta_{det}}} \tag{3.7}$$

At this stage, an additional η-correction is applied to remove biases in the reconstructed η of jets, introduced by under-measured EM-scale energies of topoclusters in poorly instrumented regions of the detector. The correction is derived as the average $\Delta\eta = \eta_{truth} - \eta_{origin}$ in $(\mathrm{E}_{truth}, \eta_{det})$-bins, parametrised as a function of η_{det} [105]. As well as removing the biases in η, the correction also brings the p_{T} response closer to 1 (improved *closure*).

Figure 3.8 displays the p_{T} response as a function of p_T^{truth} after the application of the jet energy scale corrections for R = 0.6 EM jets. It is important to note that there is still some inherent non-closure for jets with $\mathrm{p}_T^{truth} < 100$ GeV that grows as p_T^{truth} decreases. This is due to non-Gaussian and noise threshold effects causing poor fits at low p_T^{truth} [108]. Similar effects are seen when looking at the jet energy response as a function of E_{truth}.

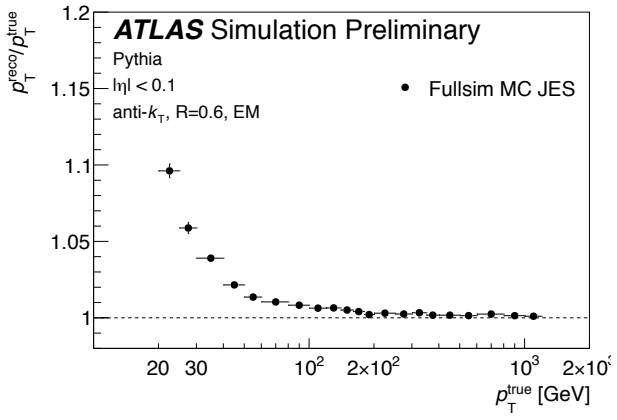


Figure 3.8: Jet p_{T} response as a function of p_T^{truth} after the application of the jet energy scale corrections (R = 0.6 EM+JES-scale jets) [108].

3.5.4 Global Sequential corrections

The Global Sequential (GS) corrections [1] form the final step in the initial 2012 ATLAS jet calibration sequence. They are a series of four sequential corrections that aim to further

improve the JER by removing multiple effects that impact the jet energy measurement, including energy loss in inactive areas of the detector, and the flavour of the parton initiating the jet. This is achieved by removing the jet response dependence on any jet property, x, that is sensitive to these effects. One jet property is used for each step, giving four steps in total. In order:

1. $\mathbf{f_{Tile0}}$ **followed by** $\mathbf{f_{LAr3}}$**:** The fraction of the jet's energy at the constituent scale deposited in the third layer of the electromagnetic calorimeter and the first layer of the hadronic calorimeter respectively. Used to improve the resolution of EM+JES jets by removing dead material effects. The fraction of energy in a given detector layer is given by Equation 3.8, where $E_{\mathrm{EM}}^{\mathrm{jet}}$ and $E_{\mathrm{EM}}^{\mathrm{layer}}$ are the jet energy and energy deposited in the layer of interest at the constituent scale (EM or LCW).

$$f_{\mathrm{layer}} = \frac{E_{\mathrm{EM}}^{\mathrm{layer}}}{E_{\mathrm{EM}}^{\mathrm{jet}}} \tag{3.8}$$

2. $\mathbf{n_{trk}}$**:** The number of tracks associated to jets is sensitive to the flavour of the jet-initiating parton. Gluon-initiated jets tend to have more particles (larger n_{trk}), while Light-Quark (LQ) initiated jets tend to have fewer particles (smaller n_{trk}) [112].

3. $\mathbf{width_{trk}}$**:** The tracking jet width, width_{trk}, is defined as the average distance between the tracks associated to the jet and the calorimeter jet axis, weighted by the track p_{T}. width_{trk} is defined in Equation 3.9, where i refers to the tracks and $\Delta R(i, \mathrm{jet}) = \sqrt{(\eta^{\mathrm{jet}} - \eta^{\mathrm{track},\, i})^2 + (\phi^{\mathrm{jet}} - \phi^{\mathrm{track},\, i})^2}$. Close to a transition region, it provides an indication of how many particles hit poorly instrumented regions.

$$\mathrm{width}_{trk} = \frac{\sum_{\mathrm{i}} p_{\mathrm{T}}^{i} \Delta R(i, \mathrm{jet})}{\sum_{\mathrm{i}} p_{\mathrm{T}}^{i}} \tag{3.9}$$

The variables used, the detector regions they are applied to, as well as the order in which they are applied, are summarized in Table 3.1. The improvement in performance does not depend on the sequence in which the corrections are applied.

$\|\eta\|$ region	Correction 1	Correction 2	Correction 3	Correction 4
$[0, 1.7]$	f_{Tile0}	f_{LAr3}	n_{trk}	width_{trk}
$[1.7, 2.5]$		f_{LAr3}	n_{trk}	width_{trk}
$[2.5, 2.7]$		f_{LAr3}		
$[2.7, 3.5]$		f_{LAr3}		

Table 3.1: Order in which jet properties are used for correction in each $|\eta|$ region[1].

For jets at the LCW+JES scale, only the track based corrections are applied, as the LCW calibration (described in Section 3.3) already includes a dead material correction, resulting in no further improvement in resolution from the calorimeter corrections.

The GS corrections are multiplicative, derived by inverting the jet response as a function of x, $C(x) = R^{-1}(x)$. The corrections are relative to the JES correction in Section 3.5.3. They are binned in $|\eta_{det}|$ and constructed as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ and jet property x, $\mathcal{F}_{\mathrm{rel}}(p_{\mathrm{T}}^{\mathrm{jet}}, x)$. The response is obtained from the mean of a Gaussian fit to the response distribution evaluated in bins of $|\eta_{det}|$, p_T^{truth} and jet property x. The function $\mathcal{F}_{\mathrm{rel}}$ is then constructed in bins of $|\eta_{det}|$ using a sequence of two Gaussian kernels: a one-dimensional Gaussian kernel followed by a two dimensional Gaussian kernel. The final correction function is defined via:

$$\mathcal{F}_{\mathrm{rel}}(p_{\mathrm{T}}^{\mathrm{jet}}, x) = \frac{\sum_{i=1}^{N_{\mathrm{bins}}} C_i w_i}{\sum_{i=1}^{N_{\mathrm{bins}}} w_i} \tag{3.10}$$

with the final two dimensional Gaussian kernel, w_i, defined to be:

$$w_i = \frac{1}{\Delta C_i^2} \times \mathrm{Gauss}\left(\frac{\log p_{\mathrm{T}}^{\mathrm{jet}} - \log \left\langle p_{\mathrm{T}}^{\mathrm{jet}} \right\rangle_i}{\sigma_{p_{\mathrm{T}}}} \oplus \frac{x - \langle x \rangle_i}{\sigma_x} \right) \tag{3.11}$$

where i denotes the index of a $(p_{\mathrm{T}}^{\mathrm{jet}}, x)$-bin, ΔC_i is the statistical uncertainty of C_i, $\left\langle p_{\mathrm{T}}^{\mathrm{jet}} \right\rangle_i$ and $\langle x \rangle_i$ are the average p_{T} and x of the jets in each bin, Gauss(x) is the amplitude of a Gaussian function with $\mu = 0$ and $\sigma = 1$, $\sigma_{p_{\mathrm{T}}}$ and σ_x are width-parameters of the Gaussian kernel and $\oplus$ denotes addition in quadrature [1]. The kernel-width parameters used were found to capture the shape of the response across η and $p_{\mathrm{T}}^{\mathrm{jet}}$, and at the same time provide stability against statistical fluctuations [1].

The procedure requires that the correction for variable x_i ($\mathcal{F}^i_{\rm rel}$) is calculated using jets to which the correction for variable x_{i-1} ($\mathcal{F}^{i-1}_{\rm rel}$) has already been applied. The jet transverse momentum after correction number i is given by [1]:

$$p^i_{\rm T} = \mathcal{F}^i_{\rm rel}(x_i) \times p^{i-1}_{\rm T} = \mathcal{F}^i_{\rm rel}(x_i) \times \mathcal{F}^{i-1}_{\rm rel}(x_{i-1}) \times p^{i-2}_{\rm T} = ... \tag{3.12}$$

3.6 Jet cleaning

In order to effectively study the impact of punch-through jets, it was important to ensure jets used were not fake, and not poorly measured due to detector effects other than jet punch-through, such as calorimeter noise spikes, or dead tile regions. Fake jets include energy deposits coming from groups of noisy cells in calorimeters, cosmic muons and beam backgrounds. A combination of *jet cleaning cuts* were used to remove such jets. The cleaning cuts used an extension of those found in Ref. [113]. A jet passing these cuts is deemed *good*.

Bad jet rejection

Bad jets are due to background or detector effects. Rejection of bad jets involves:

- **Electromagnetic coherent noise:** A noise burst occurring in the electromagnetic calorimeter can lead to a fake jet. These jets will thus have a large fraction of their energy deposited in the electromagnetic calorimeter, $\rm f_{LAr}$ [8] , originating from bad quality cells as measured by Q^{LAr}_{cell} [9]. Jets with a large fraction of energy originating from electromagnetic calorimeter cells with poor signal shape quality ($Q^{LAr}_{cell} > 4000$), a large $\rm f_{LAr}$ and a large average jet quality, $\langle Q \rangle$ [10] , are rejected if: $\rm f_{LAr3} > 0.95$

[8] $\rm f_{LAr} = \frac{E^{jet}_{LAr}}{E_{jet}}$, where E_{LAr} is the jet energy deposited in the electromagnetic calorimeter, and E_{jet} is the total measured jet energy.

[9] Q_{LAr} is computed online using five measured samples of the pulse shape from a given cell: $Q_{LAr} = \sum_{j=1}^{5}(s_j - Ag^{phys}_j)^2$, where A is the measured amplitude of the cell signal [114], s_j is the amplitude of each sample j in ADC counts and g^{phys}_j is the normalized predicted ionization shape [113].

[10] The average jet quality is defined as the energy squared weighted average of the pulse quality of the calorimeter cells (Q^{LAr}_{cell}) normalised such that $0 < \langle Q \rangle < 1$: $\langle Q \rangle = \langle \frac{Q^{LAr}_{cell}}{65336} \rangle$ [115].

and $|f_Q^{LAr}| > 0.8$ and $\langle Q \rangle > 0.8$ for jets with $|\eta_{det}| < 2.8$; OR $f_{LAr3} > 0.95$ and $|f_Q^{LAr}| > 0.8$ for jets with $|\eta_{det}| < 2.8$ [115].

- **Hadronic end-cap spikes:** Most fake jets come from sporadic noise bursts in single calorimeter cells in the HEC. These jets have large fractions of their energy deposited in the HEC, $f_{\texttt{HEC}}$ [11]. Similar to electromagnetic cells, the signal shape quality of a HEC cell is measured by $Q_{cell}^{\texttt{HEC}}$, which is calculated in the same way as Q_{cell}^{LAr}, but using HEC cells instead of electromagnetic LAr cells. Jets with a large fraction of their energy originating from HEC calorimeter cells with a poor signal shape quality ($Q_{cell}^{\texttt{HEC}} > 4000$), $f_Q^{\texttt{HEC}}$, a large $f_{\texttt{HEC}}$, and a large $\langle Q \rangle$, are rejected if: $f_{\texttt{HEC}} > 0.5$ and $|f_Q^{\texttt{HEC}}| > 0.5$ and $\langle Q \rangle > 0.8$; OR $|E_{negative}| > 60$ GeV; OR $f_{\texttt{HEC}} > 0.5$ and $|f_Q^{\texttt{HEC}}| > 0.5$ [115].

- **Non-collision backgrounds and cosmics:** Large amounts of energy can be deposited in the calorimeter outside the nominal timing window for collisions, mainly due to cosmic ray showers or beam backgrounds. The maximum energy fraction in one calorimeter layer, f_{max}, the jet time, t[12], and the jet charge fraction, f_{charge}[13] were used to reject these events if: $f_{LAr3} < 0.05$ and $f_{charge} < 0.05$ and $|\eta_{det}| < 2$; OR $f_{\rm EM} < 0.05$ and $|\eta_{det}| \geq 2$; OR $f_{max} > 0.99$ and $|\eta_{det}| < 2$; OR $|t| > 25$ns.

Ugly jet rejection

Ugly jets are poorly measured jets falling into problematic detector regions.

- **Jets extrapolated from masked cells:** Cells affected by high voltage trips and permanently offline cells are masked during jet reconstruction and their energy is estimated from the surrounding cells. Jets built from a large fraction of these masked cells with $f_{LAr3} > 50\%$ are rejected.

- **Large energy fraction in the tile-gap layers:** The calibration of the scintillator response in the gap between the hadronic tile calorimeter and the end cap is currently

[11] $f_{\texttt{HEC}} = \frac{E_{\texttt{HEC}}}{E_{jet}}$, where $E_{\texttt{HEC}}$ is the amount of energy deposited by the jet in the HEC

[12] t is computed as the energy squared cells mean time.

[13] the ratio of the sum of the $p_{\rm T}$ of tracks associated to the jets divided by the calibrated jet $p_{\rm T}$

not fully understood, thus jets depositing $>$ 50% of their energy in this region are rejected.

Rejecting jets over-corrected by masked cell correction

As mentioned above, cells affected by high voltage trips and permanently offline cells are masked during jet reconstruction. In 2012, the correction responsible for estimating the energy of these masked cells was unable to correctly handle entire dead cells. As high p_{T} jets are more collimated, they are more sensitive to this overcorrection. The presence of these over-corrected jets can bias the jet response. They are rejected by vetoing jets falling into masked tile regions. This masking is done artificially in MC to prevent biasing data/MC comparisons.

Additional pile-up rejection cuts

The **J**et **V**ertex **F**raction (**JVF**) is the fraction of summed track p_{T} for all tracks matched to a given jet and associated with the **P**rimary **V**ertex (PV), relative to the total summed p_{T} for all tracks matched to the jet [110]:

$$JVF(jet_i, PV_j) = \frac{\sum_k p_{\mathrm{T}}(track_j^{jet}, PV_j)}{\sum_n \sum_t p_{\mathrm{T}}(track_t^{jet}, PV_n)} \tag{3.13}$$

where $JVF(jet_i, PV_j)$ is the JVF defined for each given jet_j with respect to the primary vertex PV_j. The JVF is close to 1 for hard scatter jets, and close to 0 for pile-up jets. A cut on the JVF can be used to reject spurious calorimeter jets resulting from local fluctuations in pile-up activity, as well as real QCD jets originating from single pileup interactions, resulting in improved stability of the reconstructed jet multiplicity against pile-up [110].

Removal of excess overlaid pile-up events in MC

In MC, the reconstructed p_{T} spectra are not smooth at low p_{T} due to overlaid pileup events. These are a result of pileup causing jets from lower p_{T} samples to migrate to higher p_{T} bins. These jets then acquire larger Monte Carlo weights (Equation 1.10), distorting the p_{T} spectra. To remove these problematic events, it was required that $p_{\mathrm{T}}^{avg}/p_{\mathrm{T}}^{leading\ truth} < 1.4$, where $p_{\mathrm{T}}^{avg} = (p_{\mathrm{T}}^{lead}+p_{\mathrm{T}}^{subl})/2$, $p_{\mathrm{T}}^{leading\ truth}$ (p_{T}^{lead}) is the truth (reconstructed) p_{T} of the highest p_{T} jet and p_{T}^{subl} is the p_{T} of the second highest p_{T} jet.

3.7 Flavour tagging

It is necessary to use tagging methods when attempting to study effects resulting from the flavour of the jet-initiating parton. This is done differently at truth level in MC and reconstructtion level in data:

- **Truth flavour tagging:** The jet flavour is determined by geometrically matching the jet with generator level hadrons/partons: if a b-hadron is found within a cone of $\Delta R < R$ of the jet, where R is the radius parameter of the jet algorithm, the jet is labelled as a b-jet. If no b-hadron is found, the procedure is repeated for c-hadrons. If no such association can be made for a jet, the highest energy parton from the event generator before the parton shower that points to the jet determines its flavour. This definition is sufficient to study the flavour dependence of the calorimeter response to jets. Theoretical ambiguities of jet flavour assignment are irrelevant in this context [112].

- **b-jet tagging in data:** To identify b-jets, the MV1 tagger was used. MV1 is a neural-network algorithm that integrates the output of several simple tagging algorithms (SV1, JetFitter, IP3D) to exploit different features (reconstructed vertex, impact parameter) of the jet during tagging. More information of the performance of the MV1 algorithm can be found here [116, 117]. MV1 was used at a 70% efficiency operating point.

Chapter 4

Impact of Jet Punch-Through on Jet Reconstruction

Jets clustered from hadronic showers experiencing longitudinal leakage are called *punch-through* jets. It was shown in Section 1.3.3 that longitudinal shower leakage reduces the mean energy and energy resolution of objects involving hadronic showers, and so is expected to have a similar impact on energy measurements using jets.

This chapter outlines studies that test how sensitive the initial ATLAS 2012 jet energy scale calibration described in Section 3.5 is to jet punch-through effects. First, the methods used to study punch-through jets with reconstructed dectector variables are explained in Section 4.1. The investigation into the sensitivity of the initial ATLAS 2012 jet energy scale to jet punch-through effects is outlined in Section 4.2, focusing on the impact of jet punch-through on the mean jet energy response and jet energy resolution.

4.1 Identifying punch-through jets

In order to assess the impact of jet punch-through on the jet energy scale and jet energy resolution, and to develop a correction applicable to both data and Monte Carlo, it is necessary to identify reconstruction level variables that capture the dependence of the jet response on

jet punch-through. Such variables fall into two distinct categories, associated with:

- **Muon activity behind the jet:** Charged particles in the longitudinal tail of the hadronic shower penetrating into the muon spectrometer will generate muon hits and segments, N_{hits} and $N_{segments}$ respectively. The number of hits/segments present is indicative of the amount of leakage.
- **Shower profile in calorimeter:** Showers that start later in the calorimeter are more likely to experience leakage. Thus a large fraction of a jet's energy deposited in the final layers of the hadronic calorimeter, a small fraction of energy deposited in the EM calorimeter, or the shower maximum occurring in the outer layers of the hadronic calorimeter, are all suggestive a jet may be punching-through.

Initial studies outlined in Appendix B showed some dependence of the jet energy response on calorimeter variables linked to jet punch-through, however the variables that showed promise were limited to the central detector regions ($|\eta_{det}| < 1.4$). As a result, the muon activity behind a jet was used as the primary discriminator to study jet punch-through in this thesis. Variables used to quantify the muon activity behind a jet, a concrete definition of a punch-through jet based upon these variables, and an estimation of the jet punch-through probability for dijet events are outlined below. All plots in this section were made using the same data, MC and dijet, selection outlined in Section 6.2.

4.1.1 Reconstructing muon activity behind jets ($N_{segments}$)

Two main variables are used to measure the level of muon activity behind a jet: the number of muon hits, N_{hits}, and the number of muon segments, $N_{segments}$, found behind a jet in the muon spectrometer. Muon hits occur when charged particles pass through the muon spectrometer. Muon segments are partial tracks constructed from muon hits. Muon spectrometer chambers with intersecting regions of activity in the detector, such as MDTs with hits in concurrent multi-layers (see Section 2.2.4), are selected for segment reconstruction [118, 119]. Association of muon hits and segments to jets is done in two steps:

1. **Building the muon shower containers**: Segments found to be within a cone of $\Delta R < 0.4$, and hits found to be within $\Delta\eta < \pi/8$ and $\Delta\phi < \pi/8$, of an anti-k_t $R = 0.6$ jet, are filled into muon shower containers during the event reconstruction. The η and ϕ coordinates of a muon shower container correspond to the those of the anti-k_t $R = 0.6$ jet to which they are associated in this step. In Run I, muon shower containers were only reconstructed for anti-k_t $R = 0.6$ jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV at the EM+JES or LCW+JES scale. These containers had to be associated to jets with different R sizes.

2. **Association of muon shower container to a jet**: Muon shower containers are said to originate from the closest jet within a cone of $\Delta R < 0.3$. No more than one muon shower container is associated to each jet. The number of muon segments and hits within the selected container correspond to the number of muon segments and hits associated to the given jet.

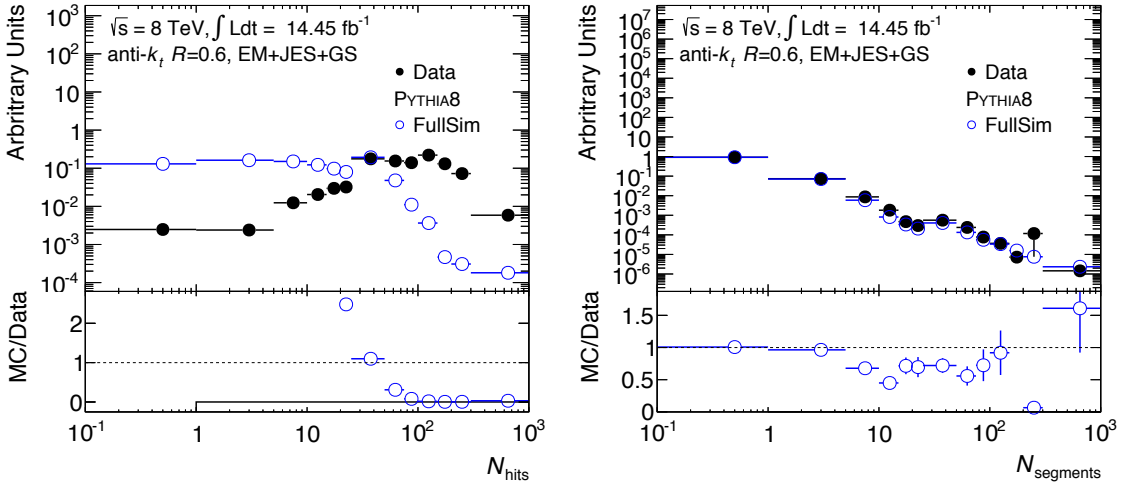


Figure 4.1: Figure comparing the shape of the N_{hits} (left) and $N_{segments}$ (right) distribution between data and MC. The bottom plot shows a ratio between MC and data.

It is important to verify any variable used for the study of jet punch-through is modelled well in MC to ensure studies using truth jets give reliable results. Figure 4.1 compares the N_{hits} and $N_{segments}$ distributions in data and MC. Reasonable shape agreement is seen between data and MC for $N_{segments}$, but poor agreement for N_{hits}, with the shape of the distribution described poorly.

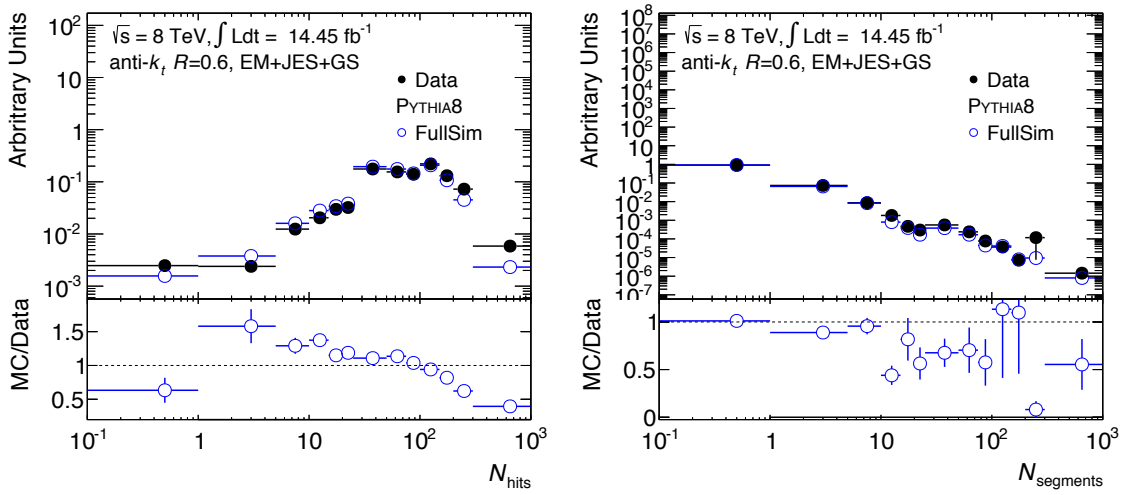


Figure 4.2: Figure comparing the shape of the N_{hits} (left) and $N_{segments}$ (right) distribution between data and overlaid MC. The bottom plot shows a ratio between MC and data.

The plots in Figure 4.2 were produced using MC samples overlaid with backgrounds from data. The full procedure used to generate these samples is described in Ref. [120]. Much better agreement is seen for both $N_{segments}$ and N_{hits}, with a drastic improvement seen in the shape agreement for N_{hits}. The impact is smaller for $N_{segments}$, likely due to inherent noise filtering present in the algorithms used to reconstruct $N_{segments}$ [118, 119]. This confirms the large shape differences seen for N_{hits} are due to poorly modelled backgrounds:

- Lack of accurate noise simulation of hits in muon chambers in MC.
- Out of time pileup in muon spectrometer not well described in MC.
- Presence of beam background in data not well simulated in MC.

As a result, N_{hits} was not used for studies in this thesis. Figure 4.3 shows the reconstructed muon segment efficiency as a function of η_{det} for data and MC. Good agreement can be seen, with a general reconstruction efficiency of above 99%. This confirms $N_{segments}$ is a well modelled, efficiently reconstructed variable.

Equation 1.7 showed the probability for leakage to occur is correlated with the energy of the shower (and thus jet) initiating parton, and the amount of material the shower passes

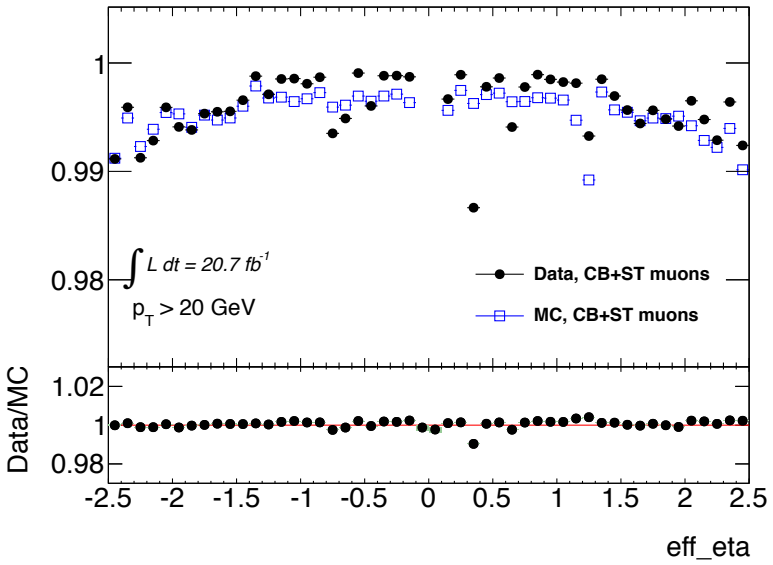


Figure 4.3: $N_{segments}$ reconstruction efficiency in data and MC as a function of η_{det} [121].

through. As a result, the jet punch-through probability is expected to be energy and η_{det} dependent, largest for high energy jets in poorly instrumented η_{det} regions.

Figure 4.4 compares the jet punch-through probability defined to be the fraction of jets with one or more particles entering the muon spectrometer measured in MC simulation [122], with plots of $N_{segments}$ vs E, η_{det} and the amount of material the jet is presumed to have traversed in units of λ_I, compared between data and MC. It is clear from the simulated probability, as expected from Equation 1.7, that the punch-through probability increases as the energy of the initiating particle increases. This is mirrored in $N_{segments}$ which increases as the jet energy increases.

Comparing the punch-through probability vs η_{det} from MC with Figure 2.7 confirms the probability is largest in poorly instrumented regions of the detector ($1.3 \leq |\eta_{det}| < 1.9$). Again, this is mirrored when looking at $N_{segments}$ vs λ_I, where $N_{segments}$ is larger in poorly instrumented regions of the detector. Spikes are also present when looking at $N_{segments}$ vs η_{det} where there is less material, and the jet punch-through probability is larger. Data and MC appear to agree relatively well in these plots.

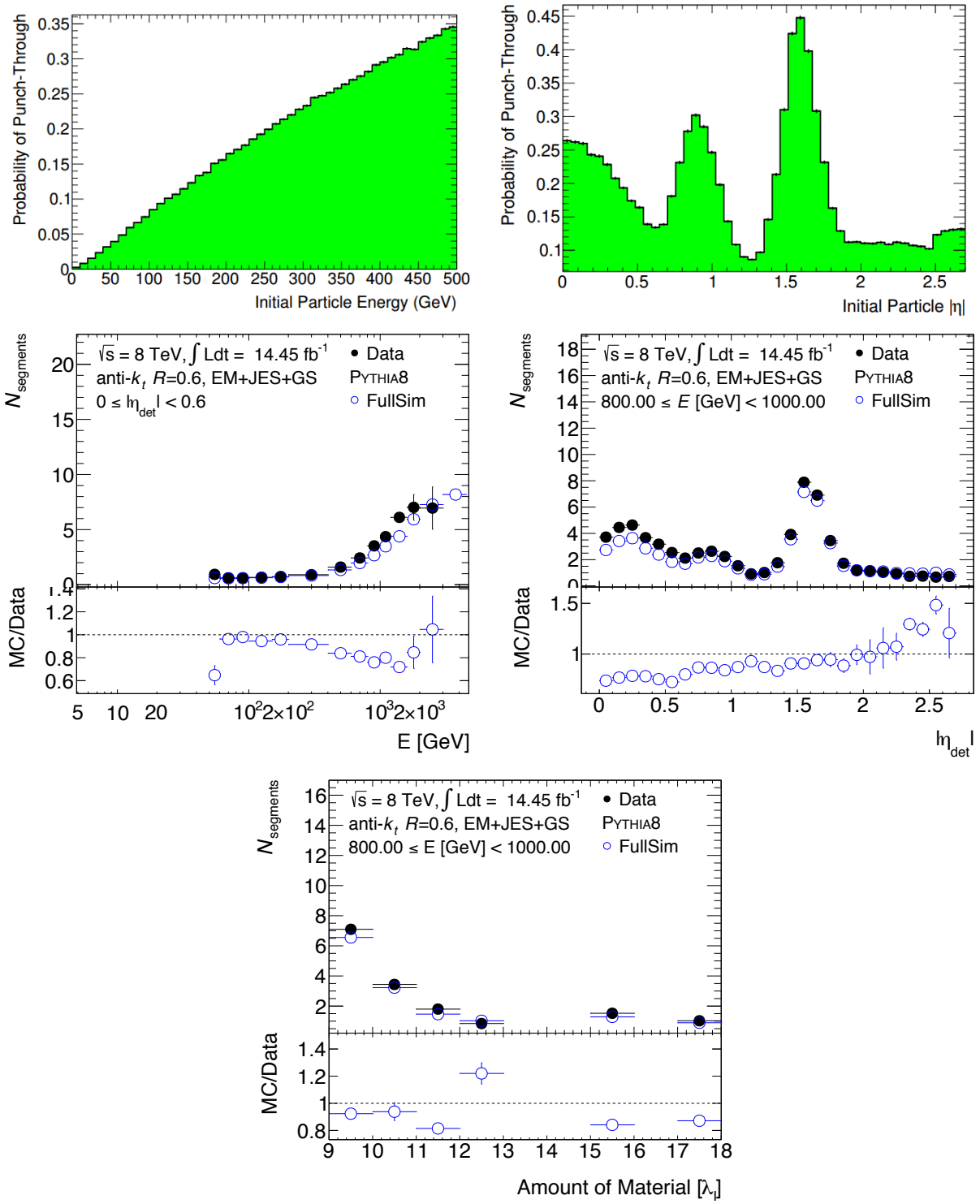


Figure 4.4: Probability of punch-through as a function of η_{det} (top-left) and E (top-right) [122], compared with $N_{segments}$ vs E (middle-left), $N_{segments}$ vs η_{det} (middle-right) and $N_{segments}$ vs λ_I (bottom). All $N_{segments}$ plots compare data with MC.

The above observations are highly suggestive that $N_{segments}$ is linked with the occurrence of jet punch-through. For this reason, $N_{segments}$ is used as the primary discriminator for jet punch-through in this thesis. The final thing to consider when using $N_{segments}$ is background processes that could produce segments behind a jet in the muon spectrometer. One key background comes from the presence of b-quarks in the sample.

The weak mixing of quark flavours is governed by the mixing angles in the **C**abibbo-**K**obayashi-**M**askawa (**CKM**) matrix [123, 124]. Thus b-quark decays to c and u quarks are constrained by the CKM up-bottom and charm-bottom mixing angles, $|V_{ub}|$ and $|V_{cb}|$ respectively. $|V_{cb}| = (40.9 \pm 1.1) \times 10^{-3}$ [100] and $|V_{ub}| = (4.15 \pm 0.49) \times 10^{-3}$ [100] are both small ($\ll 1$), resulting in b-decays to c and u quarks being heavily suppressed. As weak decays to electrons are also suppressed due to chirality, b-quarks preferentially decay into muons, which can produce hits and segments in the muon spectrometer behind the jet associated with the b-quark (b-jets). For this thesis, these muons do not constitute leakage of the hadronic shower associated with the b-jet, and so the additional hits and segments they produce constitute an additional background that must be avoided.

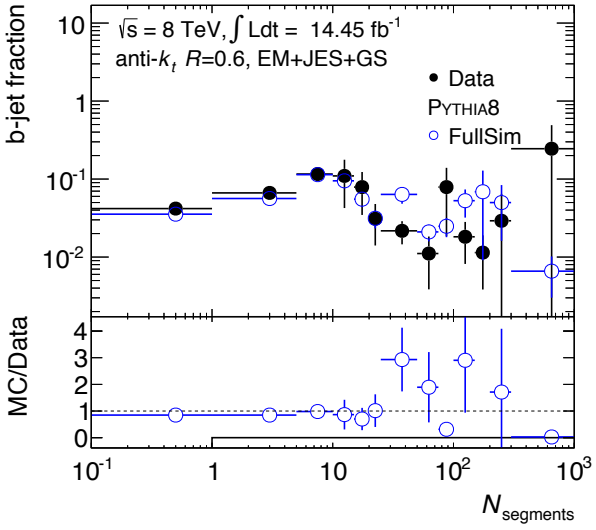


Figure 4.5: Percentage of jets tagged as b-jets using MV1 tagger as a function of $N_{segments}$. Bottom plot shows fraction in MC divided by fraction in data. Black line in ratio is line to indicate ratio of 0.

Figure 4.5 displays a data-MC comparison of the percentage of b-jets present as a function of $N_{segments}$, where a b-jet is defined as in Section 3.7. Data and MC agree fairly well. It

can be seen the fraction of b-jets appears to increase from 1-10 segments, at which point $N_{segments}$ declines until $N_{segments} \sim 20$. For $N_{segments} \geq 20$, the fraction of b-jets appears to stabilise. Real muons are expected to have between three and six segments associated with them, which suggests the observed behaviour is due to the muon produced in the b-decay. As we are looking to study the effects of real hadronic shower leakage, as opposed to muons produced from b-decays, $N_{segments} \geq 20$ is used as an operating point for many studies in this thesis. The exact requirements used to define a punch-through jet in the context of this thesis are detailed in the next section.

4.1.2 Defining a punch-through jet

In order to effectively study the impact of jet punch-through, it is important to define what is meant by a 'punch-through' jet in the context of the work described in this thesis. It was shown in Section 4.1.1 that $N_{segments}$ shows correlations expected from a variable related to a jet punching-through into the muon spectrometer. Thus, $N_{segments}$ is the main discriminator used to select and study punch-through jets in this thesis. In doing this, the following two limitations on $N_{segments}$ must be considered:

1. As mentioned in Section 4.1.1, during Run I muon shower containers were only reconstructed for jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV at the EM+JES or LCW+JES scale. As a result, $N_{segments}$ information is only available for jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV.

2. Figure 4.5 shows the fraction of b-jets stabilised when $N_{segments} \geq 20$. This suggests that in the region $N_{segments} < 20$, $N_{segments}$ cannot be used independently to discriminate between real hadronic shower leakage and b-jet decays.

Based upon these two observations, a 'punch-through jet' is defined to be a jet with $N_{segments} \geq 20$ and $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV. This because for jets with $p_{\mathrm{T}} < 50$ GeV, no $N_{segments}$ information is available. In addition, this ensures only jets experiencing real hadronic shower leakage as opposed to leakage due to b-jet decays are selected. From this point, any reference to a 'punch-through' jet refers to any jet meeting this specification.

4.1.3 Estimating a jet punch-through probability

In order to understand the prevalence of jet punch-through within the ATLAS detector at $\sqrt{s} = 8$ TeV in both data and MC, it is useful to define the fraction of punch-through jets, or jet punch-through probability, as:

$$\mathcal{P}_{punch-through} = \frac{N^{jets}(\mathrm{N}_{segments} \geq 20)}{N^{jets}(\mathrm{N}_{segments} \geq 0)}, \tag{4.1}$$

where $\mathcal{P}_{punch-through}$ is the fraction of jets with $\mathrm{N}_{segments} \geq 20$ (punch-through probability), $N^{jets}(\mathrm{N}_{segments} \geq 20)$ is the number of jets with $\mathrm{N}_{segments} \geq 20$ and $N^{jets}(\mathrm{N}_{segments} \geq 0)$ is the total number of jets.

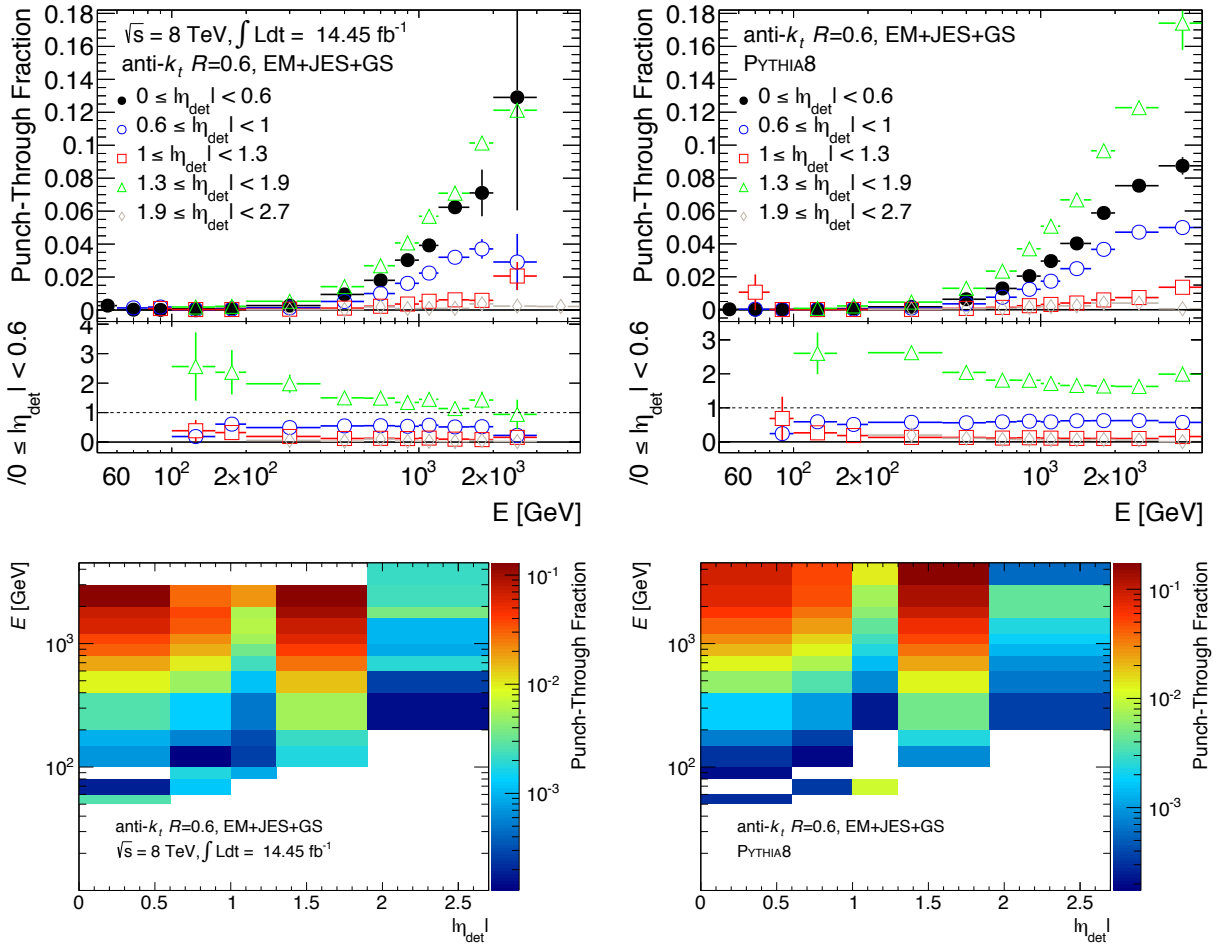


Figure 4.6: $\mathcal{P}_{punch-through}$ as a function of E in bins of η_{det} (top) and as a function of E and η_{det} (bottom) for data (left) and MC (right). The bottom plots show the ratio of the punch-through fraction in each η_{det} region to the punch-through fraction in the $0 \leq |\eta_{det}| < 0.6$.

Figure 4.6 displays the fraction of punch-through jets (or $\mathcal{P}_{punch-through}$) as a function of η_{det} and E for data and MC dijet events. It is clear that the punch-through probability increases as the jet energy increases (as expected from Figure 4.4), and is largest in the $1.3 \leq |\eta_{det}| < 1.9$ region - the poorly instrumented crack region. It ranges from near zero at low E to 10% at high E, and is lowest in the forward region - $1.9 \leq |\eta_{det}| < 2.7$.

4.2 Jet punch-through and the measured jet energy

As described in Section 4.1, $\mathrm{N}_{segments}$ is used as a measure of the muon activity behind a jet, and thus to quantify the extent to which a jet is punching-through the calorimeter. Thus the impact of jet punch-through can be studied by assessing the dependence of the jet energy scale and jet energy resolution on $\mathrm{N}_{segments}$. These studies are only done for punch-through jets as defined in Section 4.1.2. The nominal PYTHIA8 dijet Monte Carlo samples were used, with the MC dijet event selection detailed in Section 6.2 applied. Only information from the two leading jets in each event was used.

For all studies in this section, as opposed to being extracted from a fit, all mean values are calculated as the weighted mean:

$$\langle x \rangle = \frac{\sum_i w_i x_i}{\sum_i w_i} \tag{4.2}$$

where $\langle x \rangle$ is the weighted mean of x, w_i is the weight (in this case the MC weight given in Equation 1.10) and x_i is the bin centre for the i-th bin in the histogram. The RMS, for the purposes of this thesis, is defined to be the standard deviation of the distribution, calculated as:

$$RMS(x) = \sigma_x = \sqrt{\frac{1}{\mathrm{N}_{eff}} \sum_i \left(x_i - \langle x \rangle\right)^2} \tag{4.3}$$

where σ_x is the standard deviation of x, x_i is the bin centre for the i-th bin and N_{eff} is the

number of effective entries in the distribution:

$$\mathrm{N}_{eff} = \frac{\left(\sum_i^{N_{Bins}} w_i\right)^2}{\sum_j^{N_{Bins}} w_j^2} \tag{4.4}$$

4.2.1 Jet energy mis-measurement due to jet punch-through

The magnitude of energy loss due to jet punch-through is measured using two variables:

- The jet energy response as defined in Equation 3.4 is used to assess the fraction of a jet's energy lost due to jet punch-through. A jet energy response below 1 for jets with large $\mathrm{N}_{segments}$ will be a good indicator that the jet energy is being under-measured due to jet punch-through.

- The magnitude of energy lost by a jet due to jet punch-through, E_{lost}:

$$\mathrm{E}_{lost} = \mathrm{E}_{truth} - \mathrm{E}_{calo} \tag{4.5}$$

  where E_{truth} is truth energy of the jet, and E_{calo} is the reconstructed calorimeter jet energy calibrated at the EM+JES+GS or LCW+JES+GS scale. This definition is not perfect, and is sensitive to noise. The main function of this variable is to provide an alternative measure of the imbalance between the truth jet energy and the reconstructed jet energy that can be used to determine how the imbalance changes with $\mathrm{N}_{segments}$. As such, this variable will only be used to crudely assess the magnitude of energy lost due to jet punch-through in conjunction with the jet energy response.

It is important to note that both of these variables, as they use truth jets without muons and neutrinos, will not be sensitive to energy losses due to *escaped energy* (see Section 1.3.2). This is ideal, as for the purposes of this thesis, escaped energy is not considered under longitudinal hadronic shower leakage.

In Section 4.1.1, it was shown that $N_{segments}$ is correlated with the jet punch-through probability. Thus by looking at how these two variables correlate with $N_{segments}$, the extent to which energy leakage due to jet punch-through affects the jet energy scale can be assessed.

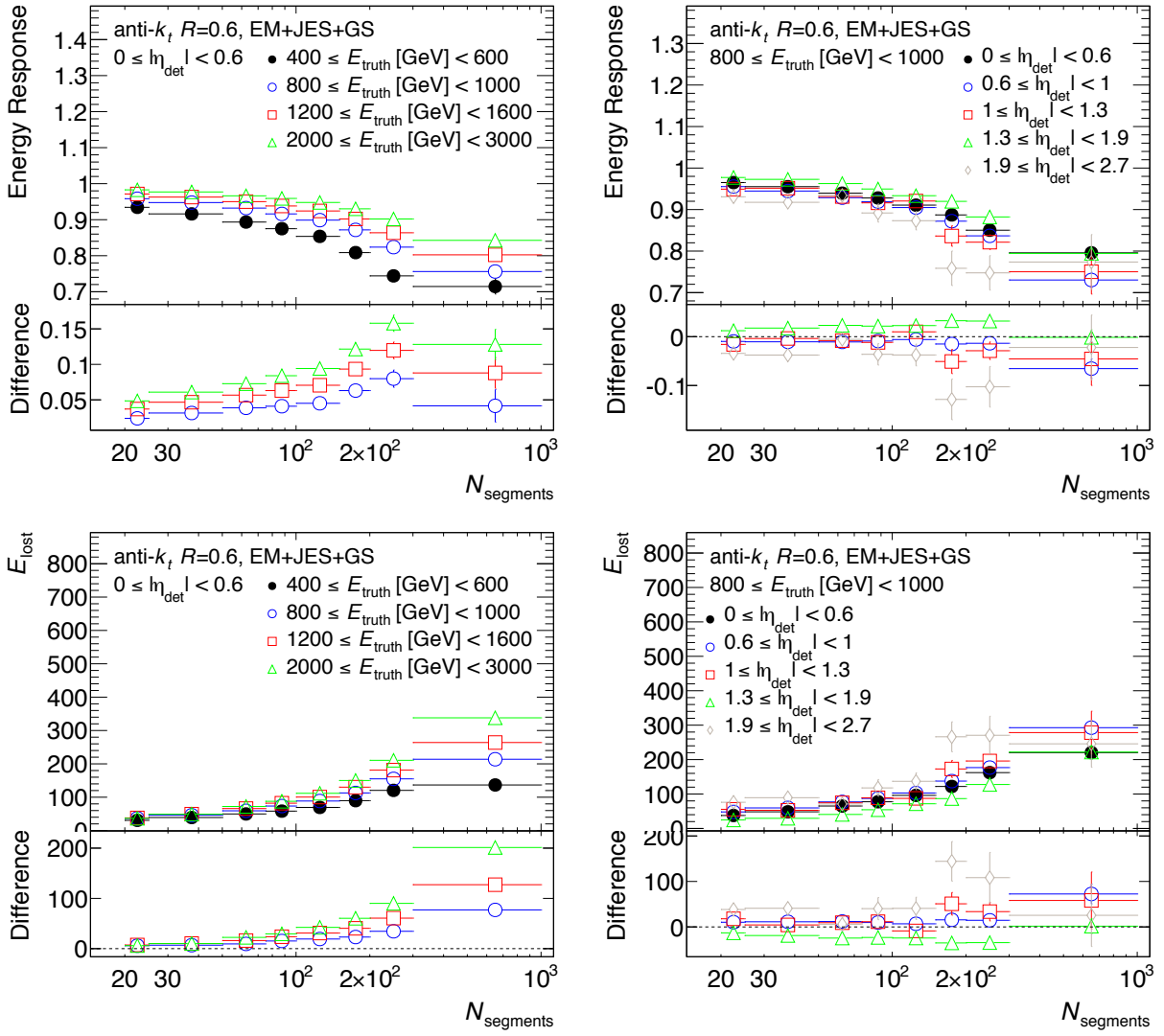


Figure 4.7: Jet energy response (top) and E_{lost} (bottom) vs $N_{segments}$ for jets in the central region ($0 \le |\eta_{det}| < 0.6$) with different E_{truth} bins (left) for jets with $800 \le E_{truth}$ [GeV] < 1000 in different $|\eta_{det}|$ regions (right). The bottom plots show $R_i - R_{nom}$, where R_i and R_{nom} are the response (top) or E_{lost} (bottom) in the i-th bin and in the $400 \le E_{truth}$ [GeV] < 600 bin (left) or in the central region (right) respectively.

Figure 4.7 shows the jet energy response and E_{lost} vs $N_{segments}$ for jets in various $|\eta_{det}|$ and E_{truth} bins. From this figure, it is clear that as $N_{segments}$ increases, the jet energy response decreases, while E_{lost} increases. However jets with the same amount of activity behind them in the muon spectrometer do not always display the same magnitude of energy

loss due to jet punch-through. Instead, the amount of energy lost as a function of $N_{segments}$ is dependant on E_{truth} and η_{det}. From the top-left plot in Figure 4.7, it is clear the fraction of energy lost decreases as E_{truth} increases, with up to 25% losses observed for jets with $200 \leq E_{truth}$ [GeV] < 400. However the bottom-left plot shows that although a smaller fraction of energy is lost by higher energy jets, the magnitude of energy loss is still larger at high E_{truth}. The top-right plot in Figure 4.7 shows the $|\eta_{det} < 1.3$ and $|\eta_{det} > 1.9$ regions appear to display the largest energy losses due to jet punch-through, even though it was observed in Figure 4.6 that jets in the region $1.3 \leq |\eta_{det}| < 1.9$ displayed the highest jet punch-through probability. This is likely due to the reduced material in the $1.3 \leq |\eta_{det}| < 1.9$ region, resulting in a larger number of low energy particles leaking into the muon spectrometer. This produces more muon segments, but with smaller energy losses, thus causing higher response jets to appear in higher $N_{segments}$ bins.

4.2.2 Effect on jet energy response tails

It has been shown that jet punch-through causes a reduction in the measured mean jet energy response. However, it is expected jet punch-through will also result in jet-by-jet fluctuations of the response, whereby punch-through jets will have a lower reconstructed energy, resulting in a jet energy response of < 1. This can artificially inflate the low response tail ($\mathcal{R} < 1$).

This is investigated by looking at the percentage of events within the response distribution that fall below a given response value, called the response threshold ('Response thr') for the purpose of these studies. By varying the response threshold and comparing the fraction of events below the threshold in different $N_{segments}$ bins, it is possible to study the impact of jet punch-through on the tails of the distribution. This is effectively comparing how the cumulative frequency distribution of the response changes with $N_{segments}$.

In Figure 4.8, the response threshold is varied from 0.2 to 1 in increments of 0.2 for an example response distribution. The response threshold is annotated by the dotted red line, with the corresponding fraction of events below this threshold represented by the shaded region. The bottom-right plot in Figure 4.8 shows the fraction of events in these shaded

regions plotted against the corresponding response threshold. As expected, as the response threshold increases, the number of events below threshold increases.

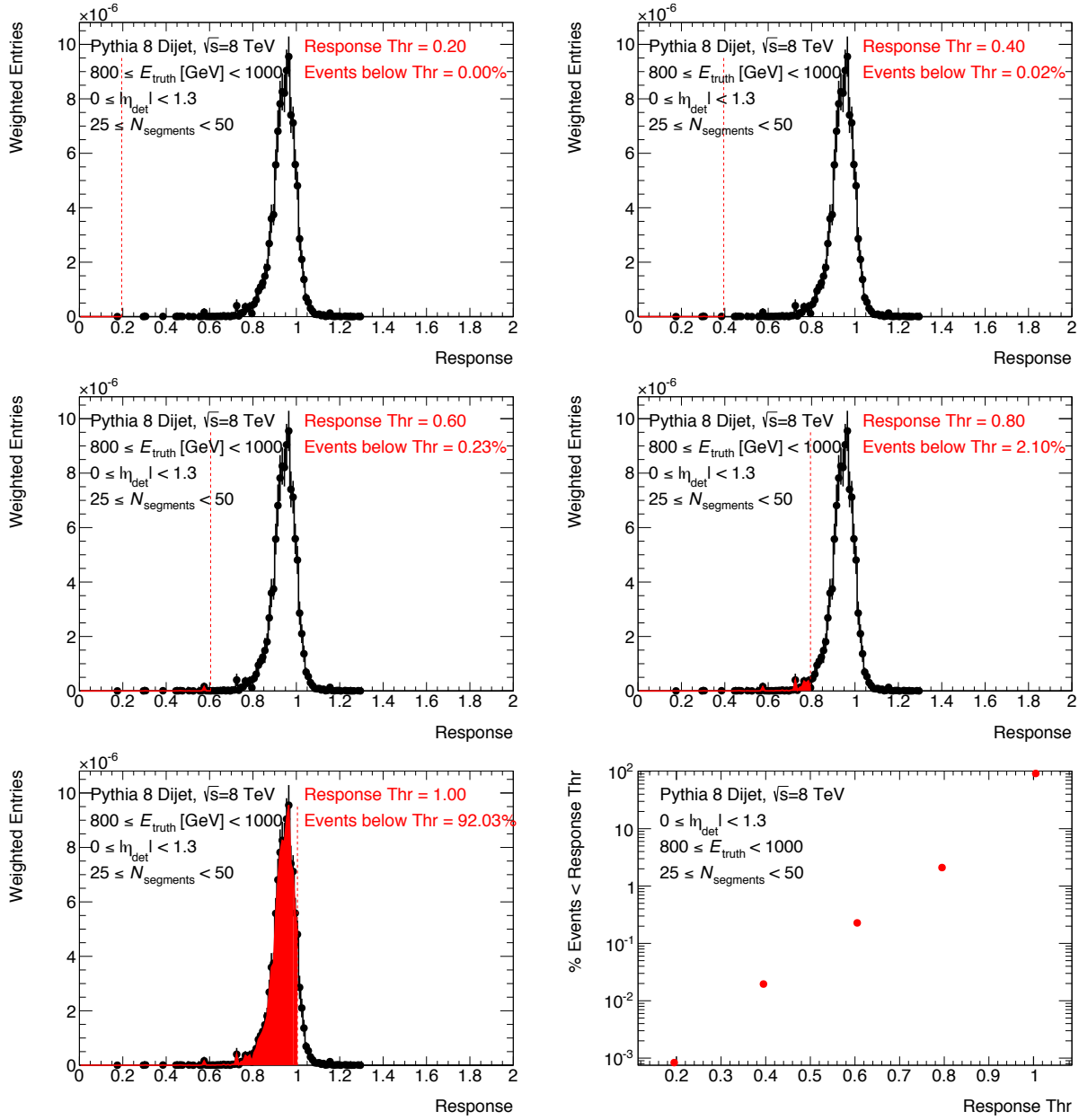


Figure 4.8: Jet energy response distribution for jets with $800 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 1000$, $25 \leq \mathrm{N}_{segments} < 50$ in the detector region $0 \leq \eta_{det} < 1.3$ with response threshold ('Response thr') defined by red line, and fraction of events below this threshold represented by the shaded region and quantified as 'Events below Thr'. Response thresholds of 0.2 (top-left), 0.4 (top-right), 0.6 (middle-left), 0.8 (middle-right) and 1.0 (bottom-left) are shown. The bottom right plot shows the fraction of events below each response threshold vs the response threshold.

Figure 4.9 shows the percentage of events below a given response threshold ('Response

thr') for different $N_{segments}$, E_{truth} and η_{det} bins. It is clear from looking at these plots that as $N_{segments}$ increases, the percentage of events in the low response tail increases, suggesting punch-through is inflating these tails. Punch-through does not appear to be having a large impact on the jet energy response tails in the $1.9 \leq |\eta_{det}| < 2.7$, with low statistics seen in this region, again expected given the punch-through fraction is much smaller in this region.

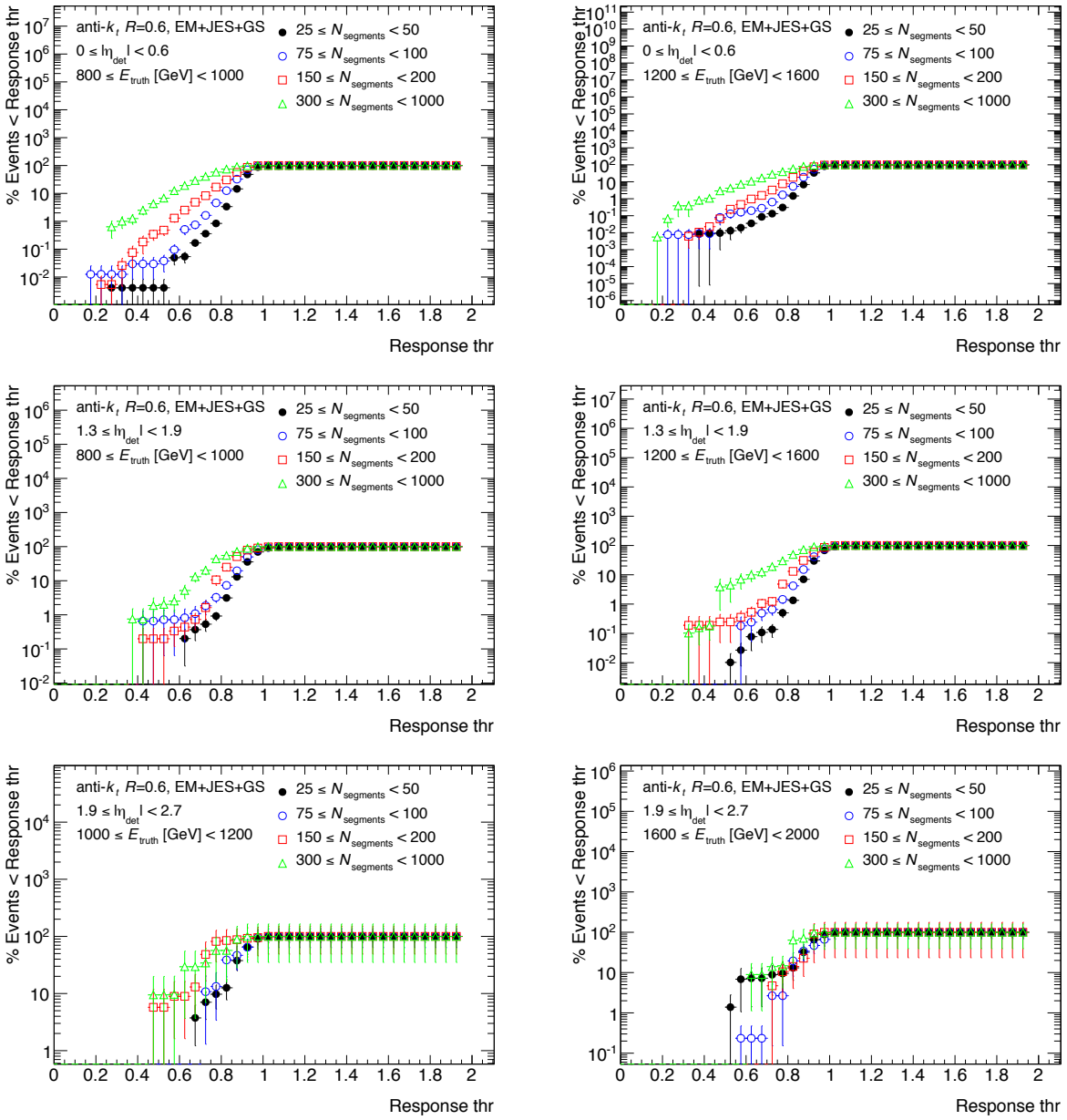


Figure 4.9: Percentage of jets below a given response threshold ('Response thr') as the response threshold increases. This is compared for jets with different $N_{segments}$ values in the regions $0 \leq |\eta_{det}| < 0.6$ (top), $1.3 \leq |\eta_{det}| < 1.9$ (middle) and $1.9 \leq |\eta_{det}| < 2.7$ (bottom), with $800 \leq E_{truth}$ [GeV] < 1000 (left) and $1200 \leq E_{truth}$ [GeV] < 1600 (right).

4.2.3 Degradation of jet energy resolution due to jet punch-through

It was seen in the previous section that energy losses due to jet punch-through increase the fluctuation in the measured jet energy, contributing to the low response tail in the jet energy response distribution, consistent with what we expect from Section 1.3.3. This effect can degrade the jet energy resolution. Thus it is important to study the impact of jet punch-through on the jet energy resolution. The jet energy resolution, r, is defined in this context in terms of the RMS of the jet energy response:

$$r = \frac{RMS(\mathcal{R})}{\langle \mathcal{R} \rangle} \tag{4.6}$$

where $RMS(\mathcal{R})$ is the RMS (as defined in Equation 4.3) of the jet energy response and $\langle \mathcal{R} \rangle$ is the mean jet energy response. In the remainder of this thesis, comparisons between jet energy resolution values are often made by taking the difference between the two resolution values in quadrature, whilst also taking into account the sign of the difference:

$$sgn(\sigma^{'} - \sigma)\sqrt{\sigma^{'2} - \sigma} \tag{4.7}$$

where $\sigma^{'}$ and σ denote the resolution values being compared, and $sgn(\sigma^{'} - \sigma)$ is the sign of the difference between $\sigma^{'}$ and σ.

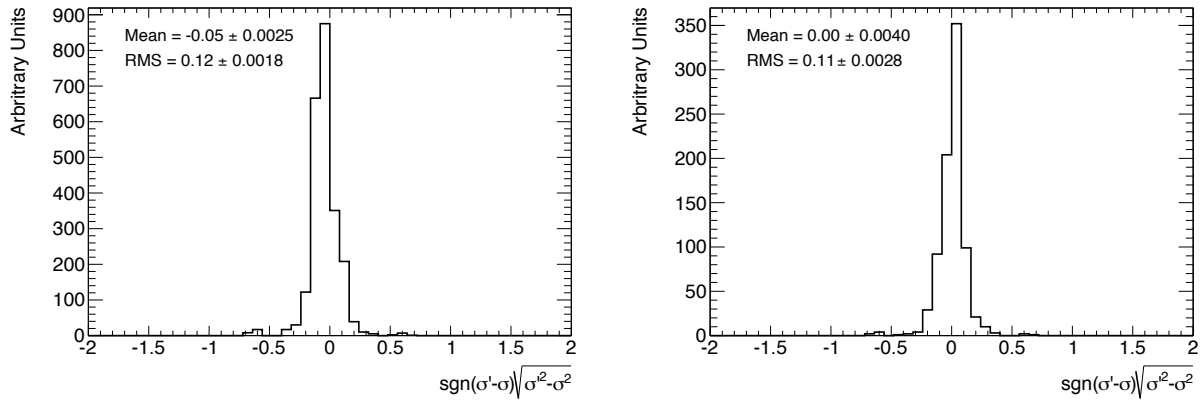


Figure 4.10: $sgn(\sigma^{'} - \sigma)\sqrt{\sigma^{'2} - \sigma}$ when comparing the resolution in different E_{truth} bins (left) and different η_{det} bins (right).

An example distribution of $sgn(\sigma' - \sigma)\sqrt{\sigma'^2 - \sigma}$, calculated when comparing the resolution in different E_{truth} bins (left) and different η_{det} bins (right), is shown in Figure 4.10. For these comparisons, this variable appears to be roughly Gaussian, with mean very close to zero and an RMS of roughly 0.12.

Figure 4.11 shows r vs $N_{segments}$ and percentage of energy loss due to jet punch-through. The percentage of energy loss due to jet punch-through is taken to be $(1 - \langle \mathcal{R}(N_{segments}) \rangle) \times 100$, where $\langle \mathcal{R}(N_{segments}) \rangle$ is the mean jet energy response in the $N_{segments}$ bin the jet in question belongs to.

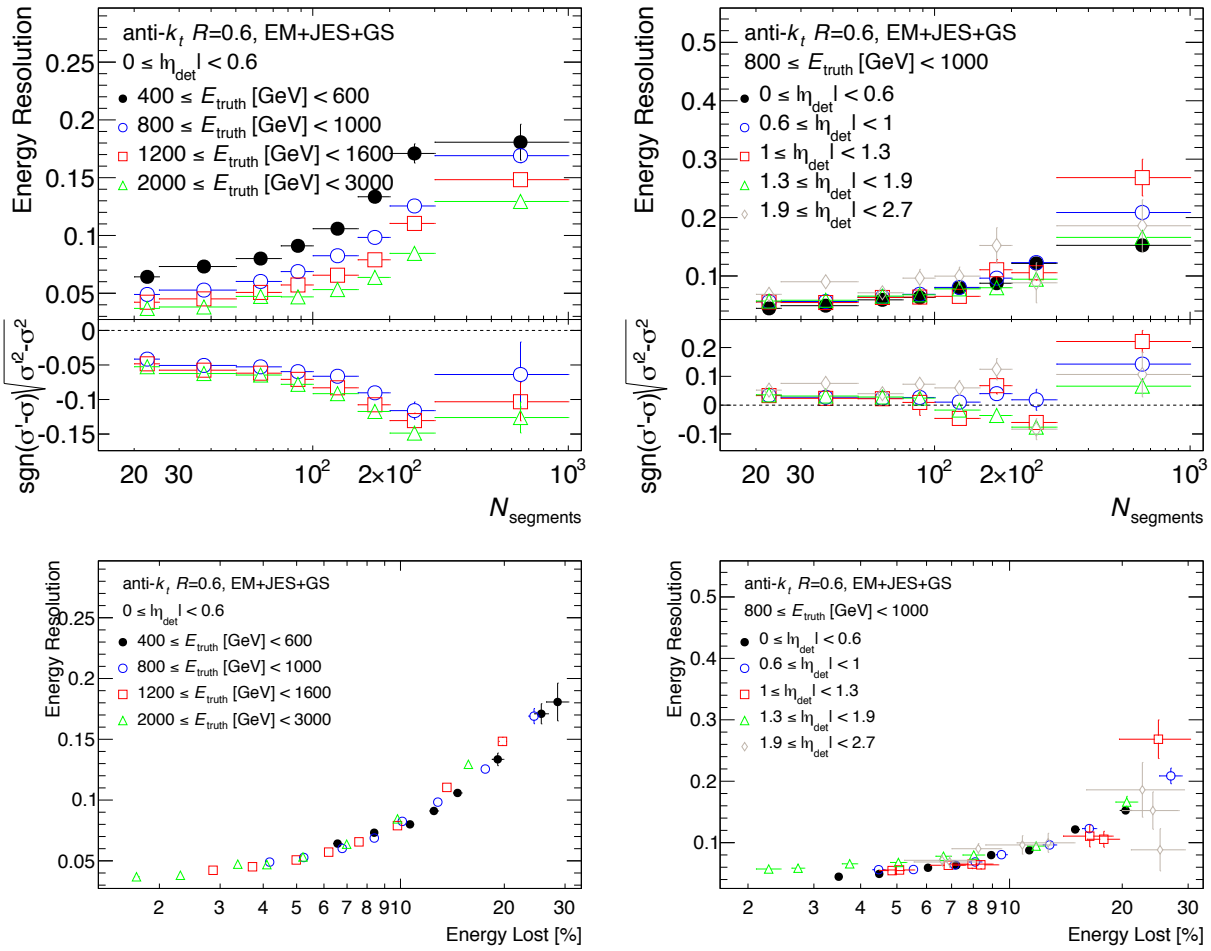


Figure 4.11: Jet energy resolution vs $N_{segments}$ (top) and the percentage of energy lost due to jet punch-through (bottom) for jets in the central region ($0 \leq |\eta_{det}| < 0.6$) with different E_{truth} (left), and jets with $800 \leq E_{truth}$ [GeV] < 1000 in different $|\eta_{det}|$ regions (right). The bottom plots show the difference in quadrature relative to jets in the $400 \leq E_{truth}$ [GeV] < 600 bin (top-left), or to jets in the central region (top-right).

It is clear that as $N_{segments}$ increases, the jet energy resolution also increases, thus de-

grading. As with the response, the effect as a function of $N_{segments}$ is worse for lower E_{truth} jets, with the resolution increasing from $\sim 6\%$ to $\sim 18\%$ - a degradation of up to 300%. The degradation as a function of $N_{segments}$ is worst in the region $1.3 \leq |\eta_{det}| < 1.9$. Although the degradation depends on $|\eta_{det}|$ and E_{truth} as a function of $N_{segments}$, when looking at the resolution vs percentage of energy loss, there appears to be very little dependency on E_{truth}. Poor statistics plague the $1.9 \leq |\eta_{det}| < 2.7$ region. Overall, a very clear degradation can be seen as the percentage of energy loss increases. These observations are consistent with those discussed in Reference [48] (see Section 1.3.3), where the resolution was seen to degrade by up to 200% due to longitudinal hadronic shower leakage.

4.2.4 Flavour dependencies

The flavour of the jet initiating parton can impact the jet constituents, changing the composition of the charged particles that may penetrate the muon spectrometer in the event of a jet punching through the calorimeter. This can impact the observed correlation between the jet energy response and $N_{segments}$. Thus it is important to study how jet flavour affects this relationship.

In this section, jets are tagged using the *truth flavour tagging* prescription in Section 3.7. Jets tagged as originating from an up, down or strange quark are labelled as **L**ight **Q**uark (**LQ**) jets, b-jets are those tagged to be originating from a b-hadron and gluon jets are those tagged to be originating from a gluon. It is important to note that the samples used were dominated by LQ-jets for jets with $E_{truth} > 400$ GeV. For jets with $E_{truth} < 400$ GeV, LQ and gluon jets are dominant, with gluon jets becoming progressively more dominant at lower E_{truth}. In addition, at high p_T^{truth} ($p_T^{truth} > 200$ GeV), the GS corrections (see Section 3.5.4) reduce the difference between the LQ and gluon jet response to below 1%, and thus very large differences in the jet response due to flavour are not expected (see Ref. [1] for more information).

Figure 4.12 shows how the jet energy response as a function of $N_{segments}$ varies for LQ, b and gluon jets in comparison with the inclusive jet sample. The comparison is made

for jets with $600 \leq E_{truth}$ [GeV] < 800 and $1600 \leq E_{truth}$ [GeV] < 2000 in the $0 \leq |\eta_{det}| < 0.6$ and $1.3 \leq |\eta_{det}| < 1.9$ regions. As the sample is dominated by LQ-jets in these E_{truth} regions, we see a clear bias of the inclusive jet response towards the LQ-jet response. In all regions, the LQ-jet response displays the strongest negative correlation with $N_{segments}$. The b-jet and gluon-jet response is generally in agreement to within errors. The difference between the LQ, b and gluon jet response is smaller at low $N_{segments}$. Overall these plots suggest that particles leaking from LQ-jets deposit fewer segments when compared with particles carrying the same amount of energy, but leaking from b and gluon jets. However, it is important to remember that the jet energy response defined here does not take into account muons and neutrinos escaping the calorimeter. Including muons and neutrinos in the truth jet collections used would cause the b-jet response to shift to lower values, a result of the escaped muon.

Figure 4.13 compares the resolution vs $N_{segments}$ for the different jet flavours in the same E_{truth} and η_{det} regions. As with the response, it can be seen the jet energy resolution of LQ-jets displays the largest sensitivity to $N_{segments}$, showing the largest degradation as $N_{segments}$ increases, while b-jets and gluon-jets generally display a more similar jet energy resolution.

Although small differences are observed in the dependence of the jet energy response and jet energy resolution on $N_{segments}$ for different flavour jets, due to the lack of efficient flavour tagging techniques at high energy in data, flavour was not taken into consideration when deriving the punch-through correction, described in Chapter 5.

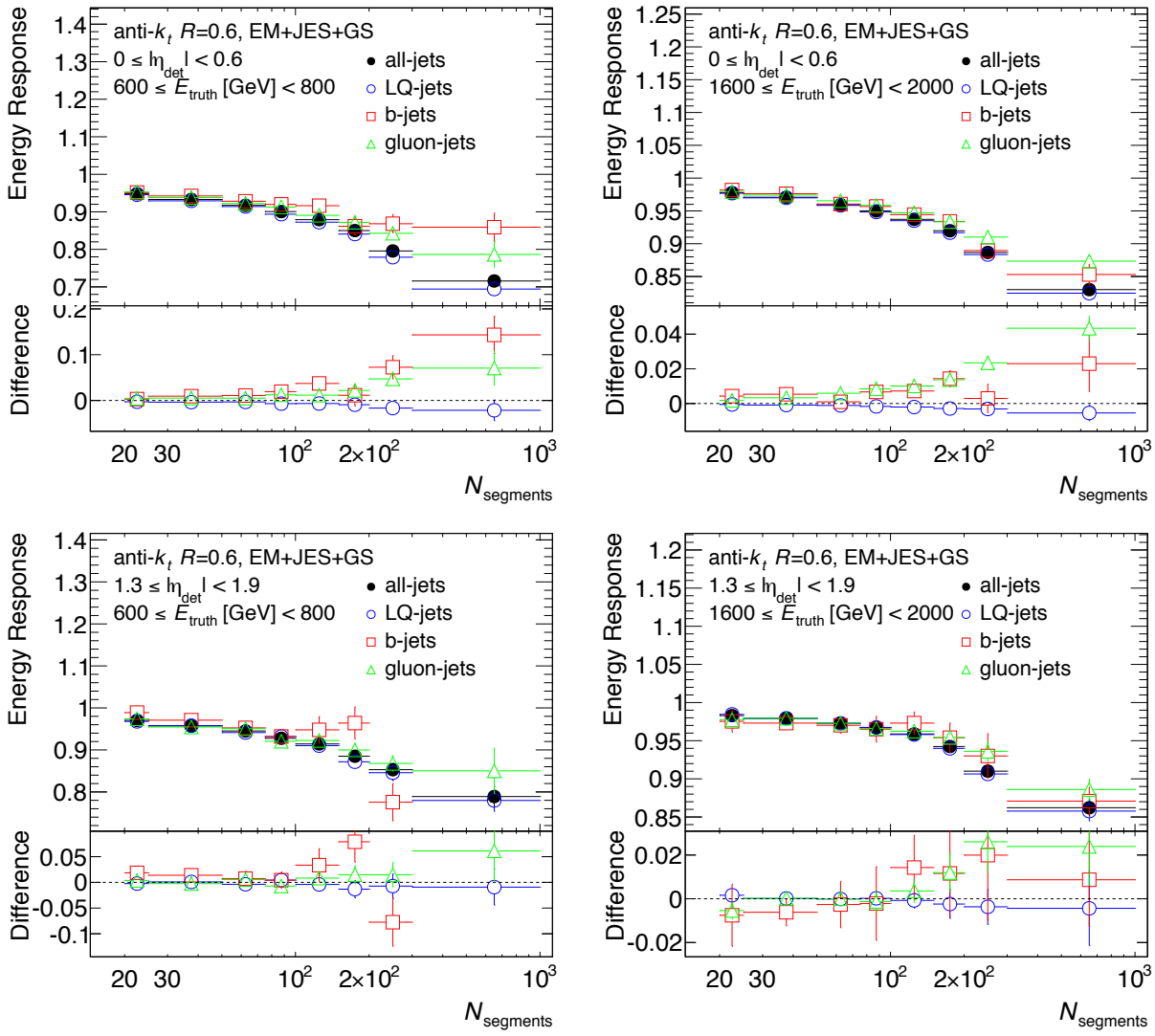


Figure 4.12: Jet energy response vs $N_{segments}$ for different flavour jets in the regions $0 \le |\eta_{det}| < 0.6$ (top) and $1.3 \le |\eta_{det}| < 1.9$ (bottom), with $600 \le E_{truth}$ [GeV] < 800 (left) and $1600 \le E_{truth}$ [GeV] < 2000 (right). The bottom plots show $R_i - R_{nom}$, where R_i is the response in the i-th bin and R_{nom} is the response for all jets.

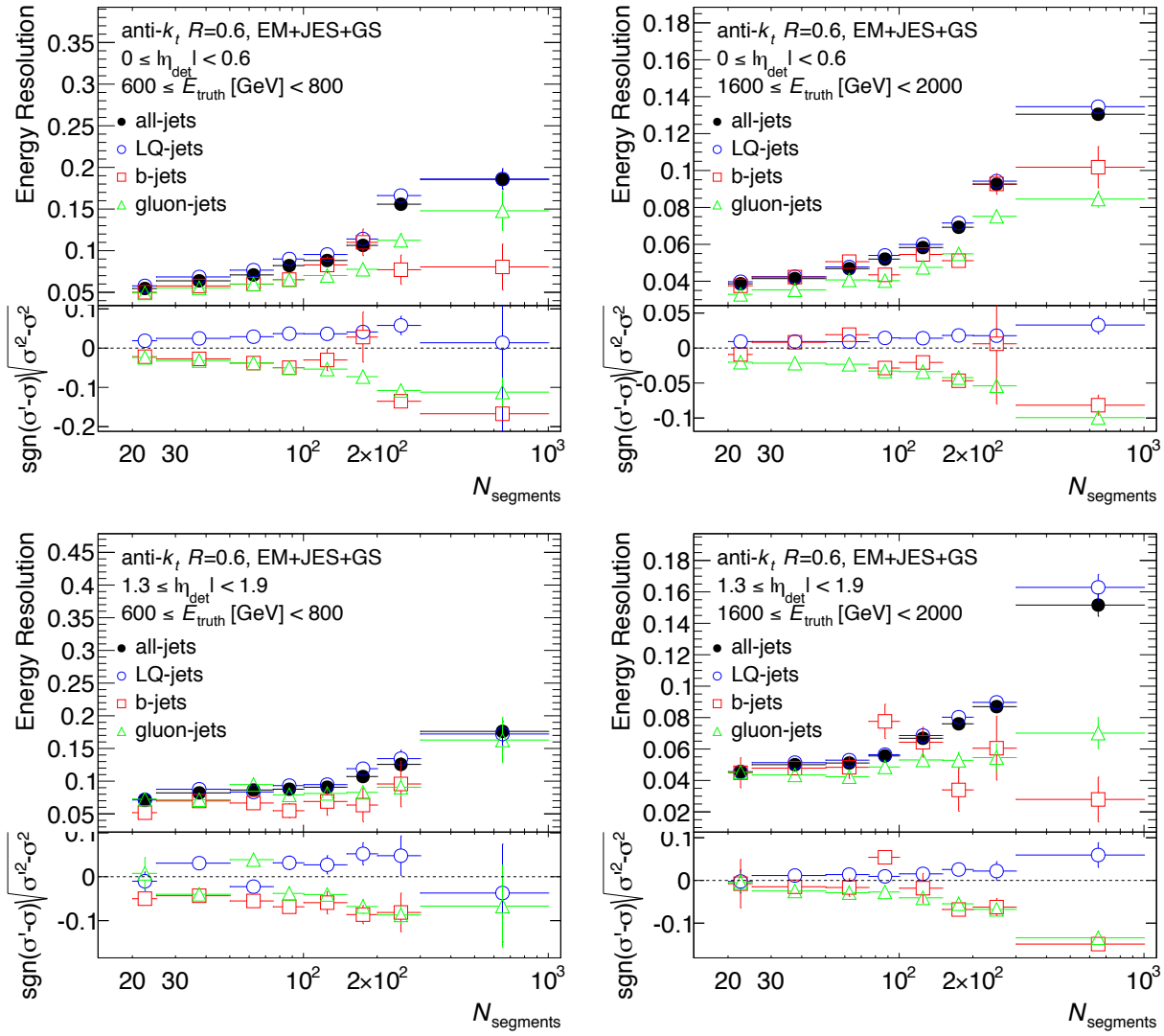


Figure 4.13: Jet energy resolution vs $N_{segments}$ for different flavour jets in the regions $0 \le |\eta_{det}| < 0.6$ (top), $1.3 \le |\eta_{det}| < 1.9$ (bottom), with $600 \le E_{truth}$ [GeV] < 800 (left) and $1600 \le E_{truth}$ [GeV] < 2000 (right). The bottom plots show the difference in quadrature for each jet flavour (σ') in comparison with all-jets (σ).

Chapter 5

Jet Punch-Through Correction

As shown in Section 4.2.1 (Appendix D), the current ATLAS EM+JES+GS (LCW+JES+GS) jet energy calibration is sensitive to effects due to jet punch-through, with a clear degredation in both the jet energy response and jet energy resolution observed as $N_{segments}$ increases.

This chapter details the derivation of a jet punch-through correction that aims to remove the observed dependancies of the jet energy response and jet energy resolution on $N_{segments}$, thus producing a jet energy scale insensitive to the effects of longitudinal hadronic shower leakage. The correction is another MC correction, applicable to both data and MC, specifically targeting punch-through jets as defined in Section 4.1.2. It is designed to be relative to the JES correction in Section 3.5.3, and applied after the GS corrections.

A detailed overview of the derivation procedure can be found in Section 5.1. Section 5.2 then describes the selection applied to jets used to derive the correction. A discussion of the binning used for the jet energy response distributions can be found in Section 5.3, while the procedure used to fit these distributions is described in Section 5.4. The process of obtaining corrections relative to the JES correction in Section 3.5.3 is explained in Section 5.5. These relative corrections are then used to form the jet punch-through correction function using a sequence of Gaussian kernels as detailed in Section 5.6. Finally, the corrections derived from the kernels, along with the final jet punch-through corrections applied to 2012 ATLAS data can be found in Section 5.7.

The jet punch-through correction was derived for anti-k_t jets, with R = 0.4 and R = 0.6, calibrated with the EM+JES+GS and LCW+JES+GS schemes, producing jets at the EM+JES+GS+PT and LCW+JES+GS+PT scales respectively, as defined by Equation 5.1.

5.1 Overview of derivation

The jet punch-through correction exploits the dependence of the jet energy response on $\mathrm{N}_{segments}$. As $\mathrm{N}_{segments}$ is just another jet property, this response dependency allows the jet punch-through correction to be constructed in a similar way to the GS corrections (described in Section 3.5.4), whereby the correction, $C(\mathrm{N}_{segments})$, is derived by inverting the jet energy response as a function of $\mathrm{N}_{segments}$ ($C(\mathrm{N}_{segments}) = \mathcal{R}^{-1}(\mathrm{N}_{segments})$). The correction is relative to the jet energy scale correction. Deriving the correction in this way enables easy integration into the initial ATLAS 2012 EM+JES+GS and LCW+JES+GS jet calibration schemes described in Section 3.5. The full procedure used to derive the jet punch-through correction is outlined below:

1. Select for real jets unaffected by detector malfunctions and other calibration related issues as described in Section 5.2.

2. Produce distributions of $\mathcal{R}$ in bins of E_{truth}, $\mathrm{N}_{segments}$ and η_{det}. The binning decisions are outlined in Section 5.3.

3. Extract $\mu = \langle\mathcal{R}\rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}}$ from a Gaussian fit to these distributions. The fitting procedure is described in Section 5.4.

4. Normalise $\langle\mathcal{R}\rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}}$ by the mean jet energy scale ($\langle\mathcal{R}\rangle_{\mathrm{JES}}|_{\eta_{det},\mathrm{E}_{truth}}$) to obtain a correction relative to the jet energy scale correction described in Section 3.5.3 ($\mathcal{F}_{\mathrm{JES}}$). This procedure is explained in Section 5.5.

5. Construct the correction function in bins of η_{det} ($\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel}}(E^{jet}, \log \mathrm{N}_{segments})|_{\eta_{det}}$) using a smoothing procedure involving a sequence of two Gaussian kernels. Section 5.6 goes into detail on this step.

The punch-through corrected energy in each η_{det} bin is then defined as:

$$\mathrm{E}_{calo}^{\mathrm{EM+JES+GS+PT}} = \frac{\mathrm{E}_{calo}^{\mathrm{EM+JES+GS}}}{\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(E^{jet}, \log \mathrm{N}_{segments})|_{\eta_{det}}} \tag{5.1}$$

where $\mathrm{E}_{calo}^{\mathrm{EM+JES+GS+PT}}$ is the reconstructed energy of the jet when taking into account energy losses due to jet punch-through, $\mathrm{E}_{calo}^{\mathrm{EM+JES+GS}}$ is the jet energy after the fourth GS correction (see Section 3.5.4) and $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(E^{jet}, \log \mathrm{N}_{segments})|_{\eta_{det}}$ is the final jet punch-through correction function. For LCW+JES+GS jets, the final corrected jet energy is $\mathrm{E}_{calo}^{\mathrm{LCW+JES+GS+PT}}$, obtained by replacing $\mathrm{E}_{calo}^{\mathrm{EM+JES+GS}}$ in the equation above with the energy of the jet at the LCW+JES+GS-scale, $\mathrm{E}_{calo}^{\mathrm{LCW+JES+GS}}$.

5.2 Jet selection

The baseline high statistic Pythia8 dijet samples detailed in Section 1.4 were used to derive the correction factors. The weights defined in Equation 1.10 were therefore applied to restore the correction jet p_{T} distribution shape. The MC cleaning cuts described in Section 3.6 were applied to remove excess overlaid pile-up events. As only MC was used, no trigger selection was applied. Events were required to contain at least one primary vertex with > 4 associated tracks.

In order to ensure only real jets unaffected by detector malfunctions were used, the jet-by-jet cleaning cuts detailed in Section 3.6 were applied. The jet isolation criteria described in Section 3.5.3 were applied to ensure only isolated jets were used. At least two jets passing these criteria were required in order for an event to be used.

Truth jets were required to have $p_{\mathrm{T}} > 7$ GeV and were matched geometrically to calorimeter jets as described in Section 3.4. In addition, only jets defined to be punch-through jets using the definition in Section 4.1.2 were used during the derivation ($p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV and $\mathrm{N}_{segments} \geq 20$). Finally, all jets used were calibrated to the EM+JES+GS or LCW+JES+GS scale.

5.3 Determination of binning

Using jets and events passing the selection criteria in the previous section, distributions of the jet energy response in bins of η_{det}, $N_{segments}$ and E_{truth} must be obtained, to which a Gaussian distribution can be fit. The mean jet energy response extracted from these fits is ultimately used to form a smooth correction as a function of $\log N_{segments}$ and E_{truth}, binned in η_{det}. The final correction function must effectively capture the dependencies of the jet energy response on $N_{segments}$, E_{truth} and η_{det} observed in 4.2.1. In order to do this, the following two requirements must be met when determining how to bin the jet energy response distributions:

1. The η_{det} binning must preserve any differences in the structure of the mean jet energy response plane in (E_{truth}, $N_{segments}$) phase space due to η_{det}, whilst also maximising the statistics available for fits.

2. The E_{truth} and $N_{segments}$ binning must provide adequate statistics for fits to converge, while avoiding the masking of finer structures present in the mean jet energy response plane in (E_{truth}, $N_{segments}$) phase space.

These requirements are met by:

1. Studying the variations in the mean jet energy response as a function of E_{truth} and $N_{segments}$ in bins of η_{det} and optimising the η_{det} binning to minimise these variations (see Section 5.3.1).

2. Optimising the E_{truth} and $N_{segments}$ binning in order to maximise the number of successful fits whilst keeping the binning as fine as possible (Section 5.3.2).

5.3.1 Preserving η_{det} dependencies

In the jet energy response studies performed in Section 4.2.1, the top-right plot in Figure 4.7 showed the jet energy response was similar for jets in the region $|\eta_{det}| < 1.3$, but different

in the $1.3 \leq |\eta_{det}| < 1.9$ and $1.9 \leq |\eta_{det} < 2.7$ regions. These regions correspond to the distinct detector regions listed in Table 5.1. As these three η_{det} bins represented distinct detector regions, and showed differences in the mean jet energy response, they were used for the derivation of the jet punch-through correction.

ATLAS Hadronic Calorimeter Detector Region	η_{det} **Range**
Barrel	$0 \leq \|\eta_{det}\| < 1.3$
Barrel-Endcap Transition (crack region)	$1.3 \leq \|\eta_{det}\| < 1.9$
Endcap (HEC only)	$1.9 \leq \|\eta_{det}\| < 2.7$

Table 5.1: Table summarising η_{det} bins and associated detector regions for jet punch-through correction derivation.

5.3.2 $N_{segments}$ and E_{truth} binning

As mentioned in Section 5.3, the $N_{segments}$ and E_{truth} binning must be chosen to be as fine as possible whilst preserving enough statistics to allow the maximum number of fits to converge.

In order to take into account the fact that we are using weighted events, the number of effective entries (N_{eff}) as defined in Equation 4.4 is used as a measure for how many events are available in each bin. Figure 5.1 shows how N_{eff} varies as a function of E_{truth} and $N_{segments}$ for the three η_{det} bins - $0 \leq \eta_{det} < 1.3$, $1.3 \leq \eta_{det} < 1.9$ and $1.9 \leq \eta_{det} < 2.7$. It is clear N_{eff} is largest in the high E_{truth}, low $N_{segments}$ region. The statistically limited regions are at low E_{truth} and high $N_{segments}$ i.e. N_{eff} decreases as $N_{segments}$ increases and E_{truth} decreases. However in the highest E_{truth} region N_{eff} begins to decline due to the limited number of jets in this region. This is compatible with what was observed in Section 4.1.3 whereby the jet punch-through probability is largest for jets at high E_{truth}. It is also important to note the derivation is very statistically limited in the $1.9 \leq \eta_{det} < 2.7$ region, where no bins have $N_{eff} > 100$. This is due to the increased detector material in this region.

From this figure, one could conclude large bin widths should be used at low E_{truth}, high $N_{segments}$, while smaller bin widths should be used at high E_{truth}, low $N_{segments}$. However as mentioned above, it is important the binning used does not mask any structures that may

be present in the mean jet energy response vs $N_{segments}$ vs E_{truth} plane. Figure 5.2 shows the mean jet energy response as a function of E_{truth} in bins of $N_{segments}$ (left) and as a function of $N_{segments}$ in bins of E_{truth} (right) for jets in each of the three detector regions.

Looking at the mean jet energy response as a function of E_{truth}, there is a clear dependence of the response on E_{truth} when binning in $N_{segments}$. As $N_{segments}$ increases, the dependence of the mean jet energy response on E_{truth} appears to grow stronger.

At low E_{truth}, the response is high in all η_{det} regions, reaching up to 60% above 1. This high mean jet energy response arises from a combination of the $p_{\mathrm{T}} > 50$ GeV cut biasing the response and the non-closure of the JES in this region (see Section 3.5.3). As E_{truth} increases, the response then decreases to below 1, before increasing again but never again reaching above 1. The turning point in the response vs E_{truth} curve occurs at between 100 and 200 GeV for the $0 \leq \eta_{det} < 1.3$ and $1.3 \leq \eta_{det} < 1.9$ regions. In the $1.9 \leq \eta_{det} < 2.7$ region this turning point occurs later ($E_{truth} \sim 600$ GeV). As mentioned in Section 4.2.1, even though the response is increasing with E_{truth}, E_{lost} is still larger at high E_{truth}.

As a function of $N_{segments}$, we see the previously observed trend of the response decreasing with increasing $N_{segments}$. The curves appear to get gradually steeper as $N_{segments}$ increases until $N_{segments} \sim 100$ after which point the gradient of the curves appears to plateau.

Based upon these observations, it is clear that even though the low E_{truth} region is lacking in statistics, large bins cannot be used as this could prevent the behaviour of the response at low E_{truth} from being captured accurately by the Gaussian kernel when forming the correction function. Instead, the binning must be fine in the low E_{truth} region, with bins getting progressively larger in the high E_{truth} region where the gradient of the curve appears to stabilise. Similarly, fine binning is necessary in the low $N_{segments}$ region to capture the change in the gradient of the curve, while at high $N_{segments}$ larger bins are sufficient to capture the observed behaviour. The final binning used is discussed in Section 5.3.3 and based on similar bin widths at log scale.

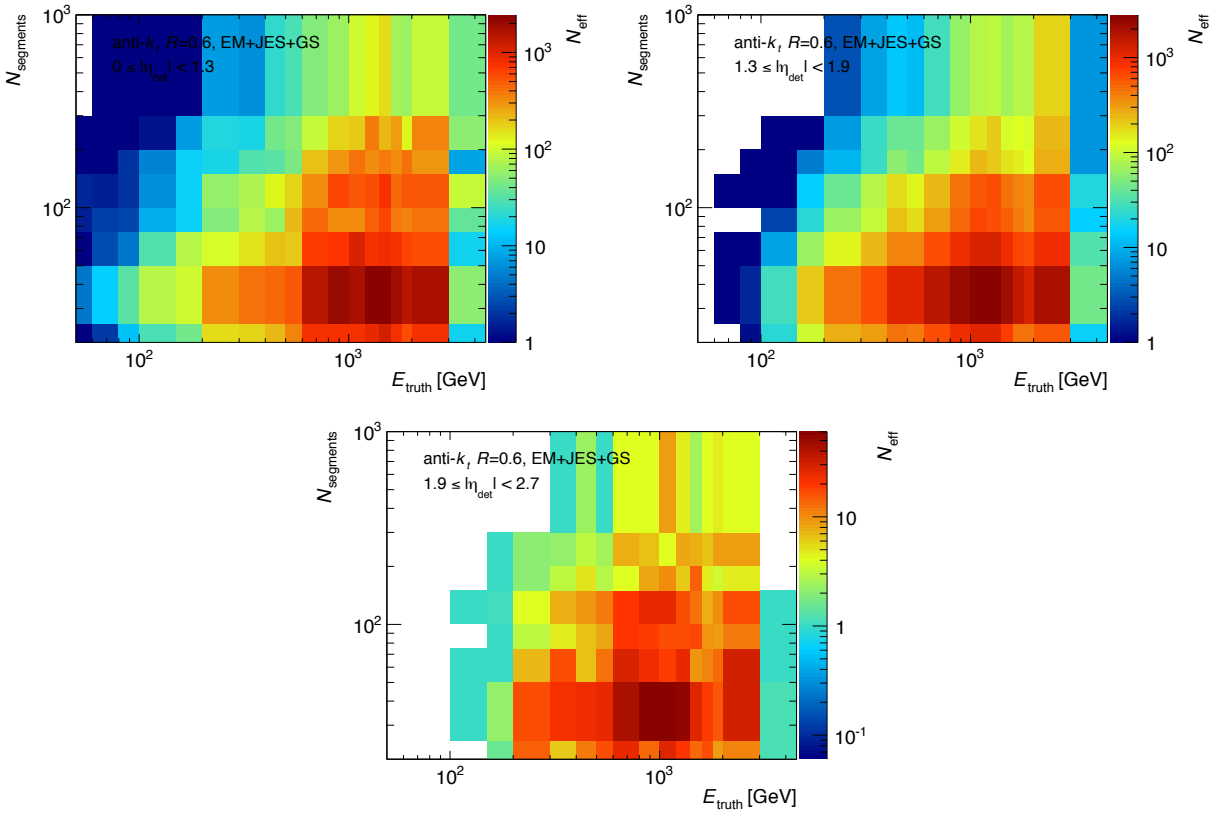


Figure 5.1: N_{eff} vs E_{truth} vs $N_{segments}$ for jets in the detector regions $0 \le \eta_{det} < 1.3$ (left), $1.3 \le \eta_{det} < 1.9$ (right) and $1.9 \le \eta_{det} < 2.7$ (bottom).

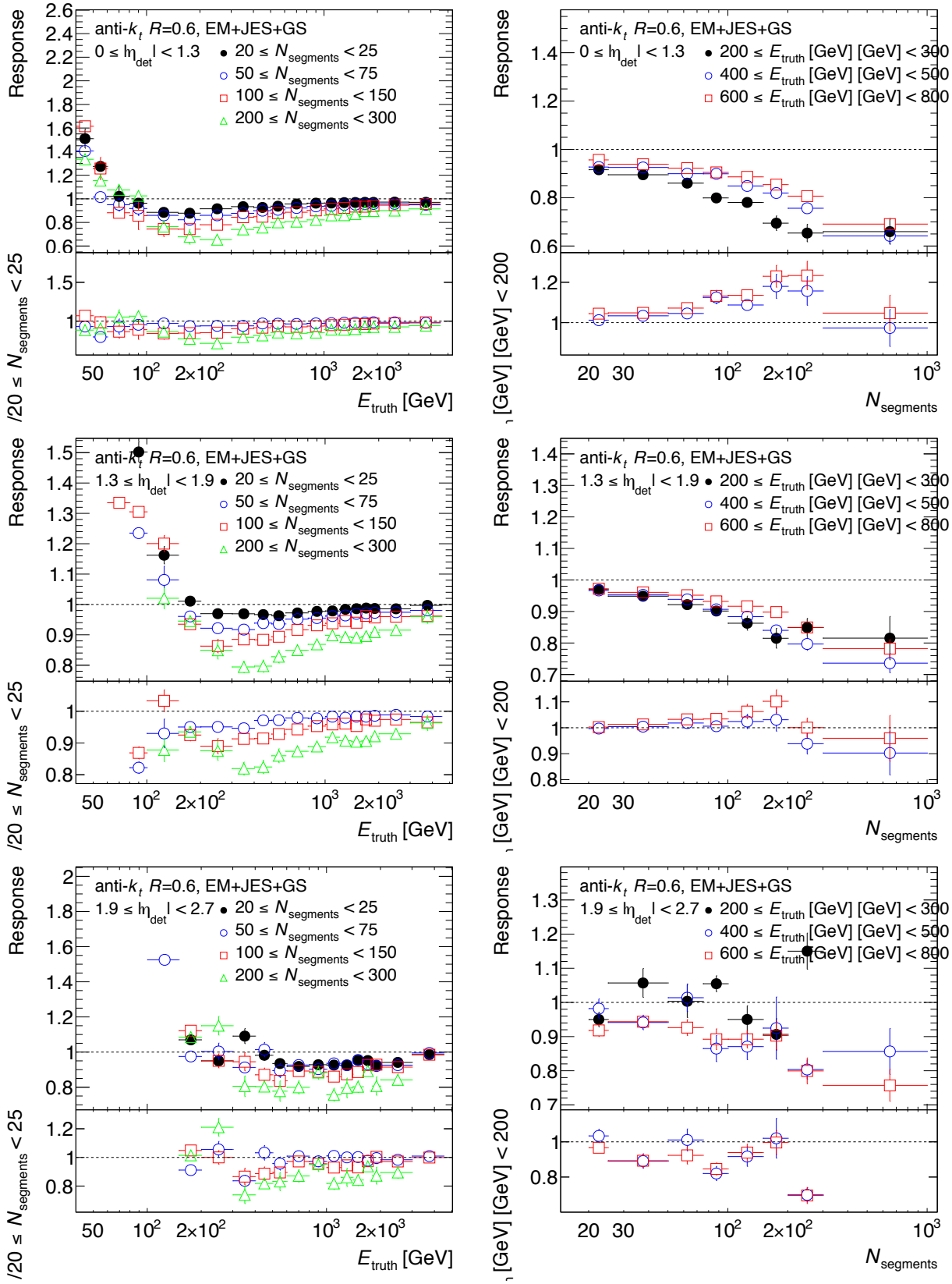


Figure 5.2: Graphs showing mean jet energy response vs E_{truth} in bins of $N_{segments}$ (left) and mean jet energy response vs $N_{segments}$ in bins of E_{truth} (right) for the regions $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom). Bottom plots display ratio to the jet energy response in the $20 \le N_{segments} < 25$ bin.

5.3.3 Final binning

Based on the studies in Sections 5.3.1 and 5.3.2, the jet energy response distributions were binned using the bins in Table 5.2. The binning is fine in the low E_{truth}, low $N_{segments}$ region, gradually getting coarser as E_{truth} and $N_{segments}$ increases, ensuring the binning does not mask structures in the response plane observed at low E_{truth} and low $N_{segments}$ (see Section 5.3.2).

$\eta_{\mathbf{det}}$	$E_{\mathbf{truth}}$ **(GeV)**	$N_{\mathbf{segments}}$
$0 \leq \lvert\eta_{det}\rvert < 1.3$	$10 \leq E_{truth} < 20$	$20 \leq N_{segments} < 25$
	$20 \leq E_{truth} < 30$	
	$30 \leq E_{truth} < 40$	$25 \leq N_{segments} < 50$
	$40 \leq E_{truth} < 50$	
	$50 \leq E_{truth} < 60$	$50 \leq N_{segments} < 75$
	$60 \leq E_{truth} < 80$	
$1.3 \leq \lvert\eta_{det}\rvert < 1.9$	$80 \leq E_{truth} < 100$	$75 \leq N_{segments} < 100$
	$100 \leq E_{truth} < 150$	
	$150 \leq E_{truth} < 200$	$100 \leq N_{segments} < 150$
	$200 \leq E_{truth} < 400$	
	$400 \leq E_{truth} < 600$	$150 \leq N_{segments} < 200$
	$600 \leq E_{truth} < 800$	
$1.9 \leq \lvert\eta_{det}\rvert < 2.7$	$800 \leq E_{truth} < 1000$	$200 \leq N_{segments} < 300$
	$1000 \leq E_{truth} < 1200$	
	$1200 \leq E_{truth} < 1600$	
	$1600 \leq E_{truth} < 2000$	$300 \leq N_{segments} < 1000$
	$2000 \leq E_{truth} < 3000$	
	$3000 \leq E_{truth} < 4000$	

Table 5.2: Table summarising binning used for derivation of correction factors.

Figure 5.3 shows N_{eff} in each of these bins. As with Figure 5.1, there are statistical limitations in general at low and high E_{truth}, at high $N_{segments}$, and in the $1.9 \leq \eta_{det} < 2.7$ region.

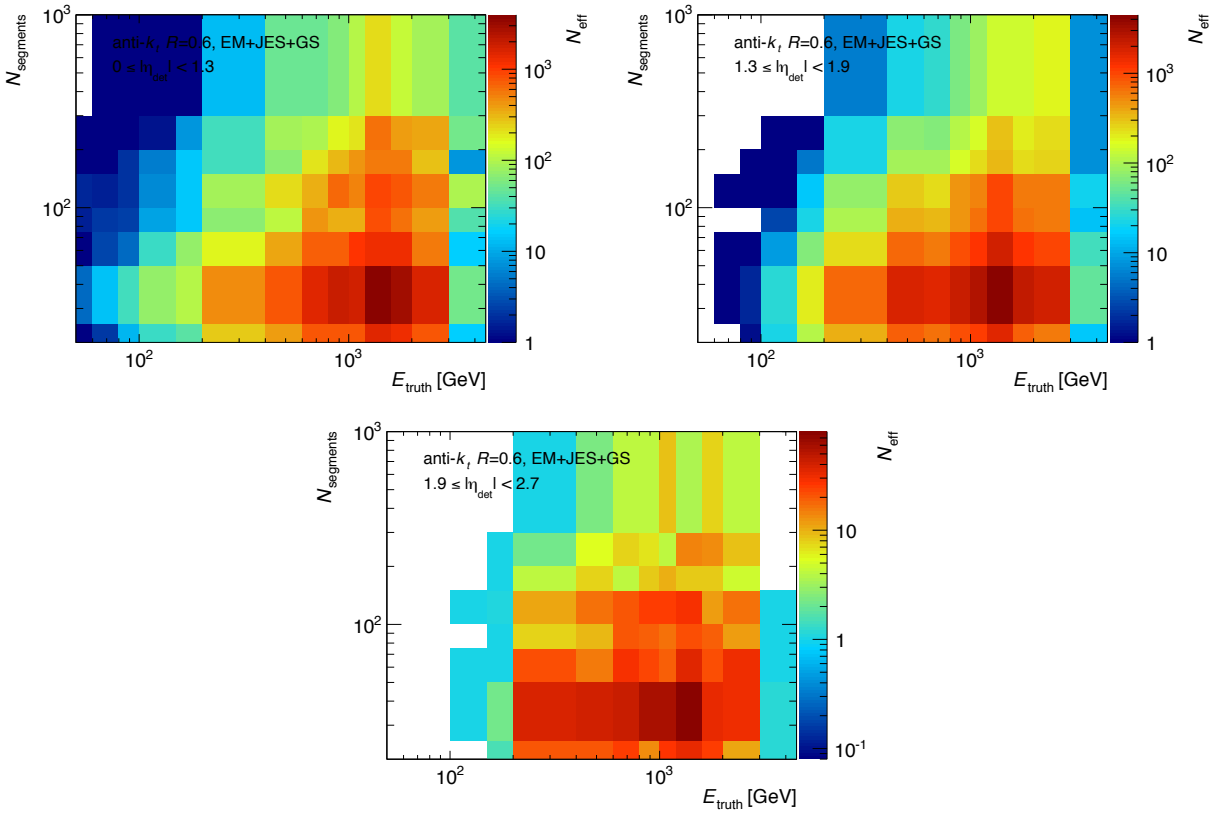


Figure 5.3: Figure showing N_{eff} in each E_{truth}, $N_{segments}$ bin in the regions $0 \le \eta_{det} < 1.3$ (left), $1.3 \le \eta_{det} < 1.9$ (right) and $1.9 \le \eta_{det} < 2.7$ (bottom).

5.4 Fitting procedure

As mentioned in Section 5.1, the third step of deriving the jet punch-through correction function involves fitting a Gaussian function to the jet energy response distribution binned in E_{truth}, $\mathrm{N}_{segments}$ and η_{det}. The mean jet energy response is extracted from this fit and used by the Gaussian kernels (see Section 5.6) to produce the correction function ($\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}$).

It was shown in Figure 3.6 that the jet energy response distribution is Gaussian when binned in η_{det} and E_{truth}. When binning in an additional third dimension ($\mathrm{N}_{segments}$), this fact also holds, as can be seen from the jet energy response distributions for jets with $25 \leq \mathrm{N}_{segments} < 50$ and $600 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 800$ in the detector region $0 \leq \eta_{det} < 1.3$ in Figure 5.4.

The initial distribution is shown on the top-left of Figure 5.4. Prior to fitting, the binning of this distribution is optimised using Scott's normal reference rule, defined in Equation 5.2 [125]. The rule calculates the optimum bin width, h_n, which limits the number of bins used in the histogram. The standard deviation of the normal distribution in question is defined by $\hat{\sigma}$ and n is the sample size (number of entries in the histogram). As we are dealing with weighted entries, $n = \mathrm{N}_{eff}$. The resultant rebinned histogram is shown on the top-right of Figure 5.4.

$$h_n = \frac{3.5\hat{\sigma}}{n^{1/3}} \tag{5.2}$$

Only the $\mu \pm 1.6\sigma$ core of the distribution is fit, where μ and σ are the mean and standard deviation of the distribution respectively. The fitting procedure uses a combination of the log-likelihood and $\tilde{\chi}^2$ minimisation fitting methods as is outlined below:

1. The initial fit is done using the log-likelihood fitting method, and is used to obtain initial values for μ and σ. The fit range is chosen on a bin-by-bin basis according to the shape of the distribution.

2. The fit range is then reset to $\mu \pm 1.6\sigma$, where μ and σ are extracted from the fit in Step

1. The distribution is then fit again, but this time using a $\tilde{\chi}^2$ minimisation method, as the log-likelihood implementation used did not calculate errors for fits to weighted distributions correctly.

3. The fit range is again reset to $\mu \pm 1.6\sigma$, this time taking μ and σ from the fit in Step 2. The distribution is then re-fit, again using the $\tilde{\chi}^2$ minimisation method.

The red line in the bottom plot of Figure 5.4 represents the resulting fit, whereby the solid red line is the fit to the $\mu \pm 1.6\sigma$ core of the distribution, with the dotted line representing the tails that are not fit. The resultant fit can be seen to be good ($\tilde{\chi}^2_{red}$ [1] ~ 1).

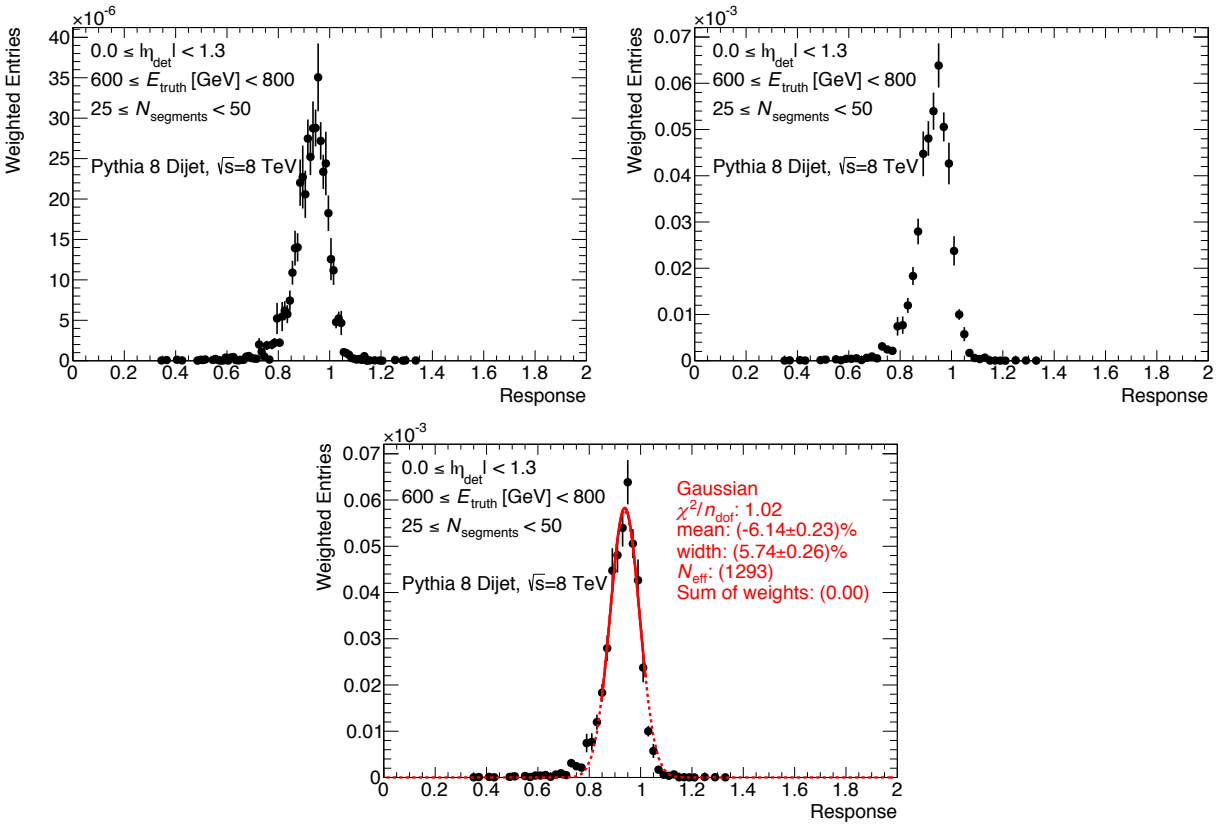


Figure 5.4: Jet energy response distribution for jets with $600 \leq \mathrm{E}_{truth}\ [\,\mathrm{GeV}] < 800$, $25 \leq \mathrm{N}_{segments} < 50$ in the detector region $0 \leq \eta_{det} < 1.3$ with no rebinning (left), rebinned with Scott's normal reference rule (right) and after the fitting procedure (bottom).

[1] $\tilde{\chi}^2_{red} = \frac{\tilde{\chi}^2_{global}}{\mathrm{n}_{dof}}$, $\mathrm{n}_{dof} = \mathrm{N}_{obs} - 2$, $\tilde{\chi}^2_{global} = \sum_{i=1}^{\mathrm{N}_{obs}} \tilde{\chi}^2_{local,i}$, $\tilde{\chi}^2_{global} = \sum_{i=1}^{\mathrm{N}_{fits}} \tilde{\chi}^2_{local,i}$, $\tilde{\chi}^2_{local} = \frac{\left(\mathrm{R}_i^{bin} - \mathrm{R}_i^{fit}\right)^2}{\sigma^2_{bin,i}}$, where i refers to the i-th response bin, R_i^{bin} is the response value at the bin centre, $\sigma_{bin,i}$ is the statistical error in this bin ($\sqrt{\sum_i^{\mathrm{N}_{entries}} w_i^2}$), R_i^{fit} is the response obtained from the fit, N_{obs} is the number of bins overlapping with the fit, and n_{dof} represents the number of degrees of freedom.

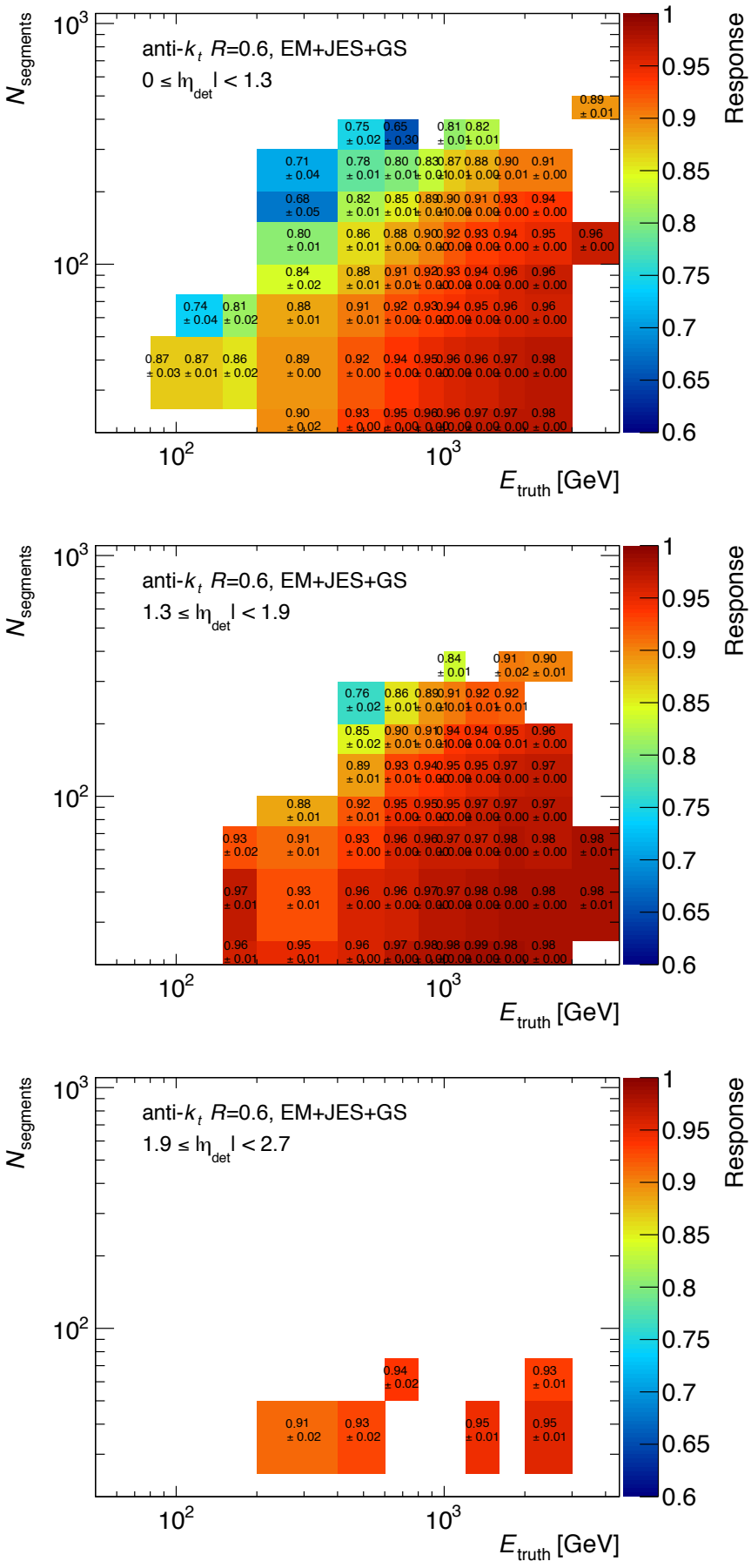


Figure 5.5: Mean jet energy response values extracted from fits along with their associated errors as a function of E_{truth} and $N_{segments}$, for jets in the detector region $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom).

Figure 5.5 shows the extracted mean responses and their associated errors for all three η_{det} bins. Only fits with a $\tilde{\chi}^2_{red} < 50$ were used. The errors (σ_{fit}) are calculated as the standard error on the mean [2]. Empty bins represent regions where there were insufficient statistics for the fits to converge. This was especially problematic in the forward region ($1.9 \leq \eta_{det} < 2.7$), where the punch-through probability is lower due to the increased detector material. As a result, there is a gap in the extracted mean response plane at $800 \leq \mathrm{E}_{truth}$ [GeV] < 1000.

It can be seen that the extracted mean jet energy response increases as E_{truth} increases and decreases as $\mathrm{N}_{segments}$ increases, as expected based upon prior studies in Section 4.2.1. Thus the jet punch-through correction will be largest for jets with low E_{truth} and high $\mathrm{N}_{segments}$- as large as 25% in some regions. In addition, the response reaches smaller values in the central η_{det} region ($0 \leq \eta_{det} < 1.3$) than in the other detector regions, therefore the correction will be largest in this region.

5.5 Converting to a relative correction

As mentioned in Section 5.1, the jet punch-through correction function is a correction relative to the jet energy scale correction. This ensures the correction leaves the mean jet energy scale unchanged while still improving the jet energy resolution. To do this, the mean jet energy response values obtained in Figure 5.5 are normalised by the mean jet energy response when integrating over all $\mathrm{N}_{segments}$ bins via Equation 5.3. Thus the relative jet energy response ($\langle \mathcal{R}_{rel} \rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}}$) is the mean jet energy response in each $\mathrm{N}_{segments}$,E_{truth},η_{det} bin ($\langle \mathcal{R} \rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}}$) relative to the mean jet energy response integrated over $\mathrm{N}_{segments}$ ($\langle \mathcal{R} \rangle_{\mathrm{JES}}|_{\eta_{det},\mathrm{E}_{truth}}$). $\langle \mathcal{R} \rangle_{JES}|_{\eta_{det},\mathrm{E}_{truth}}$ is obtained by fitting the jet energy response distributions in bins of E_{truth} and η_{det} using the same fitting procedure as described in Section 5.4.

$$\langle \mathcal{R}_{rel} \rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}} = \frac{\langle \mathcal{R} \rangle|_{\eta_{det},\mathrm{E}_{truth},\mathrm{N}_{segments}}}{\langle \mathcal{R} \rangle_{JES}|_{\eta_{det},\mathrm{E}_{truth}}} \tag{5.3}$$

[2] $\sigma_{fit} = \frac{\sigma}{\sqrt{n}}$

$\langle\mathcal{R}\rangle_{JES}|_{\eta_{det},\mathrm{E}_{truth}}$ as a function of E_{truth} in each η_{det} bin is shown in Figure 5.6. From this figure it can be seen that the mean JES ($\langle\mathcal{R}\rangle_{\mathrm{JES}}|_{\eta_{det},\mathrm{E}_{truth}}$) is $\sim$ 1 for all jets with $\mathrm{E}_{truth} > 400$ GeV. Below this point, there is an observed non-closure in the JES (deviation from a response of 1) of up to 90%. As mentioned previously, this non-closure is a combination of the JES non-closure seen in Section 3.5.3 and a bias caused by the $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV cut (see Section 4.1.2 for an explanation of this cut). The non-closure occurs at $\mathrm{E}_{truth} < 400$ GeV for jets in the region $1.9 \leq \eta_{det} < 2.7$, $\mathrm{E}_{truth} < 200$ GeV for jets in the region $1.3 \leq \eta_{det} < 1.9$, and $\mathrm{E}_{truth} < 100$ GeV for jets in the region $0 \leq \eta_{det} < 1.3$.

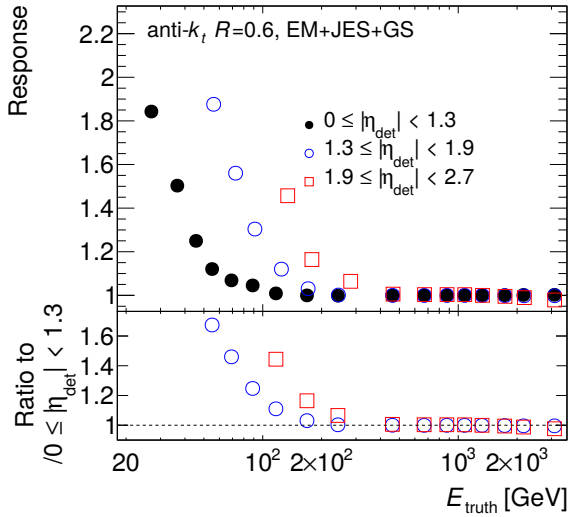


Figure 5.6: $\langle\mathcal{R}\rangle_{\mathrm{JES}}|_{\eta_{det},\mathrm{E}_{truth}}$ as a function of E_{truth} for jets in the detector region $0 \leq \eta_{det} < 1.3$, $1.3 \leq \eta_{det} < 1.9$ and $1.9 \leq \eta_{det} < 2.7$. The bottom plot displays a ratio to the response in the detector region $0 \leq \eta_{det} < 1.3$.

Normalising the response values in Figure 5.5 by those in Figure 5.6 to obtain the jet energy response relative to the mean JES ensures the jet punch-through correction function does not bias the JES by removing this inherent non-closure. However, fits were not convergent for $\mathrm{E}_{truth} < 80$ GeV, $\mathrm{E}_{truth} < 150$ GeV and $\mathrm{E}_{truth} < 200$ in the regions $0 \leq \eta_{det} < 1.3$, $1.3 \leq \eta_{det} < 1.9$ and $1.9 \leq \eta_{det} < 2.7$ respectively (seen in Figure 5.5). As a result, the impact of the non-closure will be at most $5 - 10\%$ in the lowest E_{truth} bins in each η_{det} bin, leaving the majority of bins unaffected.

The relative jet energy response along with its error is displayed in Figure 5.7 as a function of $\mathrm{N}_{segments}$ and E_{truth} in each of the three η_{det} regions.

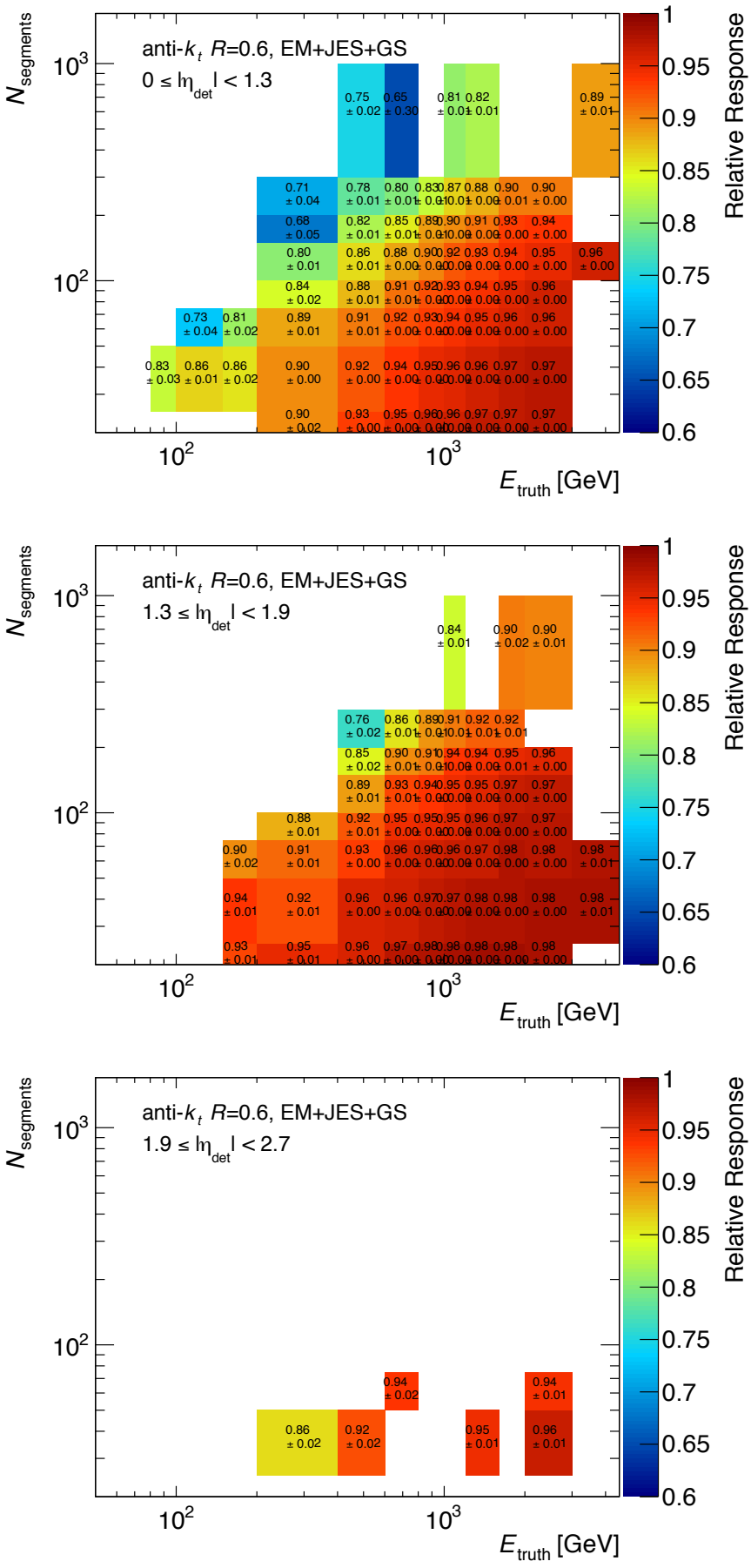


Figure 5.7: Jet energy response relative to mean JES, calculated via Equation 5.3, along with their associated errors, shown as a function of E_{truth} and $N_{segments}$, for jets in the region $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom).

As expected, only the low E_{truth} region is affected by the normalisation. The observed dependence in Figure 5.5 of the jet energy response on E_{truth}, $N_{segments}$ and η_{det} remains unchanged, with the response being lowest in the central η_{det} region ($0 \leq \eta_{det} < 1.3$), for the low E_{truth}, high $N_{segments}$ region of phase space.

5.6 Obtaining a smooth correction function

As the parametric form of the response as a function of (E_{truth}, $N_{segments}$) is not known, non-parametric methods must be used to obtain a smooth correction function. One method that can be used to achieve this is bilinear interpolation [126] of the relative response values shown in Figure 5.7 as a function of (E_{truth}, $N_{segments}$). However, bilinear interpolation is sensitive to local fluctuations, which can result in statistical fluctuations in the relative response as a function of (E_{truth}, $N_{segments}$) becoming exaggerated, distorting the shape of the correction function. An alternative approach is to use a non-parametric regression technique that allows the global shape of the response curve to be taken into account when constructing the correction function. This is achieved by using a smoothing technique known as *Gaussian kernel regression*.

5.6.1 Overview of Gaussian kernel regression

Given a set of n observations of the random variable Y as a function of the random variable X, regression aims to determine the relationship between Y and X:

$$Y = m(X) \tag{5.4}$$

where m is a mathematical function. *Kernel regression* can be used to determine this relationship when the mathematical form of m is not known. This involves determining the conditional expectation of Y given that $X = x$ using the Nadaraya-Watson estimator, $\hat{m}_\alpha(x)$. The estimator is constructed as a locally weighted average of all n observations of Y as a

function of X, using a kernel[3] with bandwidth α, $K_\alpha(x - X_i)$, as a weighting function. The bandwidth is a parameter of the smoothing whose significance will be explained below. The estimator is defined as [128]:

$$\hat{m}_\alpha(x) = \frac{\sum_{i=1}^{n} K_\alpha(x - X_i) Y_i}{\sum_{i=1}^{n} K_\alpha(x - X_i)} \tag{5.5}$$

where Y_i and X_i are the Y and X values respectively of the i-th observation (data point). The chosen kernel is a Gaussian kernel [129] (hence *Gaussian* kernel regression):

$$K_\alpha(x - X_i) = e^{-\frac{(x - X_i)^2}{2\alpha^2}} \tag{5.6}$$

The Gaussian kernel acts as a similarity function [130], characterising the distance between x and X_i. The kernel has value 1 if x and X_i are identical and falls to zero as the distance between x and X_i increases. As a result, the Gaussian kernel will give larger weights to Y_i observations at X_i values closer to the x position $\hat{m}_\alpha(x)$ is being estimated at, and smaller weights to those observations further away. In this way, the estimated value of $\hat{m}_\alpha(x)$ is most dependant on nearby observations, with a small dependence on the global shape of Y as a function of X.

The *bandwidth* or *smoothing parameter*, α, determines the width of the Gaussian kernel, and thus how quickly the kernel weight falls to zero as the distance between x and X_i increases. As a result, the bandwidth controls how strongly $\hat{m}_\alpha(x)$ is weighted towards nearby observations over the global shape of the Y vs X curve, and thus how smooth the estimated curve is. If the bandwidth is too large, under-fitting occurs, resulting in an estimated curve that does not capture the important features of the relationship between Y and X. However if the bandwidth is too small, over-fitting can occur, causing statistical fluctuations to become over-exaggerated. Therefore the bandwidth must be optimised to avoid over and under fitting.

[3] A kernel is defined to be a non-negative weighting function that integrates to one and has mean zero [127].

Equation 5.5 can only be used to estimate the random variable as a function of one variable (X), however the jet energy response needs to be constructed as a function of two variables ($N_{segments}$ and E_{truth}). In order to estimate Y as a function of a d-dimensional vector of variables X, where $X = (X_1, ..., X_d)^\intercal$, $K_\alpha(x - X_i)$ in Equation 5.5 must be replaced by a multiplicative kernel, $\mathcal{K}_\alpha(x - X_i)$ [131]:

$$\mathcal{K}_\alpha(x - X_i) = \prod_{j=1}^{d} e^{-\frac{(x-X_{ij})^2}{2\alpha_j^2}} \tag{5.7}$$

where X_{ij} and α_j are the X_i and bandwidth values for the j-th variable. Thus different bandwidths must be provided for each variable Y is being parametrised as a function of, resulting in d bandwidth parameters. The resulting d-dimensional estimator is given by:

$$\hat{m}_\alpha(x) = \frac{\sum_{i=1}^{n} \mathcal{K}_\alpha(x - X_i) Y_i}{\sum_{i=1}^{n} \mathcal{K}_\alpha(x - X_i)} \tag{5.8}$$

For binned data, Y_i and X_i represent the mean Y and mean X values respectively in the i-th bin, with n being equal to the total number of bins, N_{bins}. In this case, the error on Y_i, ΔY_i, can be taken into account by applying a normalisation of $\frac{1}{\Delta Y_i^2}$ to the kernels in Equations 5.6 and 5.7 to give the normalised kernels in Equations 5.9 and 5.10 respectively. This normalisation ensures Y_i values with large errors have smaller weights, and thus a smaller impact on $\hat{m}_\alpha(x)$.

$$K_\alpha(x - X_i) = \frac{1}{\Delta Y_i^2} e^{-\frac{(x-X_i)^2}{2\alpha^2}} \tag{5.9}$$

$$\mathcal{K}_\alpha(x - X_i) = \frac{1}{\Delta Y_i^2} \prod_{j=1}^{d} e^{-\frac{(x-X_{ij})^2}{2\alpha_j^2}} \tag{5.10}$$

5.6.2 Methodology for deriving correction function

The correction as a function of (E^{jet}, $\mathrm{N}_{segments}$) in bins of η_{det}, $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(E^{jet},\ \mathrm{N}_{segments})|_{\eta_{det}}$, where E^{jet} is the reconstructed jet energy at the EM+JES+GS or LCW+JES+GS scale, is obtained from the relative response values in Figure 5.7 via a three step procedure:

1. Perform smoothing of the relative response in the $\mathrm{N}_{segments}$ dimension only, in bins of η_{det} and E_{truth}, using $\hat{m}_\alpha(x)$ from Equation 5.5 with the normalised kernel from Equation 5.9. Here, X is $\log \mathrm{N}_{segments}$, and Y represents the relative response values from Figure 5.7, where X_i, Y_i and ΔY_i are the mean $\log \mathrm{N}_{segments}$ ($\log\langle \mathrm{N}_{segments}\rangle_i$), relative response and error on the relative response (σ_{fit}) respectively in the i-th $\mathrm{N}_{segments}$ bin. x is therefore the $\log \mathrm{N}_{segments}$ value at which the relative response is to be estimated. The smoothing parameter, α, was chosen to avoid biasing the response towards statistical fluctuations. This step produces the relative response as a function of $\mathrm{N}_{segments}$ in bins of η_{det} and E_{truth}, $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(\log \mathrm{N}_{segments})|_{\eta_{det},\mathrm{E}_{truth}}$.

2. Convert from bins of E_{truth} to E^{jet} via Equation 5.11 to give $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(\langle \log \mathrm{N}_{segments}\rangle_i)|_{\eta_{det},E^{jet}}$. Here, $\langle E^{jet}\rangle_i$ and $\langle \mathrm{E}_{truth}\rangle_i$ are the mean E^{jet} and mean E_{truth} values respective, and $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(\log \langle \mathrm{N}_{segments}\rangle_i)$ is the relative response obtained by evaluating the function from Step 1 at the mean $\mathrm{N}_{segments}$, all in the i-th (E_{truth}, $\mathrm{N}_{segments}$) bin.

$$\left\langle E^{jet}\right\rangle_i = \mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(\log \langle \mathrm{N}_{segments}\rangle_i)\, \langle \mathrm{E}_{truth}\rangle_i \tag{5.11}$$

3. Perform a two-dimensional smoothing of the relative response as a function of (E^{jet}, $\mathrm{N}_{segments}$), in bins of η_{det}, to give the final correction function, $\mathcal{F}_{\mathrm{rel}}^{\mathrm{PT}}(E^{jet},\ \mathrm{N}_{segments})|_{\eta_{det}}$. This step is performed using $\hat{m}_\alpha(x)$ from Equation 5.8 with the normalised kernel from Equation 5.10. Here, X, and thus X_i, are two dimensional vectors:

$$X = \begin{bmatrix} \log \mathrm{N}_{segments} \\ \log E^{jet} \end{bmatrix},\ X_i = \begin{bmatrix} \log\langle \mathrm{N}_{segments}\rangle_i \\ \log\langle E^{jet}\rangle_i \end{bmatrix} \tag{5.12}$$

where $X_{i1} = \log\langle \mathrm{N}_{segments} \rangle_i$ and $X_{i2} = \log\langle E^{jet} \rangle_i$ are the mean $\mathrm{N}_{segments}$ and mean E^{jet} values respectively in the i-th (E^{jet}, $\mathrm{N}_{segments}$)-bin. As this step involves smoothing as a function of two variables, the kernel takes two smoothing parameters as input - one for $\mathrm{N}_{segments}$ ($\alpha_{\mathrm{N}_{segments}}$) and one for E^{jet} ($\alpha_{E^{jet}}$). As with the smoothing parameter in Step 1, both $\alpha_{\mathrm{N}_{segments}}$ and $\alpha_{E^{jet}}$ were chosen to ensure the response was not biased towards statistical fluctuations. Y is the relative response function from Step 2 ($\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel}}(\log \mathrm{N}_{segments})|_{\eta_{det},E^{jet}}$), where Y_i is this function evaluated at $\langle \log \mathrm{N}_{segments} \rangle_i$ in the E^{jet} bin containing $\langle \log E^{jet} \rangle_i$ ($\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel}}(\langle \log \mathrm{N}_{segments} \rangle_i)|_{\eta_{det},E^{jet}}$). ΔY_i is then calculated as:

$$\Delta Y_i = \sqrt{\sum_{j}^{\mathrm{N}_i^{fits}} \left(\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel,nom}}(\log \mathrm{N}_{segments}) - \mathcal{F}^{\mathrm{PT}}_{\mathrm{rel,j}}(\log \mathrm{N}_{segments}) \right)^2} \tag{5.13}$$

where $\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel,nom}}(\log \mathrm{N}_{segments})$ is the nominal relative response function obtained after Step 1 and $\mathcal{F}^{\mathrm{PT}}_{\mathrm{rel,j}}(\log \mathrm{N}_{segments})$ is a shifted relative response function obtained by performing Step 1 using a modified version of the response vs $\mathrm{N}_{segments}$ curve, whereby the fit in the j^{th} $\mathrm{N}_{segments}$-bin is shifted by $+1\sigma_{fit}$.

In regions at the edges of the (E_{truth}, $\mathrm{N}_{segments}$) phase space, where no correction could be derived due to a lack of successful fits, the correction function was frozen to be the correction derived in the nearest bin with information. This freezing procedure involved:

1. Iterating over all $\mathrm{N}_{segments}$ bins from low $\mathrm{N}_{segments}$ to high $\mathrm{N}_{segments}$. When reaching areas of phase-space where the estimator failed to estimate a relative response value due to lack of statistics, use the relative response value in the prior (lower) $\mathrm{N}_{segments}$ bin. This iteration is performed over all E^{jet} and η_{det} bins separately.

2. Repeat Step 1, but this time iterate over all E^{jet} bins from low to high E^{jet}, instead of from low to high $\mathrm{N}_{segments}$. This iteration is performed over all $\mathrm{N}_{segments}$ and η_{det} bins separately.

5.7 Results

The correction function after the application of the one-dimensional Gaussian kernel can be found in Section 5.7.1, while the final correction function derived using the two-dimensional Gaussian kernel can be found in Section 5.7.2. To assess how well the derived correction function resembles the original relative response values in Figure 5.7, the $\tilde{\chi}^2_{red}$ [4] is used.

5.7.1 After one-dimensional kernel

Figure 5.8 compares the response curves before/after the one-dimensional kernel for jets in the region $0 \leq |\eta_{det}| < 1.3$ with $200 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 400$ and $1200 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 1400$. The yellow bands represent the error on the smoothed curve as estimated via Equation 5.13. It can be seen that the linear smoothed curve is compatible with the original relative response to within errors.

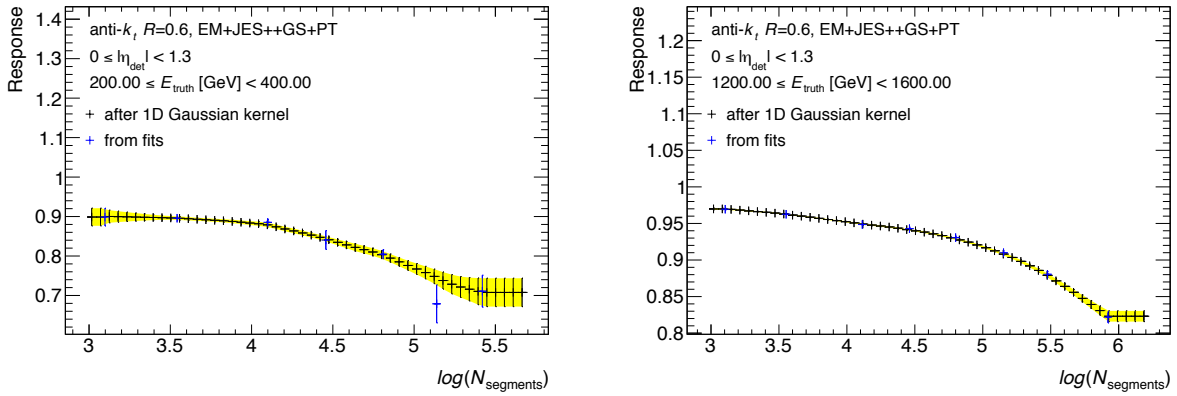


Figure 5.8: Response as a function of $\log \mathrm{N}_{segments}$ curves before/after one-dimensional kernel application for jets in the region $0 \leq \eta_{det} < 1.3$ with $200 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 400$ (left) and $1200 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 1600$ (right).

Figure 5.9 shows the correction function after applying the one-dimensional kernel as a function of $\log \mathrm{N}_{segments}$ and E^{jet} in all three η_{det} regions. Empty bins represent regions where the kernel produced no information due to a lack of successful fits. This is espe-

[4] This is defined as in[1]. However in this case, i is the index of a (E^{jet}, $\log \mathrm{N}_{segments}$)-bin, R^{bin} is the relative response as extracted and calculated from the fits, R_i^{fit} is the response calculated using the derived correction function, $\sigma_{bin,i}$ is the error on the relative response from Figure 5.7 and N_{obvs} represents the number of successful fits.

cially noticeable in the $1.9 \leq \eta_{det} < 2.7$ region where no correction is available in the $450 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 500$ region. Comparing Figure 5.9 with Figure 5.7, it can be seen that the response as a function of $\log \mathrm{N}_{segments}$ is smoother after the application of the linear kernel.

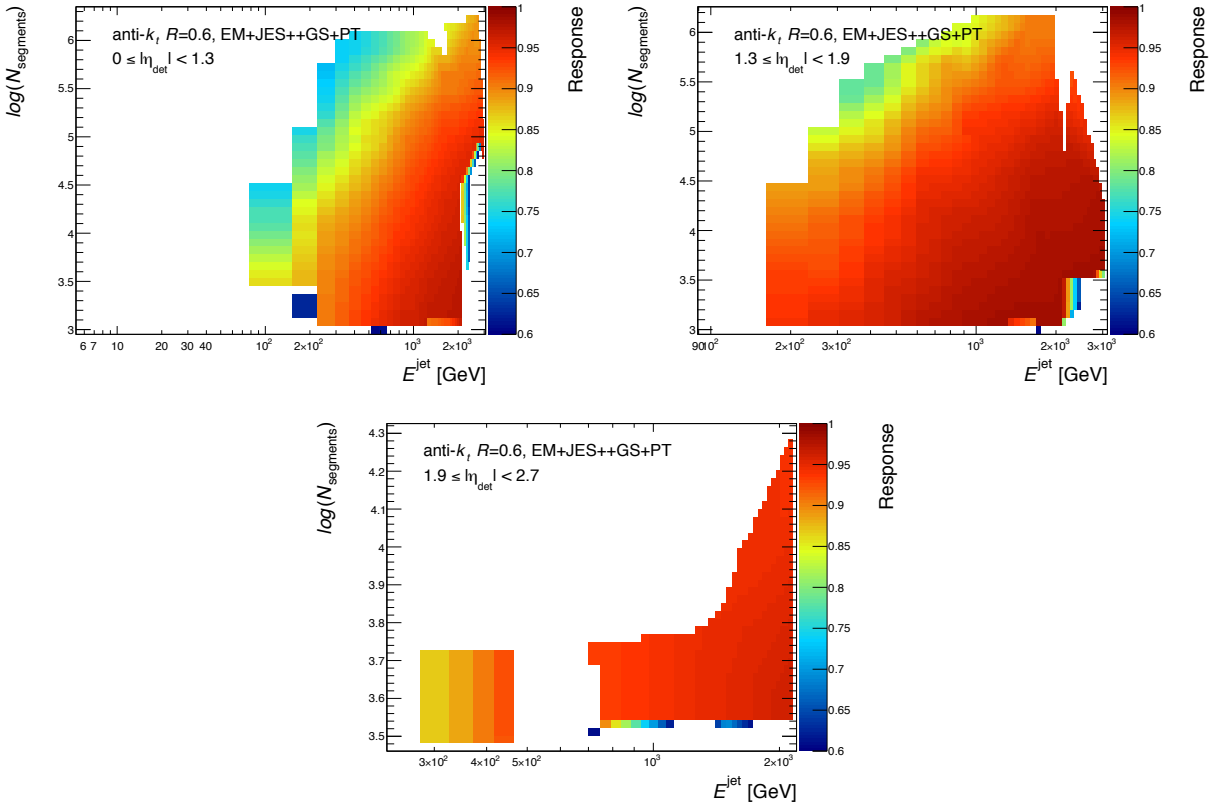


Figure 5.9: Jet response as a function of E^{jet} and $\log \mathrm{N}_{segments}$ for jets with $0 \leq \eta_{det} < 1.3$ (left), $1.3 \leq \eta_{det} < 1.9$ (right) and $1.9 \leq \eta_{det} < 2.7$ (bottom).

$\tilde{\chi}^2_{local}$ [5] is shown as a function of E^{jet} and $\log \mathrm{N}_{segments}$ in each η_{det} bin in Figure 5.10. $\tilde{\chi}^2_{red}$ is also summarised on each plot. The $\tilde{\chi}^2_{red}$ can be seen to be large (> 1) in the $0 \leq \eta_{det} < 1.3$ and $1.3 \leq \eta_{det} < 1.9$ regions. In the $0 \leq \eta_{det} < 1.3$ it appears $\tilde{\chi}^2_{local}$ is very large in the high E^{jet} region, whereas $\tilde{\chi}^2_{local}$ is large in the low E^{jet} region for $1.3 \leq \eta_{det} < 1.9$. In both of these regions, the large $\tilde{\chi}^2_{local}$ is due to the lack of successful fits in a particular E_{truth} region. This is amplified by the fact the correction has not yet been frozen and extended into these regions.

The low $\tilde{\chi}^2_{red}$ value (< 1) in the $1.9 \leq \eta_{det} < 2.7$ region suggest the errors in this

[5] This is defined as in[1] with the substitutions from[3].

region have been over-estimated. This comes from the lack of successful fits in the region as a whole, resulting in the interpolated response after applying the one-dimensional kernel being incredibly close to the original values obtained from the fits. Thus even though the errors were observed to be small for the relative response in Figure 5.7, the $\tilde{\chi}^2_{red}$ is still small.

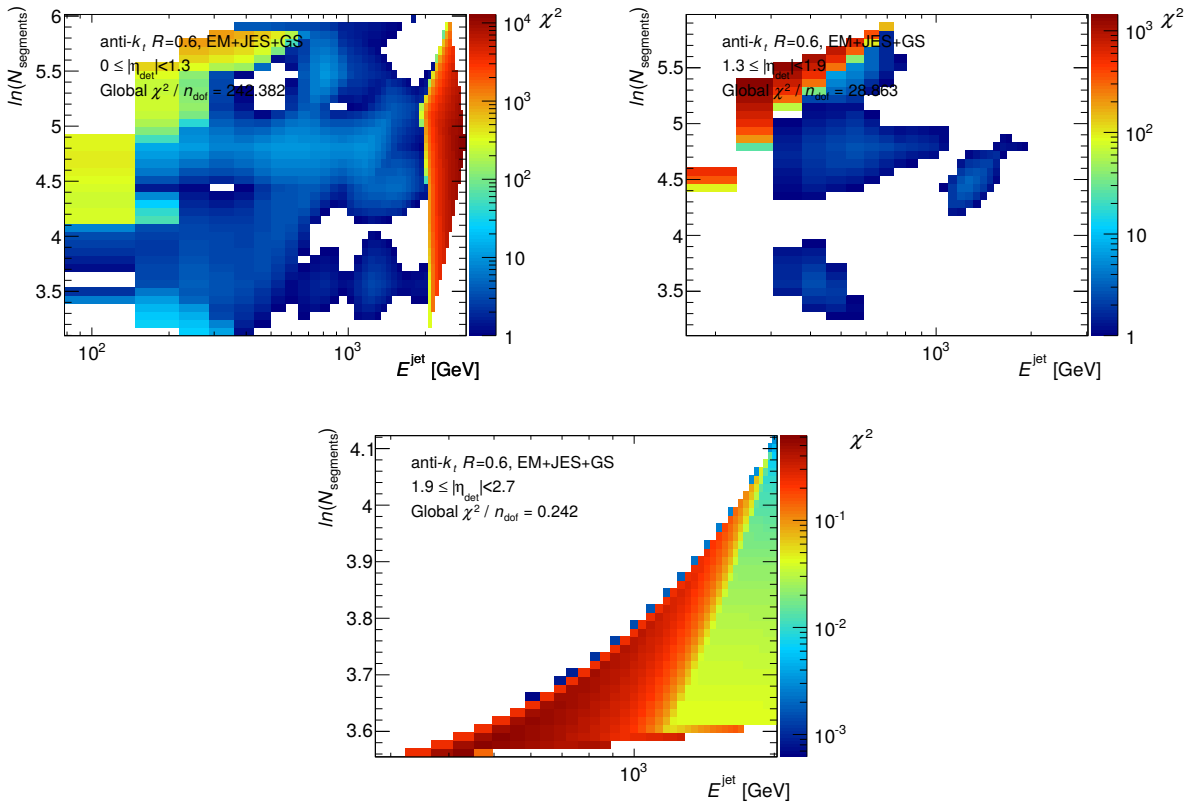


Figure 5.10: $\tilde{\chi}^2_{local}$ before/after one-dimensional kernel as a function of $\log \mathrm{N}_{segments}$ and E^{jet} for jets with $0 \le \eta_{det} < 1.3$ (left), $1.3 \le \eta_{det} < 1.9$ (right) and $1.9 \le \eta_{det} < 2.7$ (bottom).

5.7.2 Final jet punch-through correction function

The final correction as a function of E^{jet} and $\log \mathrm{N}_{segments}$ after the two dimensional kernel is shown in Figure 5.11 for each of the three η_{det} regions.

The correction function decreases with $\mathrm{N}_{segments}$ and increases with E^{jet}, which is as expected. The correction function is relatively smooth as a function of $\log \mathrm{N}_{segments}$ and E^{jet} in the $0 \le \eta_{det} < 1.3$ and $1.3 \le \eta_{det} < 1.9$ regions. Comparing Figure 5.11 with

Figure 5.9, it can be seen that the empty bins present in Figure 5.9 have now been filled out through a combination of the two-dimensional kernel estimating response values for empty regions of phase space, and the correction being frozen at the edges of the (E^{jet}, $N_{segments}$) phase space. In the $1.9 \leq \eta_{det} < 2.7$ region, the correction function displays a distinct boundary between the high and low E^{jet} regions at $E^{jet} \sim 350$ GeV. This is a result of the empty region created by the lack of fits in the $800 \leq E_{truth}$ [GeV] < 1000 bin observed in Figure 5.5.

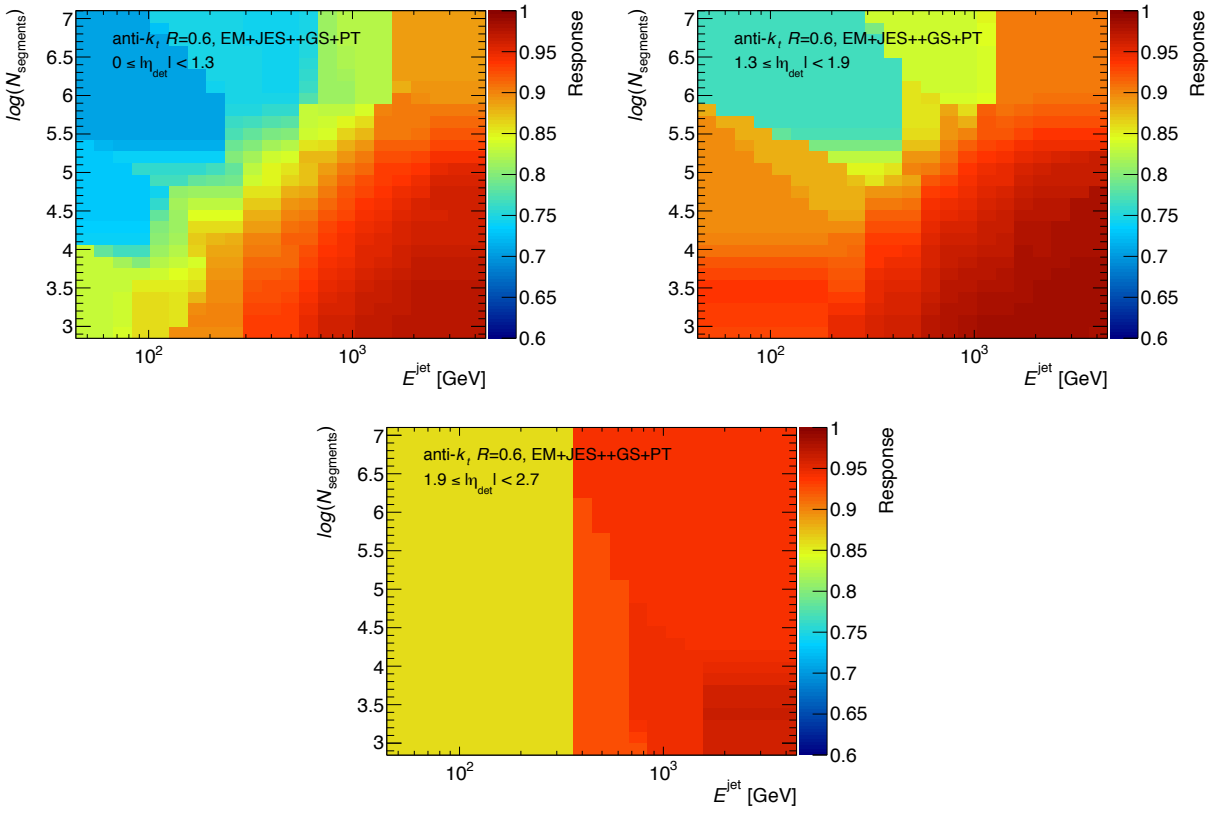


Figure 5.11: Final correction function vs E^{jet} vs $N_{segments}$ for jets with $0 \leq \eta_{det} < 1.3$ (left), $1.3 \leq \eta_{det} < 1.9$ (right) and $1.9 \leq \eta_{det} < 2.7$ (bottom).

Figure 5.12 shows the local $\tilde{\chi}^2_{red}$ values as a function of E^{jet} and $N_{segments}$ in each η_{det} bin. The global $\tilde{\chi}^2_{red}$ is also summarised on each plot.The $\tilde{\chi}^2_{red}$ can be seen to be 1.342 in the $0 \leq \eta_{det} < 1.3$, 0.537 in the $1.3 \leq \eta_{det} < 1.9$ region and 0.017 in the $1.9 \leq \eta_{det} < 2.7$ region. Comparing these values with those after the application of the one-dimensional in Figure 5.10, it can be seen the $\tilde{\chi}^2_{red}$ is reduced in all regions.

For the $0 \leq \eta_{det} < 1.3$ and $1.3 \leq \eta_{det} < 1.9$ regions, this suggests the two-dimensional

kernel provides a much more stable fit. For the $1.9 \leq \eta_{det} < 2.7$ region, the very small $\tilde{\chi}^2_{red}$ suggests over-estimated errors, as with the $\tilde{\chi}^2_{red}$ observed for the one-dimensional kernel (Figure 5.10). However, the errors on the relative response in this region were seen to be small in Figure 5.7. Thus the conclusion is the same as with the one-dimensional kernel in this region - the small $\tilde{\chi}^2_{red}$ is due to the lack of convergent fits in this region, resulting in the derived correction function, on average, being almost identical to the input relative response values.

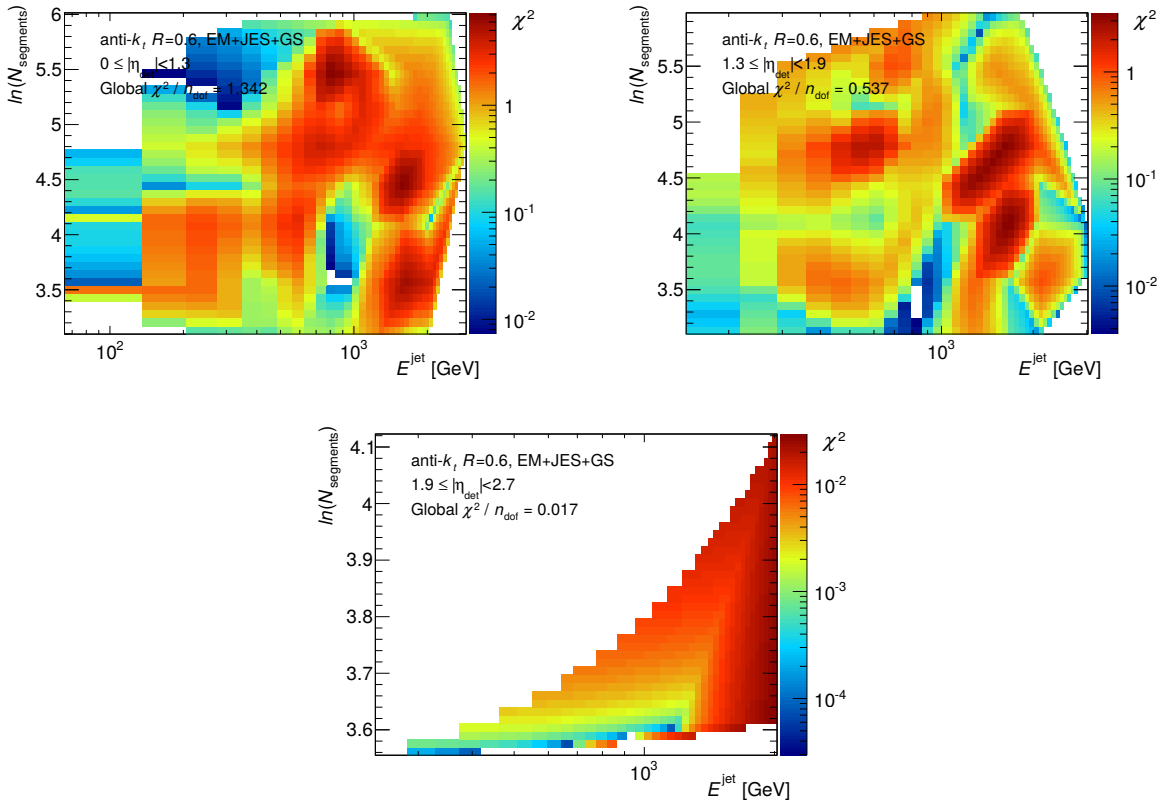


Figure 5.12: $\tilde{\chi}^2$ comparing mean jet energy response from fits with response from final correction function after two-dimensional kernel as a function of E^{jet} and $N_{segments}$ for jets with $0 \leq \eta_{det} < 1.3$ (left), $1.3 \leq \eta_{det} < 1.9$ (right) and $1.9 \leq \eta_{det} < 2.7$ (bottom).

Chapter 6

Data Validation and Systematic Uncertainty

Chapter 5 described the derivation of the jet punch-through correction function. The correction was derived using MC, and is designed to be applied to both MC and data. The validity of the correction function in data is dependant on the correlation between the jet response and $N_{segments}$ being well modelled. Any differences due to modelling must be taken into account as an uncertainty on the correction function.

This chapter details the method used to validate the jet punch-through correction function in data, from which a systematic uncertainty on the jet punch-through correction is estimated. This involves assessing how well MC can reproduce: (1) Shape of the input $N_{segments}$ distributions and (2) Dependance of the jet response on $N_{segments}$.

In data, of course, only reconstructed information is available. Therefore, a jet response cannot be constructed as in Equation 3.4. As a result, in-situ balance techniques must be used to assess the response of the jet energy scale in data, and thus the impact of jet punch-through on the jet energy scale. Section 6.1 details the in-situ dijet balance technique used to do this, while Section 6.2 details the selection criteria used to select well balanced dijet events. The trigger strategy used in data is explained in Section 6.3.

A comparison of the shape of the input $N_{segments}$ distributions between data and MC can be found in Section 6.4. The impact of jet punch-through on the jet energy scale in data and MC is then compared in Section 6.5. Inconsistencies in modelling are used to derive an uncertainty on the jet punch-through correction function as described in Section 6.6. The final uncertainties on the jet punch-through correction function can be found in Section 6.6.2.

6.1 Dijet balance method

As mentioned previously, in data the jet response cannot be calculated as in Equation 3.4 due to the lack of truth information. Instead, a tag-and-probe technique exploiting the p_{T} balance in dijet events was used to form a reconstructed jet response. This reconstructed jet response captures the imbalance in a dijet system with one fully contained jet and one punch-through jet, and is thus sensitive to jet punch-through effects.

Dijet events were selected by requiring the two highest p_{T} jets to be back-to-back using the selection criteria in Section 6.2. The tag (reference) jet was defined to be a well calibrated reference jet fully contained within the calorimeter with $N_{segments} = 0$. The probe jet was defined to be a punch-through jet with $N_{segments} \geq 20$. A schematic representation of the topology of such an event is displayed in Figure 6.1.

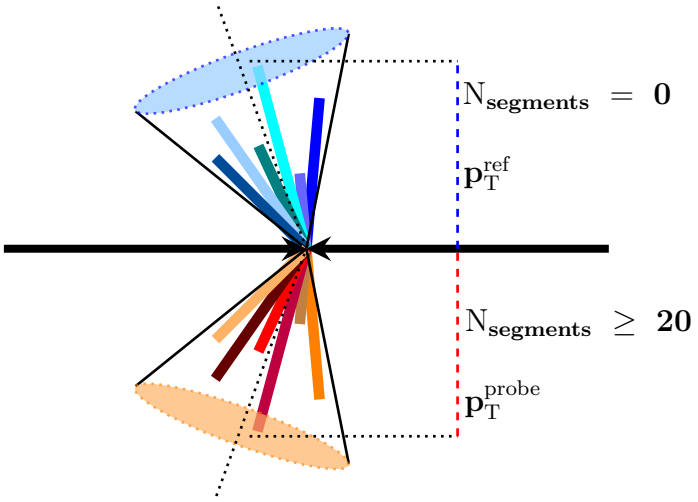


Figure 6.1: Schematic representation of the reference and probe jets in the dijet topology, balanced in transverse momentum.

The mean reconstructed jet response as a function of $\mathrm{N}_{segments}^{probe}$ ($\mathcal{R}(\mathrm{N}_{segments}^{probe})$), defined in Equation 6.1, was used to assess the balance. Here, $p_{\mathrm{T}}^{\mathrm{probe}}(\mathrm{N}_{segments} \geq 20)$ is the p_{T} of the probe jet, while $p_{\mathrm{T}}^{\mathrm{ref}}(\mathrm{N}_{segments} = 0)$ is the p_{T} of the reference jet. Thus comparing the reconstructed response between data and MC provides an avenue to verify the effects of jet punch-through on the jet energy scale were well modelled. Section 6.5 goes into detail on how the mean reconstructed response was measured.

$$\left\langle \mathcal{R}(\mathrm{N}_{segments}^{probe}) \right\rangle = \left\langle \frac{p_{\mathrm{T}}^{\mathrm{probe}}(\mathrm{N}_{segments} \geq 20)}{p_{\mathrm{T}}^{\mathrm{ref}}(\mathrm{N}_{segments} = 0)} \right\rangle \tag{6.1}$$

The uncertainty on the jet punch-through correction aims to account for any miss-modelling in the way jet punch-through affects the jet energy scale. To do this, the uncertainty on the jet punch-through correction function was taken to be the difference between the data and MC mean reconstructed response as a function of $\mathrm{N}_{segments}$, E^{jet} and η_{det}. Jets used in the balance were calibrated at the EM+JES+GS+PT and LCW+JES+GS+PT scale. This ensured any additional data/MC differences that could have been introduced by the application of the jet punch-through correction were also incorporated into the uncertainty, which could then also be interpreted as the uncertainty on the final scale after the application of the correction. Section 6.6 goes into detail on how the uncertainty is derived.

It is important to note the balance can also be assessed using the asymmetry variable ($\mathcal{A}$). Studies comparing the asymmetry variable with the reconstructed response can be found in Appendix C.

6.2 Event selection

Events were required to pass the data quality requirements outlined in Section 2.4. Events were also rejected if either of the leading jets, or any other jet with p_{T} greater than 30% of the p_{T} of the sub-leading jet, failed the jet cleaning requirements in Section 3.6.

In order to select for well-balanced dijet events, additional kinematic cuts were applied.

The event was required to have at least two jets, with at least two tracks coming from the primary vertex (as defined in Section 2.2.2). The two leading jets were required to have $\Delta\phi > 170°$[1]. They were also required to be isolated via the reconstructed jet isolation criteria described in Section 3.5.3. If there was a third jet (J_3), a cut on the Jet Vertex Fraction (see Section 3.6) was used to reduce the effect of in-time and out-of-time pile-up:

- $p_{\mathrm{T}}^{J_3}$ was required to be smaller than max(12 GeV, $0.25p_{\mathrm{T}}^{\mathrm{avg}}$)
- If $|\eta_{J_3}| < 2.5$, $|\mathrm{JVF}_{\mathrm{J_3}}| > 0.25$ was required.

To ensure we were only looking at probe jets with the jet punch-through correction applied, a cut of $p_{\mathrm{T}}^{\mathrm{probe}} > 50$ GeV was used. In addition, all jets were required to be within the muon spectrometer acceptance region ($\eta_{det} < 2.7$).

The binning defined in Section 5.3.3 that was used to derive the correction function was also used to validate it in data. This was done to minimise any differences that could occur due to bin convolutions. However, instead of binning in E_{truth}, p_{T}^{probe} was used. Some p_{T}^{probe} bin merging was done in the high p_{T}^{probe} region due to the lack of statistics in data. This resulted in nine $p_{\mathrm{T}}^{\mathrm{probe}}$ (GeV) bins: 50-60, 60-80, 80-100, 100-200, 200-600, 600-1000, 1000-1600, 1600-3000, 3000-4500.

The same dijet selection criteria applied to data was applied to MC. No trigger selection was applied to MC, however the additional cleaning cuts specified in Section 3.6 were applied to remove additional overlaid pileup jets. The trigger selection applied to data was described in Section 6.3.

Finally, for the reconstructed response studies in Section 6.5 that utilise the tag-and-probe method described in Section 6.1, events must have one fully contained calorimeter jet ($\mathrm{N}_{segments} = 0$) and one punch-through jet ($\mathrm{N}_{segments} \geq 20$).

A total of 14.45 $\mathrm{fb}^{-1}$of data collected by ATLAS during 2012 were used for these studies. For MC, the nominal Pythia8 dijet samples detailed in Section 1.4 were used.

[1] $\Delta\phi = |\phi_{lead} - \phi_{subl}|$, where ϕ_{lead} is the azimuthal angle of the leading jet and ϕ_{subl} is the azimuthal angle of the sub-leading jet.

6.3 Trigger strategy

The trigger selection in data made use of a combination of single jet triggers. Triggers are characterised by their efficiency vs $p_{\mathrm{T}}^{\mathrm{jet}}$ curves (see Figure 2.14) and are increasingly prescaled in order to ensure the available bandwidth of 200 Hz was not exceeded. Distinct $p_{\mathrm{T}}^{\mathrm{jet}}$ regions were identified in which single jet triggers had an efficiency of greater than 99% (their plateau region). The lowest prescaled trigger in each $p_{\mathrm{T}}^{\mathrm{jet}}$ region for EM+JES-scale jets is shown in Table 6.1. Please see Section D.4.1 for the equivalent LCW+JES-scale $p_{\mathrm{T}}^{\mathrm{jet}}$ regions.

$p_{\mathrm{T}}^{\mathrm{jet}}$ **Range**	**Lowest Prescale Trigger With Efficiency** > 99%
$47 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 59	`EF_j25_a4tchad`
$59 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 99	`EF_j35_a4tchad`
$99 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 135	`EF_j55_a4tchad`
$135 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 185	`EF_j80_a4tchad`
$185 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 241	`EF_j110_a4tchad`
$241 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 273	`EF_j145_a4tchad`
$273 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 341	`EF_j180_a4tchad`
$341 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 411	`EF_j220_a4tchad`
$411 < p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] ≤ 460	`EF_j280_a4tchad`
$p_{\mathrm{T}}^{\mathrm{jet}}$ [GeV] > 460	`EF_j145_a4tchad_ht700_L2FS`

Table 6.1: Table summarising p_{T} ranges and the lowest prescaled trigger with efficiency greater than 99% in each region. `EF` refers to event filter, `j` signifies jet, the subsequent number signifies the $p_{\mathrm{T}}^{\mathrm{jet}}$ threshold, `a4thchad` signifies anti-k_t R = 0.4 at the EM+JES scale fully reconstructed from topoclusters were used, `ht700` signifies a $\sum \mathrm{E}_T > 150$ GeV cut, while `L2FS` signifies the trigger chain used the L2 Full Scan algorithm (see Section 2.2.5).

For this analysis, the $p_{\mathrm{T}}^{\mathrm{jet}}$ regions the leading and sub-leading jets belong to were found, and the lowest prescaled triggers from each $p_{\mathrm{T}}^{\mathrm{jet}}$ region combined. This resulted in the event being accepted if at least one of the triggers passed. If both jets fell into the same $p_{\mathrm{T}}^{\mathrm{jet}}$ region, they were assigned identical triggers, and thus only one trigger was used.

In order to account for the trigger prescales, a weight calculated using Equation 6.2 was applied. Here, w_{jk} is the applied weight, PS is the prescale of the j-th or k-th trigger, whereby j and k are the index of the trigger used for the leading and sub-leading jet respect-

ively. The weight is based upon the inclusive weighting strategy found in reference [132].

$$w_{jk} = \frac{PS_j PS_k}{PS_j + PS_k - 1} \tag{6.2}$$

Figure 6.2 shows the $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution, split by triggers, before and after the weights were applied, showing the shape of the $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution is restored. See Section 2.2.5 for more information on the ATLAS trigger system, single jet triggers and trigger efficiency curves.

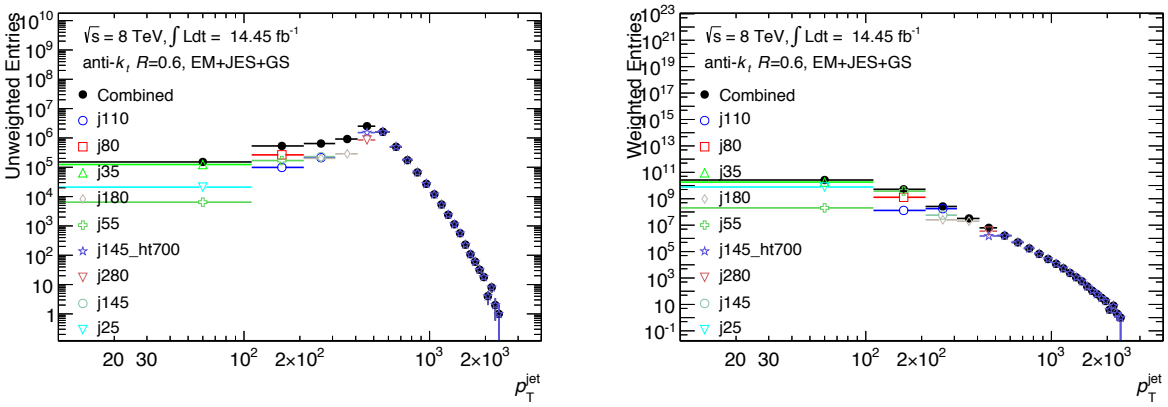


Figure 6.2: $p_{\mathrm{T}}^{\mathrm{jet}}$ distributions before (left) and after (right) applying the prescale weights defined in Equation 6.2 for each trigger in Table 6.1 and combined.

6.4 $\mathrm{N}_{segments}$ distribution comparison

As mentioned at the beginning of this chapter, it is important to verify the shape of the input $\mathrm{N}_{segments}$ distribution is comparable between data and MC when binning in both p_{T} and $\mathrm{N}_{segments}$. In Figure 4.2 when comparing the $\mathrm{N}_{segments}$ distribution in data and MC for all jets, MC deviated from data by up to 50% in the high $\mathrm{N}_{segments}$ region.

Figure 6.3 compares the shape of the $\mathrm{N}_{segments}$ distribution in each of the three η_{det} regions for jets with $600 \leq p_{\mathrm{T}}^{ref}$ [GeV] < 1000. In the $0 \leq \eta_{det} < 1.3$ region, the discrepancy between data and MC grows with $\mathrm{N}_{segments}$, reaching up to 70% in the high $\mathrm{N}_{segments}$ region. For the $1.3 \leq \eta_{det} < 1.9$ region, differences of up to $\sim 20\%$ can be seen for jets with $\mathrm{N}_{segments} \leq 100$, however this discrepancy disappears for jets with $\mathrm{N}_{segments} > 100$. In

both of these regions, MC has more events than data at very low $N_{segments}$ and fewer events at high $N_{segments}$. As observed during the derivation of the correction function in Chapter 5, there is a lack of statistics in the $1.9 \leq \eta_{det} < 2.7$ region at high $N_{segments}$, however at low $N_{segments}$ (< 5) the distributions show relatively good agreement.

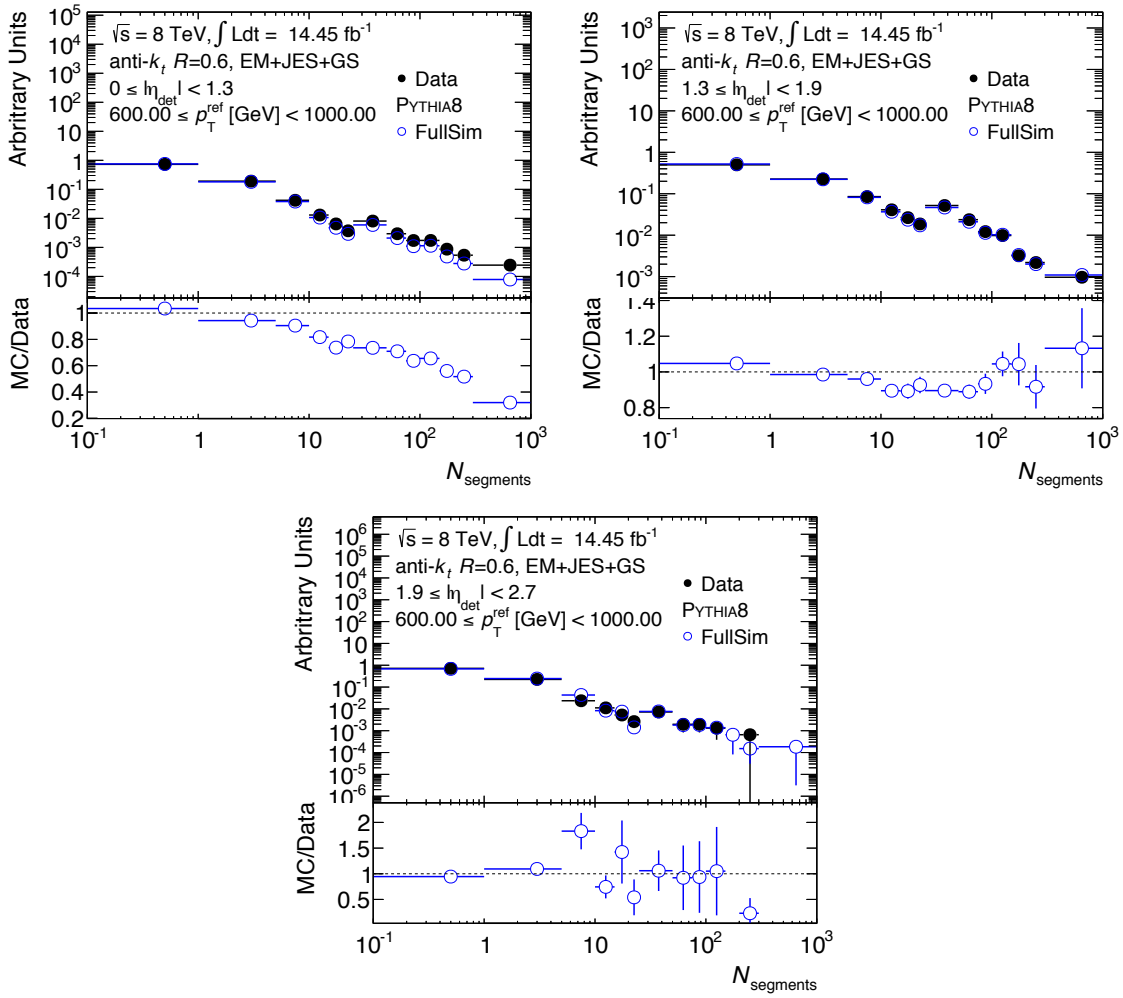


Figure 6.3: Figure comparing the shape of the $N_{segments}$ distribution between data and MC for jets with $600 \leq p_{\mathrm{T}}^{ref}$ [GeV] < 1000 in the $0 \leq \eta_{det} < 1.3$ (left), $1.3 \leq \eta_{det} < 1.9$ (right) and $1.9 \leq \eta_{det} < 2.7$ (bottom). The bottom plots show the ratio between MC and data.

Comparing Figure 6.3 with Figure 4.2, it appears the large discrepancy in the $0 \leq \eta_{det} < 1.3$ region at high $N_{segments}$ is offset by the better agreement in the $1.3 \leq \eta_{det} < 1.9$ region. Although the discrepancies in the $0 \leq \eta_{det} < 1.3$ region are observed to be noticeable in the high $N_{segments}$ region, the impact of $N_{segments}$ on the response in data and MC must be compared to assess how this difference in modelling may affect the correction function.

6.5 Assessing dijet imbalance due to jet punch-through

The reconstructed response as defined in Equation 6.1 is used to compare the impact of jet punch-through on the JES between data and MC. The mean reconstructed response is obtained from Gaussian fits to the reconstructed response distribution in bins of $N_{segments}$, p_T^{probe} and η_{det}.

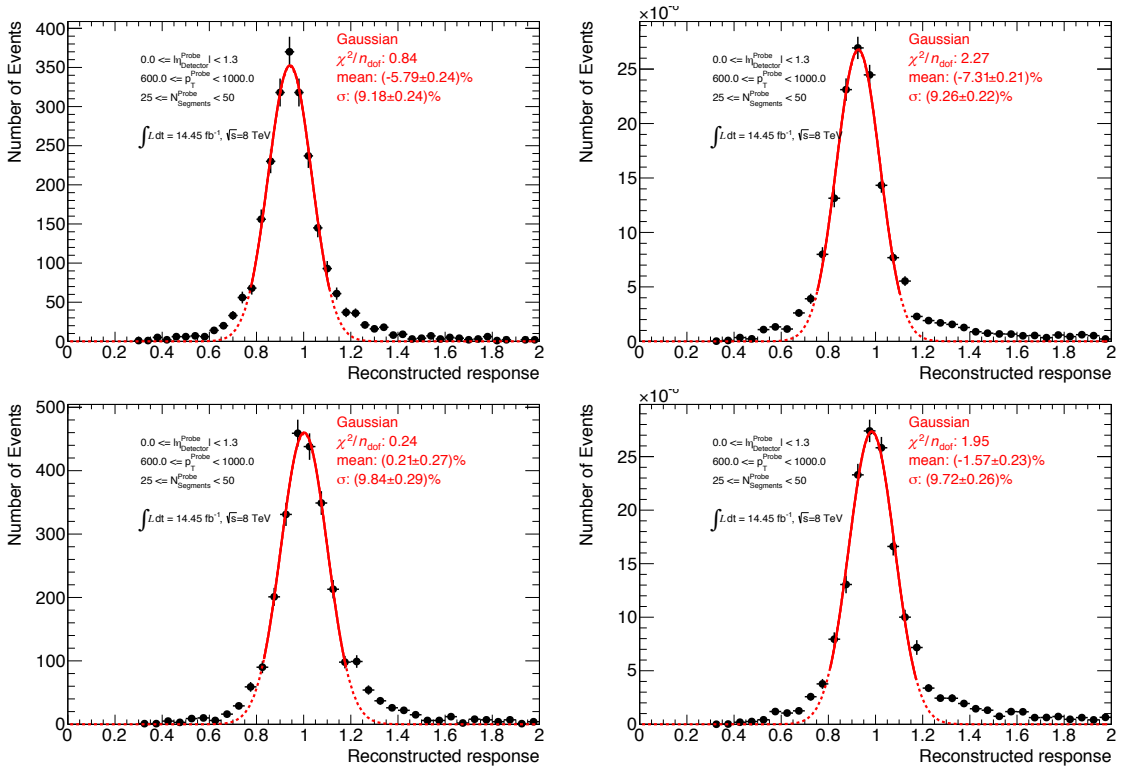


Figure 6.4: Examples of fits to reconstructed response distributions for jets with $600 \leq p_T^{probe}$ [GeV] < 1000, $25 \leq N_{segments}^{probe} < 50$ in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied, in data (left) and MC (right).

The fitting procedure used was identical to the one used to derive the correction function (described in Section 5.4). However, it is important to note here that the fits at high $N_{segments}$ are very statistically limited in all regions, a result of requiring events to have one punch-through jet ($N_{segments} \geq 20$) and one non-punch-through jet ($N_{segments} = 0$). To account for this, the quality of fit requirement was loosened by removing the limitation of $\tilde{\chi}^2_{red} < 50$, and instead requiring $|mean_{fit} - mean_{hist}| \leq (0.5 * \sigma_{hist})$, where $mean_{fit}$ is the mean extracted from the fit, $mean_{hist}$ is the histogram mean calculated from Equation 4.2

and σ_{hist} is the standard deviation calculated from Equation 4.3.

Example fits to the reconstructed response distribution for jets with $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000, $25 \leq \mathrm{N}_{segments}^{probe} < 50$ and $0 \leq |\eta_{det}| < 1.3$, for data and MC, with and without the punch-through correction applied, are shown in Figure 6.4. From these plots, it can be seen that the reconstructed response distributions are Gaussian before and after the application of the correction. The $\tilde{\chi}^2_{red}$ values are ~ 1, suggesting reasonable fits.

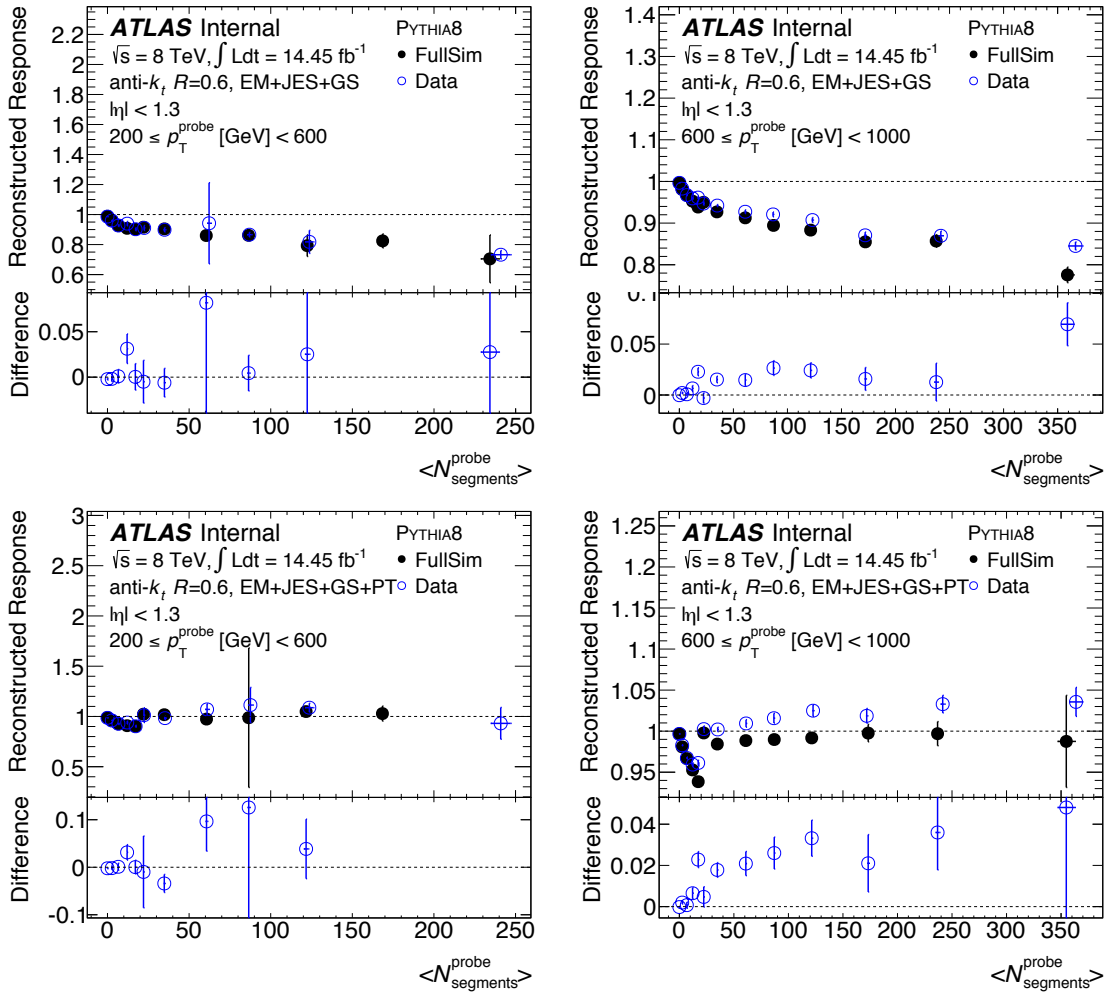


Figure 6.5: Reconstructed response as a function of $\mathrm{N}_{segments}^{probe}$ compared between data and MC for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 600 (left) and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000 (right) in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied. The bottom plot shows the difference between data and MC.

Figure 6.5 compares the reconstructed response as a function of $\mathrm{N}_{segments}^{probe}$ between data and MC for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 600 and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000 in the central η_{det} region ($0 \leq \leq |\eta_{det}| < 1.3$). The curve is shown before and after the

application of the jet punch-through correction.

Before the correction is applied, the reconstructed response decreases as $N_{segments}$ increases. The degradation is less severe for jets at higher p_T, where the reconstructed response can be seen to degrade by as much as 30% ($\sim$ 15%) in the highest $N_{segments}$ region for jets with $200 \leq p_T^{probe}$ [GeV] < 600 ($600 \leq p_T^{probe}$ [GeV] < 1000). Thus it can be concluded that the reconstructed response and truth response (Equation 3.4) behave in a similar way.

After the correction is applied, the reconstructed response is corrected back to within a couple of percent of 1 in regions it is applied ($N_{segments} \geq 20$), suggesting any imbalance due to jet punch-through has been largely removed. The $200 \leq p_T^{probe}$ [GeV] < 600 region is statistically limited in the high $N_{segments}$ region in data with fits producing large errors.

Before the correction, the data/MC differences appear to be roughly compatible with zero for jets with $200 < p_T^{probe}$ [GeV] < 600. The differences are slightly larger after the correction, but still within a few percent of 0 when taking into account uncertainties. For jets with $600 \leq p_T^{probe}$ [GeV] < 1000 at high $N_{segments}$, the differences reach $\sim$ 7% before the correction. This is reduced to $\sim$ 5% after the correction.

From these plots, it can be concluded that the effect of jet punch-through on the jet energy scale is reasonably well modelled. Any differences seen here are taken into account through an uncertainty on the correction function as described in Section 6.6.

6.6 Derivation of an uncertainty

An uncertainty on the jet punch-through correction is necessary to take into account the data/MC differences seen in the reconstructed response curves in Figure 6.5. The uncertainty is constructed as a function of ($N_{segments}$, p_T^{probe}) in bins of η_{det}, analogous to the jet punch-through correction being constructed as a function of ($N_{segments}$, E_{truth}) in bins of η_{det}.

Uncertainties are derived for anti-k_t jets, with R = 0.4 and R = 0.6, calibrated with the EM+JES+GS+PT and LCW+JES+GS+PT schemes.

6.6.1 Constructing the uncertainty

The uncertainty is calculated as the absolute difference between the data and MC mean reconstructed response as a function of $\mathrm{N}_{segments}$, in η_{det} and $p_{\mathrm{T}}^{\mathrm{probe}}$ bins. This is defined in Equation 6.3, where $\Delta\mathcal{R}_{reco}$ is the uncertainty on the jet punch-through correction function, $\langle\mathcal{R}_{reco}^{MC}\rangle$ is the reconstructed response in MC, $\langle\mathcal{R}_{reco}^{data}\rangle$ is the reconstructed response in data and $|_{\eta_{det},p_{\mathrm{T}}^{\mathrm{probe}},\mathrm{N}_{segments}}$ signifies these values are in bins of (η_{det}, $p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$). Thus for Figure 6.5, the uncertainties are obtained by taking the absolute values of the differences in the bottom plots in order to obtain symmetric uncertainties.

$$\Delta\mathcal{R}_{reco}|_{\eta_{det},p_{\mathrm{T}}^{\mathrm{probe}},\mathrm{N}_{segments}} = |\langle\mathcal{R}_{reco}^{MC}\rangle|_{\eta_{det},p_{\mathrm{T}}^{\mathrm{probe}},\mathrm{N}_{segments}} - \langle\mathcal{R}_{reco}^{data}\rangle|_{\eta_{det},p_{\mathrm{T}}^{\mathrm{probe}},\mathrm{N}_{segments}}| \qquad (6.3)$$

The uncertainty obtained in (η_{det}, $p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$) bins is displayed in Figure 6.6. Empty bins represent regions where no uncertainty was derived due to the lack of successful fits.

The uncertainty increases as p_{T} decreases, and in general increases as $\mathrm{N}_{segments}$ increases. The uncertainty reaches up to 13% in the $0 \leq \eta_{det} < 1.3$ region, 10% in the $1.3 \leq \eta_{det} < 1.9$ region and 8% in the $1.9 \leq \eta_{det} < 2.7$ region. It is important to note however the error on the uncertainty is large for the larger uncertainties in the $0 \leq \eta_{det} < 1.3$ and $1.3 \leq \eta_{det} < 1.9$ regions. This likely due to low statistics in these regions. Also, it is clear (as with the derivation of the correction) that the $1.9 \leq \eta_{det} < 2.7$ region is statistically limited, with uncertainties only derived for two bins in this region.

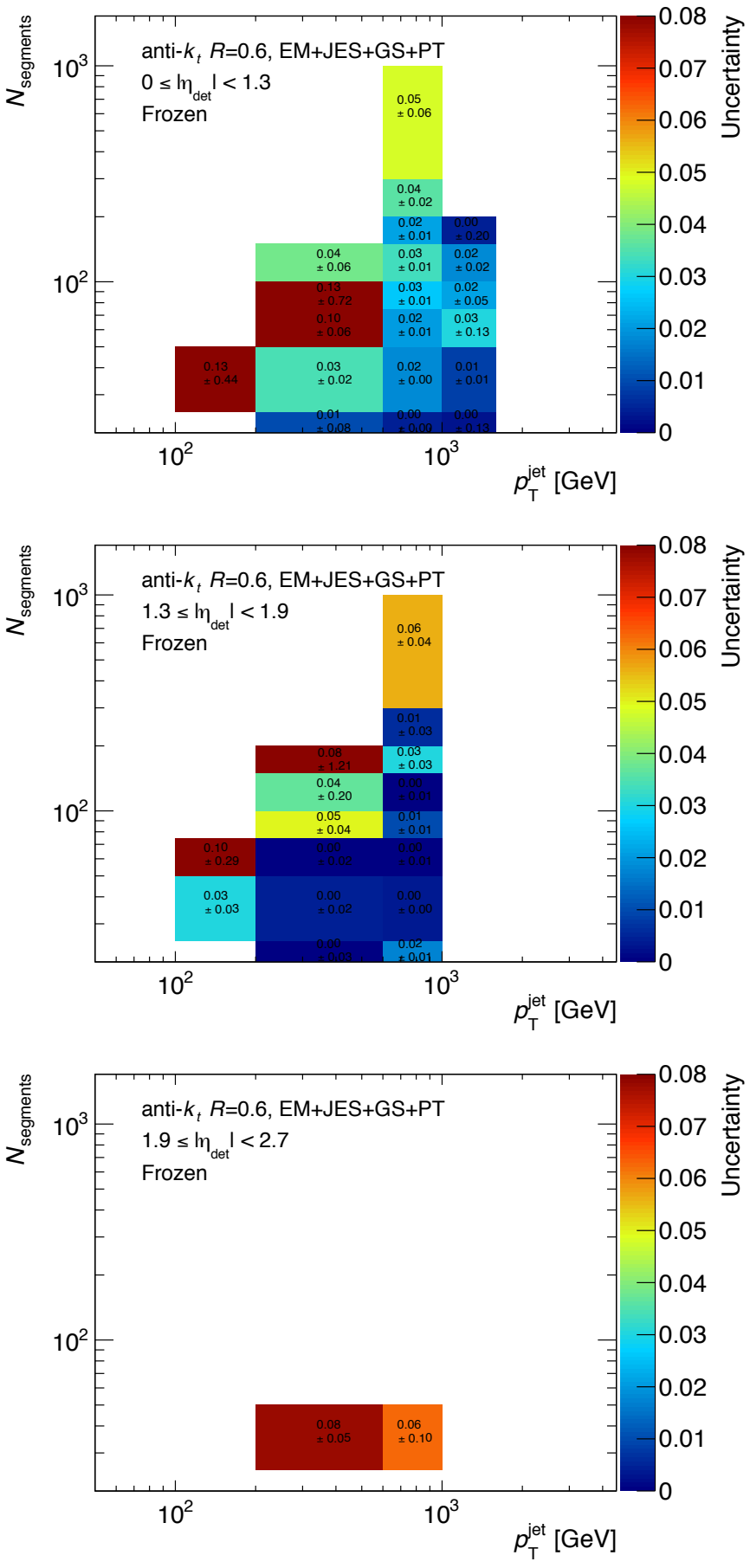


Figure 6.6: Uncertainty vs $N_{segments}$ vs p_T^{jet} for jets with $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom).

To construct the uncertainty function, the same sequence of Gaussian kernels used to derive the correction function in Section 5.6 was used. For the uncertainty however, the kernels are a function of ($\log p_{\mathrm{T}}^{\mathrm{probe}}$, $\log \mathrm{N}_{segments}$) as opposed to ($\log E^{jet}$, $\log \mathrm{N}_{segments}$). As with the correction function derivation, $\sigma_{p_{\mathrm{T}}^{\mathrm{probe}}}$ and $\sigma_{\mathrm{N}_{segments}}$ were optimised to capture the shape of the uncertainty across η_{det} and ($p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$), whilst providing stability against statistical fluctuations. The uncertainties shown in Figure 6.6 were used as inputs (C_i) to the one-dimensional Gaussian kernel to form an uncertainty as a function of ($p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$) in bins of η_{det}. For regions of phase space limited by a lack of successful fits, three approaches were taken:

1. Freeze the uncertainty function at the value in the nearest ($p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$) bin with information, using the same freezing procedure mentioned at the end of Section 5.6.

2. Allow the derived uncertainty function to extrapolate the uncertainty into regions with no information.

3. As the $1.9 \leq \eta_{det} < 2.7$ regions suffered from low statistics and thus very few fits (seen in Figure 6.6), a weighted average was calculated via Equation 6.4 using the available uncertainties. This weighted average was then taken as a flat uncertainty for the entire region. In Equation 6.4, $\bar{x}$ is the weighted average taken to be the flat uncertainty, ${w_i}^2$ is the inverse square of the error on the uncertainty and x_i is the uncertainty. It is also important to note that for anti-k_t $R = 0.4$ jets, no fits were successful in the $1.9 \leq \eta_{det} < 2.7$ regions. As a result, the derived anti-k_t $R = 0.6$ uncertainties in this region were used for anti-k_t $R = 0.4$ jets.

$$\bar{x} = \frac{\sum_{i=0}^{n} w_i x_i}{\sum_{i=0}^{n} w_i} \tag{6.4}$$

[2] $w_i = \frac{1}{\sigma_i^2}$

6.6.2 Final uncertainties

The final uncertainties as a function of ($p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$) for each of the three η_{det} bins are shown in Figure 6.7. Both the frozen and extrapolated uncertainties are shown.

It can be seen the uncertainties are relatively smooth as a function of ($p_{\mathrm{T}}^{\mathrm{probe}}$, $\mathrm{N}_{segments}$). For the $1.9 \leq \eta_{det} < 2.7$ region, the uncertainty is the weighted average as described above, and is thus the same for the frozen and extrapolated methods. This gives a flat uncertainty at $\sim 7.5\%$. The extrapolated uncertainties reach up to 12% ($\sim$10%), much higher values than the frozen uncertainties that reach up to $\sim 6\%$ ($\sim 4.5\%$) in the $0 \leq |\eta_{det}| < 1.3$ ($1.3 \leq |\eta_{det}| < 1.9$) region. The uncertainties grow larger as $\mathrm{N}_{segments}$ increases and $p_{\mathrm{T}}^{\mathrm{probe}}$ decreases. Thus the uncertainty is large in regions of phase space with a large jet punch-through correction.

The $\tilde{\chi}^2_{red}$ is 0.33 (0.32) in the $0 \leq |\eta_{det}| < 1.3$ region and 0.433 (0.430) in the $1.3 \leq |\eta_{det}| < 1.9$ region for the frozen (extrapolated) uncertainties. No $\tilde{\chi}^2_{red}$ was calculated for the $1.9 \leq \eta_{det} < 2.7$ region as a weighted average was used here. These $\tilde{\chi}^2_{red}$ values are small (< 1) for both the frozen and extrapolated uncertainties, although slightly larger for the frozen ones. The small $\tilde{\chi}^2_{red}$ values are due to the limited statistics resulting in large errors on the uncertainties in some regions.

As the frozen uncertainties are observed to be smaller and with slightly better $\tilde{\chi}^2_{red}$ values, these were used as the nominal uncertainties on the jet punch-through correction function. The uncertainties have negligible impact on the total JES uncertainty (as will be shown in Section 8.2).

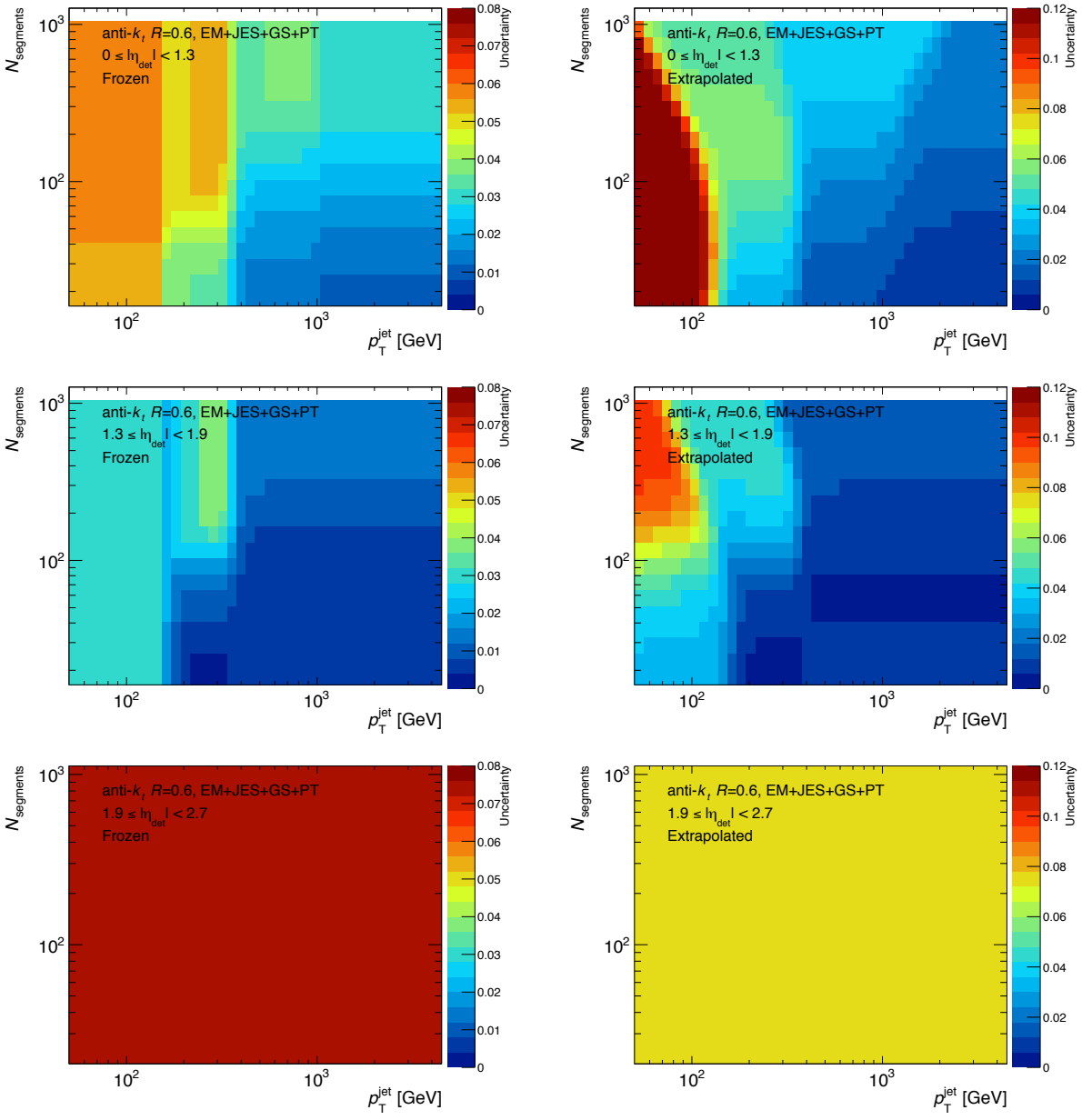


Figure 6.7: Final uncertainty a function of $N_{segments}$ and p_T^{jet} for jets with $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom). Plots on the left show the frozen uncertainty, while plots on the right show the extrapolated uncertainty.

Chapter 7

Performance of the Jet Punch-Through Correction

It was shown in Chapter 6 that the jet punch-through correction removes the dijet imbalance due to jet punch-through in both MC and data, correcting the in-situ response as a function of $N_{segments}$ back to 1. This chapter investigates how the jet punch-through correction impacts jet energy reconstruction in MC in Section 7.1, followed by an assessment of how the correction affects a search for new physics in Section 7.2.

All performance studies in this chapter used the nominal Pythia8 samples, with the selection cuts used to derive the correction factors described in Section 5.2.

7.1 Impact on jet energy reconstruction

The impact of the jet punch-through correction on jet energy reconstruction is assessed by looking at the affect on the jet energy scale in Section 7.1.1, the resultant change in the jet energy response tails in Section 7.1.2, and corresponding effect on the jet energy resolution in Section 7.1.3.

7.1.1 Impact on jet energy scale

The aim of the jet punch-through correction was to remove the dependence of the jet energy response on $N_{segments}$, while leaving the mean jet energy scale in η_{det} and E_{truth} bins unchanged. Studies in this section will confirm the correction meets these specifications.

From Figure 7.1, it can be seen that the correction does not affect the Gaussian nature of the jet energy response distribution, with the correction mainly impacting the low-response tail. The higher E_{truth} bin appears to have been affected more than the low E_{truth} bin. This effects suggests the correction is behaving as designed, accounting for jet energy losses due to jet punch-through by pushing the response of jets in the low response tail back to 1, and having a larger impact on the RMS of the response of higher E_{truth} jets, that have a larger jet punch-through probability. A more thorough study of the impact of the correction on the jet energy response tails can be found in Section 7.1.2.

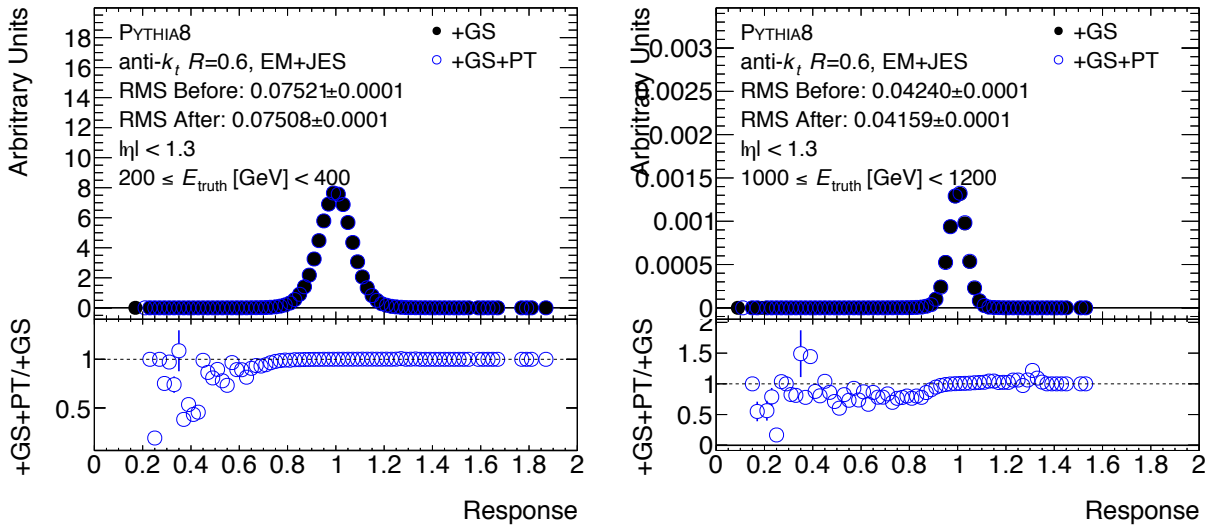


Figure 7.1: Jet energy response distribution integrated over $N_{segments}$ compared before and after the jet punch-through correction, for jets in the $0 \leq |\eta_{det}| < 1.3$ region with $200 \leq E_{truth}$ [GeV] < 400 (left) and $1000 \leq E_{truth}$ [GeV] < 1200 (right). The bottom plots show the bin-by-bin ratio of entries after the correction to before the correction.

Figure 7.2 shows the jet energy response[1] as a function of $N_{segments}$ in E_{truth} and η_{det} bins. It is clear from these plots that the jet punch-through correction restores the jet

[1]For all studies in this section, the jet energy response was obtained from a fit to the jet energy distribution as described in Section 5.4. Cuts used are identical to those in Section 5.2.

energy response to 1, suggesting energy mis-measurements due to jet punch-through that $N_{segments}$ is sensitive to are largely removed.

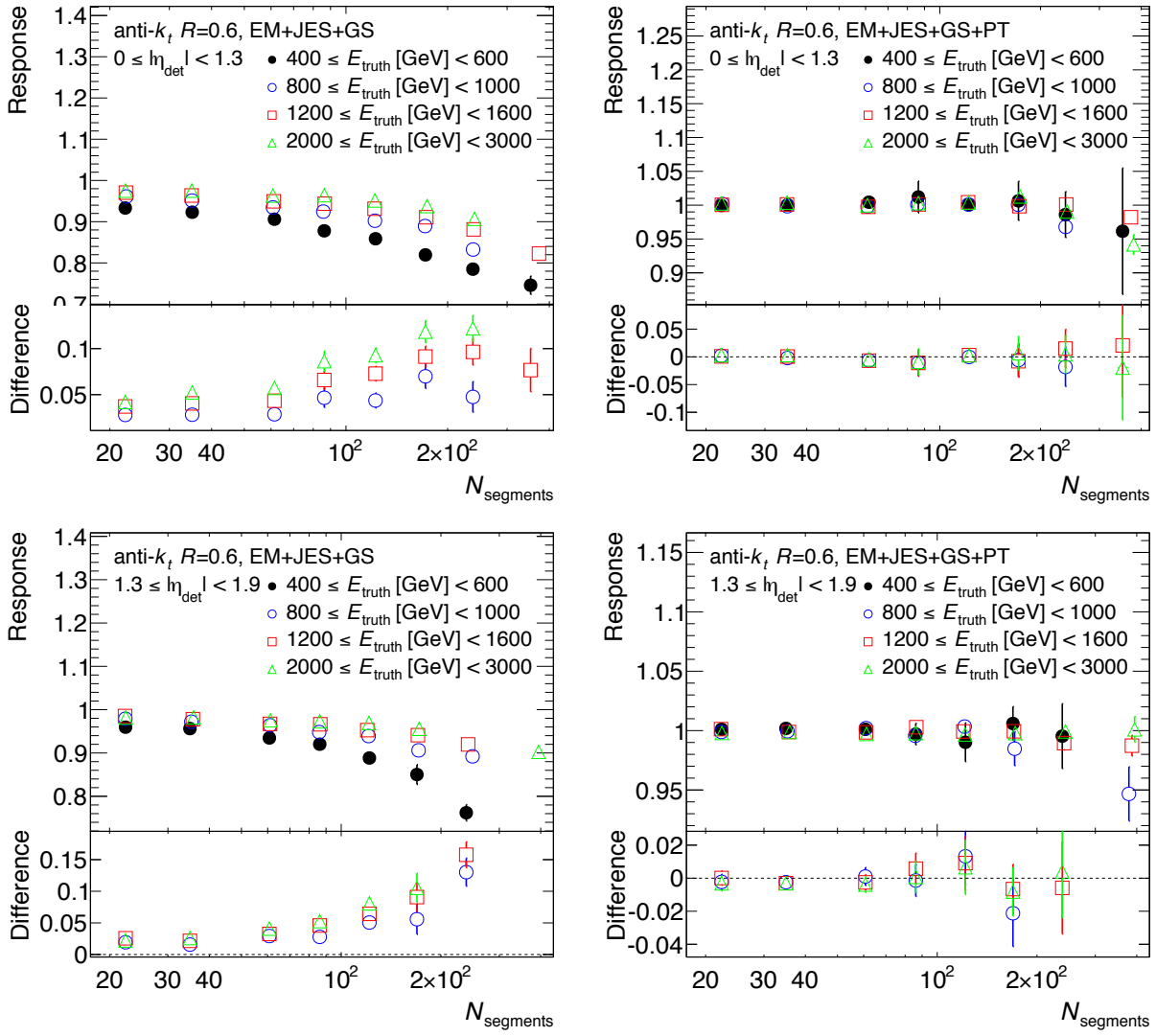


Figure 7.2: Jet energy response vs $N_{segments}$ before (left) and after (right) the jet punch-through correction, for jets in the $0 \leq |\eta_{det}| < 1.3$ region (top) and $1.3 \leq |\eta_{det}| < 1.9$ region (bottom). The bottom plots show the difference in the response relative to the $400 \leq E_{truth}$ [GeV] < 600 bin.

Figure 7.3 shows the jet energy response as a function of E_{truth} when integrating over all $N_{segments}$ bins compared before (solid circles) and after (empty circles) the application of the jet punch-through correction, in each of the three η_{det} bins used to derive the corrections. The mean jet energy response is almost identical before and after the correction, making it very difficult to discern the empty circles from the solid circles. The difference is observed to be small - below 0.3% in all bins. This difference appears to grow with E_{truth}, as expected

from the fact the fraction of punch-through jets grows with E_{truth} (see Figure 4.6).

Thus it can be concluded the correction is working as designed - removing the jet energy scale dependence on $N_{segments}$, as seen in Figure 7.2, whilst also keeping the mean jet energy scale response unchanged, seen in Figure 7.3.

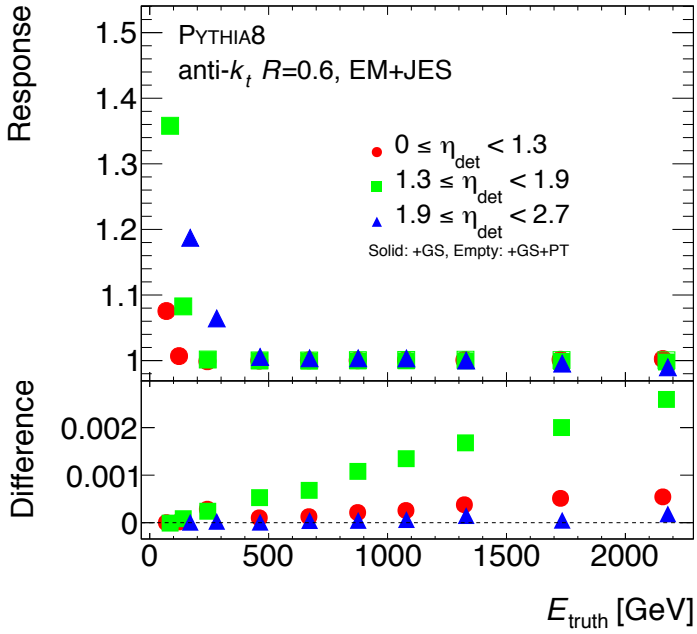


Figure 7.3: Jet energy response vs E_{truth} before (solid circles) and after (empty circles) the application of the punch-through correction for jets in each η_{det} region. The bottom plot shows the difference in the response before and after the application of the correction.

7.1.2 Effect of correction on jet energy response tails

In order to understand the impact of the jet punch-through correction on the jet energy resolution, it is first important to investigate the impact of the correction on the low response tails in the jet energy response distributions, observed to grow with $N_{segments}$ in Section 4.2.2. Figure 7.4 compares the percentage of events below a given response thresholds for jets with and without the jet punch-through correction applied. The comparison is shown for jets in the $0 \le \eta_{det} < 1.3$, $100 \le N_{segments} < 150$ bin at high ($200 \le E_{truth}$ [GeV] < 400) and low ($1000 \le E_{truth}$ [GeV] < 1200) truth energy. These plots are analogous to those in Figure 4.9.

It can be seen the jet punch-through correction reduces the percentage of events in the tails in both bins. This effect is common in all η_{det}, E_{truth} and $N_{segments}$ bins. Combined with the removal of the dependence of the jet energy response on $N_{segments}$, these effects help to reduce the low response tail in the mean JES response, as was seen in Figure 7.1.

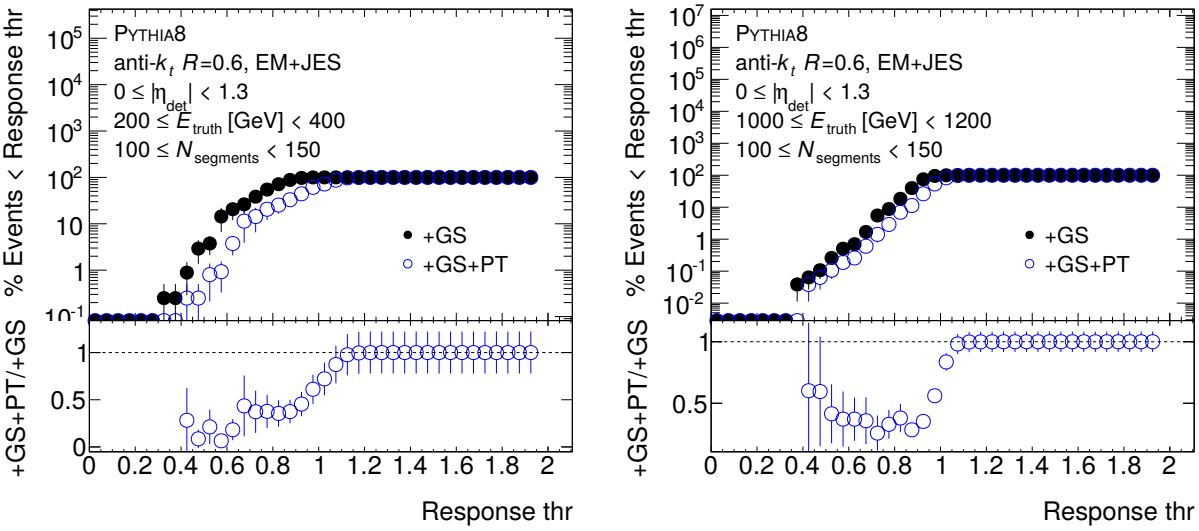


Figure 7.4: Percentage of events below a given response threshold value compared for jets with and without the jet punch-through correction, in the $0 \le \eta_{det} < 1.3$, $100 \le N_{segments} < 150$ bin, with $200 \le E_{truth}$ [GeV] < 400 (left) and $1000 \le E_{truth}$ [GeV] < 1200 (right). The bottom plots show the ratio of the percentage of events below the threshold before and after the correction.

An effective way of analysing the impact of the correction on the jet energy response tails as a function of E_{truth} is to analyse how the -2σ and -1σ quantile positions of the jet energy response distribution are affected by the correction. Figure 7.5 shows the position of the -2σ and -1σ histogram quantiles, and position of the quantiles relative to the mean response (calculated as in Equation 4.2), for the jet energy response distribution as a function of E_{truth}, before and after the correction, in each of the three η_{det} regions. The correction pushes the -2σ and -1σ quantiles to a higher response value, with the shift in the position of the quantiles growing with E_{truth}, and largest in the $1.3 \le \eta_{det} < 1.9$ region (a shift of up to 15%). This shows the largest impact is in regions of phase space with the largest fraction of punch-through jets.

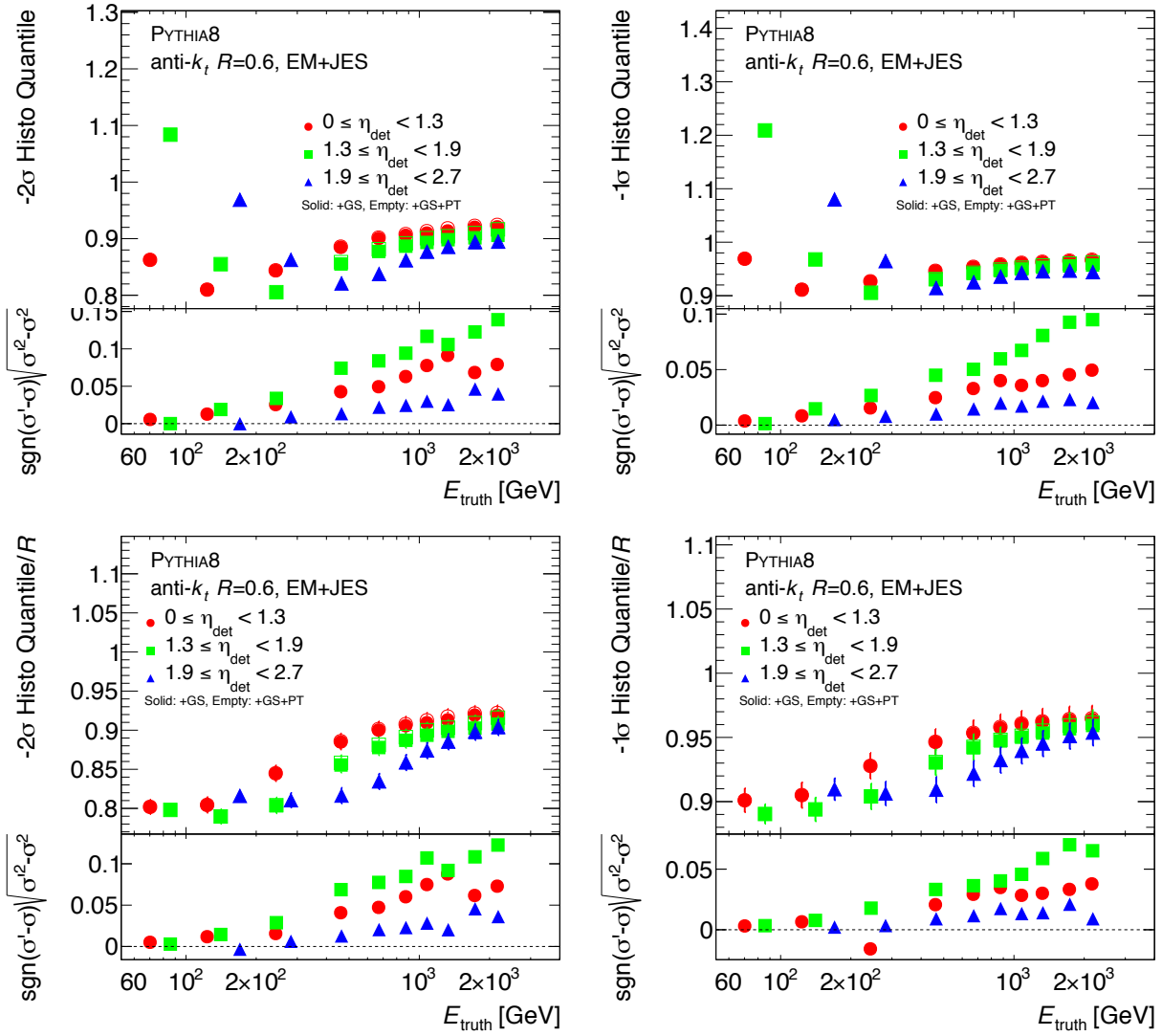


Figure 7.5: -2σ and -1σ response distribution quantile positions (top) relative to the histogram mean response (bottom) as a function of E_{truth} in each of the three η_{det} bins, before (solid circles) and after (empty circles) the punch-through correction is applied. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

The -2σ and -1σ fit quantile positions for the jet energy response distribution are shown in Figure 7.6, along with the position of the quantiles relative to the mean response extracted from the fit. The fit was performed in the same way as described in Section 5.4, with only the $\mu \pm 1.6\sigma$ core of the distribution being fit. As a result, the fit does not adequately capture the shape of the tails of the underlying distribution. However, in general, the correction also pushes the fit quantile positions to higher response values, although the shift is not as large as compared with that of the underlying distribution seen in Figure 7.5.

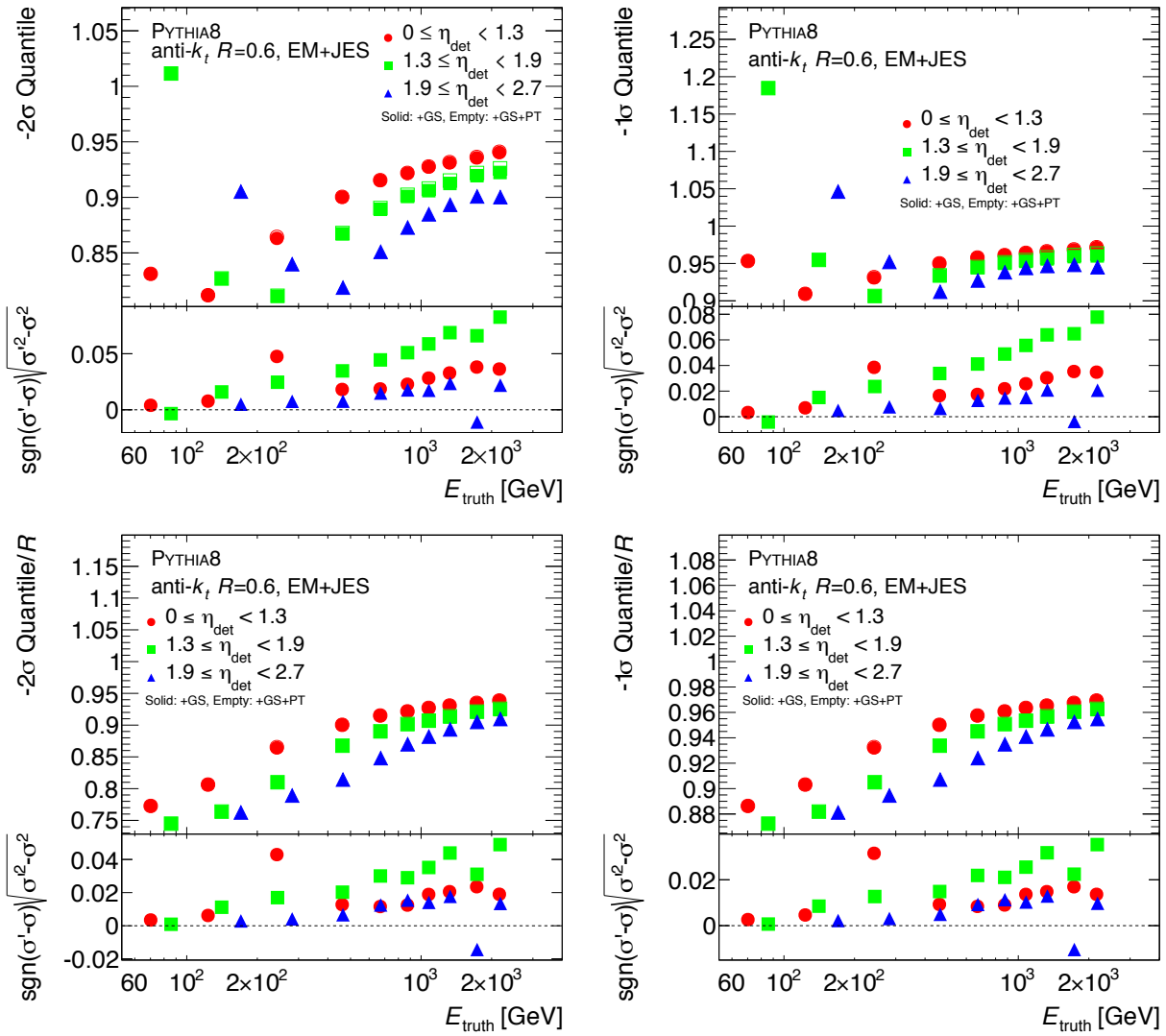


Figure 7.6: -2σ and -1σ response distribution quantile positions extracted from fits (top) relative to the fit mean response (bottom) as a function of E_{truth} in each of the three η_{det} bins, before (solid circles) and after (empty circles) the punch-through correction is applied. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

The -2σ and -1σ fit quantile positions in Figure 7.6 for jets with $1600 \leq \mathrm{E}_{truth}\,[\,\mathrm{GeV}] < 2000$ in the $1.9 \leq \eta_{det} < 2.7$ region appear to display a degradation that was not observed when looking at the histogram quantiles in Figure 7.5. Figure 7.7 displays the fitted jet energy response distribution for these jets along with a direct comparison of the distributions. It is clear that the correction has had very little impact on this distribution, with no noticeable impact on the shape of the distribution. Additionally, even though a degradation is observed when using the fits, an improvement is seen in the RMS, whereby the RMS is reduced, albeit by a very small amount, from 0.05039 before the correction to 0.0506 after.

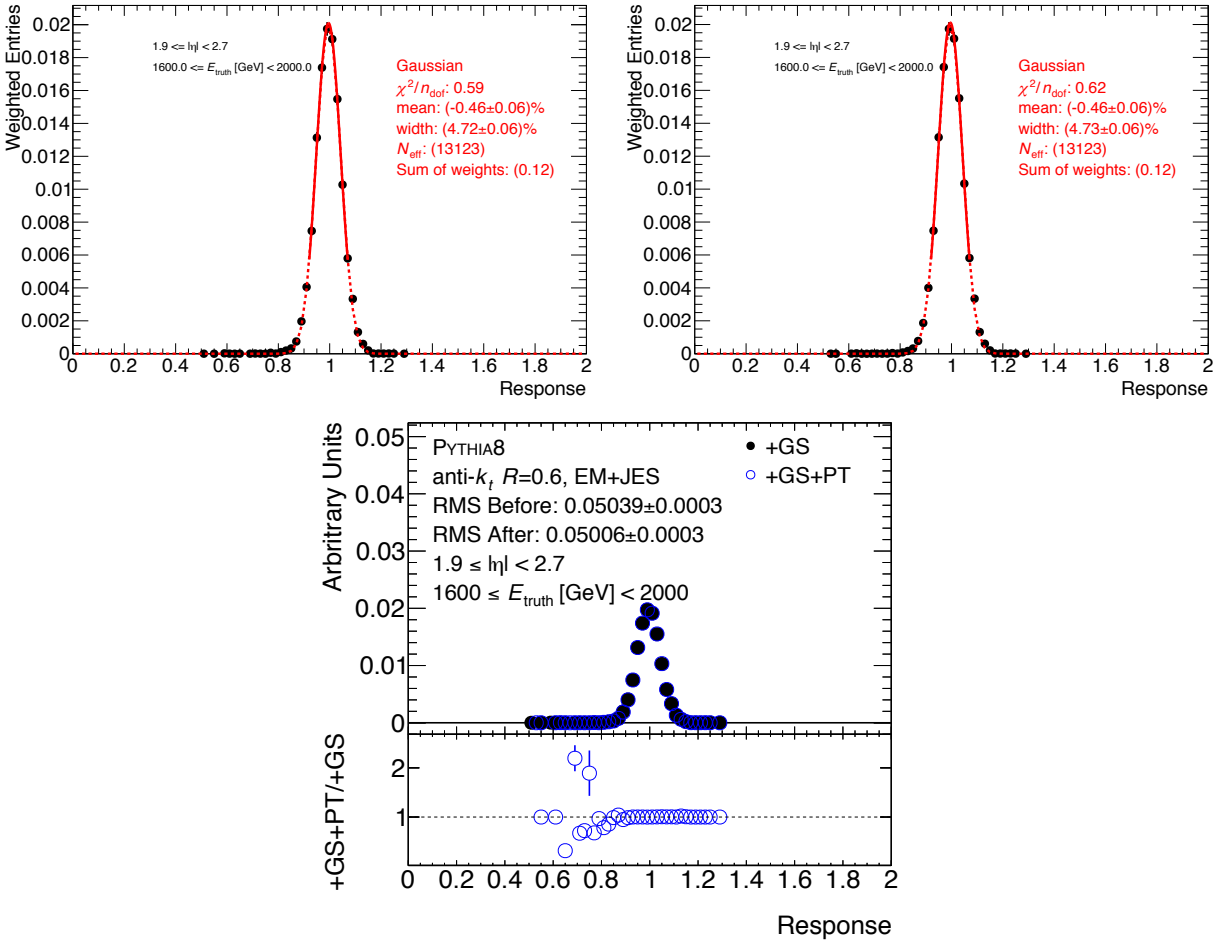


Figure 7.7: The top two plots show the fitted jet energy response distribution for jets with $1.6 \leq \mathrm{E}_{truth}\,[\,\mathrm{TeV}] < 2$ in the $1.9 \leq \eta_{det} < 2.7$ region, before (left) and after(right) the application of the jet punch-through correction. The bottom plot shows a direct comparison of the distributions, along with a bin-by-bin ratio in the lower part of the plot.

7.1.3 Impact on jet energy resolution

The reduction in the size of the low jet energy response tails seen in Section 7.1.2 is expected to reduce the overall fluctuations in the jet energy response, therefore improving the jet energy resolution. Figure 7.8 shows the jet energy resolution, described in terms of the RMS in Equation 4.6, before and after the application of the jet punch-through correction. The resolution is shown as a function of $N_{segments}$, for different E_{truth} bins in each η_{det} region. These plots are analagous to those in Figure 4.11.

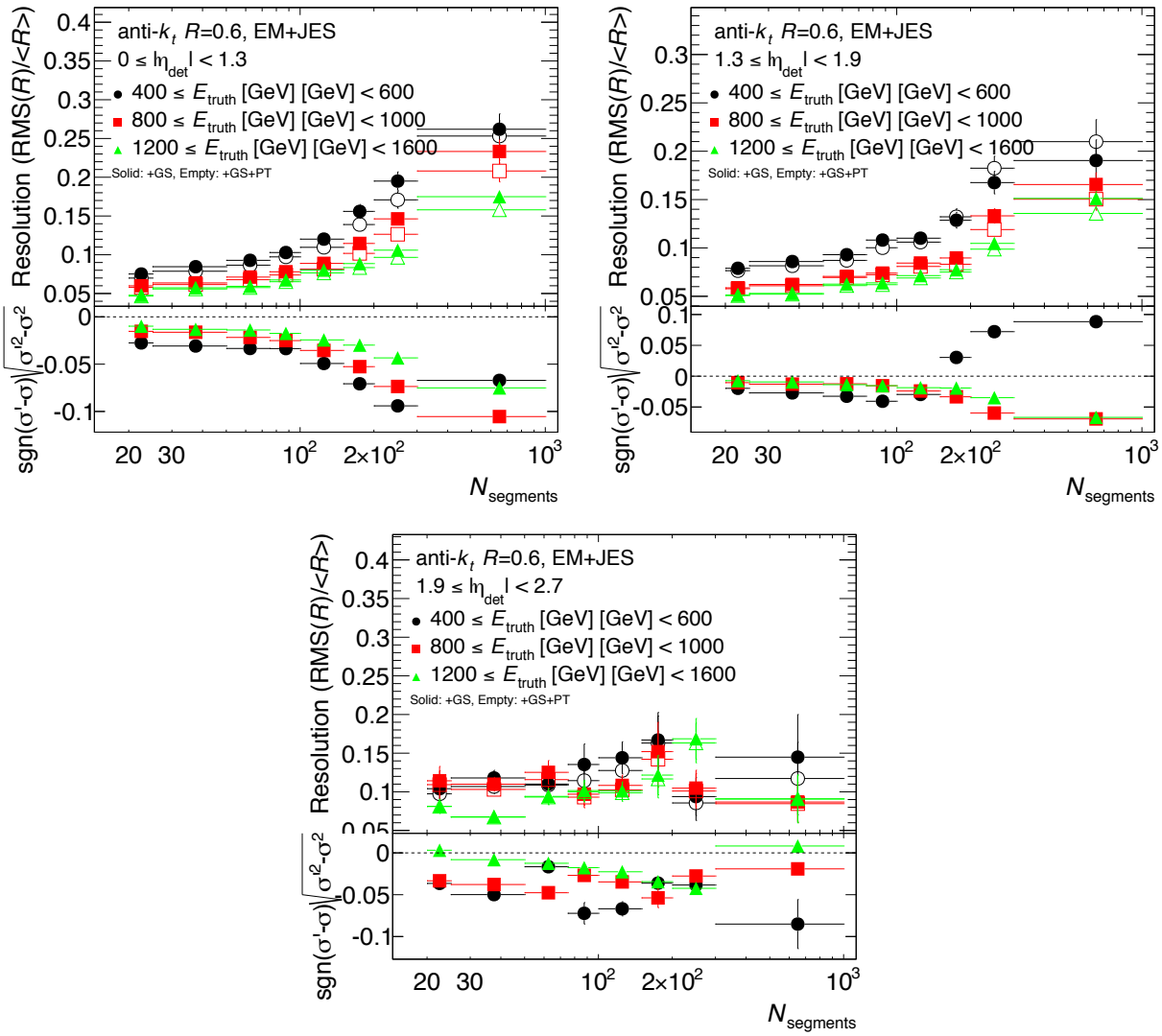


Figure 7.8: The RMS($\mathcal{R}$) divided by the mean histogram response (left) as a function of $N_{segments}$ in each of the three η_{det} regions for various E_{truth} bins, before (solid circles) and after (empty circles) the application of the jet punch-through correction. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

In the $0 \leq \eta_{det} < 1.3$ and $1.3 \leq \eta_{det} < 1.9$ regions, it can be seen that the correction, in general, improves the jet energy resolution. The improvement appears to be marginally larger as E_{truth} increases, reaching up to $\sim 10\%$. Although an improvement in the jet energy resolution is observed in most regions, the dependence on $N_{segments}$ is not completely removed. An apparent degradation is seen in the $1.3 \leq \eta_{det} < 1.9$, $400 \leq E_{truth}$ [GeV] < 600, high $N_{segments}$ region, however, low statistics in this region result in large errors on the RMS. The response distributions for these two bins are shown in Figure 7.9, from which the apparent degradation can be seen to be caused by a large low response tail still present, while the mean of the distribution has been shifted back to 1.

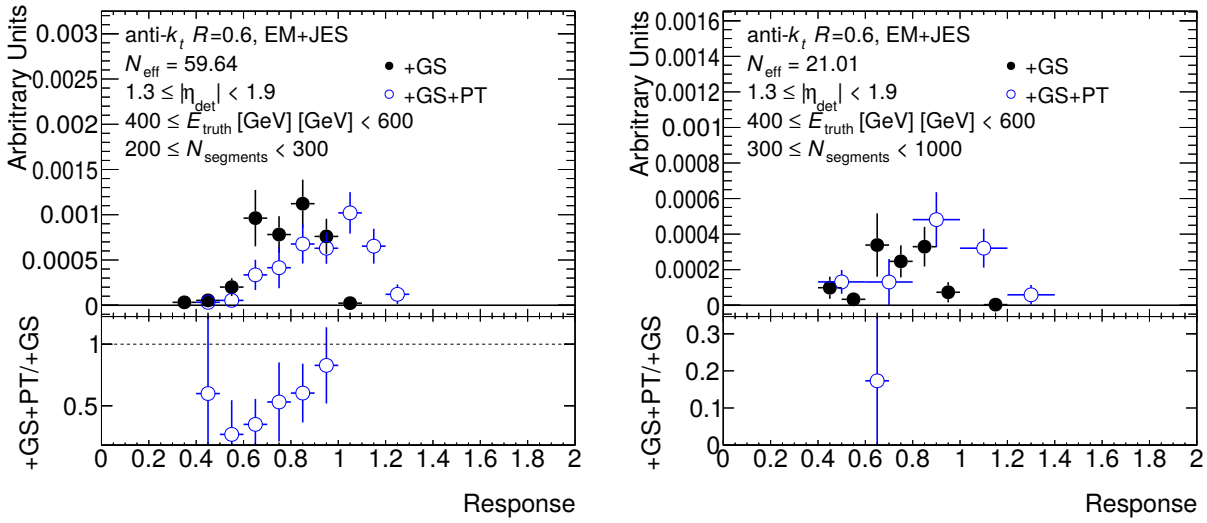


Figure 7.9: The jet energy response distributions in the $1.3 \leq |\eta_{det}| < 1.9$ region for jets with $400 \leq E_{truth}$ [GeV] < 600 in the bins $200 \leq N_{segments} < 300$ (left) and $300 \leq N_{segments} < 1000$ (right), before and after the application of the jet punch-through correction. The bottom plots show the ratio of events after to before the correction.

In the $1.9 \leq \eta_{det} < 2.7$ region, low statistics prevent any clear conclusions from being made, as was found when investigating the effect of jet punch-through on the jet energy resolution in Figure 4.11.

Figure 7.10 shows the jet energy resolution when integrating over $N_{segments}$, defined in terms of the RMS, and extracted as σ from a Gaussian fit to the response distribution performed using the procedure described in Section 5.4.

The jet punch-through correction is observed to improve the jet energy resolution. When

looking at the resolution in terms of the RMS, the improvements clearly grow with E_{truth}, and are largest in the $1.3 \leq \eta_{det} < 1.9$ region, reaching up to $\sim 1.6\%$. This is consistent with the shift in quantile positions observed in Section 7.1.2. The improvement in the resolution extracted from the fit can be seen to follow the same trend in general, however the improvement observed is smaller. This is again due to the fit only being performed for the $\mu \pm 1.6\ \sigma$ core of the response distribution, preventing it from adequately capturing the full effect of the correction on the low response tail. In addition, although no improvement in the resolution could be seen in Figure 7.8 for jets in the $1.9 \leq \eta_{det} < 2.7$ region, the improved statistics when integrating over $N_{segments}$ makes it possible to see a clear improvement in Figure 7.10.

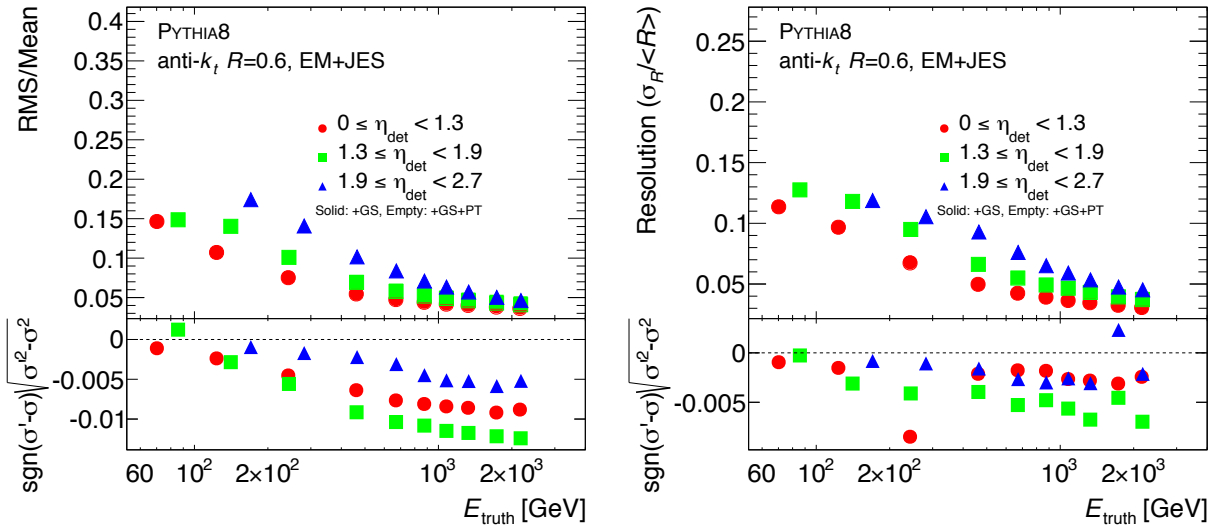


Figure 7.10: The RMS($\mathcal{R}$) divided by the mean histogram response (left) and the jet energy resolution extracted from the fit (right) as a function of E_{truth} in each of the three η_{det} regions, before (solid circles) and after (empty circles) the application of the jet punch-through correction. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

7.2 Impact on Dijet Mass Resonances search

One of the simplest searches that can be constructed is a search for high mass resonances in the invariant dijet mass (m_{jj}) distribution in dijet events, where m_{jj} is the invariant mass

of the vectorial sum of the four-momenta of the two highest p_{T} jets in the event:

$$m_{jj} = \sqrt{m_{j1}^2 + m_{j2}^2 + 2(E_{j1}E_{j2} - |p_{j1}||p_{j2}|\cos(\theta_{j1j2}))} \tag{7.1}$$

where m_{j1}, E_{j1}, and $|p_{j1}|$ are the mass, energy and magnitude of the three-momentum of the highest p_T jet in the event ($j1$), whilst m_{j2}, E_{j2}, and $|p_{j2}|$ are the mass, energy and magnitude of the three-momentum of the second highest p_T jet in the event ($j2$), and θ_{j1j2} is the angle between the three-momenta of the highest and second highest p_T jets in the event.

The hypothesis for the search is that many new physics models describe new particles or excitations created as S-channel resonances that can produce dijet final states. Processes from Quantum Chromodynamics produce a smooth, rapidly falling, m_{jj} spectrum. Thus, if the resonances produced by the new particles or excitations are sufficiently narrow, they will appear as bumps in the dijet mass distribution over the smooth background, that can be searched for.

This search relies on the jet energy scale being well calibrated at high energy, the region of phase space most affected by jet punch-through. As a result, an important test of the jet punch-through correction is to assess the impact of the correction on the obtained dijet mass spectrum for this analysis.

Reference [3] describes the dijet mass resonances search in detail and shows the final limits set on the existence of new particles when using the full 2012 ATLAS dataset, amounting to 20.3fb^{-1} of data.

7.2.1 Overview of analysis

The analysis follows the following steps:

1. *Preparation of invariant dijet mass distribution:* This is formed in data and Monte Carlo by selecting dijet events and constructing the m_{jj} distribution.

2. *Search Phase*: This stage involves searching for resonances above the QCD background in the m_{jj} distribution. The expected background is defined by a four-parameter fit function of the form [3]:

$$f(x) = p_1(1-x)^{p_2}x^{p_3+p_4\ln x} \tag{7.2}$$

This function is fit to the m_{jj} distribution obtained from data in Step 1. Statistical tests are then performed in order to find the largest bump and calculate the significance in terms of the deviation of the bump from the background fit. A bump with 5σ significance would indicate the discovery of a new particle. This technique and its sensitivity to certain benchmark models is tested in Monte Carlo simulation, after which the technique is further validated using a random sub-set of the full dataset (amounting to 5.07fb^{-1}).

3. *Limit setting phase*: In the event no significant deviation from background is found, *limits* (constraints) on New Physics phenomena are set using the current data. This involves using simulated signals for New Physics models as benchmarks. Limits are set for two scenarios: (1) New Physics benchmark models and (2) Generic resonances in the form of Gaussian and Breit-Wigner signals.

The impact of the correction on the dijet mass distributions and the excited quark benchmark signal samples is assessed.

7.2.2 Jet and event selection

The analysis uses the nominal Pythia8 MC samples, and 20.3fb^{-1}of data. It is performed using anti-k_t R = 0.6 jets. Events were required to pass the data quality requirements outlined in Section 2.4. In addition, events were rejected if either of the leading jets, or any other jet with p_{T} greater than 30% of the p_{T} of the sub-leading jet, fail the jet cleaning requirements in Section 3.6.

In order to select for well-balanced dijet events, additional kinematic cuts were applied. The event was required to have at least two jets, with at least two tracks coming from the primary vertex (as defined in Section 2.2.2). The rapidity, y [2] of two leading jets were required to be within $|y| < 2.8$. The two leading jets were required to have $p_{\mathrm{T}} > 50$ GeV, and have a combined invariant dijet mass of $m_{jj} > 250$ GeV. Finally, events must satisfy $|y^*| = \frac{1}{2}|y_{lead} - y_{sublead}| < 0.6$, where y_{lead} and $y_{sublead}$ are the rapidities of the leading and subleading jet respectively. The y^* cut favours central collisions where the dijet mass is expected to be largest, helping to enhance the signature of exotic phenomena.

No trigger selection was applied to Monte Carlo. In data, in order to increase the number of events in the dijet mass distribution by an order of magnitude, events are collected using a combination of two single jet trigger streams - the *normal* stream, and the *delayed* stream. The delayed trigger stream includes events that were recorded after the normal stream due to constraints on the reconstruction time.

7.2.3 Impact on simulated excited quark signal samples

One such New Physics benchmark model tested by the dijet mass resonances search is the existence of excited u and d quarks (q^*) that are one possible manifestation of quark compositeness. Details on the simulation of these samples can be found in Reference [3].

Example signal m_{jj} distributions for q^* with masses $m_{q^*} = 1250$ GeV and $m_{q^*} = 3750$ GeV before (GSC) and after (PT) the application of the jet punch-through correction are shown in Figure 7.11. Distinct resonances in the m_{jj} distributions can be seen representing the excited quark particles. From the ratio, it is clear the punch-through correction appears to be increasing the number of events at high mass.

[2] $y = \frac{1}{2} \ln \frac{E+p_T}{E-p_T}$

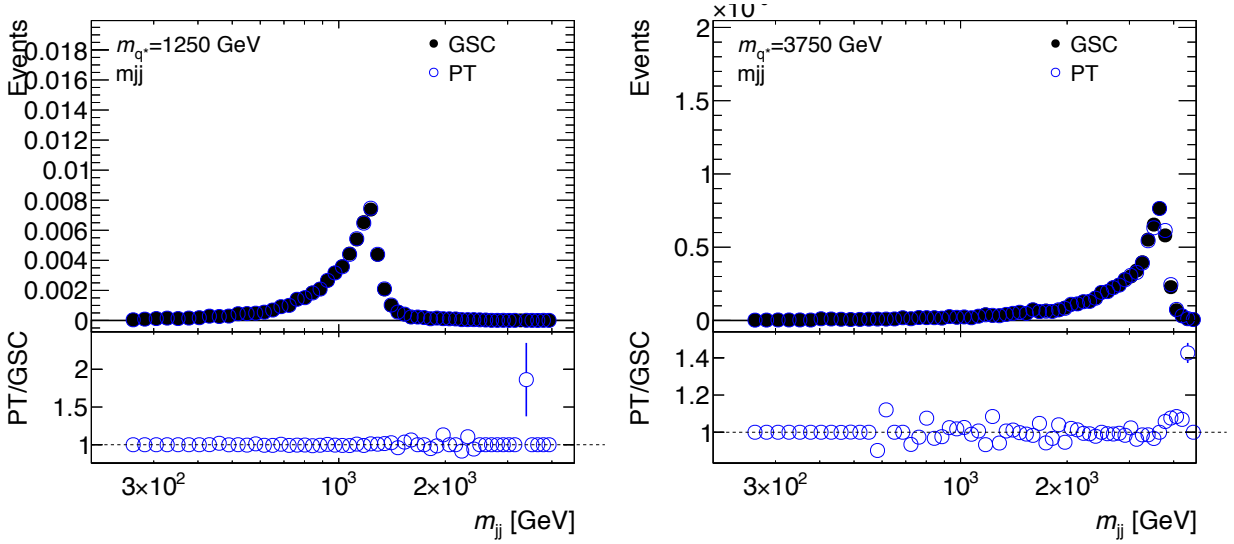


Figure 7.11: Signal m_{jj} distributions for q^* with masses $m_{q^*} = 1250$ GeV (left) and $m_{q^*} = 3750$ GeV (right) before (GSC) and after (PT) the application of the jet punch-through correction. Ratio plot shows the number of events after the punch-through correction divided by the number of events before the punch-through correction.

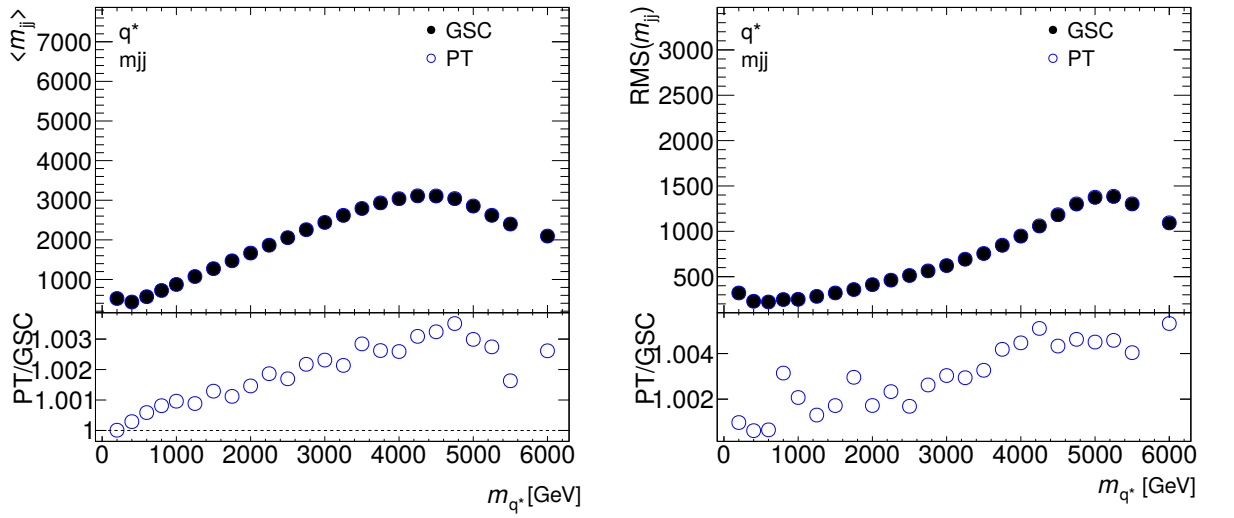


Figure 7.12: Mean dijet mass, $\langle m_{jj} \rangle$, (left) and RMS(m_{jj}) (right) for various m_{q^*} values, before (GSC) and after(PT) the punch-through correction is applied. Ratio shows $\langle m_{jj} \rangle$ (left) or RMS(m_{jj}) after applying the punch-through correction divided by the value before the correction is applied.

Figure 7.12 shows the mean dijet mass, $\langle m_{jj} \rangle$, (left) and RMS(m_{jj}) (right) for various m_{q^*} values, before (GSC) and after (PT) the punch-through correction is applied. It is clear from this figure that the correction increases $\langle m_{jj} \rangle$ and RMS(m_{jj}), with larger increases seen

at higher values of m_{q^*}. The increases seen are very small, between 0.1 and 0.3 % for m_{q^*}, and 0.2 and 0.5 % for RMS(m_{jj}). These increases are expected, as the correction will be pushing up m_{jj} in events affected by punch-through, increasing the higher mass m_{jj} tails as seen in Figure 7.11, increasing both $\langle m_{jj} \rangle$ and RMS(m_{jj}). The increase seen is small due to the limited $N_{segments}$ range to which the correction is applied.

7.2.4 Impact on dijet mass distribution in data

The final m_{jj} distribution (left), and the percentage of events above some m_{jj} threshold against the threshold (right) in data are shown in Figure 7.13 before (GSC) and after (PT) applying the jet punch-through correction. It is clear the correction increases the number of events in the high mass tail, suggesting it is doing as intended and increasing the number of high-mass events available that could potentially aid in the discovery of a new resonance. The fraction of events in the high mass tail can be seen to increase by up to 30%.

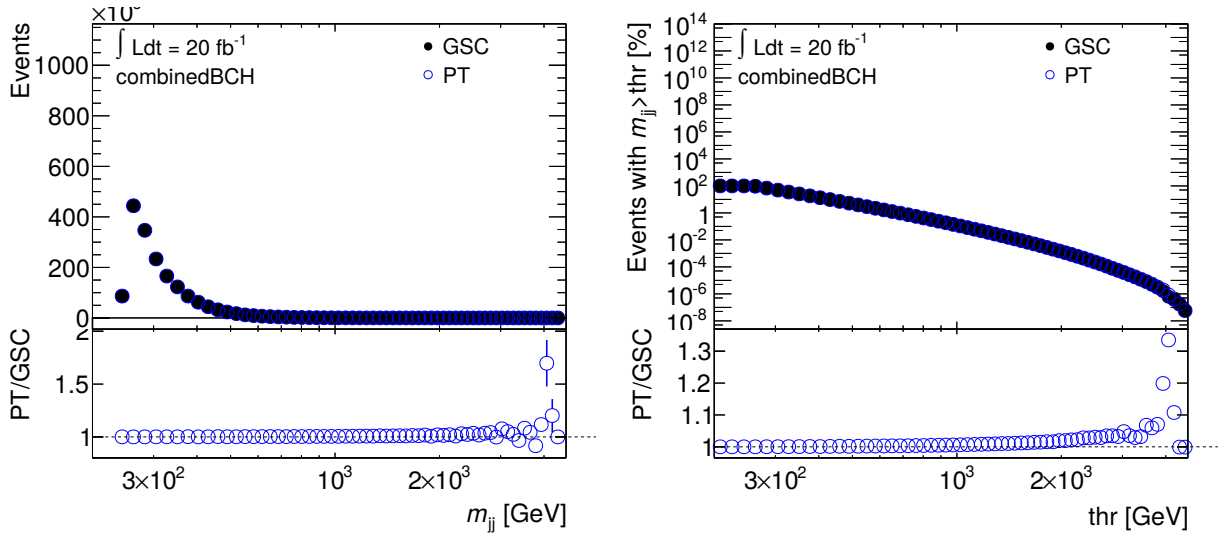


Figure 7.13: The final m_{jj} distribution (left), and the percentage of events above some m_{jj} threshold against the threshold (right) in data. Ratio shows number of events (left) and percentage of events above some m_{jj} threshold (right) after the correction divided by the value before the correction.

7.2.5 Future studies

Whilst the studies in Sections 7.2.3 and 7.2.4 allowed the impact of the jet punch-through correction on the distributions involved in the dijet mass resonances analysis to be assessed, better measures of the impact of the correction are available. One such measure would be to assess the impact of the correction on the significance obtained when performing the Search Phase of the analysis detailed in Section 7.2.1 on a QCD background sample injected with the simulated signal from Section 7.2.3. This would involve running the Search Phase for QCD background obtained from data injected with simulated signal with and without the jet punch-through correction applied, and then assessing the change in the significance of the resulting bump in the dijet mass distribution. A large improvement in the observed significance would indicate the correction could provide large benefits to the analysis. Due to time constraints, this study could not be performed, but is important to consider for the future.

Chapter 8

Final ATLAS 2012 Jet Energy Scale

The jet punch-through correction is an MC based jet energy response correction dependant on a jet property ($N_{segments}$), similar to the GS corrections described in Section 3.5.4. This enabled it to be easily intergated into the final 2012 ATLAS jet energy scale calibration as the fifth and final step in the GS correction sequence.

This chapter details the final 2012 ATLAS jet energy scale calibration after the inclusion of the jet punch-through correction derived in this thesis. Section 8.1 gives an overview of the additions to the calibration outlined in Section 3.5 that result in the final ATLAS 2012 jet energy scale. The uncertainties on the final ATLAS 2012 jet energy scale are shown in Section 8.2. Lastly, the final ATLAS 2012 jet energy resolution is detailed in Section 8.3.

As the jet punch-through correction was integrated into the GS corrections for the final ATLAS 2012 jet energy calibration, the EM+JES+GS and LCW+JES+GS scales in this chapter refer to the scales after the application of the jet punch-through correction.

8.1 Final ATLAS 2012 jet energy scale

The final ATLAS jet calibration scheme in 2012 was an extension of the initial calibration described in Section 3.5, involving the same Monte Carlo corrections with the addition of the

jet punch-through correction as part of the GS corrections, followed by an additional set of data-driven corrections derived using the full 2012 ATLAS dataset with $\int \mathcal{L}\, dt = 20\ \text{fb}^{-1}$. The data-driven corrections aim to reduce the differences observed in the jet energy scale response between data and the Monte Carlo used to derive the corrections. The final calibration scheme is outlined in Figure 8.1.

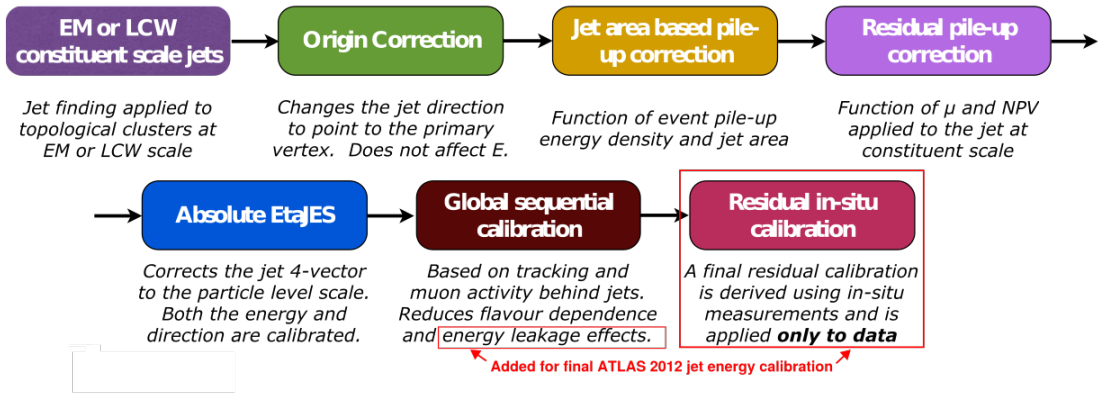


Figure 8.1: Final calibration chain for ATLAS 2012 jets [108].

8.1.1 Final 2012 Global Sequential corrections

The final 2012 GS corrections included the punch-through correction developed in this thesis as the fifth step, with the first four steps and their derivation procedure as described in Section 3.5.4. The derivation of the punch-through correction integrated into the GS derivation was identical to that described in Section 5. As shown in Figure 8.2, the final 2012 GS corrections improve the jet p_{T} resolution. The punch-through correction is applied from the upper green triangles to the lower blue triangles, however no clear improvements are seen from this figure. Improvements due to the jet punch-through correction are not visible in this plot due to a couple of main reasons:

- The p_{T} resolution in Figure 8.2 is shown as a function of $p_{\text{T}}^{\text{truth}}$, whereas the improvements seen in Section 7.1.3 were for the jet energy resolution as a function of E_{truth}.

- The resolution in Figure 8.2 is being measured as the standard deviation of a fit to the jet p_{T} response distribution, divided by the mean jet p_{T} response. This fit is performed in the same way as described in Section 5.4, only being done to the $\mu \pm 1.6\sigma$ core of the distribution. As a result, this method does not capture the full effect of the correction on the tails of the response distribution - the place where the punch-through correction has the most impact.

The reduction in differences in the response of quark-initiated and gluon-initiated jets also results in a reduction in the jet energy scale flavour based uncertainties (shown in Ref.[1]).

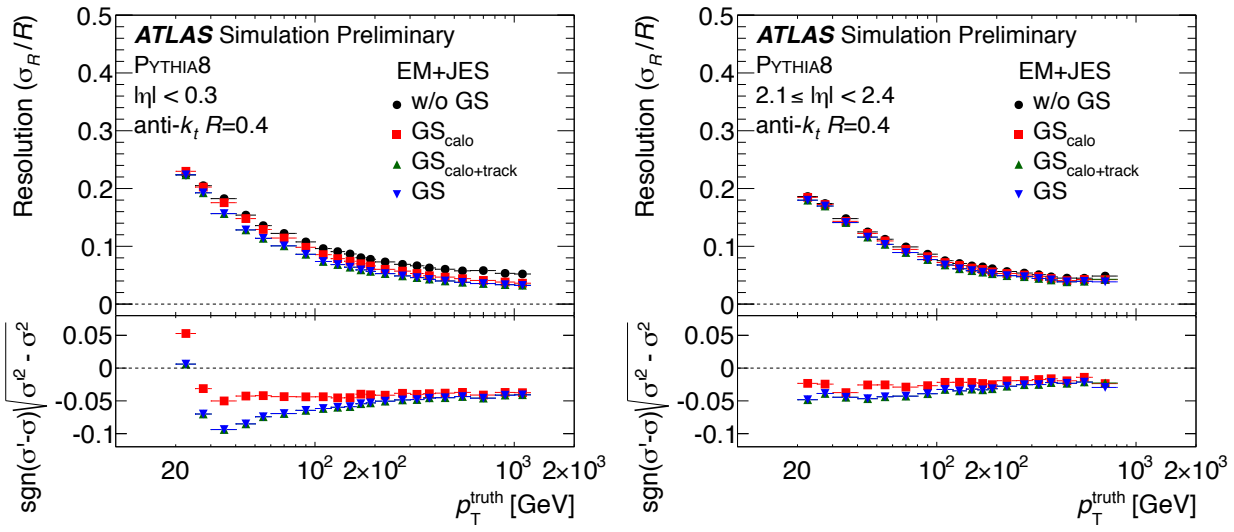


Figure 8.2: Jet transverse momentum resolution as a function of $p_{\mathrm{T}}^{\mathrm{truth}}$ in the nominal Pythia8 Monte Carlo sample for $|\eta| < 0.3$ (left) and $2.1 < |\eta| < 2.4$ (right). Markers are shown after the EM+JES calibration without GS corrections (black circles), with calorimeter based GS corrections only (red squares), with the four calorimeter and tracking based GS corrections only (upper green triangles) and all five GS corrections including the jet punch-through correction (lower blue triangles). Bottom plots show the difference in quadrature between the resolution with the GS corrections (σ') and the resolution at the EM+JES scale without the GS corrections (σ)[1].

The final 2012 GS corrections including the jet punch-through correction, were validated in-situ in data by exploiting the p_{T} balance in dijet events. The method used was independent of the one described in Chapter 5. The full data validation procedure for the final 2012 GS corrections is detailed in Appendix E. The final 2012 GS corrections were published in Ref. [1].

8.1.2 Residual in-situ calibration

The final step in the final 2012 calibration chain is the residual in-situ calibration, only applied to data. It is derived in data, and designed to improve the data-MC agreement.

In-situ techniques employing the balance of physics objects in the transverse plane are used to derive the corrections, whereby the balanced is assessed through the ratio:

$$\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{ref}\rangle_{data}/\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{ref}\rangle_{MC}, \tag{8.1}$$

where $p_{\mathrm{T}}^{\mathrm{jet}}$ and p_{T}^{ref} are the reconstructed p_{T} values of the jet and balance object (photons, Z bosons or other jets) respectively. The ratio represents the compatibility of the reconstructed p_{T} response in data ($\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{ref}\rangle_{data}$) with that in MC ($\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{ref}\rangle_{MC}$), and defines the residual correction. The closer the ratio is to 1, the better the data-MC agreement.

Firstly, the p_{T} balance in dijet events is exploited to apply an η-intercalibration to remove residual pseudorapidity dependent differences in the jet response. This is achieved by equalising the average p_{T} of forward jets ($0.8 \leq |\eta_{det}| < 4.5$) to the p_{T} of balancing jets in the central region ($|\eta_{det}| < 0.8$). These corrections are generally below 2% [133].

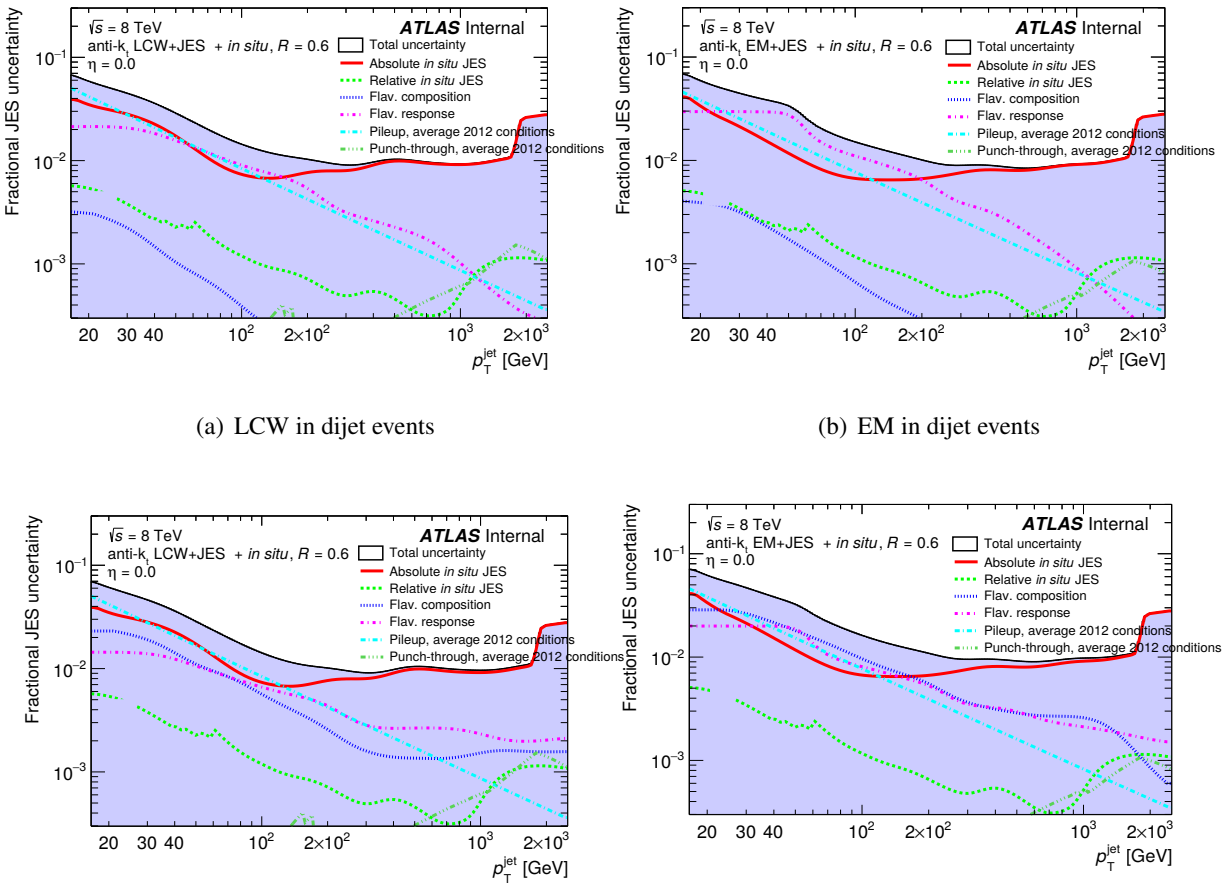
Following this, the balance of Z bosons, photons and well calibrated low-p_{T} jets (multi-jet) against an ordinary jet is used to derive in-situ JES corrections for jets with $|\eta_{det}| < 0.8$. This is done for jets with $20 \leq p_{\mathrm{T}} \leq 200$ GeV (Z +jet), $30 \leq p_{\mathrm{T}} \leq 800$ GeV (γ+jet) and $300 \leq p_{\mathrm{T}} \leq 1700$ GeV (multi-jet). These measurements are combined via a weighted average based on a $\tilde{\chi}^2$-minimisation to give a combined measurement of the response ratio, used for the residual in-situ correction [133], while also providing a source of uncertainty on the jet energy scale (see Section 8.2).

The ratio from these in-situ methods was cross-checked using a method whereby the jet energy scale is estimated from the response of single hadrons (*single hadron response*) [101, 134]. This method also provides an extension of the insitu-methods to higher energies ($p_{\mathrm{T}} > 1500$ GeV). The two methods were found to be in agreement.

8.2 Jet energy scale uncertainty

A jet energy scale uncertainty is derived for jets calibrated to the final 2012 jet energy scale with $p_{\mathrm{T}}^{\mathrm{jet}} > 15$ GeV and $|\eta_{det}| < 4.5$. The uncertainty is defined in terms of 66 independent parameters (nuisance parameters) originating from multiple sources. From the insitu-methods described in Section 8.1.2, the balance methods contribute 22 systematic and 34 statistical sources, related to methodology, detector description, physics modelling and a mix of detector and modelling effects. The single-hadron response contributes a single high p_{T} uncertainty at $p_{\mathrm{T}} > 1500$ GeV. The η-intercalibration contributes 2 sources - one systematic and one statistical. The pile-up corrections described in Section 3.5.2 contribute 4 sources - a μ, N_{PV} and p_{T} dependent uncertainty, and an uncertainty associated with the miss-modelling of ρ, where ρ is as defined in Section 3.5.2 and used in Equation 3.5. 2 sources are attributed to jet flavour differences, one associated with the degree of knowledge of the flavour composition of jets in an event, and another associated with generator differences in the response for gluon-initiated jets. The final source of uncertainty comes from the jet punch-through correction, determined in Chapter 6.

The final uncertainties on the EM+JES+GS and LCW+JES+GS scales after the application of the insitu calibrations are shown as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ for central jets in Figure 8.3, and for forward jets in Figure 8.4 [135]. The total jet energy uncertainty is dependent on how well the composition of an event is known, due to the increase in the flavour uncertainty in events with unknown composition. In the central region, the jet punch-through uncertainty can be seen to be small (< 0.2 %), and only visible in the high $p_{\mathrm{T}}^{\mathrm{jet}}$ region ($p_{\mathrm{T}}^{\mathrm{jet}} > 600$GeV), contributing a negligible amount to the overall uncertainty. In the forward region, the uncertainty can be seen to be larger, reaching up to ~ 0.6 % in the higher p_{T} region, but is still not the dominant uncertainty.



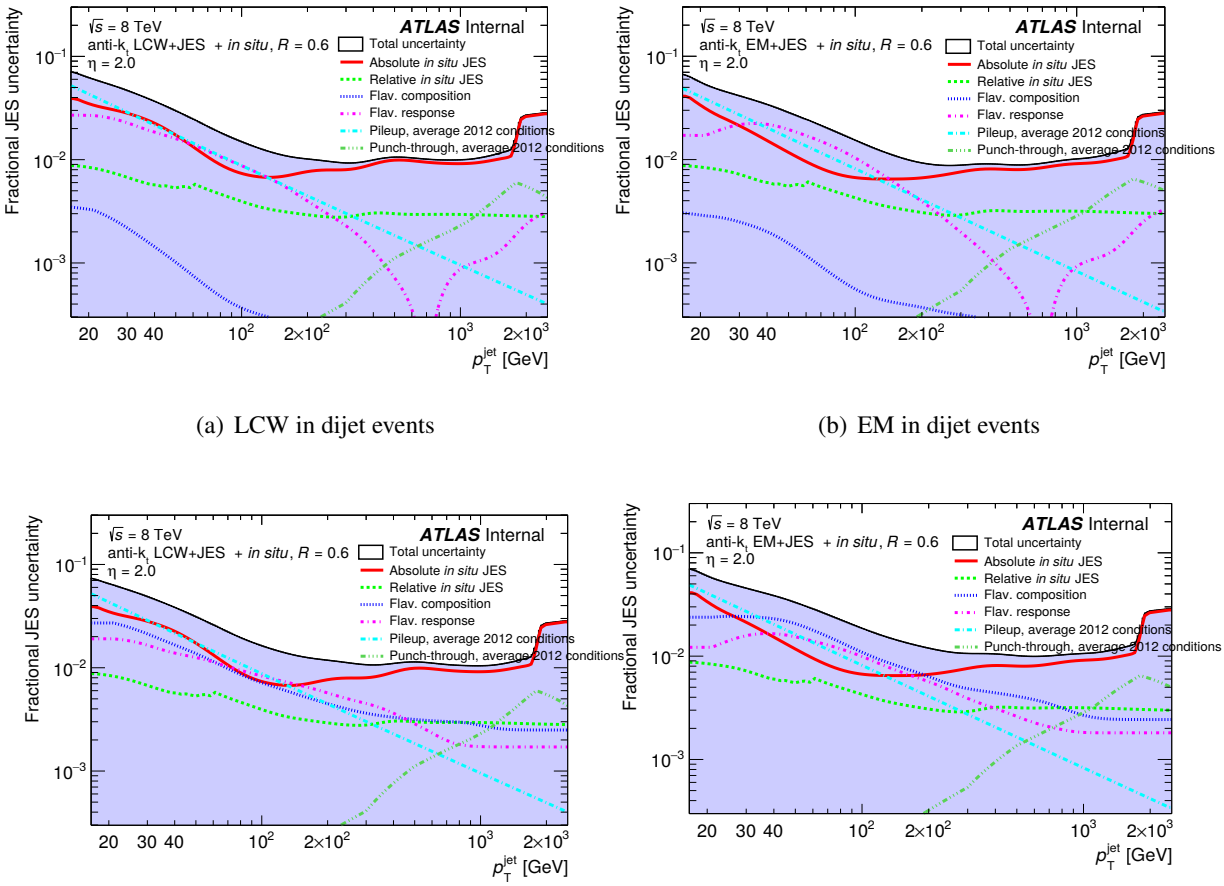
(a) LCW in dijet events

(b) EM in dijet events

(c) LCW in unknown composition events

(d) EM in unknown composition events

Figure 8.3: The total jet energy scale uncertainty as a function of p_{T} for central R = 0.6 jets. Two flavour compositions are shown, that for dijet events, where the quark/gluon composition is taken from Monte Carlo and an associated uncertainty from generator comparisons, and that for an unknown flavour composition (assuming 50:50 quark:gluon jets with a 100% uncertainty). Absolute in situ JES refers to the uncertainty arising from V+jet and multi-jet measurements whilst Relative in situ JES refers to the uncertainty arising from the dijet η-intercalibration. Punch-through refers to the uncertainty on the final (muon-based) stage of the global sequential correction. [108, 135]



(a) LCW in dijet events

(b) EM in dijet events

(c) LCW in unknown composition events

(d) EM in unknown composition events

Figure 8.4: The total jet energy scale uncertainty as a function of p_{T} for forward R = 0.6 jets. Two flavour compositions are shown, that for dijet events, where the quark/gluon composition is taken from Monte Carlo and an associated uncertainty from generator comparisons, and that for an unknown flavour composition (assuming 50:50 quark:gluon jets with a 100% uncertainty). Absolute in situ JES refers to the uncertainty arising from V+jet and multi-jet measurements whilst Relative in situ JES refers to the uncertainty arising from the dijet η-intercalibration. Punch-through refers to the uncertainty on the final (muon-based) stage of the global sequential correction. [108, 135]

8.3 Jet energy resolution

As mentioned briefly in Section 3.4, the energy of a reconstructed jet can fluctuate with respect to the initial energy of the initiating object. This occurs due to effects mentioned in Section 3.5 that affect the energy measurement of a jet, including statistical fluctuations inherent to the energy measurement in calorimeters, the finite jet area, pile-up radiation being captured within the jet, jet punch-through, etc. Most of these effects lead to a Gaussian

distribution, shown in Figure 3.6. An exception to this is jet punch-through, which affects the tails of the jet response, as shown in this thesis.

Fluctuations in the jet energy are measured in terms of the (fractional) jet energy resolution, parametrised as a function of three terms:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C, \tag{8.2}$$

where N parametrises the effect of noise (electronic and pile-up), S parametrises the stochastic effect arising from the sampling nature of the calorimeters and C is a p_{T} independent constant term. In MC, $\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}$ can be parametrised in terms of the width of the jet response distribution (see Section 3.4). In data, for the majority of the p_{T} spectrum, the resolution is determined from the width of the p_{T} balance of jets to well measured photons or Z bosons, and extended through higher p_{T} through dijet events, using the in-situ methods described in Section 8.1.2.

In both MC and data, the jet energy resolution is largest at low p_{T}, improving as p_{T} increases. The jet energy resolution values for EM+JES+GS (LCW+JES+GS) anti-k_t $R = 0.6$ jets are outlined in this paragraph. In MC, for jets with $|\eta_{det}| < 0.3$, the resolution is $\sim 18\%$ (21%) at $\mathrm{p}_T^{truth} = 30$ GeV, decreasing to 3% (3%) at $\mathrm{p}_T^{truth} = 1000$ GeV. The MC resolution displays $\sim 10\%$ ($\sim 5\%$) fluctuations as a function of η_{det}, with the best resolution observed at $2 \leq |\eta_{det}| < 3$. In data, at $p_{\mathrm{T}} \sim 30$ GeV, the combined resolution is $\sim 20\%$, decreasing to $\sim 4\%$ ($\sim 4\%$) at $p_{\mathrm{T}} \sim 1000$ GeV [108]. A fit to the resolution as a function of p_{T} was performed in data using the parametrisation in Equation 8.2. The extracted values for the parameters N, S and C are displayed in Table 8.1.

Jet punch-through effects are expected to contribute to the constant term, however at the time of writing, studies to test the magnitude of this contribution had not yet been performed. This is something that will be investigated in the future.

EM+JES $R = 0.6$			LCW+JES $R = 0.6$		
N	S	C	N	S	C
$\lvert\eta\rvert < 0.8$					
4.34 ± 0.93	0.67 ± 0.08	0.030 ± 0.003	5.50 ± 0.99	0.66 ± 0.12	0.026 ± 0.004
$0.8 < \lvert\eta\rvert < 1.2$					
4.06 ± 0.93	0.76 ± 0.10	0.031 ± 0.003	5.40 ± 0.98	0.78 ± 0.15	0.032 ± 0.005
$1.2 < \lvert\eta\rvert < 2.1$					
3.96 ± 0.91	0.56 ± 0.14	0.042 ± 0.007	5.7 ± 1.0	0.62 ± 0.16	0.031 ± 0.006
$2.1 < \lvert\eta\rvert < 2.8$					
3.41 ± 0.84	0.48 ± 0.27	0.049 ± 0.012	5.2 ± 1.0	0.51 ± 0.38	0.028 ± 0.019

Table 8.1: The fitted values of N, S and C in the combined fit to the jet energy resolution for $R = 0.6$ jets at both the EM+JES and LCW+JES scales [108].

Summary, Conclusions and Outlook

Measuring the kinematics of hadronic final states is important not just for precision measurements of QCD predictions, but also for searches for New Physics, such as the Dijet Mass Resonances search described in Section 7.2. Hadronic jets are indispensable tools in this process, allowing measurements of the kinematic properties of hadronic showers produced by hadronic final states to be made. As a result, in order for searches and precision measurements involving hadronic final states to be successful, the energy of hadronic jets must be accurately measured. One key consideration when measuring the energy of hadronic showers, and thus jets, is longitudinal hadronic shower leakage outside of the calorimeter, also known as jet punch-through. This is a stochastic phenomenon that is dependant on the energy of the shower initiating parton and the amount of material the shower passes through. Studies using test beam data showed longitudinal shower leakage decreased the measured mean shower energy by 10%, worsening the resolution by up to 30% (shown in Section 1.3.3).

In order to study the impact of jet punch-though on the initial jet energy scale calibration described in Section 3.5, Section 4.1 established that the number of muon segments found behind a jet in the muon spectrometer was a good indicator of jet punch-through. A punch-through jet was thus defined to have $N_{segments} \geq 20$ in order to distinguish between b-jets and real jet punch-through. The fraction of punch-through jets was confirmed in the same section to be larger for jets with higher energy passing through regions of the detector with less material, reaching up to 10%.

Studies in Section 4.2 then showed that even after calibrating jets with the EM+JES+GS

(LCW+JES+GS) calibration scheme discussed in Section 3.5, jet punch-through still had an impact on the jet energy response and jet energy resolution. The jet energy response was observed to decrease by up to 30% as $N_{segments}$ increased, with the decrease smaller in the higher energy region, but corresponding to a larger energy loss. An increase in the fraction of events in the low response tail due to jet punch-through was shown to cause the jet energy resolution to worsen by up to 15%, consistent with the test beam data. The resolution degradation as a function of $N_{segments}$ was worse for higher energy jets, as expected. However, as a function of the percentage of energy lost, the degradation in the energy resolution was shown to be independent of jet energy. In addition, comparing the impact of punch-through on different jet flavours showed that light-quark jets are the most sensitive to energy losses due to jet punch-through, and also experience the largest degradation in jet energy resolution. These observations motivated the need for a correction, the jet punch-through correction, to remove the effect of jet punch-through on the initial 2012 ATLAS jet energy scale.

Chapter 4 outlined the derivation procedure used to derive the jet punch-through correction from Monte Carlo simulation. This was achieved by extracting the mean response in bins of η_{det}, E_{truth} and $N_{segments}$ using a Gaussian fit, followed by the use of Gaussian kernels to obtain a smooth correction function as a function of $N_{segments}$ and E_{truth} in bins of η_{det}, with values ranging from 0.6 to 0.98. The largest corrections were observed in the central region $|\eta_{det}| < 1.3$ in the low energy, high $N_{segments}$ region. The correction was validated in data in Chapter 6 by assessing the dijet imbalance in a dijet system with a punch-through jet balancing a fully contained jet, as a function of $N_{segments}$ of the punch-through jet. The dijet imbalance due to jet punch-through was observed to be consistent between data and Monte Carlo to within a few percent, with the correction largely removing the imbalance observed as a function of $N_{segments}$. Two sets of systematic uncertainties on the correction were derived as the difference in the in-situ response between data and Monte Carlo - frozen uncertainties and extrapolated uncertainties. The frozen uncertainties were observed to be lower, and taken to be the nominal uncertainties on the correction, growing with $N_{segments}$, but decreasing with $p_{\mathrm{T}}^{\mathrm{jet}}$, reaching 8% in the $1.9 \leq |\eta_{det}| < 2.7$ region. These uncertainties had near negligible impact on the overall jet energy scale uncertainties (Section 8.2).

The performance of the correction was assessed in Chapter 7, where the dependence of the jet energy response on $N_{segments}$ was observed to be largely removed by the correction. The correction was also observed to shrink the low response tails, pushing the negative 1 and 2 σ quantile positions to higher response values, resulting in the jet energy resolution being improved by up to 1.6% in the highest energy regions. However, the correlation between the increasing energy resolution and an increase in $N_{segments}$ was not completely removed, suggesting the effects of jet punch-through on the jet energy resolution were not completely negated. These improvements were shown to have a positive impact on the dijet mass distribution, used by the dijet mass resonances search to look for new, high-mass resonances, whereby the correction increased the fraction of events in the high mass tail in data by over 30% in some regions. The mean and RMS of signal dijet mass distributions were also increased. As a result of the positive impact of the jet punch-through correction, it was integrated into the final ATLAS 2012 jet energy scale as the fifth and final step of the GS corrections, mentioned in Chapter 8.

In conclusion, this thesis described the derivation, data validation, determination of systematic uncertainties and assessment of the performance of a new correction designed to remove the effects of longitudinal hadronic shower leakage, also known as jet punch-through, from the ATLAS jet energy scale and jet energy resolution. $N_{segments}$ was identified as an observable linked to jet punch-through, which the jet energy response was shown to be sensitive to. The correction derived to remove this dependence reached values of 0.6% (a 40% correction) in the lowest E_{truth} regions with an uncertainty of up to 8% in the $1.9 \leq |\eta_{det}| < 2.7$ region that had near negligible impact on the overall jet energy scale uncertainty. The correction was shown to remove the dependence of the jet energy response on $N_{segments}$, improving the jet energy resolution by up to 1.6% in the highest energy regions. These effects consequently increased the fraction of events in the high mass tail in data by over 30%, whilst also increasing the mean and RMS of signal dijet mass distributions, useful in setting limits/constraints on the search for New Physics.

As a result of the work in this thesis, the jet punch-through correction was included in the

final 2012 ATLAS jet energy calibration as the fifth step in the Global Sequential calibration. Although only results for anti-k_t jets with R = 0.6 calibrated with the EM+JES+GS scheme are described in the main chapters of this thesis, similar results were found for anti-k_t jets with R = 0.4 and R = 0.6, calibrated with the EM+JES+GS and LCW+JES+GS schemes, resulting a correction derived for all four of these jet collections. Results for anti-k_t R = 0.6 jets calibrated with the LCW+JES+GS scheme are described in Appendix D.

In Run II, with the LHC operating at $\sqrt{s} = 13$ TeV, the jet punch-through correction will be even more crucial due to produced jets reaching higher energies with greater punch-through probabilities. At present, the correction is being derived using the procedure outlined in this thesis, however, as shown, the effects of jet punch-through on the jet energy resolution were not completely removed by the correction. In addition, only small improvements were observed in relation to the dijet mass resonances signal samples. These effects are largely a result of the $N_{segments}$ definition, and limiting the correction to $N_{segments} \geq 20$ due to the presence of b-jets in the low $N_{segments}$ region. Therefore in order to improve the performance of the correction as Run II progresses, $N_{segments}$ will be ghost associated to jets in an attempt to reduce backgrounds and thus allow the $N_{segments}$ cut to be reduced by further pin-pointing the b-jet threshold. Augmenting the correction with calorimeter information (see Appendix B) will further improve the correction, and allow an extension of the correction to regions of the detector with no muon spectrometer ($|\eta_{det}| > 2.7$).

As well as improving the correction, additional studies on the performance of the correction could also be performed. These include further studies on the impact of the jet punch-through correction on searches for New Physics, specifically looking at the impact on the sensitivity of searches and the limits set. It would also be useful to perform a more thorough analysis of the impact of jet punch-through on the constant term of the jet energy resolution, including a parametrisation of how energy losses due to jet punch-through impact the overall jet energy resolution. This could be compared to Equation 1.9, and also be used to gain a better understanding of the impact of longitudinal hadronic shower leakage on jet energy measurements in general.

Appendix A

Backgrounds to $N_{segments}$

This appendix discusses the background contributions to $N_{segments}$.

A.1 Muons

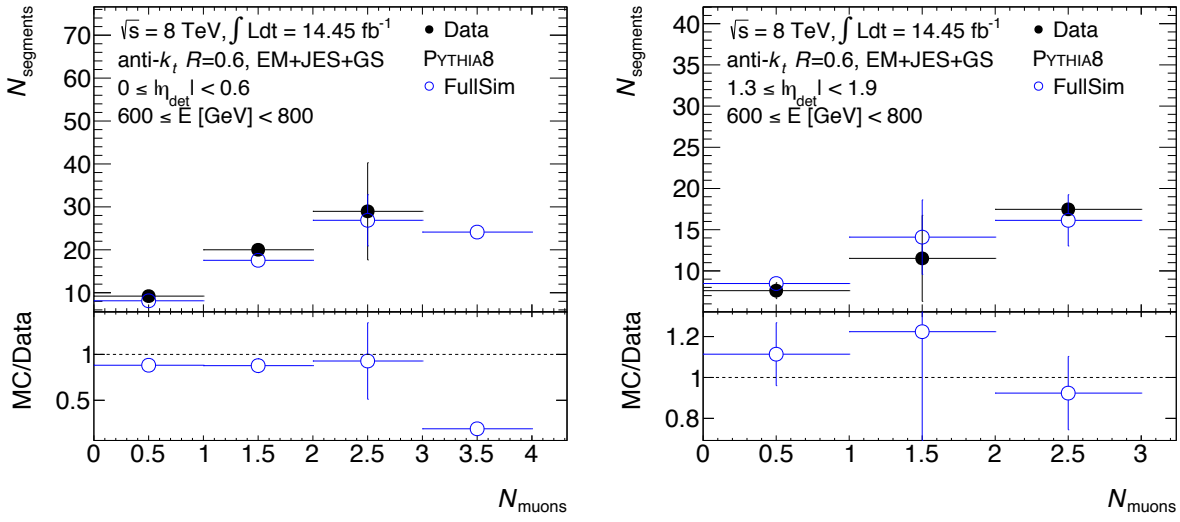


Figure A.1: Data/MC comparison of $\langle N_{segments} \rangle$ vs N_{muons} for jets with $600 \leq E[GeV] < 800$ in the region $0 \leq \eta_{det} < 0.6$ (left) and $1.3 \leq \eta_{det} < 1.9$.

Unsurprisingly, muons can affect the number of muon segments reconstructed behind a jet in the muon spectrometer. Figure A.1 shows the average number of segments reconstructed behind a jet as the number of muons clustered into the jet (N_{muons}) increases. It is clear

that $N_{segments}$ increases with N_{muons}, as one may expect. This increases is consistent in all η_{det} regions. A clear correlation between $N_{segments}$ and the jet energy response was still observed however, suggesting the background from muons does not have a large impact on the correlation between $N_{segments}$ and the jet energy response.

A.2 Close-by jets

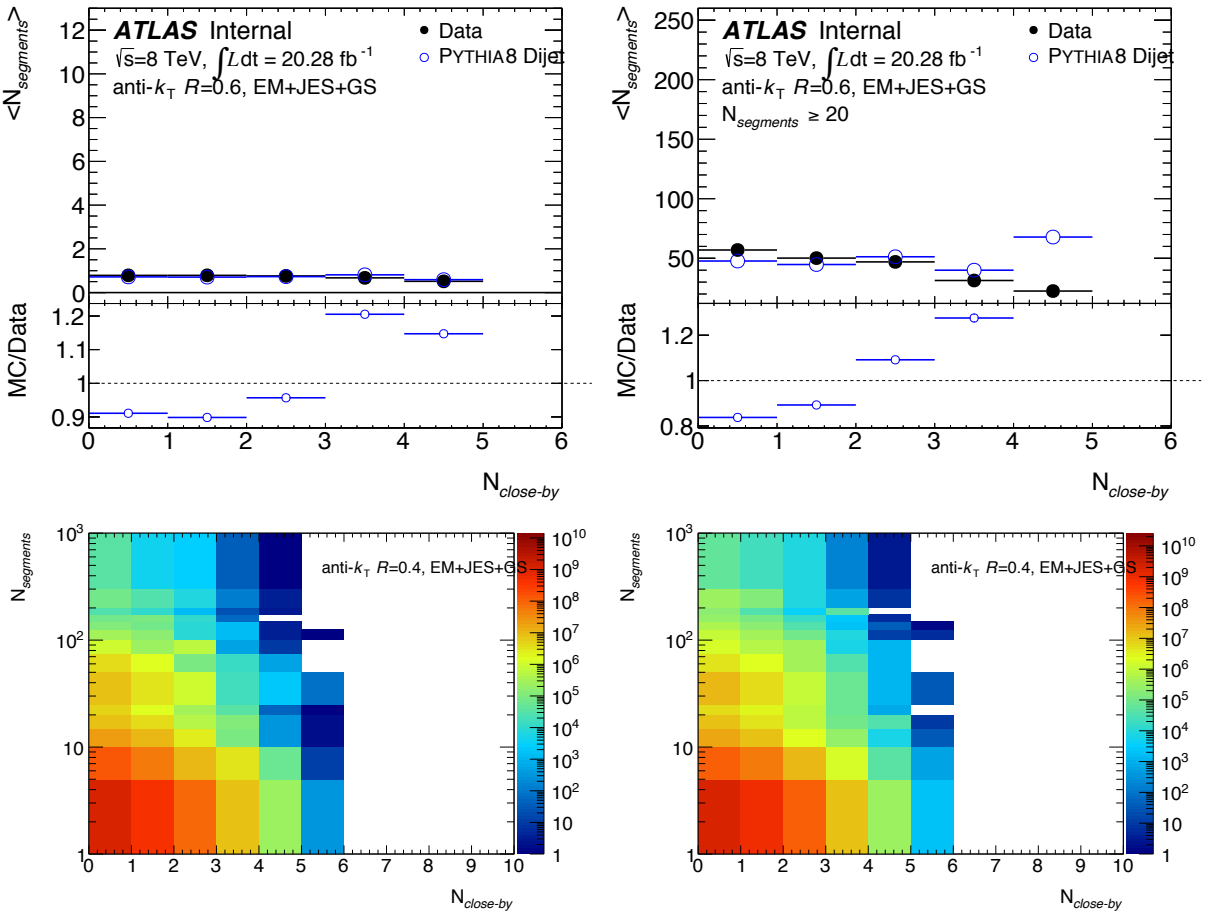


Figure A.2: Data/MC comparison of $\langle N_{segments} \rangle$ (top) and $N_{segments}$ vs $N_{close-by}$ for jets: integrated over all $N_{segments}$ (left) and with $N_{segments}$ *ge* 20 (right).

During reconstruction, a geometric matching of segments to anti-k_T R=0.6 EM+JES jets is done. Any jets close-by to the matching jet would also have a small punch-through probability, and could thus affect the number of muon segments clustered into the MuonSp-Container object. To investigate this, figure A.2 shows $N_{segments}$ as a function of the number

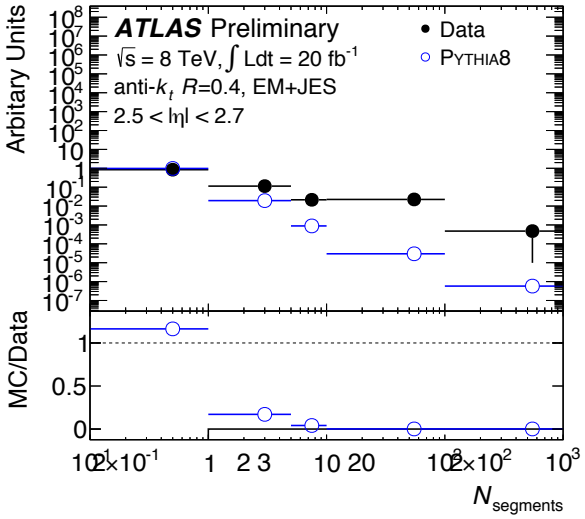
of close-by jets ($N_{close-by}$). A close-by jet is any jet with $p_T > 50\ GeV$ found to be within $\Delta R < 1.5 * R_{jet}$[1] where R_{jet} is the jet radius. The selection used for these studies corresponds to the di-jet selection described in Sec. 6.2.

From figure A.2 (a) we very little dependency of $N_{segments}$ on $N_{close-by}$ when integrating over $N_{segments}$. In figure A.2 (b) we see a slight negative correlation in data when looking at jets with $N_{segments} \geq 20$, however MC shows no clear dependency. Both plots suggest $N_{segments}$ is relatively unaffected by the presence of close-by jets. This is most likely due to the probability of having two close-by jets within the same event punching through the calorimeter being incredibly small.

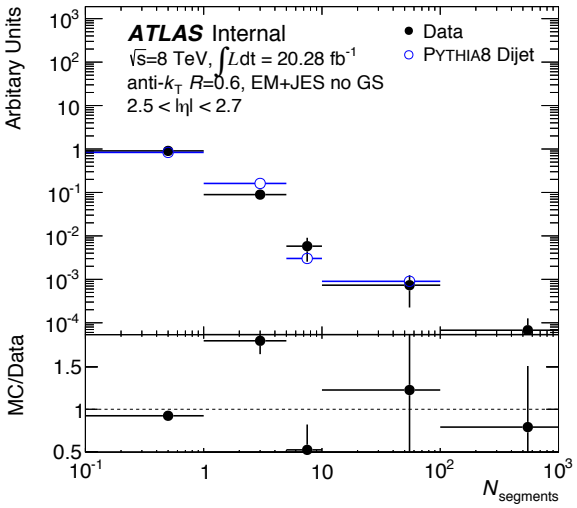
A.3 Modelling at edge of muon spectrometer

Figure A.3 shows $N_{segments}$ distributions for jets with $2.5 < |\eta| < 2.7$ in data and Pythia8 Monte Carlo simulation. Four different jet types are shown. The statistics in data is small, but a general trend of having more $N_{segments}$ in simulation is observed.

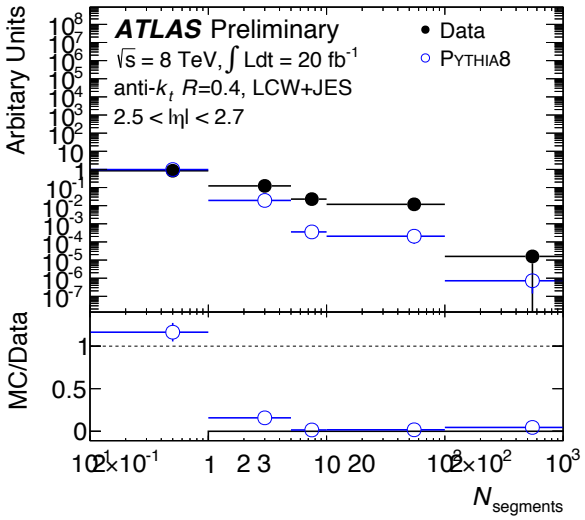
[1] ΔR is the sum in quadrature of the difference in pseudorapidity and the difference in azimuth: $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$



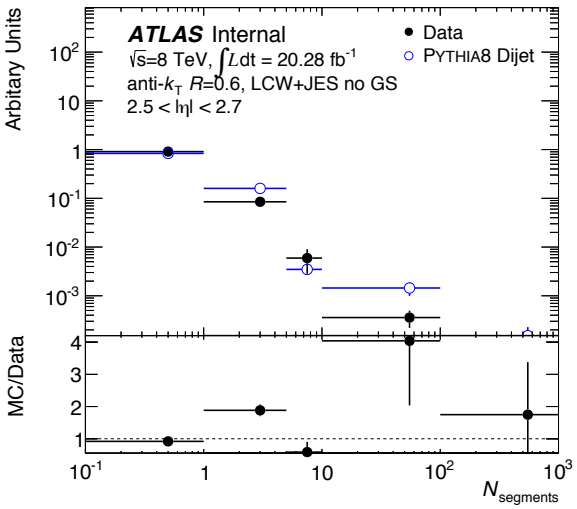
(a) anti-k_t $R = 0.4$, EM+JES



(b) anti-k_t $R = 0.6$, EM+JES



(c) anti-k_t $R = 0.4$, LCW+JES



(d) anti-k_t $R = 0.6$, LCW+JES

Figure A.3: Distribution of $N_{segments}$ for jets inclusive in p_T and $2.5 < |\eta| < 2.7$ in the data and PYTHIA8 Monte Carlo simulation. Four different jet collections are shown.

Appendix B

Calorimeter Studies

This appendix describes studies utilising calorimeter information on the energy profile of jets in the calorimeter to study jet punch-through, mentioned briefly in Section 4.1.

B.1 Calorimeter variables sensitive to jet punch-through

A jet that begins showering later in the calorimeter is more likely to punch-through. Therefore the following variables that relate the jet's energy deposition profile in the calorimeter can be used to study jet punch-through:

1. $\mathbf{f_{Tile2}}$**:** Fraction of the jet's energy deposited in the final (third) layer of the hadronic barrel calorimeter. A large f_{Tile2} could indicate a punch-through jet. This variable is limited to the region $|\eta_{det}| \leq 1.4$.

2. $\mathbf{f_{ECAL}}$**:** Fraction of the jet's energy deposited in the electromagnetic calorimeter. A small f_{ECAL} could indicate a punch-through jet. This variable is limited to $|\eta_{det}| \leq 3.2$.

3. $\mathbf{S_{MAX}}$**:** The calorimeter layer in which the maximum energy per unit interaction length was deposited[1]. The correspondence between the sampling layer numbers

[1]The energy per unit interaction length in calorimeter layer L, $E(L)$, is calculated as $E(L)[GeV/\lambda_I] = \frac{E_{layer}[GeV]}{d_{layer}[\lambda_I]}$, where E_{layer} is the energy deposited in the layer, and d is the depth of the calorimeter layer L in units of interaction length (λ_I). Please see Table B.1 for the depth of each calorimeter layer used.

and calorimeter layers represented by S_{MAX} and are displayed in Table B.1. Currently this variable is only studied in the region $|\eta_{det}| < 0.6$, taking values $\mathrm{S_{MAX}} = [0, 1, 2, 3, 12, 13, 14]$.

Sampling Layer Number	**Calorimeter Layer**
0	Electromagnetic Barrel PreSampler.
1	Electromagnetic Barrel First Layer.
2	Electromagnetic Barrel Second Layer.
3	Electromagnetic Barrel Third Layer.
4	Electromagnetic Endcap PreSampler.
5	Electromagnetic Endcap First Layer.
6	Electromagnetic Endcap Second Layer.
7	Electromagnetic Endcap Third Layer.
8	Hadronic Endcap First Layer.
9	Hadronic Endcap Second Layer.
10	Hadronic Endcap Third Layer.
11	Hadronic Endcap Fourth Layer.
12	Tile Barrel Calorimeter First Layer (Cells A1-A10 in figure 2.9).
13	Tile Barrel Calorimeter Second Layer (Cells BC1-BC9 in figure 2.9).
14	Tile Barrel Calorimeter Third Layer (Cells D0-D3 in figure 2.9).
15	Tile Gap Cell C10 (see figure 2.9).
16	Tile Gap Cell D4 (see figure 2.9).
17	Tile Gap Scintillators (Cells E1-E4 in figure 2.9).
18	Extended Tile Barrel Calorimeter First Layer (Cells A12-A16 in figure 2.9).
19	Extended Tile Barrel Calorimeter Second Layer (Cells B11-B15 in figure 2.9).
20	Extended Tile Barrel Calorimeter Third Layer (Cells D5-D6 in figure 2.9).
21	FCAL First Layer.
22	FCAL Second Layer.
23	FCAL Third Layer.

Table B.1: Table summarising correspondence between sampling layer number and calorimeter layer. Please refer to Section 2.2.3 for more information on the ATLAS calorimeter structure.

B.2 Comparisons with data

Figure B.1 compares the shape of the f_{ECAL} (top), f_{Tile2} (middle) and S_{MAX} (bottom) distributions between data and MC, for jets with $100 \leq \mathrm{E} < 150$ GeV (left) and $800 \leq \mathrm{E} < 1000$ GeV (right) in the region $|\eta_{det}| < 0.6$.

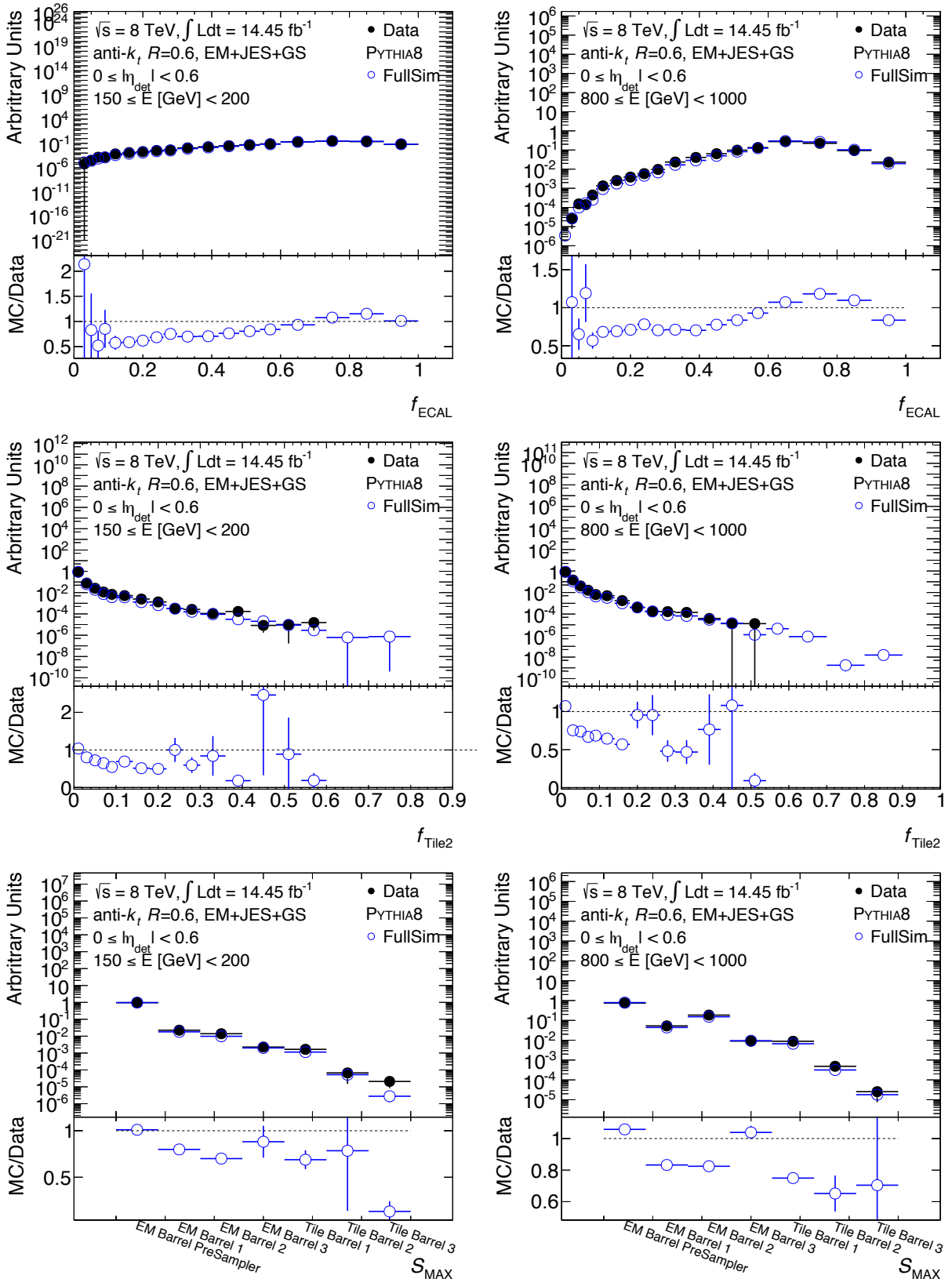


Figure B.1: Distributions of the f_{ECAL} (top), f_{Tile2} (middle) and S_{MAX} (bottom) variables for jets with $100 \leq E < 150$ GeV (left) and $800 \leq E < 1000$ GeV (right) in the region $|\eta_{det}| < 0.6$, in the data and the Pythia8 Monte Carlo simulation. Bottom plots display the MC/Data ratio.

For f_{ECAL}, the agreement is within 50% at low fractions ($f_{ECAL} < 0.2$), with the agreement improving as f_{ECAL} increases, and discrepancies observed to be similar for both energy regions. f_{Tile2} displays good agreement at very low values, with increased discrepancies observed as f_{Tile2} increases. Discrepancies in f_{Tile2} are slightly reduced in the $800 \leq E < 1000$ GeV bin compared with the $100 \leq E < 150$ GeV bin. S_{MAX} displays good agreement in the ECAL presampler and the third ECAL layer. Agreements for the other ECAL layers are within 25% for jets with $100 \leq E < 150$ GeV and 20% for jets with $800 \leq E < 1000$ GeV. Data and MC agree to within 30% in the HCAL layers in both energy regions, with the exception of the third HCAL layer for jets with $100 \leq E < 150$ GeV, where large differences are observed.

Overall the agreement is observed to be satisfactory, with no drastic differences in shape observed between data and MC.

B.3 Correlations with E_{truth}

In order to determine how the calorimeter variables mentioned in Section B.1 correlate with jet punch-through, it is important to determine how these variables depend on the jet energy and η_{det}. Figure B.2 displays f_{ECAL} (top-left), f_{Tile2} (top-right) and S_{MAX} (bottom) variables vs E_{truth}. Each variable is displayed in the η_{det} regions it is valid within.

It is clear that as E_{truth} increases, f_{ECAL} decreases in all η_{det} regions. This is the correlation one may expect to see if f_{ECAL} is an indicator for jet punch-through, whereby the jet punch-through probability increases with E_{truth}, and a punch-through jet is expected to have a lower f_{ECAL} due to the shower developing later in the calorimeter, thus f_{ECAL} is expected to decrease as E_{truth} increases, as observed. f_{ECAL} is observed to be consistent in the regions $|\eta_{det}| < 1.3$ and $|\eta_{det}| > 1.3$ for $E_{truth} > 90$ GeV, with a lower f_{ECAL} observed in the $|\eta_{det}| > 1.3$ region. Thus f_{ECAL} does not seem to clearly correlate with detector material, which is desired for a variable useful in studying jet punch-through.

f_{Tile2} increases with E_{truth} in the $|\eta_{det}| < 1$ region. However in the $1 \leq |\eta_{det}| < 1.3$

region the correlation is not as clear, with f_{Tile2} decreasing in the region $40 \leq E_{truth} < 200$ GeV, however f_{Tile2} then increases with E_{truth} for jets with $E_{truth} \geq 200$ GeV. Thus in the high E_{truth} region, where the jet punch-through probability is expected to be highest, we see f_{Tile2} increases, which suggests the increasing f_{Tile2} could possibly be linked to an increased punch-through probability at higher E_{truth}.

S_{MAX} is only shown for $|\eta_{det}| < 0.6$. It is clear that for $E_{truth} > 40$ GeV, as S_{MAX} increases as E_{truth} increases, consistent with the idea that S_{MAX} could correlate with jet punch-through, with the shower maximum moving to outer detector layers as E_{truth} increases.

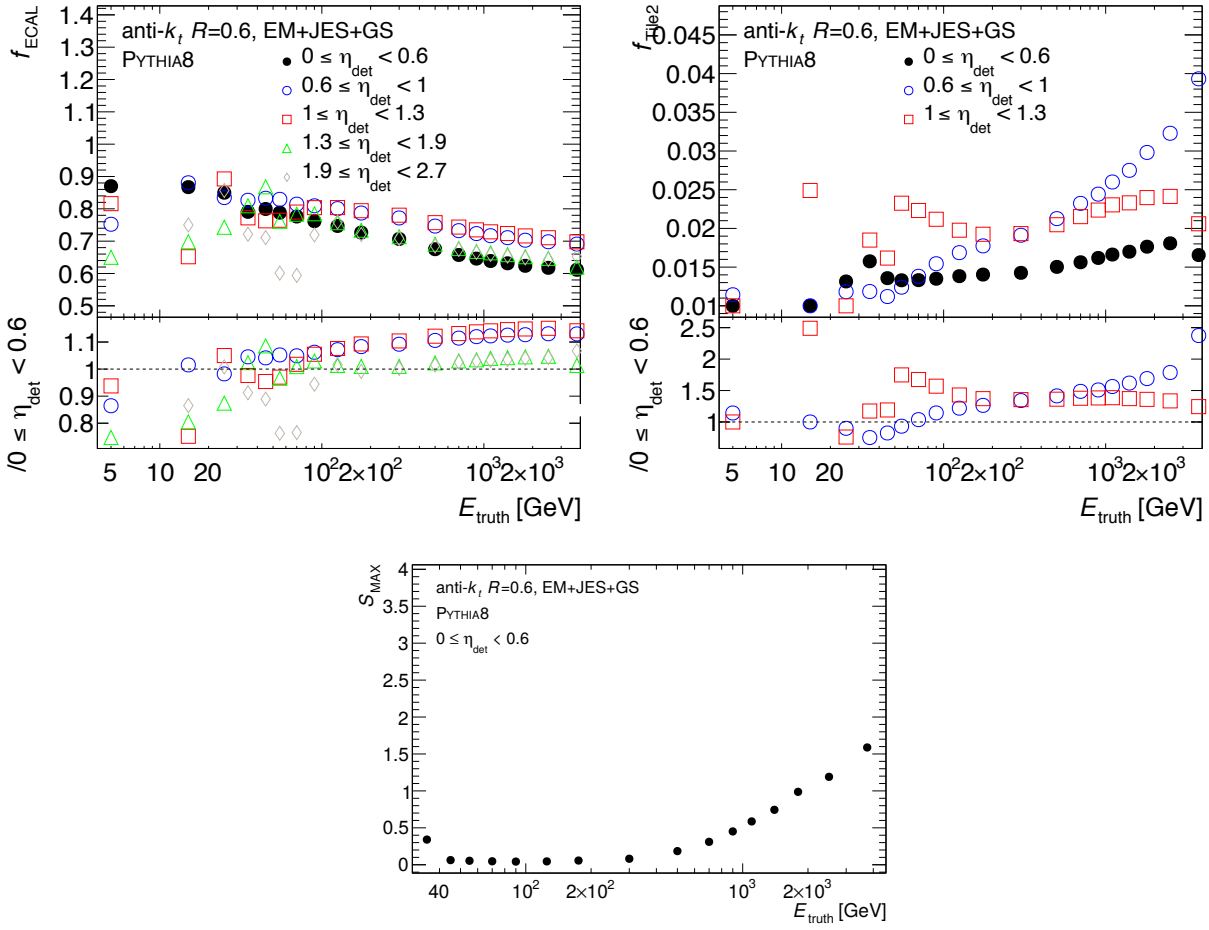


Figure B.2: f_{ECAL} (top-left), f_{Tile2} (top-right) and S_{MAX} (bottom) variables vs E_{truth}.

B.4 Correlations with muon variables

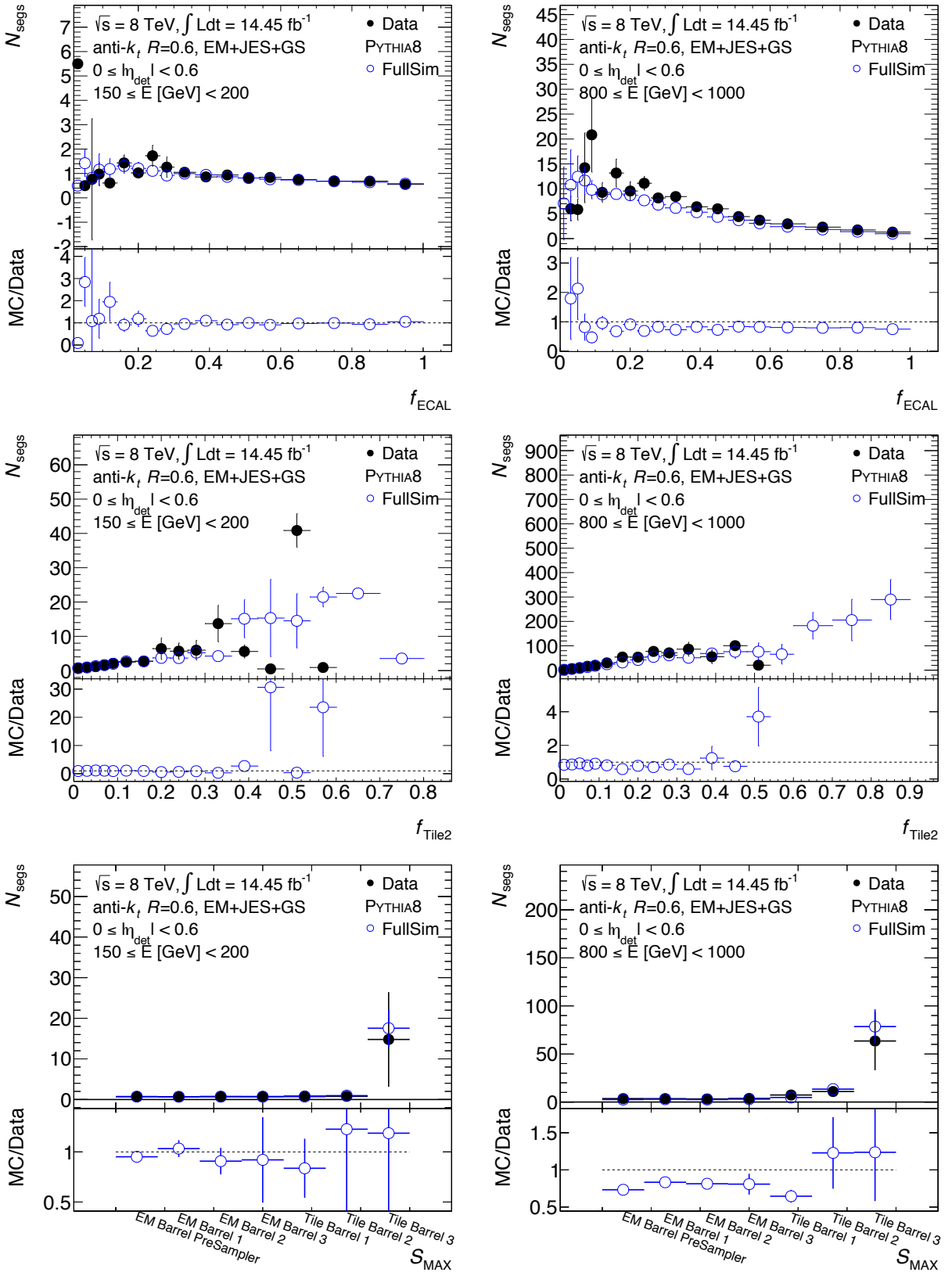


Figure B.3: f_{ECAL} (top), f_{Tile2} (middle) and S_{MAX} (bottom) variables vs $N_{segments}$ for jets with $100 \leq E < 150$ GeV (left) and $800 \leq E < 1000$ GeV (right) in the region $|\eta_{det}| < 0.6$ in the data and the Pythia8 Monte Carlo simulation. Bottom plots show the MC/Data ratio.

It was shown in this thesis that a large amount of muon activity ($N_{segments}$) behind a jet is indicative a jet has punched through the calorimeter, resulting in energy losses. Thus $N_{segments}$ is expected to be correlated with f_{ECAL}, f_{Tile2} and S_{MAX}. Figure B.3 shows $N_{segments}$ vs f_{ECAL}, f_{Tile2} and S_{MAX} for jets with $100 \leq E < 150$ GeV and $800 \leq E < 1000$ GeV in the region $|\eta_{det}| < 0.6$, compared between data and MC.

$N_{segments}$ can be seen to increase as f_{Tile2} increases and f_{ECAL} decreases, and increases as S_{MAX} moves to outer detector layers, peaking for jets with S_{MAX} in the final layer of the hadronic calorimeter. The correlations grow stronger with higher E_{truth}. Coupled with the observations in Section B.1, this is highly suggestive that jets with a high f_{Tile2}, low f_{ECAL} and S_{MAX} in the final layer of the hadronic calorimeter, are more likely to punch-through the calorimeter, as increased muon activity is observed behind them. Reasonable agreement is observed between data and MC. Similar observations were made in all other η_{det} regions these variables were studied in, with larger numbers of segments seen in regions of the detector with less material.

B.5 Dependence of response on calorimeter variables

In the previous section it was observed that jets with a high f_{Tile2}, low f_{ECAL} and low S_{MAX} have a larger $N_{segments}$ behind them, suggesting these jets are more likely to punch-through the calorimeter. It is also important to test how the jet energy response, and thus the measurement of the jet energy, is correlated with these variables. Figure B.4 shows the response vs f_{Tile2}, f_{ECAL} and S_{MAX} in the $|\eta_{det}| < 0.6$ and $1 \leq |\eta_{det}| < 1.3$ regions, before and after the application of the muon based ($N_{segments}$) jet punch-through correction. Only the $|\eta_{det}| < 0.6$ region is shown for S_{MAX} as this variable is not currently constructed in the other detector regions.

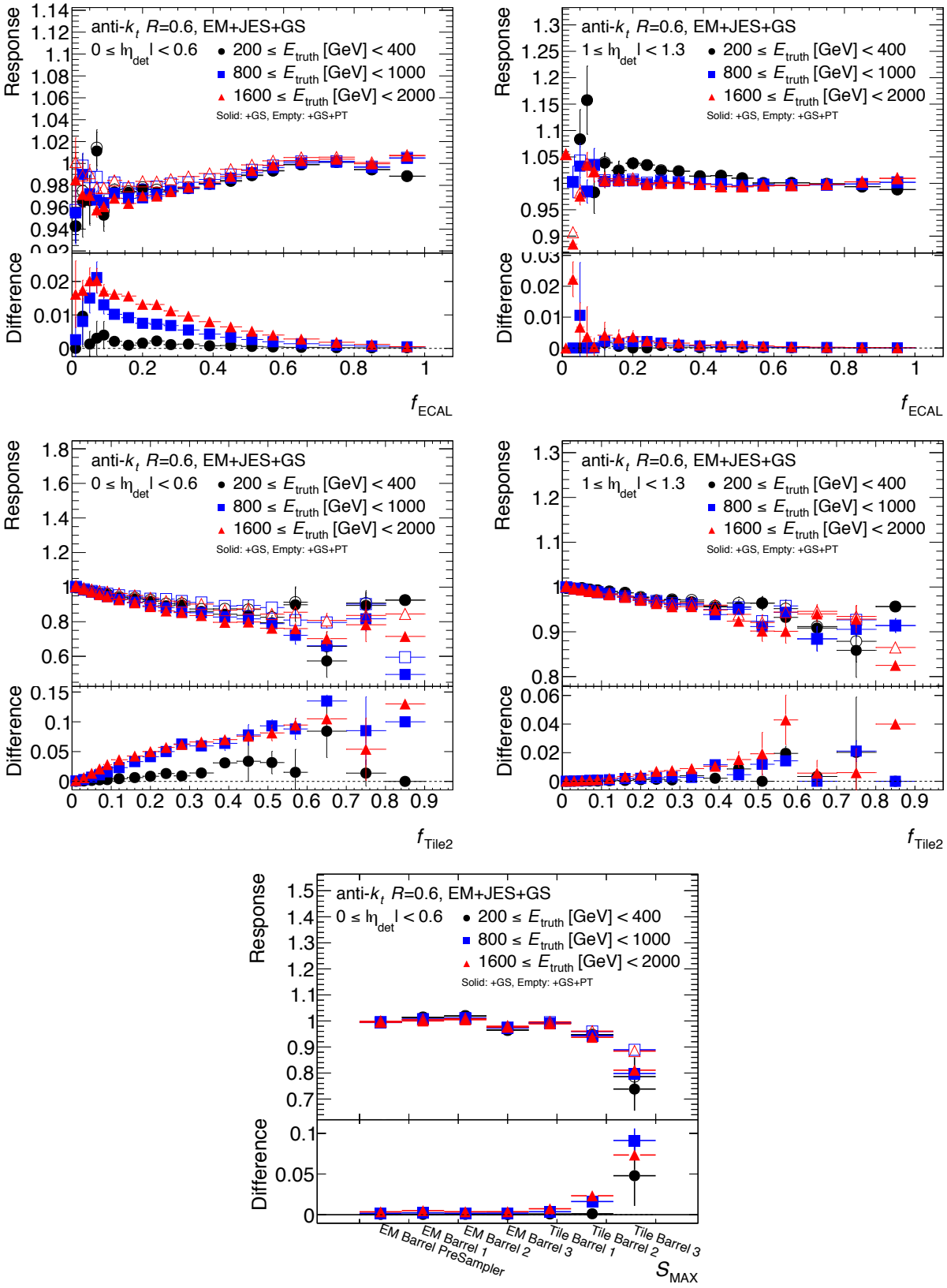


Figure B.4: Response vs f_{ECAL} (top), f_{Tile2} (middle) and S_{MAX} (bottom) variables for jets with $|\eta_{det}| < 0.6$ (left) and $1 \le |\eta_{det}| < 1.3$ (right) in data and PYTHIA8 Monte Carlo simulation. Response values are shown before and after the application of the jet punch-through correction. Bottom plots show the difference in the response before and after the correction.

In the $|\eta_{det}| < 0.6$ region, the response increases as f_{ECAL} increases from ~ 0.96 at $f_{ECAL} \sim 0.1$ to 1 at $f_{ECAL} \sim 0.6$, showing the response degrades as f_{ECAL} decreases. This could be due to the higher punch-through probability at low f_{ECAL}, as suggested by the higher muon activity in this region. This is further suggested by the effect of applying the punch-through correction, whereby the jet response at low f_{ECAL} is increased by the correction by up to 2%, with the largest increases seen at low f_{ECAL}. The cut on $N_{segments} \geq 20$ could be preventing the response from being completely restored to 1 in this region. In the $1 \leq |\eta_{det}| < 1.3$ region, no clear correlation is seen with f_{ECAL} in general, aside from the increase in the jet response as f_{ECAL} decreases for jets with $200 \leq E_{truth}\ [GeV] < 400$.

For f_{Tile2}, we see a clear degradation in the jet energy response as f_{Tile2} increases, with a reduction of 30-40% observed in the $|\eta_{det}| < 0.6$ region for jets with $f_{Tile2} \sim 0.6$. The increased $N_{segments}$ in this region highly suggests this degradation is linked to jet punch-through, whereby the jet punch-through correction pushes the response in the high f_{Tile2} region back towards 1. The impact of the correction increases with f_{Tile2}, improving the response by up to 10% (2%) for higher E_{truth} jets in the highest f_{Tile2} region, with $|\eta_{det}| < 0.6$ ($1 \leq |\eta_{det}| < 1.3$). As with f_{ECAL}, the correlation is not completely removed, likely due to the $N_{segments} \geq 20$ restriction on the correction.

It is clear when looking at the response vs S_{MAX}, the response degrades as S_{MAX} approaches the outer calorimeter layers, with a degradation of 20-30% for jets with S_{MAX} in the final layer of the hadronic calorimeter. Again the high value of $N_{segments}$ observed in this region suggests this degradation is linked to jet punch-through, whereby a 5-10% improvement in the response is observed when applying the jet punch-through correction.

For all calorimeter variables, very little dependence is seen on E_{truth} (agree in general to within 1%), unlike the correlation between the response and $N_{segments}$ which showed a clear dependence on E_{truth}. The only E_{truth} based differences observed are at low f_{ECAL} in the $1 \leq |\eta_{det}| < 1.3$ where jets with $200 \leq E_{truth}\ [GeV] < 400$ show a higher response.

These plots suggest jets with low f_{ECAL}, high f_{Tile2} and S_{MAX} in the outer calorimeter layers experience a degradation in their measured energy due to jet punch-through. The

lack of correlation between the response and f_{ECAL} in the region $1 \leq |\eta_{det}| < 1.3$, along with the restriction of f_{Tile2} and S_{MAX} to $|\eta_{det}| < 1.3$ and $|\eta_{det}| < 0.6$ respectively prevented the iteration of the jet punch-through correction in this thesis from exploiting the observed correlations between calorimeter variables associated with jet punch-through and the jet response, instead opting to focus on $N_{segments}$. However, future iterations of the jet punch-through correction will aim to incorporate the dependencies seen here, in order to improve the impact of the jet punch-through correction on the jet energy resolution and analyses.

B.6 Shower shape studies

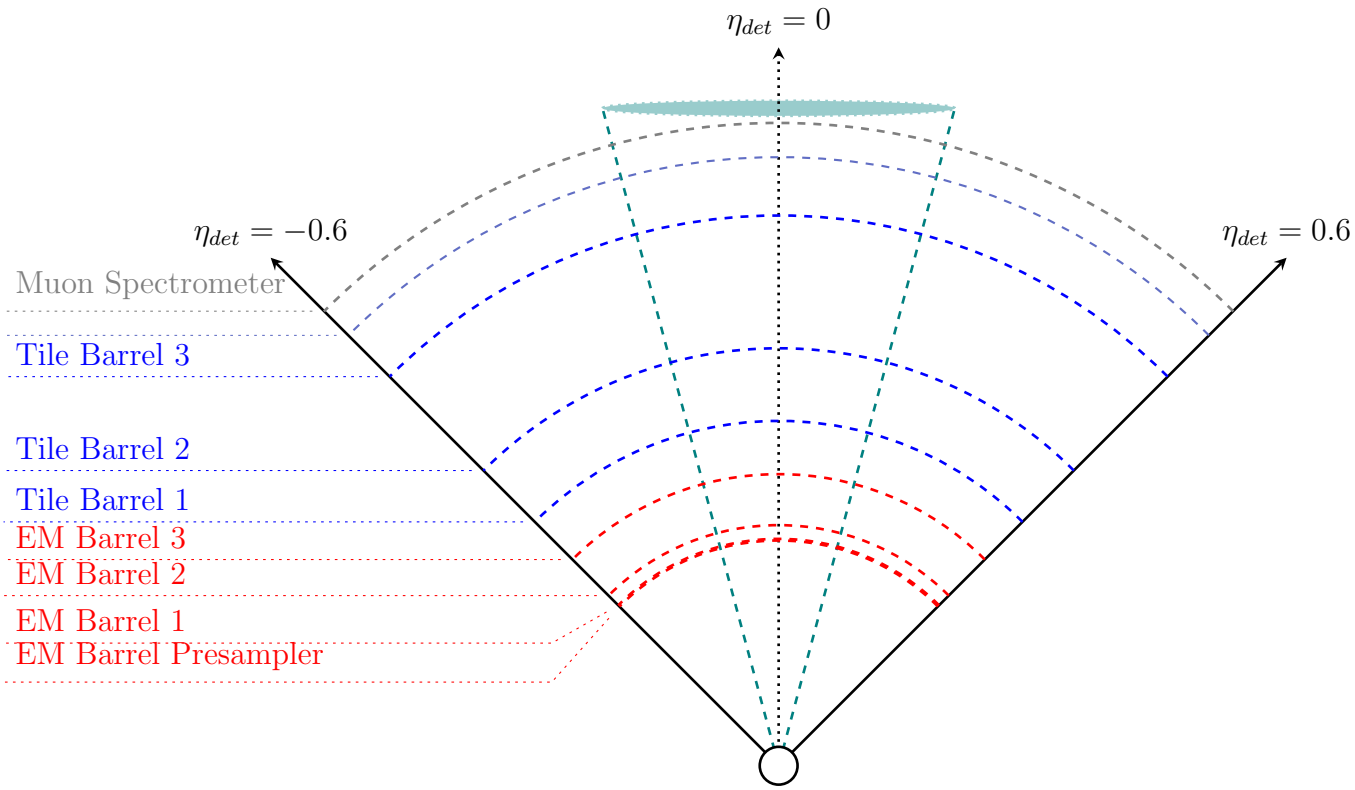


Figure B.5: Schematic representation of the electromagnetic (EM) and hadronic tile calorimeter layers a jet in the region $|\eta_{det}| < 0.6$ passes through.

As jet punch-through is related to how the hadronic shower develops in the calorimeter, it is useful to study the shape of the shower in the calorimeter in relation to the amount of muon activity found behind the jet in the calorimeter. By doing this, a method to

correlate shower shape with muon activity behind the jet could be found, allowing the punch-through correction to be extended to regions of the detector without a muon spectrometer ($|\eta_{det}| > 2.7$). This section will detail preliminary studies that attempt to study how the shower shape is related to $N_{segments}$ in the region $|\eta_{det}| < 0.6$, with the aim to extrapolate this to the region $|\eta_{det}| > 2.7$ in the future. The calorimeter layers a jet in the $|\eta_{det}| < 0.6$ region passes through are shown in Figure B.5.

To study how the shower shape affects the muon activity behind a jet ($N_{segments}$), the fraction of the jet's energy deposited per unit interaction length, $\frac{f_{layer}}{d_{layer}[\lambda_I]}$, in each of these calorimeter layers, is plotted, along with the $N_{segments}$ and the energy of the jet (E). For MC, E_{truth} and the jet energy response are also plotted. These plots are then split based upon which calorimeter layer contains the maximum energy deposition per unit interaction length (S_{MAX}), which is equivalent to binning depending on which layer has the maximum fraction of energy per unit interaction length (as E^{jet} is a constant factor for all layers when looking at a single jet). Figure B.6 shows these plots for data while Figure B.7 shows these plots for MC.

In both MC and data, as observed previously, $N_{segments}$ grows larger as S_{MAX} approaches later layers in the calorimeter. Further studies will need to be performed in the future to fully understand the shower shapes observed for jets with S_{MAX} in the later layers in the calorimeter.

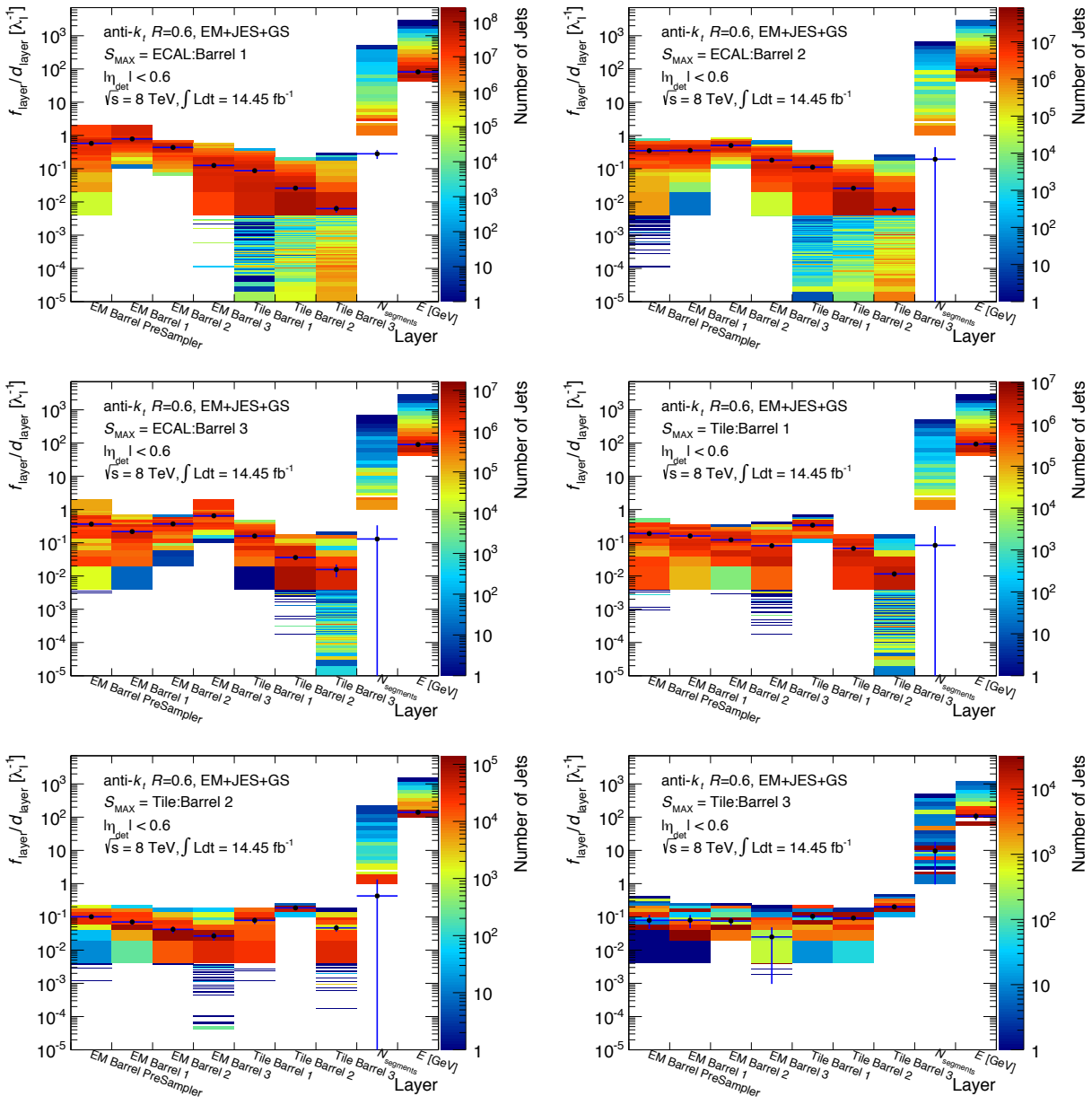


Figure B.6: Fraction of jet's energy per unit interaction length deposited in each calorimeter layer, binned in S_{MAX}, for data. Points indicate mean value. MC is scaled to 14.45 fb^{-1}.

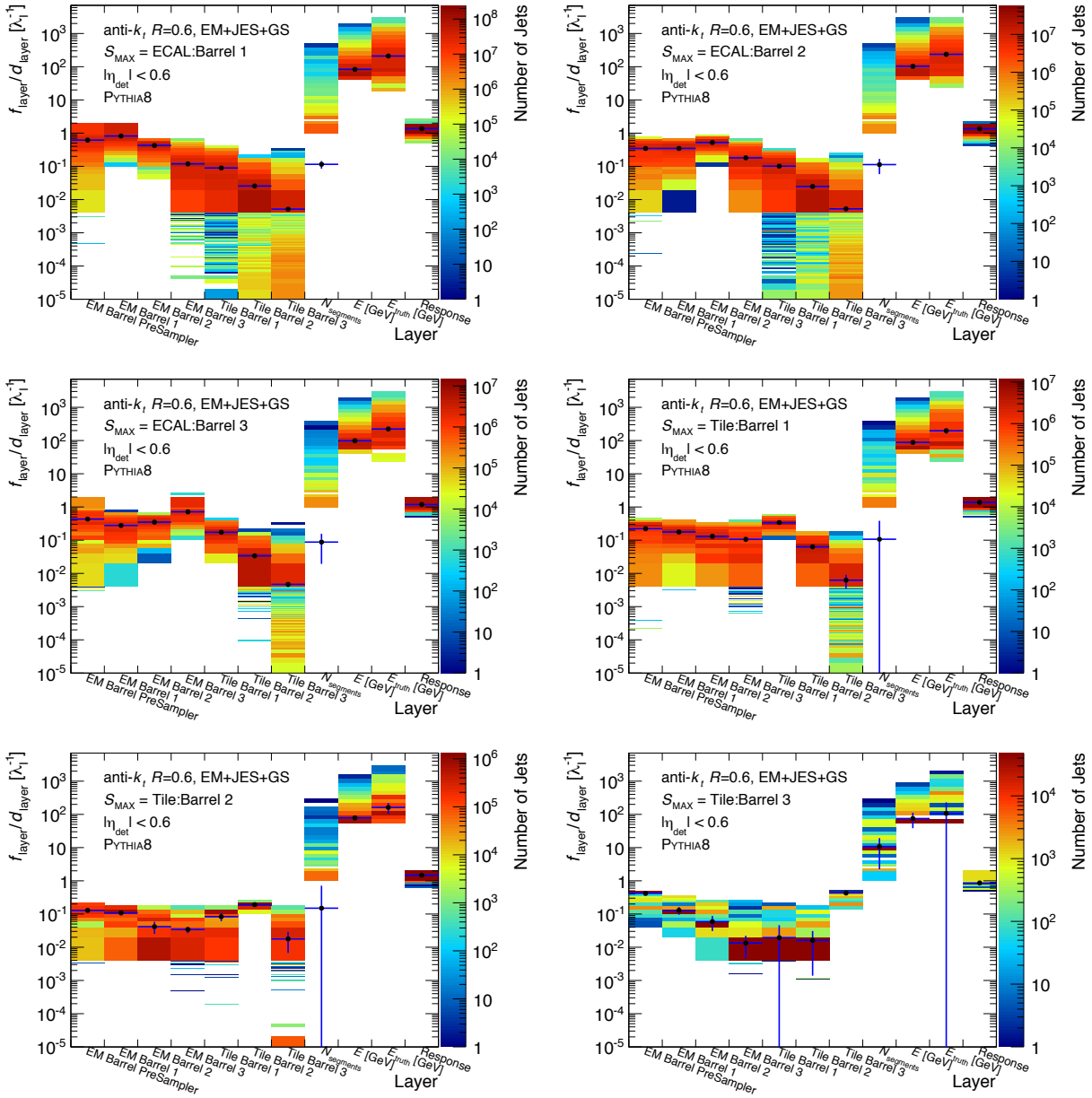


Figure B.7: Fraction of jet's energy per unit interaction length deposited in each calorimeter layer, binned in S_{MAX}, for MC. Points indicate mean value. MC is scaled to 14.45 fb^{-1}.

Appendix C

Data Validation: Asymmetry Studies

As mentioned in Section 6.1, the imbalance in a dijet event can also be assessed via the asymmetry variable, as defined in Equation C.1, where $p_{\mathrm{T}}^{\mathrm{avg}}$ [1] is the average p_{T} of the two jets in the dijet balance.

$$\mathcal{A}(\mathrm{N}_{segments}) = \frac{p_{\mathrm{T}}^{\mathrm{probe}}(\mathrm{N}_{segments} \geq 20) - p_{\mathrm{T}}^{\mathrm{ref}}(\mathrm{N}_{segments} = 0)}{p_{\mathrm{T}}^{\mathrm{avg}}(\mathrm{N}_{segments})}, \tag{C.1}$$

The mean reconstructed response can be approximated to:

$$\langle \mathcal{R}(\mathrm{N}_{segments}) \rangle = \frac{1 + \langle \mathcal{A}(\mathrm{N}_{segments}) \rangle / 2}{1 - \langle \mathcal{A}(\mathrm{N}_{segments}) \rangle / 2}. \tag{C.2}$$

Comparisons of the mean reconstructed response calculated from Equations 6.1 and C.2 are shown in this appendix as a cross-check with the reconstructed response values obtained directly in Section 6.5. Section C.1 shows example Gaussian fits to some asymmetry distributions, Section C.2 shows some example asymmetry vs $\mathrm{N}_{segments}$ curves and Section C.3 compares the reconstructed response as calculated from the reconstructed asymmetry using Equation C.2 with the reconstructed response obtained directly.

[1] $p_{\mathrm{T}}^{\mathrm{avg}} = (p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}})/2$

C.1 Asymmetry fits

As with the mean reconstructed response, the mean reconstructed asymmetry is obtained from a Gaussian fit to the mean reconstructed asymmetry distributions in bins of $N_{segments}$, p_T^{probe} and η_{det}. The reconstructed asymmetry distributions are fitted in the same way as the reconstructed response distributions as outlined in Section 6.5.

Figure C.1 shows some examples of fits to the reconstructed asymmetry distribution for jets with $600 \leq p_T^{probe}$ [GeV] < 1000 and $25 \leq N_{segments}^{probe} < 50$ in the $0 \leq |\eta_{det}| < 1.3$ region, before and after applying the jet punch-through correction. To aid the fits, the reconstructed asymmetry is shifted by +1 when constructing these distributions.

As with the reconstructed response in Figure 6.5, the reconstructed asymmetry distributions are observed to be Gaussian before and after the jet punch-through correction is applied. Again, the fits appear to be relatively good with $\tilde{\chi}^2_{red} \sim 1$.

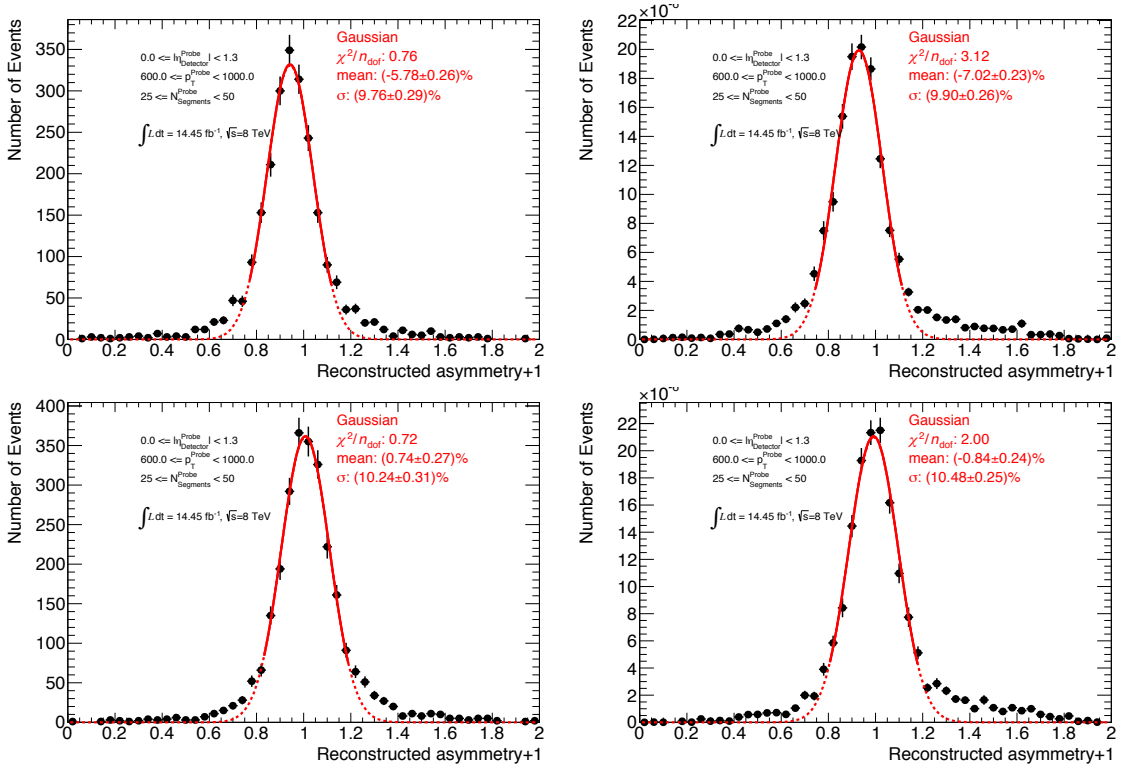


Figure C.1: Examples of fits to reconstructed asymmetry distributions for jets with $600 \leq p_T^{probe}$ [GeV] < 1000, $25 \leq N_{segments}^{probe} < 50$ in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied, in data (left) and MC (right).

C.2 Dependence of asymmetry on $N_{segments}$

The reconstructed asymmetry as a function of $N_{segments}$ is compared between data and MC in Figure C.2 for jets with $200 \leq p_T^{probe}$ [GeV] < 600, $600 \leq p_T^{probe}$ [GeV] < 1000 and $0 \leq |\eta_{det}| < 1.3$ before and after the jet punch-through correction is applied. The reconstructed asymmetry displays a clear dependency on $N_{segments}$, whereby the imbalance increases as $N_{segments}$ increases. The imbalance appears to be smaller at higher p_T^{probe}. This is compatible with what was seen with the truth and reconstructed response.

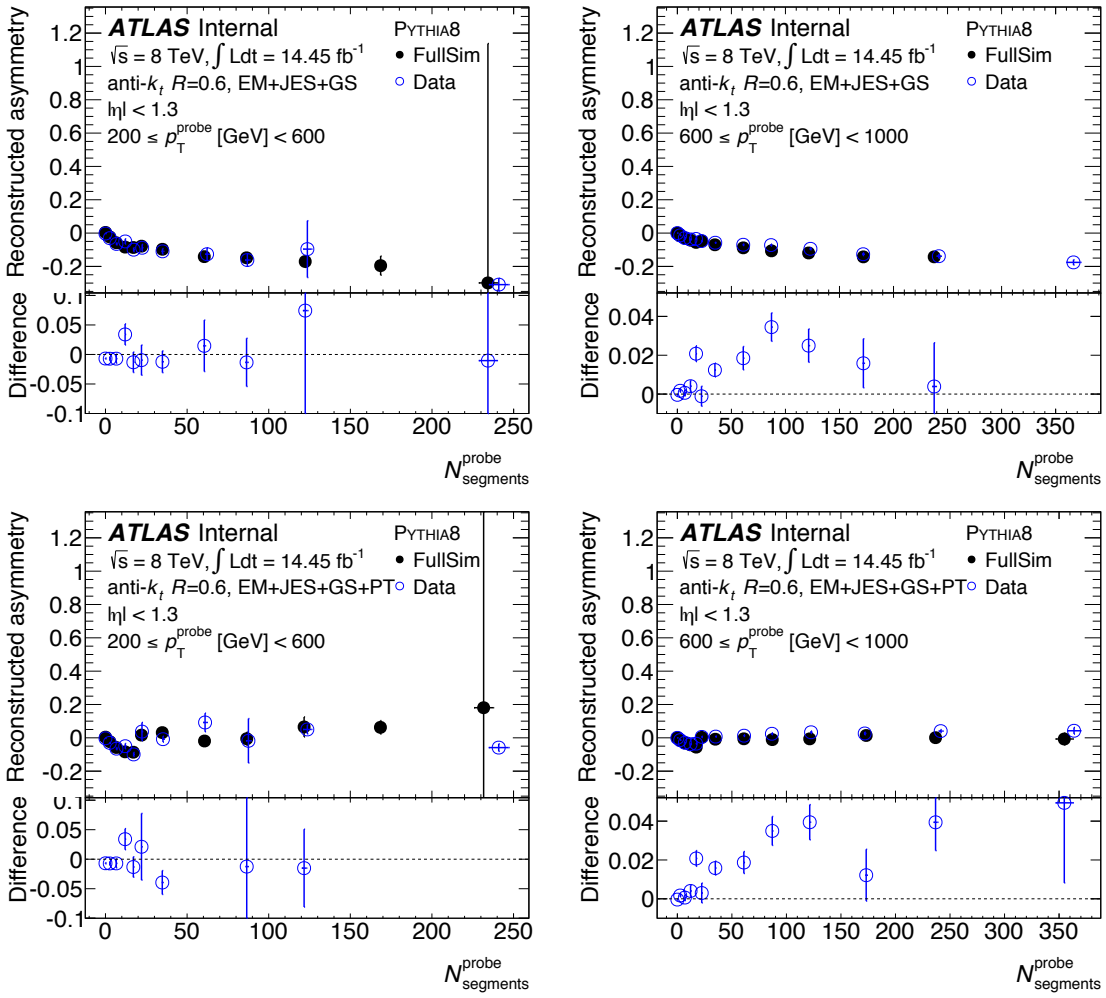


Figure C.2: Reconstructed asymmetry as a function of $N_{segments}^{probe}$ compared between data and MC for jets with $200 \leq p_T^{probe}$ [GeV] < 600 (left) and $600 \leq p_T^{probe}$ [GeV] < 1000 (right) in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied. The bottom plot shows the difference between data and MC.

The data/MC differences in the $200 \leq p_T^{probe}$ [GeV] < 600 region are within a

couple of percent of zero to within errors. In the $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000 region, the differences are generally observed to be small ($< 5\%$). The differences appear to be lower after the punch-through correction when compared with those in Figure 6.5 in the $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 600 region, but similar in the $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000 region. As with the reconstructed response, this suggests the affect of jet punch-through on the jet energy scale is reasonably well modelled.

C.3 Comparison with reconstructed response

The reconstructed response can be obtained from the asymmetry via Equation C.2. Figure C.3 shows plots comparing the response obtained from the asymmetry ('Response (A)') with the response obtained directly from the reconstructed response distributions shown in Figure 6.5 ('Response'). The comparison shown is for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 600, $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}$ [GeV] < 1000 and $0 \leq |\eta_{det}| < 1.3$ for both data and MC.

It is clear the response obtained from both methods is compatible to within errors, confirming the reconstructed response values measured in Section 6.5 are reliable. For the uncertainties on the jet punch-through correction in Chapter 6, the reconstructed response is obtained directly as opposed to from the asymmetry as it was found the reconstructed response fits had better $\tilde{\chi}^2_{red}$ values. This can be seen when comparing Figures 6.4 and C.1, whereby three out of the four fits have $\tilde{\chi}^2_{red}$ closer to 1 for the reconstructed response fits. As mentioned before, the correction is working well - leaving the $\mathrm{N}_{segments} < 20$ region untouched and restoring the asymmetry and reconstructed response to 1 for jets with $\mathrm{N}_{segments} \geq 20$.

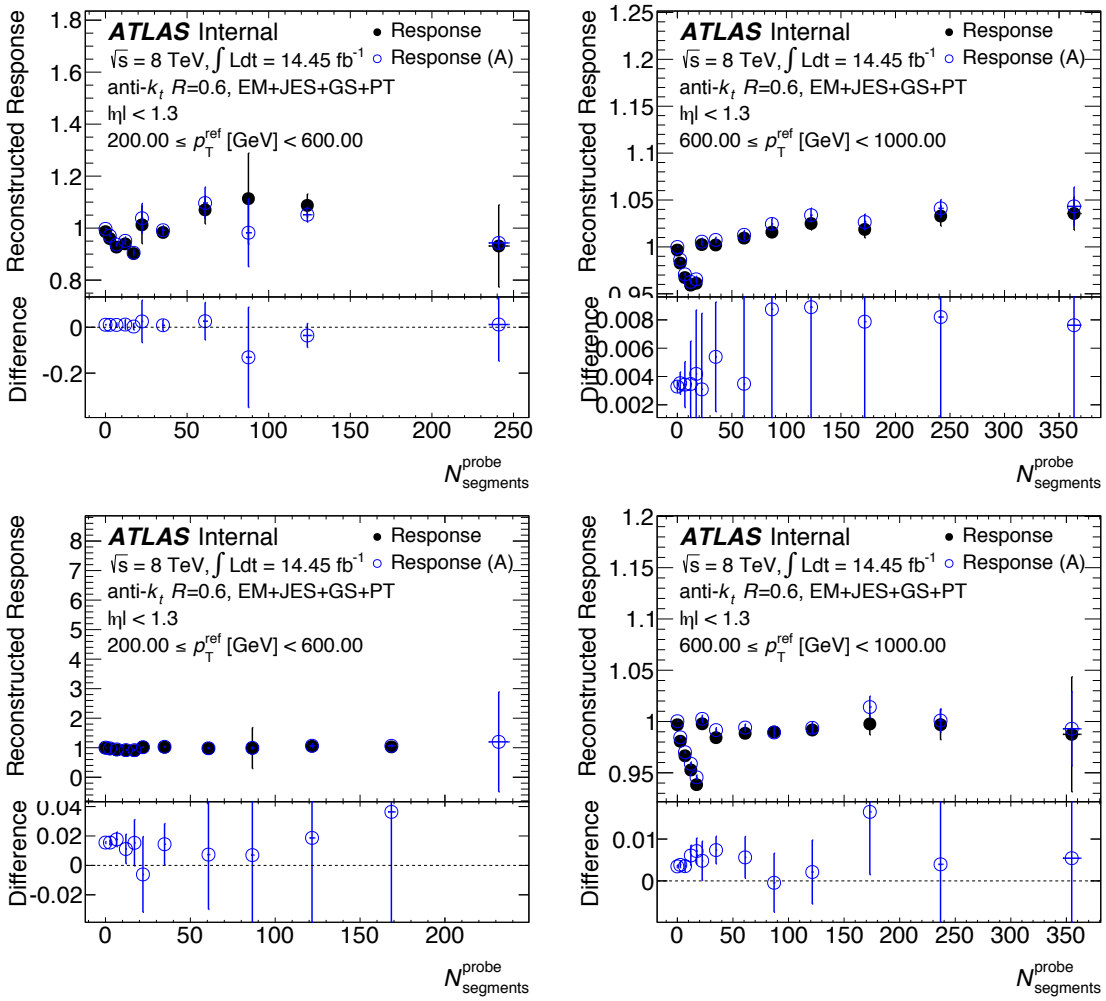


Figure C.3: Reconstructed response as a function of $N^{probe}_{segments}$ calculated directly and from the reconstructed asymmetry for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [\mathrm{GeV}] < 600$ (left) and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [\mathrm{GeV}] < 1000$ (right) in the $0 \leq |\eta_{det}| < 1.3$ region for data (top) and MC (bottom). The bottom plot shows the difference between the reconstructed response calculated via the asymmetry with the reconstructed response obtained directly.

Appendix D

Punch-through studies with LCW jets

This appendix contains jet punch-through studies pertaining to the fraction of LCW jets that experience jet punch-through in Section D.1 and the effect of jet punch-through on the LCW+JES+GS jet energy scale and resolution in Section D.2, as well as outlining the derived LCW jet punch-through correction function in Section D.3, uncertainties on the correction function in Section D.4, and performance of the correction in D.5. Although plots are only shown for jets with radius R = 0.6 in this appendix, the correction was derived for both R = 0.4 and R = 0.6, EM+JES+GS and LCW+JES+GS jets.

D.1 Punch-through probability

Figure D.1 shows the fraction of punch-through jets as defined in Section 4.1 as a function of E and $|\eta_{det}|$. It is clear that as with EM+JES+GS jets, LCW+JES+GS jets with at high E, passing through $|\eta_{det}|$ regions with less material, are more likely to punch-through the calorimeter, reaching as high as 15 (20) %.

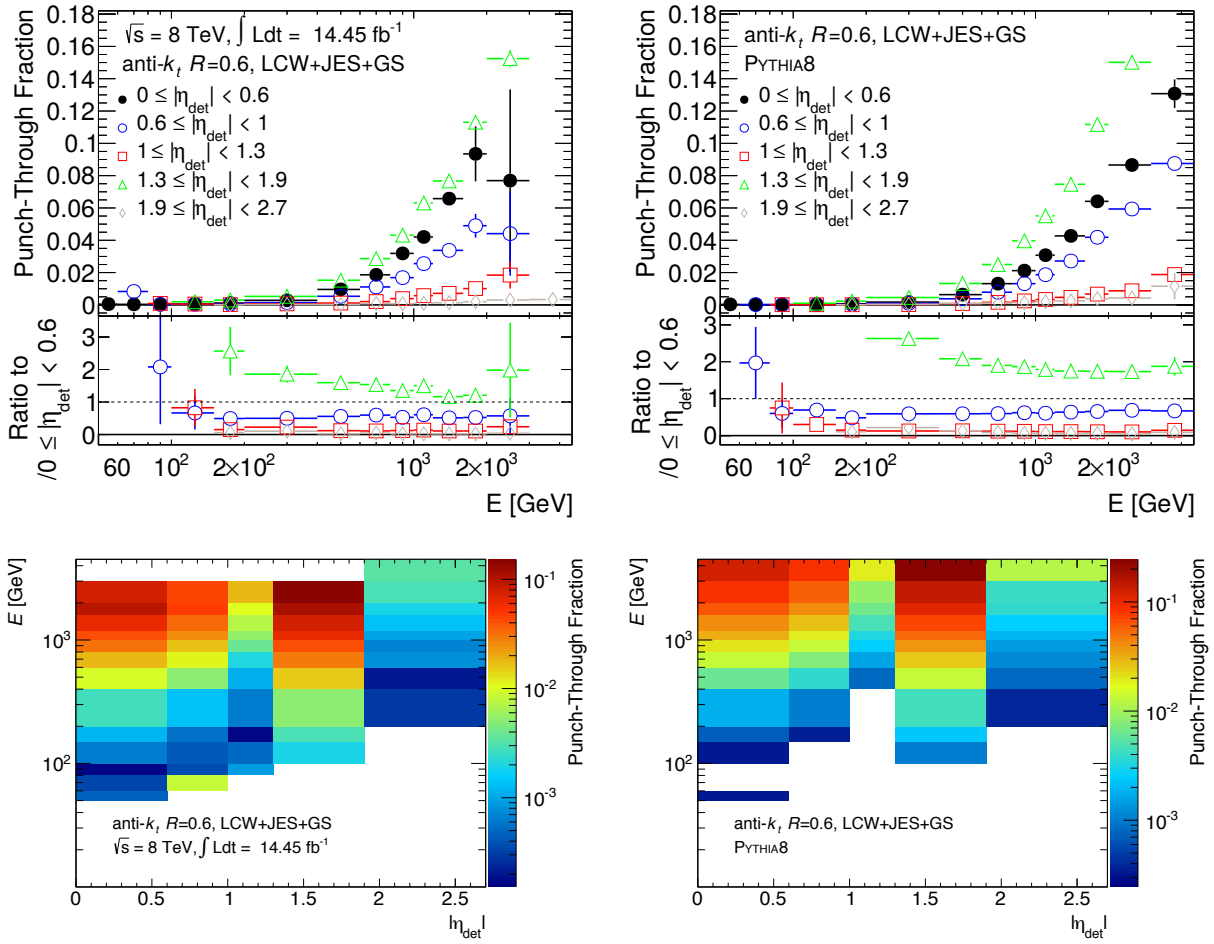


Figure D.1: Fraction of LCW jets punching-through as a function of E and η_{det} for data (left) and MC (right). The bottom plots show the ratio of the punch-through fraction in each η_{det} region to the punch-through fraction in the $0 \leq |\eta_{det}| < 0.6$.

D.2 General energy studies

Here are the LCW+JES+GS equivalents to those in Chapter 4. Sections D.2.1, D.2.2 and D.2.4 contain plots related to Sections 4.2.1, 4.2.2 and 4.2.3 respectively.

D.2.1 Degradation of the jet energy scale due to jet punch-through

Figure D.2 shows the energy response and E_{lost} as a function of $N_{segments}$. It is clear that even though the LCW correction has a correction sensitive to jet punch-through (see Section 3.3), there is still a dependence of the LCW+JES+GS calibration scheme on $N_{segments}$ and thus jet punch-through. The degradation is more severe for low E_{truth} jets - up to 30% - compatible

with what was observed for EM+JES+GS jets in Figure 4.7.

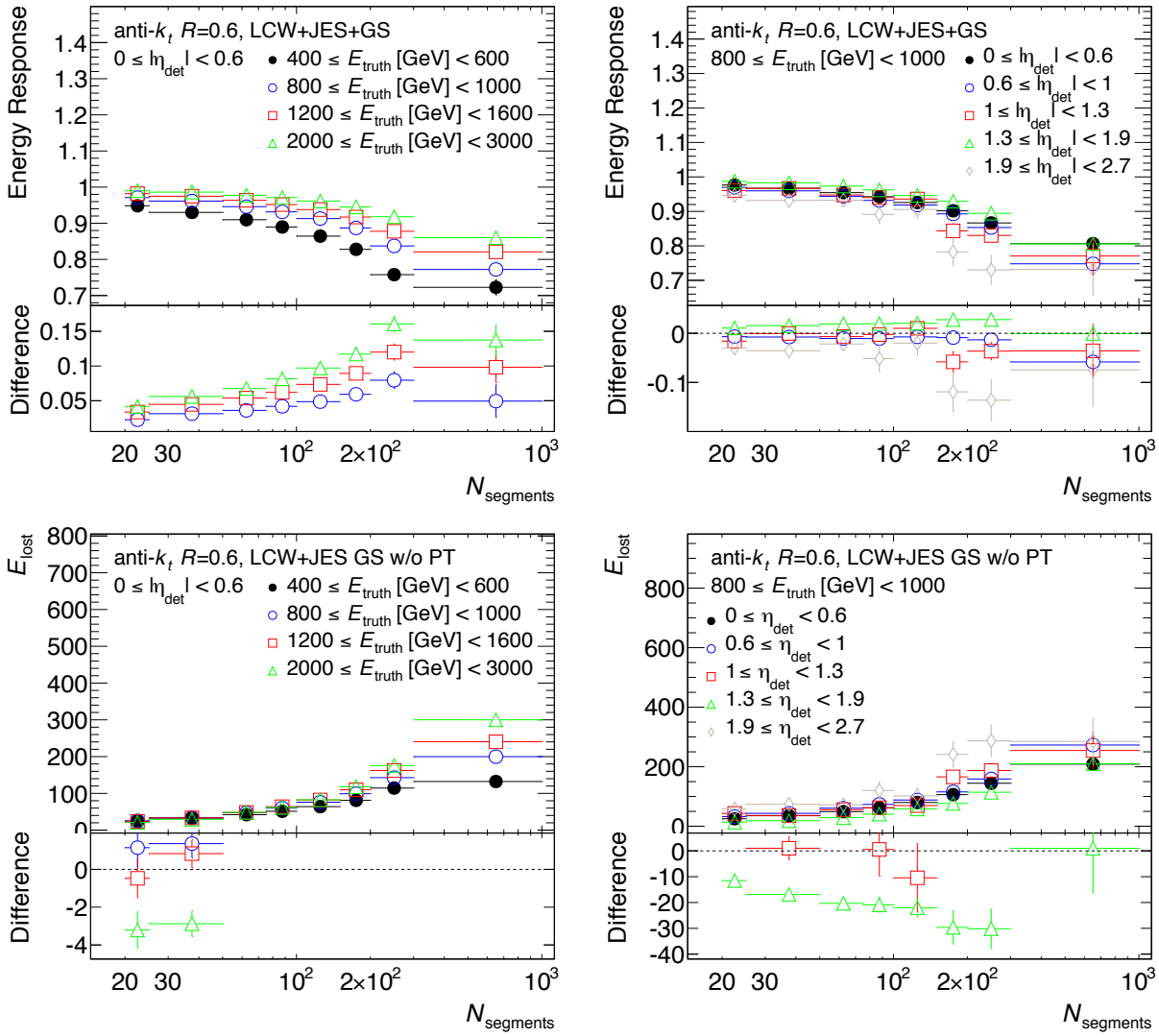


Figure D.2: Jet energy response (top) and E_{lost} (bottom) vs $N_{segments}$ for jets in the central region ($0 \le |\eta_{det}| < 0.6$) with different E_{truth} bins (left) for jets with $800 \le E_{truth}$ [GeV] < 1000 in different $|\eta_{det}|$ regions (right). The bottom plots show $R_i - R_{nom}$, where R_i and R_{nom} are the response (top) or E_{lost} (bottom) in the i-th bin and in the $400 \le E_{truth}$ [GeV] < 600 bin (left) or in the central region (right) respectively.

D.2.2 Response tails

Figure D.3 shows the percentage of jets below a response threshold vs $N_{segments}$, from which it is clear that jet punch-through increases the fraction of jets in the LCW+JES+GS-scale low response tail, as was seen for EM+JES+GS-scale jets (Figure 4.9).

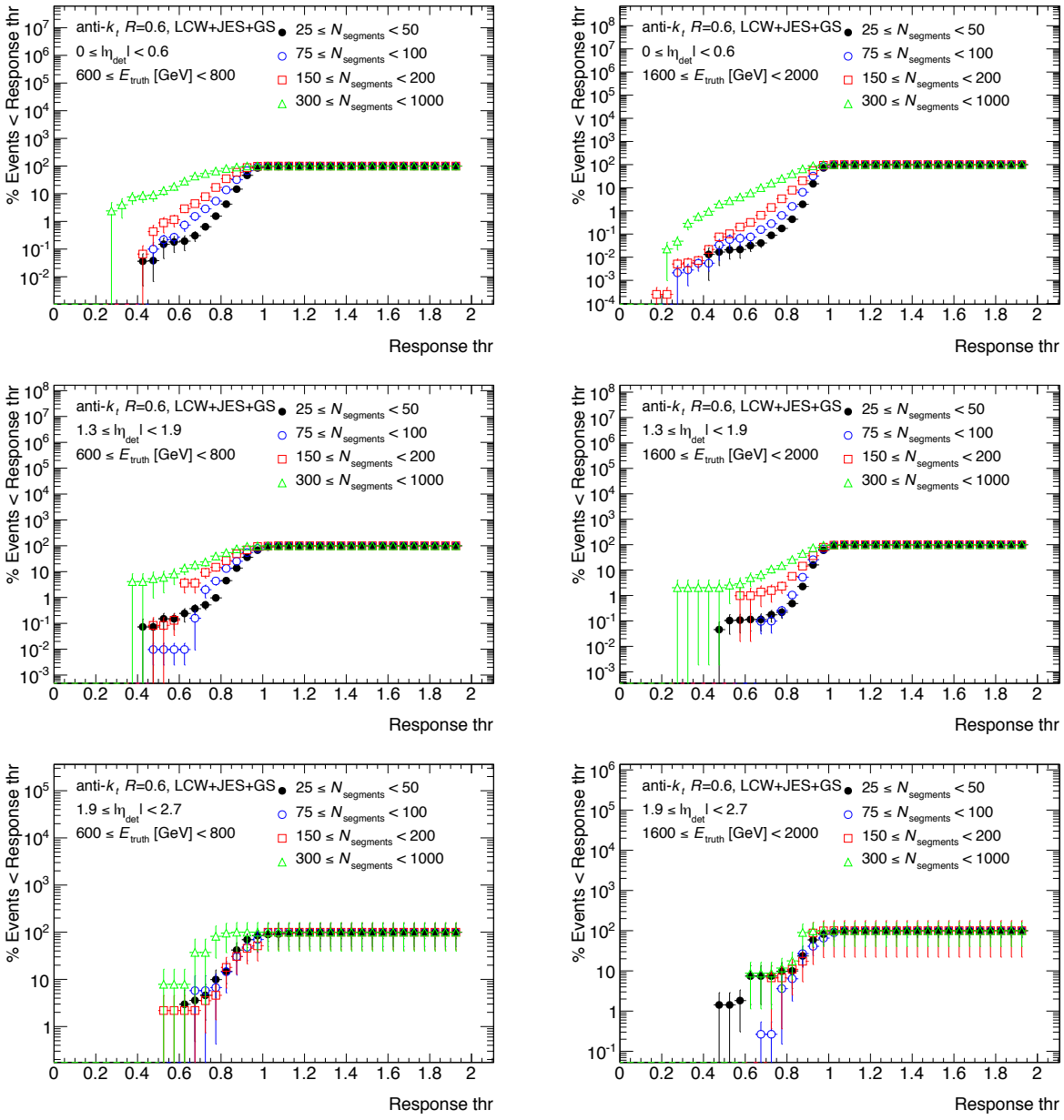


Figure D.3: Percentage of jets below a given response threshold ('Response thr') as the response threshold increases. This is compared for jets with different $N_{segments}$ values in the regions $0 \le |\eta_{det}| < 0.6$ (top), $1.3 \le |\eta_{det}| < 1.9$ (middle) and $1.9 \le |\eta_{det}| < 2.7$ (bottom), with $800 \le E_{truth}$ [GeV] < 1000 (left) and $1200 \le E_{truth}$ [GeV] < 1600 (right).

D.2.3 Degradation of jet energy resolution

The jet energy resolution vs $N_{segments}$ and energy loss % due to jet punch-through is shown in Figure D.4. A degradation in the LCW+JES+GS-scale jet energy resolution due to an increase in $N_{segments}$, and thus jet punch-through, is observed, consistent with EM+JES-scale jets (Figure 4.11). The degradation as a function of $N_{segments}$ is worst for low E_{truth} jets in the $1.3 \leq \eta_{det} < 1.9$ region, however as a function of energy loss %, very little variation is seen as a function of E_{truth}. However, for jets in the $1.9 \leq \eta_{det} < 2.7$ region, the low jet punch-through fraction limits the available statistics, preventing a clear degradation of the jet energy resolution as energy loss increases from being seen.

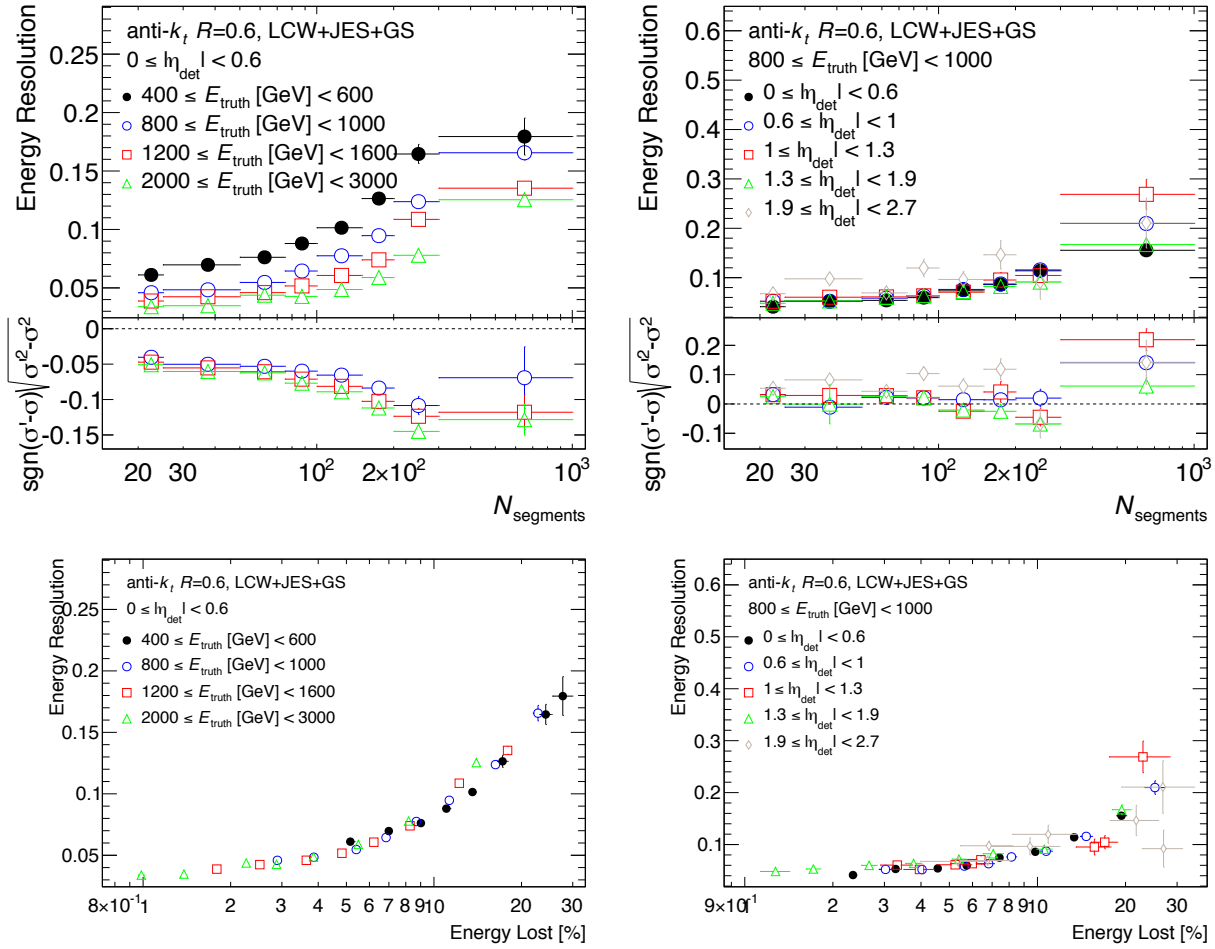


Figure D.4: Jet energy resolution vs $N_{segments}$ (top) and the percentage of energy lost due to jet punch-through (bottom) for jets in the central region ($0 \leq |\eta_{det}| < 0.6$) with different E_{truth} (left), and jets with $800 \leq E_{truth}$ [GeV] < 1000 in different $|\eta_{det}|$ regions (right). The bottom plots show the difference in quadrature relative to jets in the $400 \leq E_{truth}$ [GeV] < 600 bin (top-left), or to jets in the central region (top-right).

D.2.4 Flavour differences for LCW jets

A comparison between the jet energy response vs $N_{segments}$ for LQ, gluon, b and all-jets is shown in Figure D.5. As mentioned in Section 4.2.4, the samples used are LQ dominated at high E_{truth}, and so the inclusive jet response is heavily weighted towards the LQ-jet response. As with EM+JES+GS jets (Figure 4.12), LQ-jets show the largest degradation in the jet energy response as $N_{segments}$ increases, with the degradation of gluon and b-jets generally being compatible to within errors.

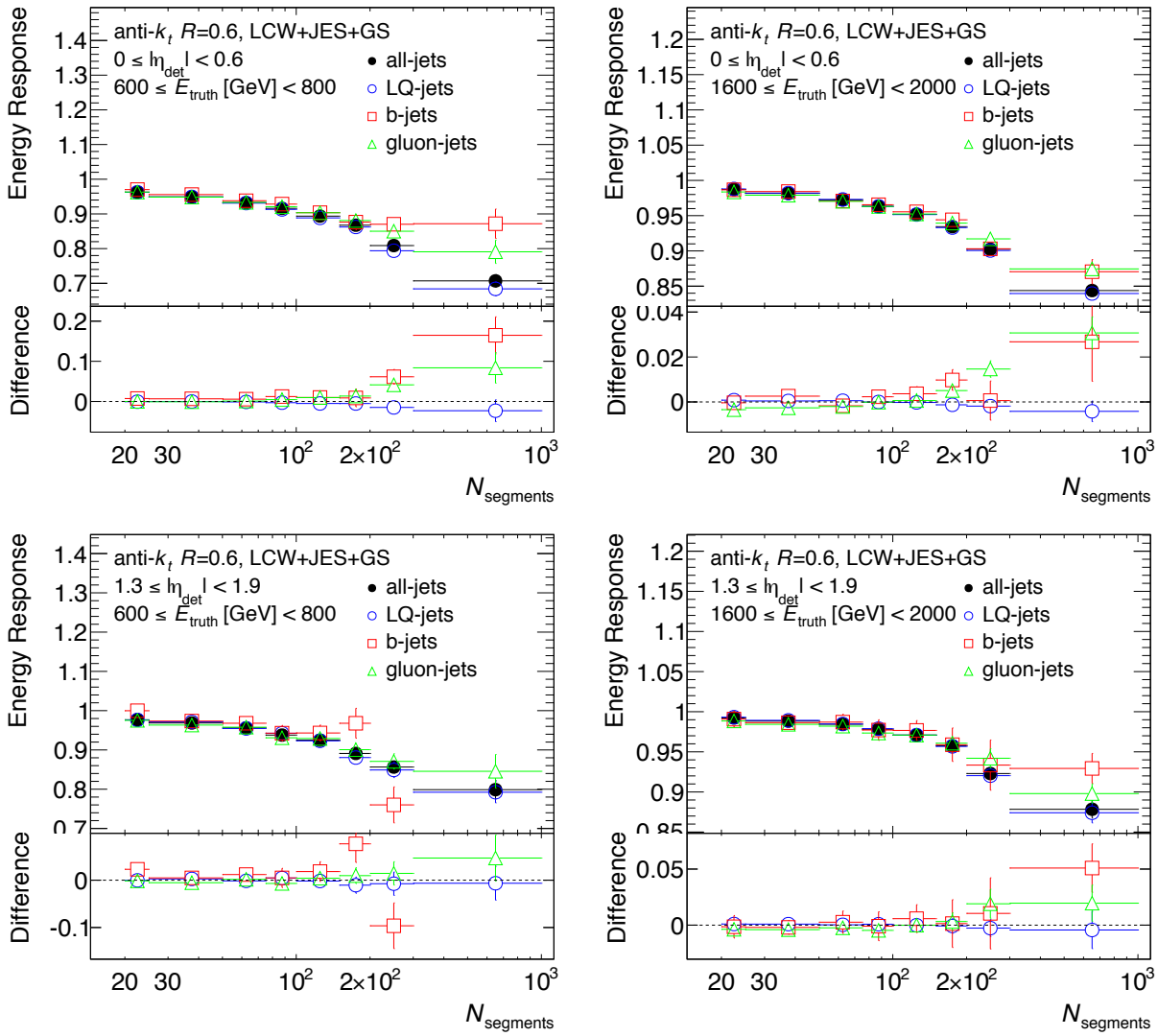


Figure D.5: Jet energy response vs $N_{segments}$ for different flavour jets in the regions $0 \le |\eta_{det}| < 0.6$ (top) and $1.3 \le |\eta_{det}| < 1.9$ (bottom), with $600 \le E_{truth}\,[GeV] < 800$ (left) and $1600 \le E_{truth}\,[GeV] < 2000$ (right). The bottom plots show $R_i - R_{nom}$, where R_i is the response in the i-th bin and R_{nom} is the response for all jets.

Figure D.6 compares the degradation in the jet energy resolution as a function of $N_{segments}$ for LQ, gluon, b and all-jets. Once again, as with EM+JES-scale jets (Figure 4.13), in general, LQ-jets display the largest degradation in the resolution as $N_{segments}$ increases, with a similar degradation observed for gluon-jets and b-jets.

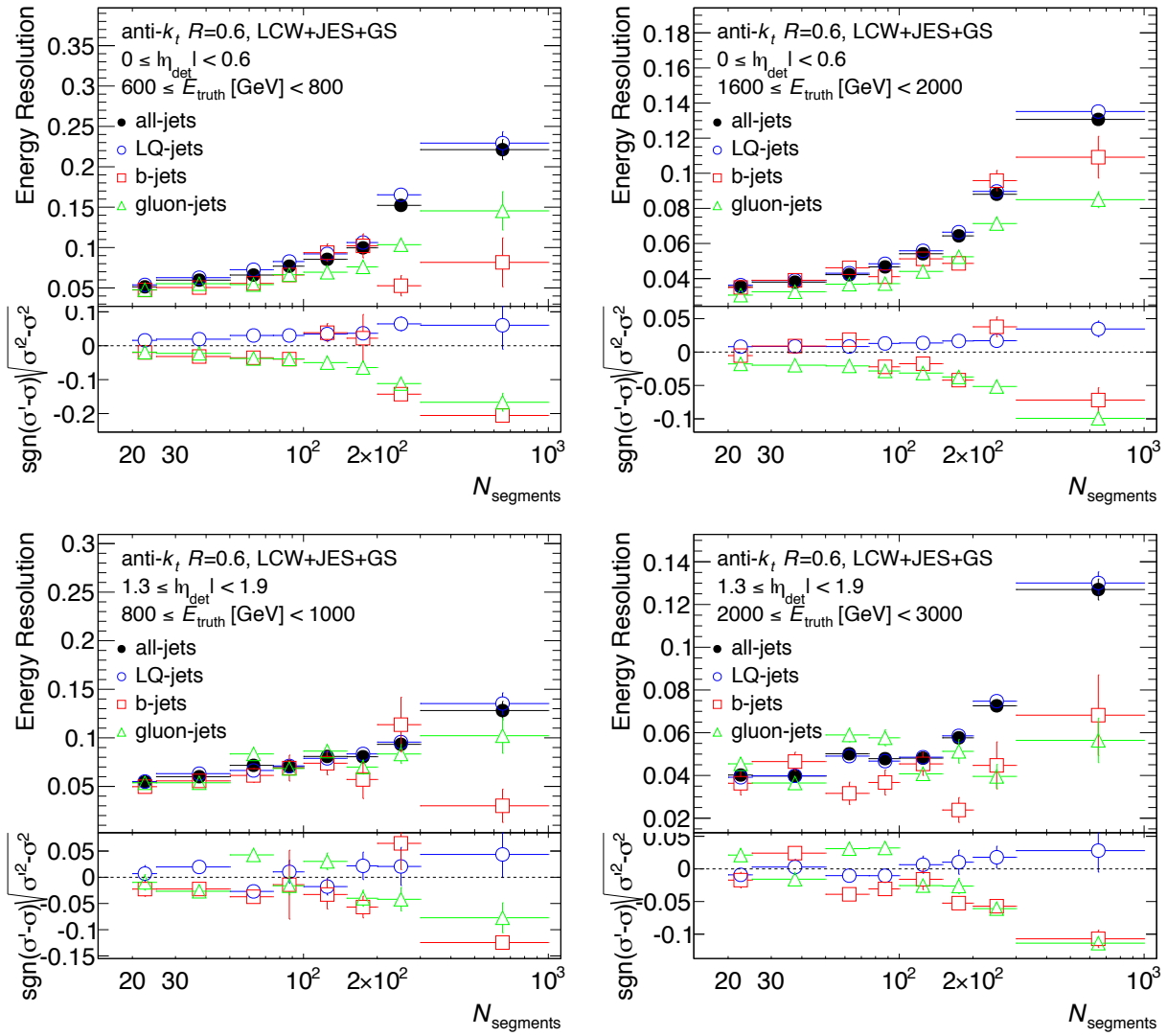


Figure D.6: Jet energy resolution vs $N_{segments}$ for different flavour jets in the regions $0 \le |\eta_{det}| < 0.6$ (top), $1.3 \le |\eta_{det}| < 1.9$ (bottom), with $600 \le E_{truth}$ [GeV] < 800 (left) and $1600 \le E_{truth}$ [GeV] < 2000 (right). The bottom plots show the difference in quadrature for each jet flavour (σ') in comparison with all-jets (σ).

D.3 LCW jet punch-through correction

This section displays the LCW+JES+GS-scale jet punch-through correction factors derived using the method explained in Chapter 5. The mean jet energy response for LCW+JES+GS-scale jets extracted from the fitting procedure described in Section 5.4 is displayed in Figure D.7 (see Figure 5.4 for the equivalent EM+JES-scale jet plots). The extracted mean response values display the previously observed degradation of the jet energy response as $N_{segments}$ increases, with a larger degradation observed for lower E_{truth} jets. In the $1.9 \leq |\eta_{det}| < 2.7$ region, very few fits converged due to the lack of statistics, as was the case for EM+JES-scale jets.

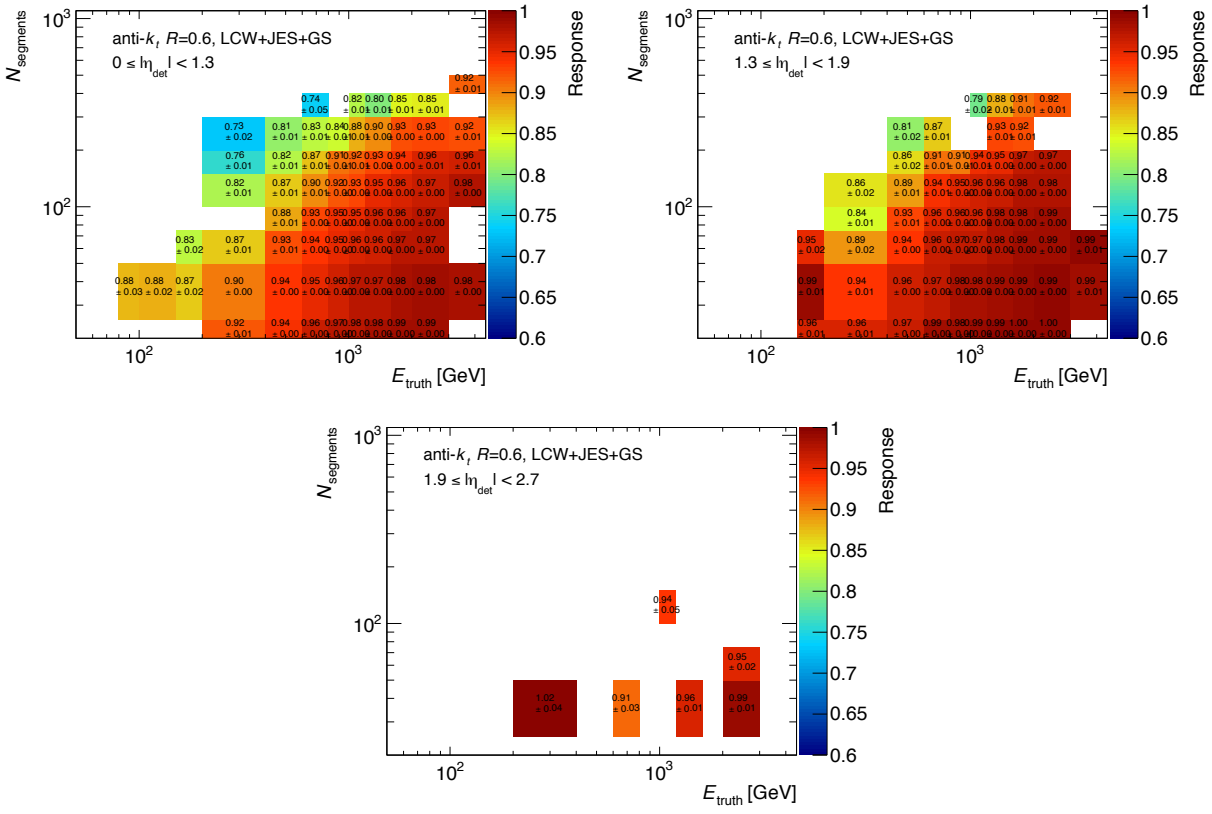


Figure D.7: Mean jet energy response values extracted from fits along with their associated errors as a function of E_{truth} and $N_{segments}$, for jets in the detector region $0 \leq \eta_{det} < 1.3$ (left), $1.3 \leq \eta_{det} < 1.9$ (right) and $1.9 \leq \eta_{det} < 2.7$ (bottom).

The final correction function derived after applying the one and two dimensional Gaussian kernels described in Section 5.6 is displayed for each of the three η_{det} regions in Figure D.8.

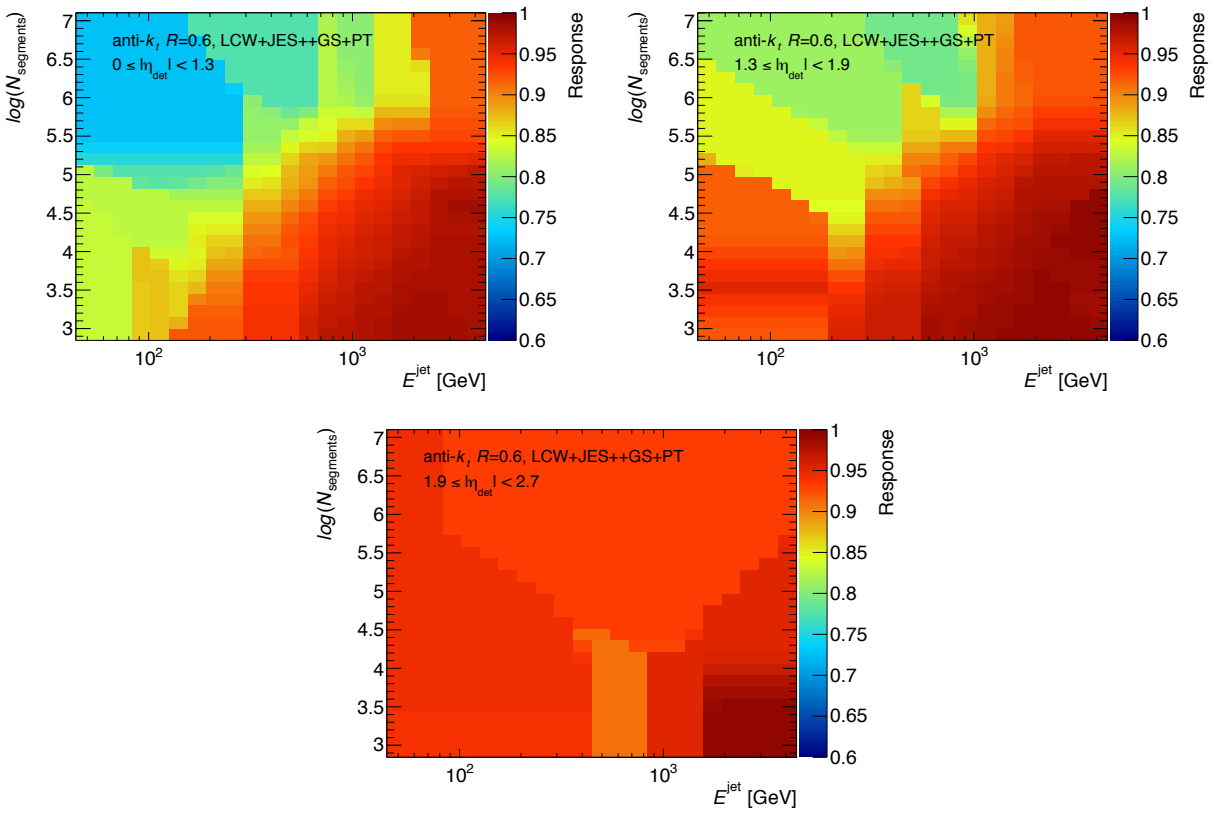


Figure D.8: Final correction function vs E^{jet} vs $N_{segments}$ for jets with $0 \le \eta_{det} < 1.3$ (left), $1.3 \le \eta_{det} < 1.9$ (right) and $1.9 \le \eta_{det} < 2.7$ (bottom).

D.4 Systematic uncertainty on correction for LCW jets

The uncertainties on the LCW+JES+GS-scale jet punch-through correction were derived using the method described in Chapter 6. The triggers used are detailed in Section D.4.1, while the LCW+JES+GS-scale punch-through uncertainties are displayed in Section D.4.2.

D.4.1 Trigger strategy

The trigger strategy for the dijet balance method used to estimate the systematic uncertainties for the jet punch-through correction is described in Section 6.3. Table D.1 shows the $p_{\mathrm{T}}^{\mathrm{jet}}$ regions and the associated fully efficient single jet triggers for LCW+JES+GS-scale jets.

$p_{\mathrm{T}}^{\mathrm{jet}}$ **Range**	**Lowest Prescale Trigger With Efficiency** $>$ 99%
$45 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 61$	EF_j25_a4tchad
$61 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 97$	EF_j35_a4tchad
$97 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 121$	EF_j55_a4tchad
$121 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 161$	EF_j80_a4tchad
$161 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 217$	EF_j110_a4tchad
$217 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 269$	EF_j145_a4tchad
$269 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 329$	EF_j180_a4tchad
$329 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 421$	EF_j220_a4tchad
$421 < p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] \leq 533$	EF_j280_a4tchad
$p_{\mathrm{T}}^{\mathrm{jet}}\ [GeV] > 533$	EF_j145_a4tchad_ht700_L2FS

Table D.1: Table summarising p_{T} ranges and the lowest prescaled trigger with efficiency greater than 99% in each region. See Table D.1 for the trigger naming scheme.

D.4.2 Results

In this section, the LCW+JES+GS-scale plots for the response curves in Section 6.5 are displayed in Section D.4.3, the LCW+JES+GS-scale punch-through uncertainties before the application of the Gaussian kernels are displayed in Section D.4.4 and the final uncertainties are displayed in Section D.4.5.

D.4.3 Reconstructed response curves

Figure D.9 shows the reconstructed response vs $N_{segments}$ before and after applying the jet punch-through correction for LCW+JES+GS-scale jets. As with EM+JES+GS jets (see Figure 6.5), the dependence of the reconstructed response on $N_{segments}$ seen before the application of the correction is removed by the jet punch-through correction, with the reconstructed response corrected back towards 1. The data and MC reconstructed response values agree to within 5% both before and after the application of the correction, with the disparity growing with $N_{segments}$.

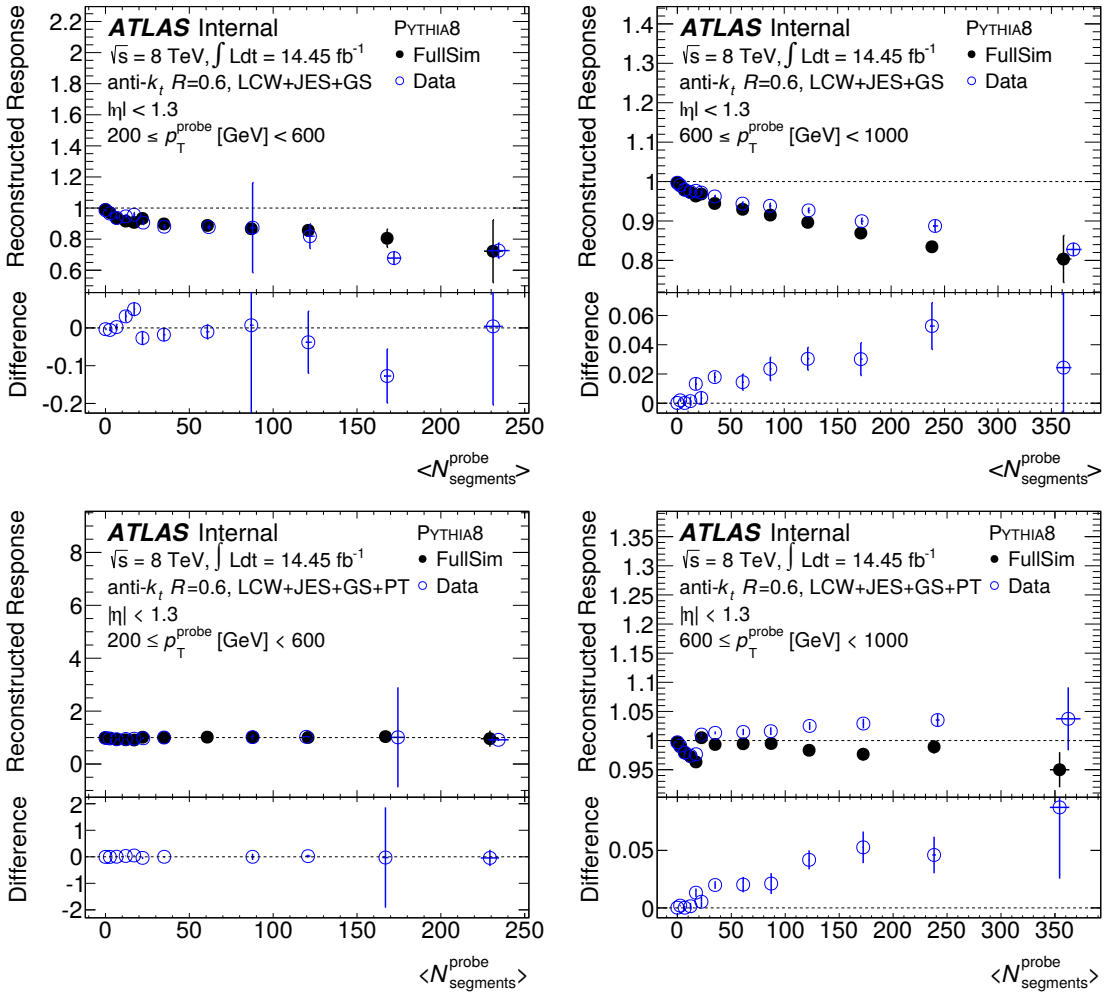


Figure D.9: Reconstructed response as a function of $N_{segments}^{probe}$ compared between data and MC for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}\,[GeV] < 600$ (left) and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}\,[GeV] < 1000$ (right) in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied. The bottom plot shows the difference between data and MC.

Figure D.10 shows the reconstructed asymmetry as a function of $N_{segments}$ for LCW+JES+GS-scale jets (see Figure C.2 for EM+JES-scale jets) before and after the application of the jet punch-through correction. It can be seen the correction removes the increase in the reconstructed asymmetry as $N_{segments}$ increases. Data and MC agree to within 5%.

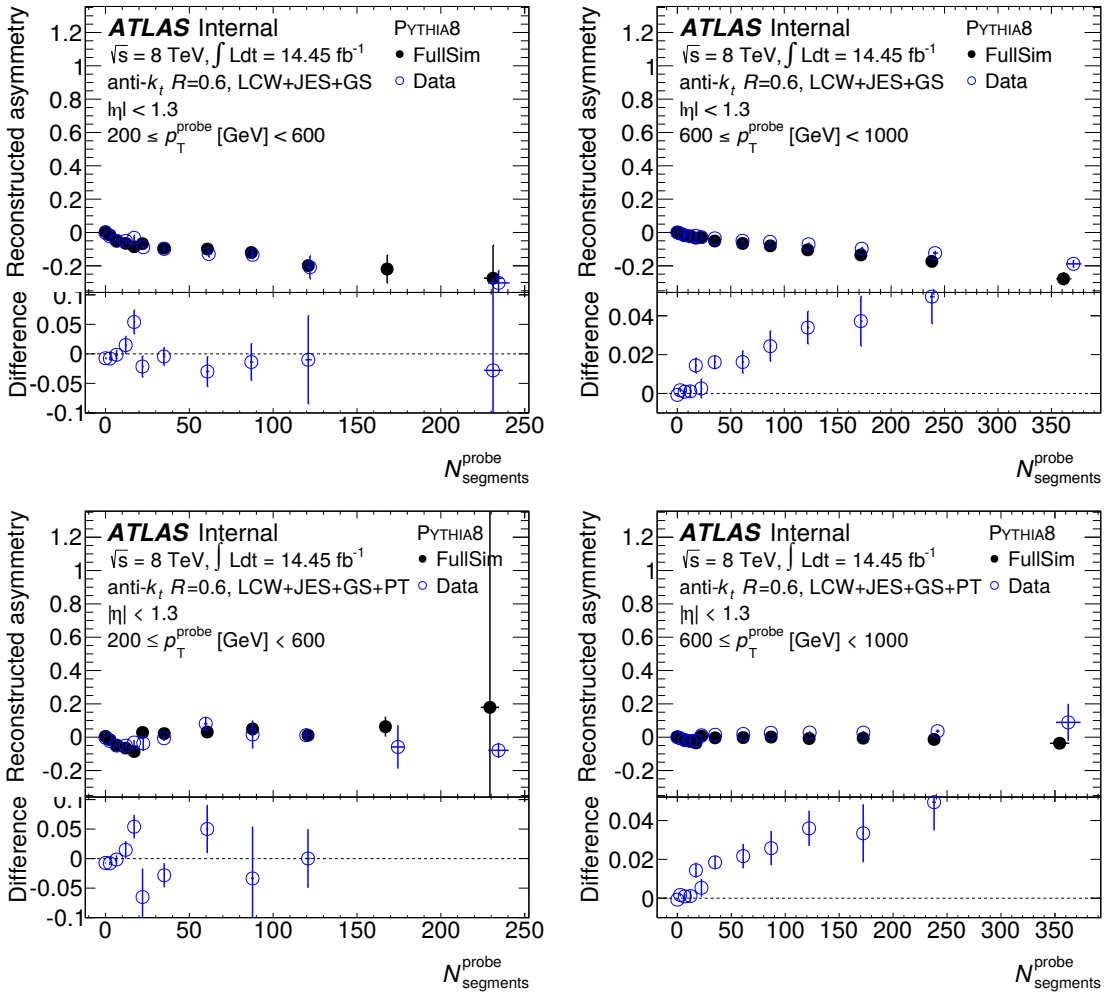


Figure D.10: Reconstructed asymmetry as a function of $N_{segments}^{probe}$ compared between data and MC for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [GeV] < 600$ (left) and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [GeV] < 1000$ (right) in the $0 \leq |\eta_{det}| < 1.3$ region before (top) and after (bottom) the jet punch-through correction applied. The bottom plot shows the difference between data and MC.

The reconstructed response obtained from the reconstructed asymmetry and from fits to the reconstructed response are compared in Figure D.11 for data and MC, after the application of the jet punch-through correction. Again we can see the reconstructed response is corrected back to 1 by the jet punch-through correction, and as with EM+JES-scale jets in Figure C.3, the reconstructed response obtained from both methods for LCW+JES+GS-

scale jets agrees well, to within a couple of percent.

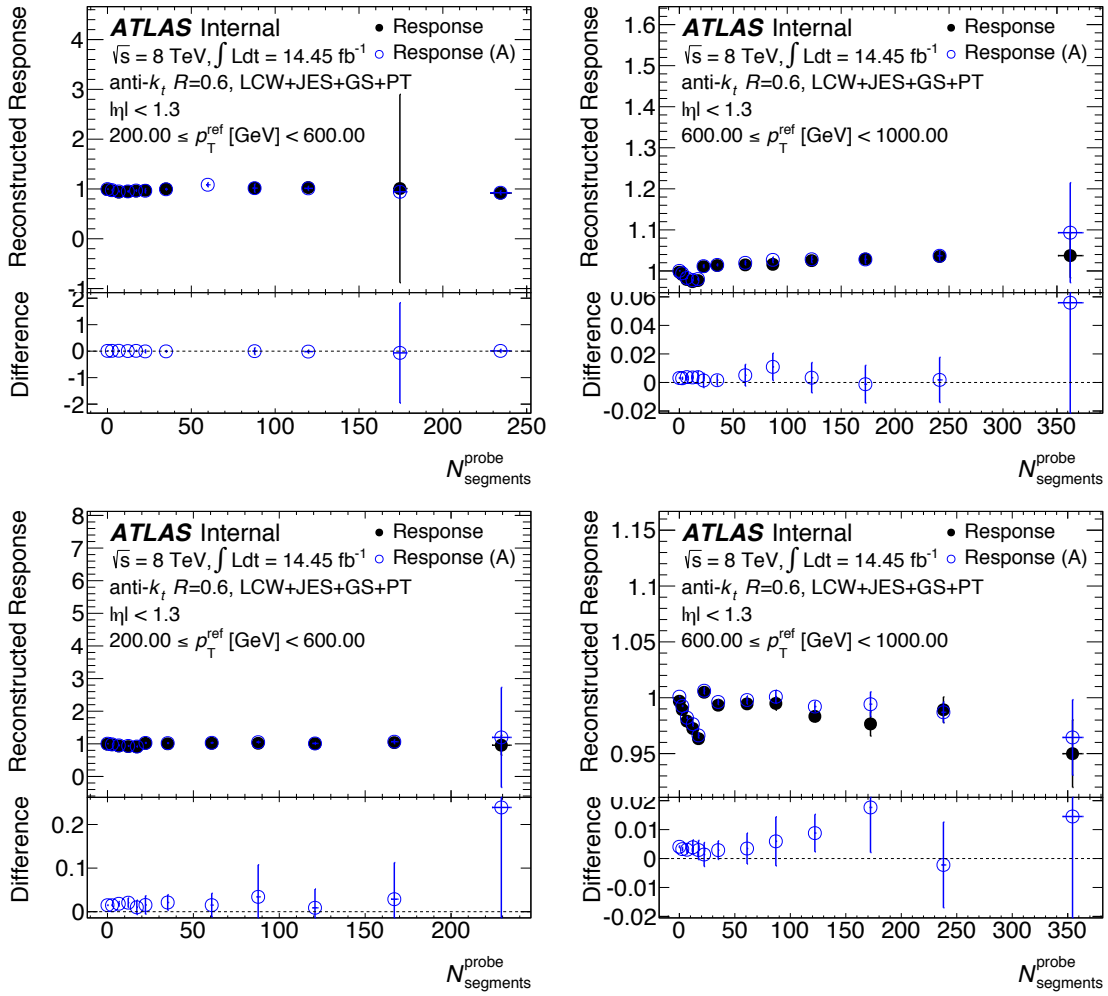


Figure D.11: Reconstructed response as a function of $N^{probe}_{segments}$ calculated directly and from the reconstructed asymmetry for jets with $200 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [GeV] < 600$ (left) and $600 \leq p_{\mathrm{T}}^{\mathrm{probe}}\ [GeV] < 1000$ (right) in the $0 \leq |\eta_{det}| < 1.3$ region for data (top) and MC (bottom). The bottom plot shows the difference between the reconstructed response calculated via the asymmetry with the reconstructed response obtained directly.

D.4.4 Uncertainties from fits

The uncertainties on the LCW+JES+GS-scale jet punch-through correction obtained by comparing the mean reconstructed response extracted from fits for data and MC is shown in Figure D.12 (see Figure 6.6 for EM+JES-scale jets), as a function of $N_{segments}$ and $p_{\mathrm{T}}^{\mathrm{jet}}$ in all three η_{det} regions. The uncertainties can be seen to reach 9% in the $0 \leq \eta_{det} < 1.3$ region, increasing as $p_{\mathrm{T}}^{\mathrm{jet}}$ decreases as $N_{segments}$ increases. This is the same trend as was seen

for EM+JES-scale jets.

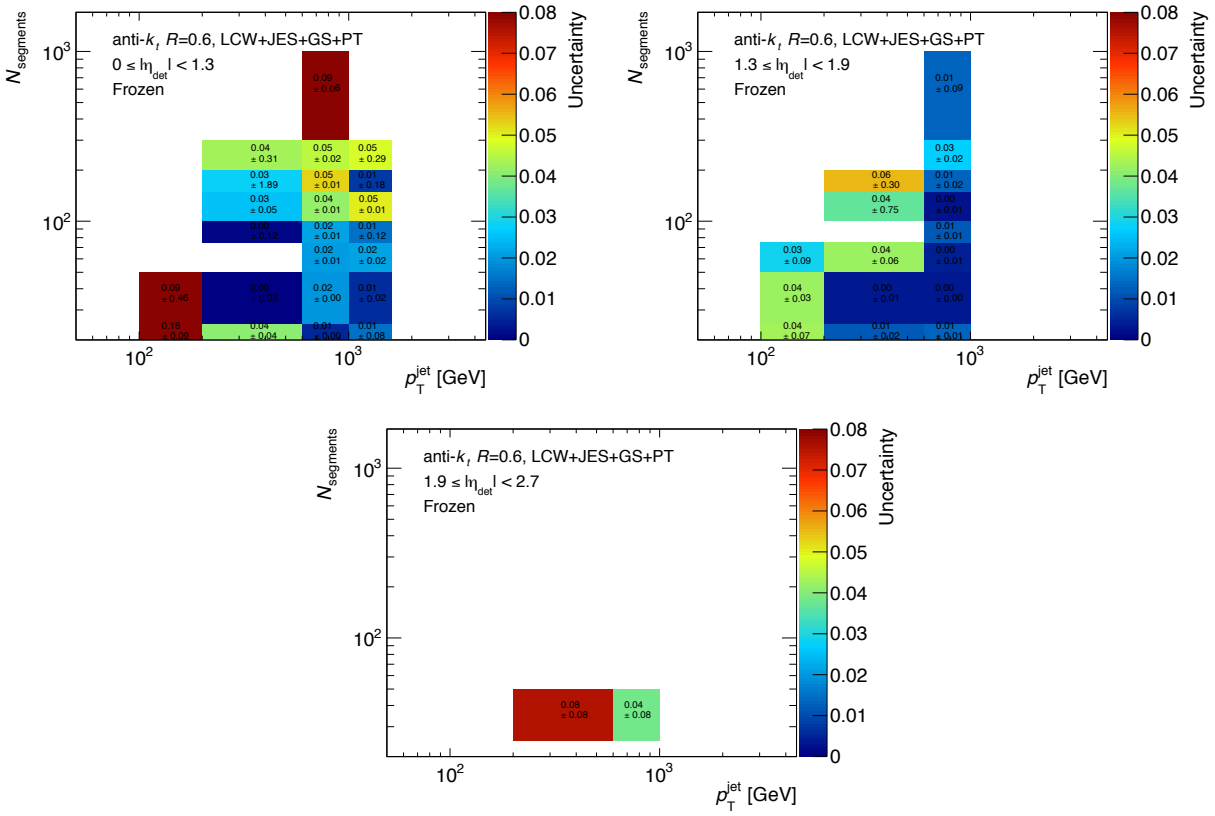


Figure D.12: Uncertainty vs $N_{segments}$ vs p_T^{jet} for jets with $0 \le \eta_{det} < 1.3$ (left), $1.3 \le \eta_{det} < 1.9$ (right) and $1.9 \le \eta_{det} < 2.7$ (bottom).

D.4.5 Final LCW jet punch-through uncertainties

The final jet punch-through uncertainties, after the application of the one and two dimensional Gaussian kernels, for LCW+JES+GS-scale jets, are shown in Figure D.13 for all three η_{det} regions. Both the frozen and extrapolated uncertainties are shown. As with the EM+JES-scale jets (see Figure 6.7), the uncertainty in the $1.9 \le \eta_{det} < 2.7$ region is the average of the two uncertainty values in Figure D.12, weighted by their error, calculated to be $\sim 5\%$. The extrapolated uncertainties are seen to be largest, reaching 12% in the lowest p_T^{jet}, $N_{segments}$, $0 \le \eta_{det} < 1.3$ bin, while the frozen uncertainties reach 8% in the same region.

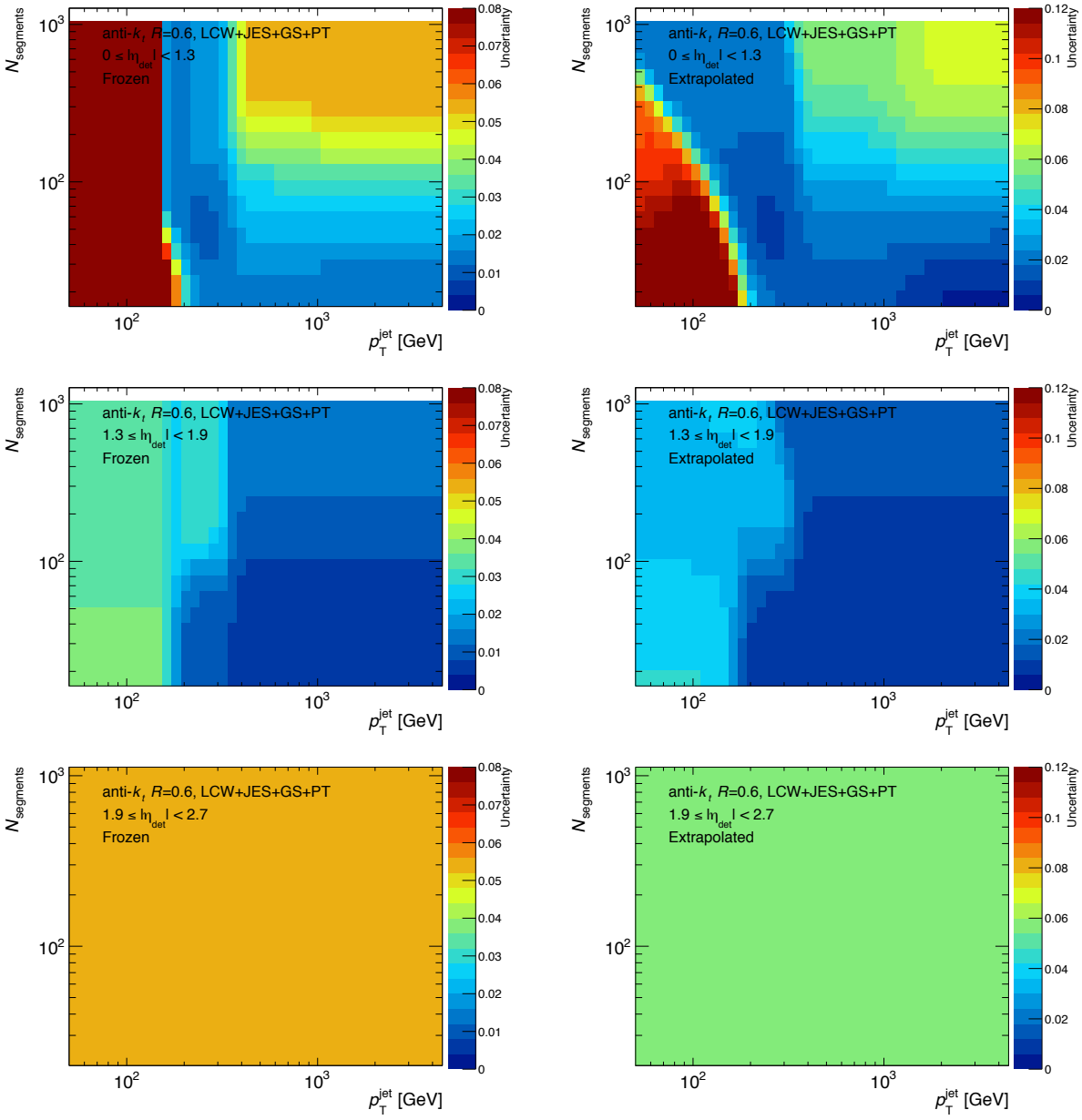


Figure D.13: Final uncertainty a function of $N_{segments}$ and p_T^{jet} for jets with $0 \le \eta_{det} < 1.3$ (top), $1.3 \le \eta_{det} < 1.9$ (middle) and $1.9 \le \eta_{det} < 2.7$ (bottom). Plots on the left show the frozen uncertainty, while plots on the right show the extrapolated uncertainty.

D.5 Performance of LCW punch-through correction

This section discusses the performance of the LCW jet punch-through correction in terms of its impact on the fully corrected LCW jet energy scale and jet energy resolution. The plots are the LCW equivalent to those shown in Chapter 7.

D.5.1 Impact on the jet energy scale

From Figure D.19, it can be seen that the jet energy response distribution for LCW jets is still Gaussian after the application of the jet punch-through correction, with the correction mainly impacting the low response tail (see Section D.5.2 for more detailed studies).

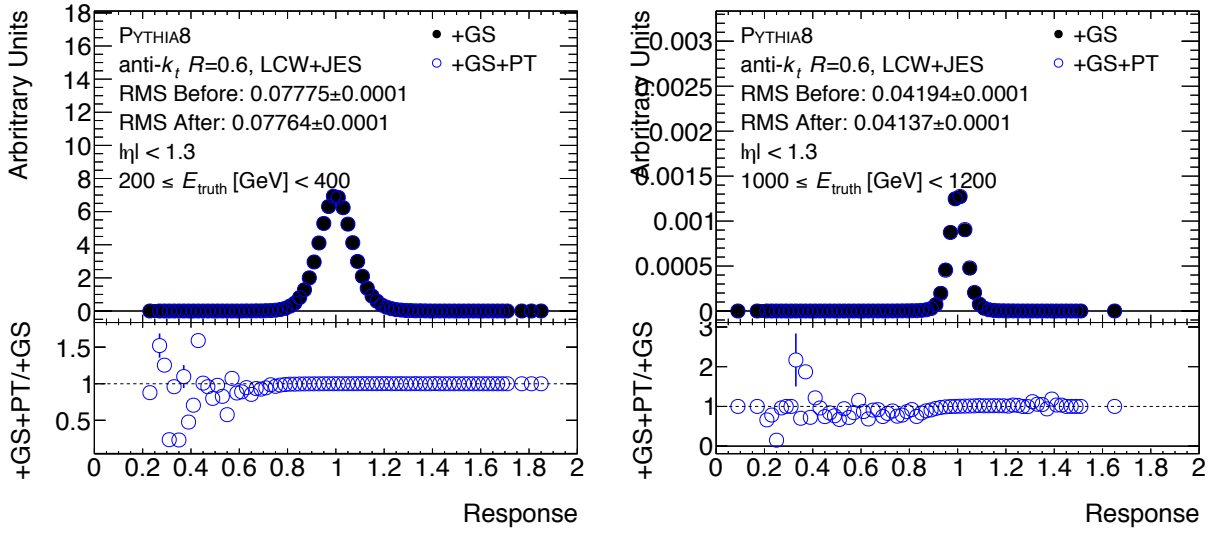


Figure D.14: Jet energy response distribution integrated over $N_{segments}$ compared before and after the jet punch-through correction, for jets in the $0 \leq |\eta_{det}| < 1.3$ region with $200 \leq E_{truth}\ [GeV] < 400$ (left) and $1000 \leq E_{truth}\ [GeV] < 1200$ (right). The bottom plots show the bin-by-bin ratio of entries after the correction to before the correction.

The effect of the jet punch-through correction on the dependence of the jet energy response on $N_{segments}$ for LCW jets is shown in Figure D.15. The correction works reasonably well, largely removing the dependence of the jet energy response on $N_{segments}$.

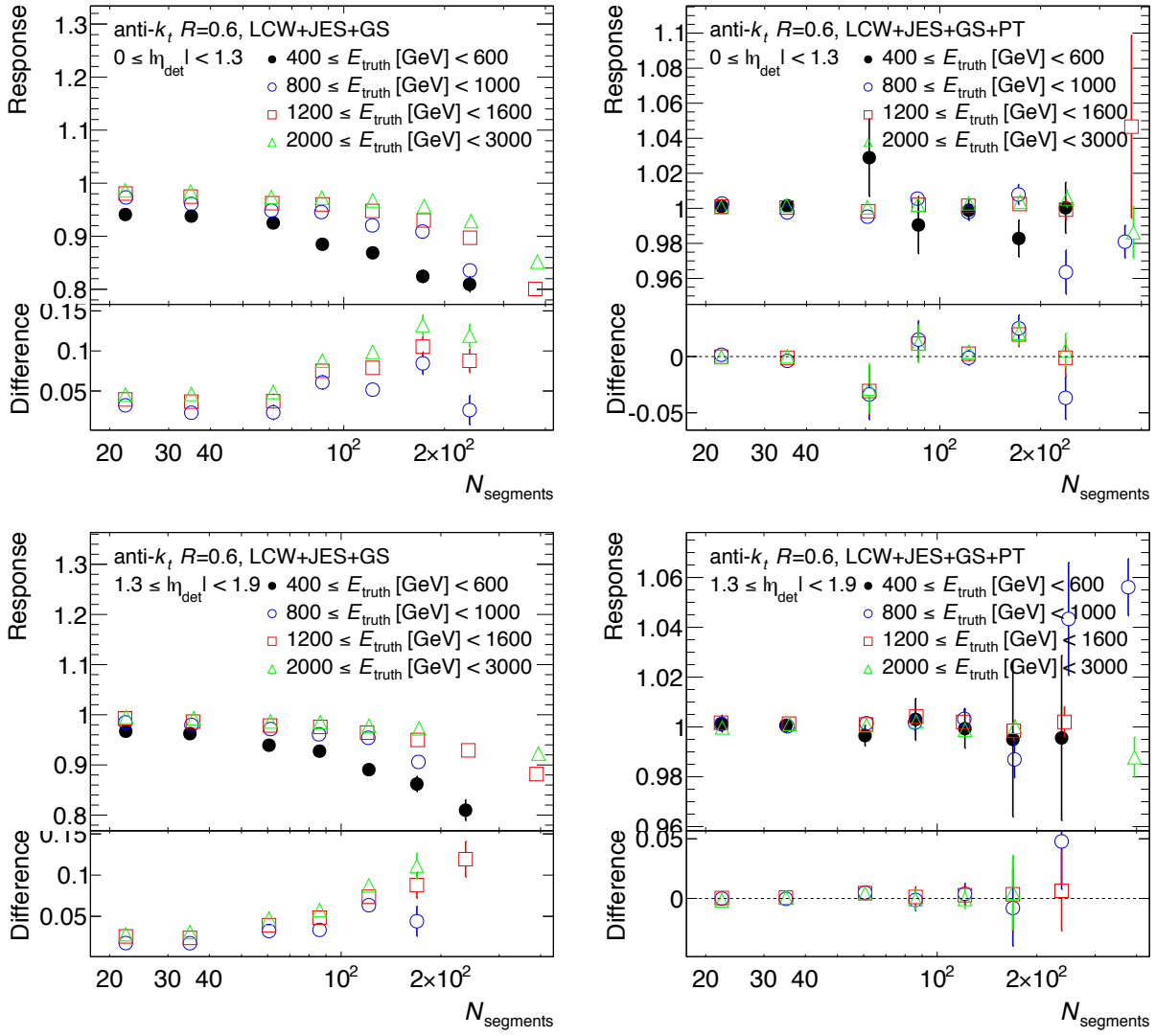


Figure D.15: Jet energy response vs $N_{segments}$ before (left) and after (right) the jet punch-through correction, for jets in the $0 \le |\eta_{det}| < 1.3$ region (top) and $1.3 \le |\eta_{det}| < 1.9$ region (bottom). The bottom plots show the difference in the response relative to the $400 \le E_{truth}\ [GeV] < 600$ bin.

Finally, Figure D.16 compares the jet energy response integrated over $N_{segments}$ as a function of E_{truth} for each of the three η_{det} regions, before and after the application of the jet punch-through correction. It is clear from this figure that the jet punch-through correction has a negligible affect on the mean JES ($< 0.1\%$). As mentioned previously in this thesis, the jet punch-through correction aims to remove the dependence of the jet energy scale on $N_{segments}$, whilst also leaving the mean JES binned in η_{det} and E_{truth} the same. Thus it can be concluded the jet punch-through correction works as designed for LCW+JES+GS-scale jets, as it also does for EM+JES-scale jets.

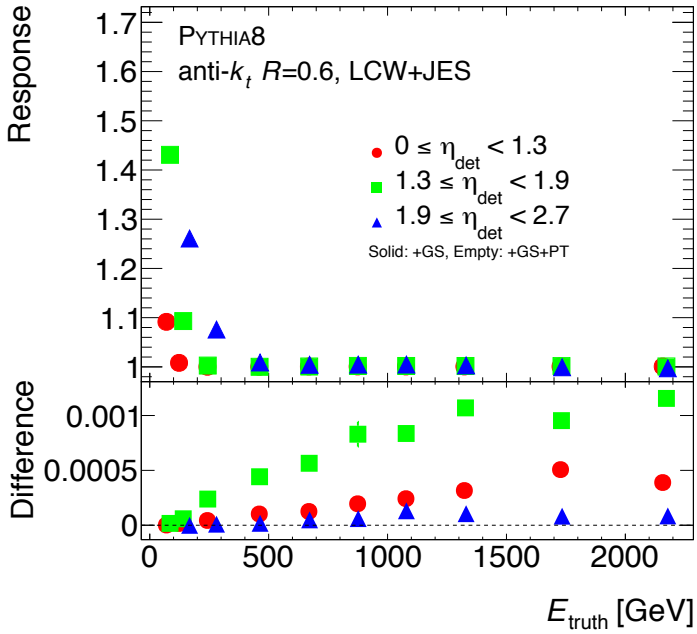


Figure D.16: Jet energy response vs E_{truth} before (solid circles) and after (empty circles) the application of the punch-through correction for jets in each η_{det} region. The bottom plot shows the difference in the response before and after the application of the correction.

D.5.2 Impact on jet energy response tail

The percentage of LCW jets below a response threshold before and after the application of the jet punch-through correction is show in Figure D.19, from which it can be seen the percentage of jets in the low response tail is reduced by the correction.

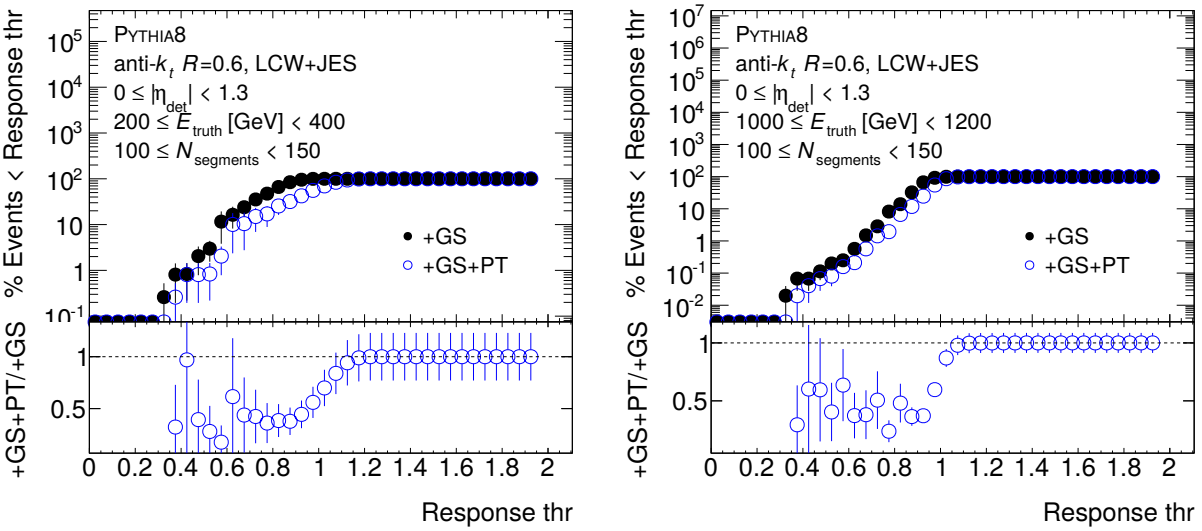


Figure D.17: Percentage of events below a given response threshold value compared for jets with and without the jet punch-through correction, in the $100 \le \mathrm{N}_{segments} < 150$ bin, with $200 \le \mathrm{E}_{truth}\ [GeV] < 400$ (left) and $1000 \le \mathrm{E}_{truth}\ [GeV] < 1200$ (right). The bottom plots show the ratio of the percentage of events below the threshold before and after the correction.

The -2σ and -1σ quantile positions of the response distribution histograms are shown in Figure D.19, before and after the application of the correction, as a function of E_{truth}, in each of the three η_{det} regions. The correction can be seen to push the position of the -2σ and -1σ quantiles to higher response values, with the shift growing as E_{truth} increases, and largest in the $1.9 \le \eta_{det} < 2.7$ region.

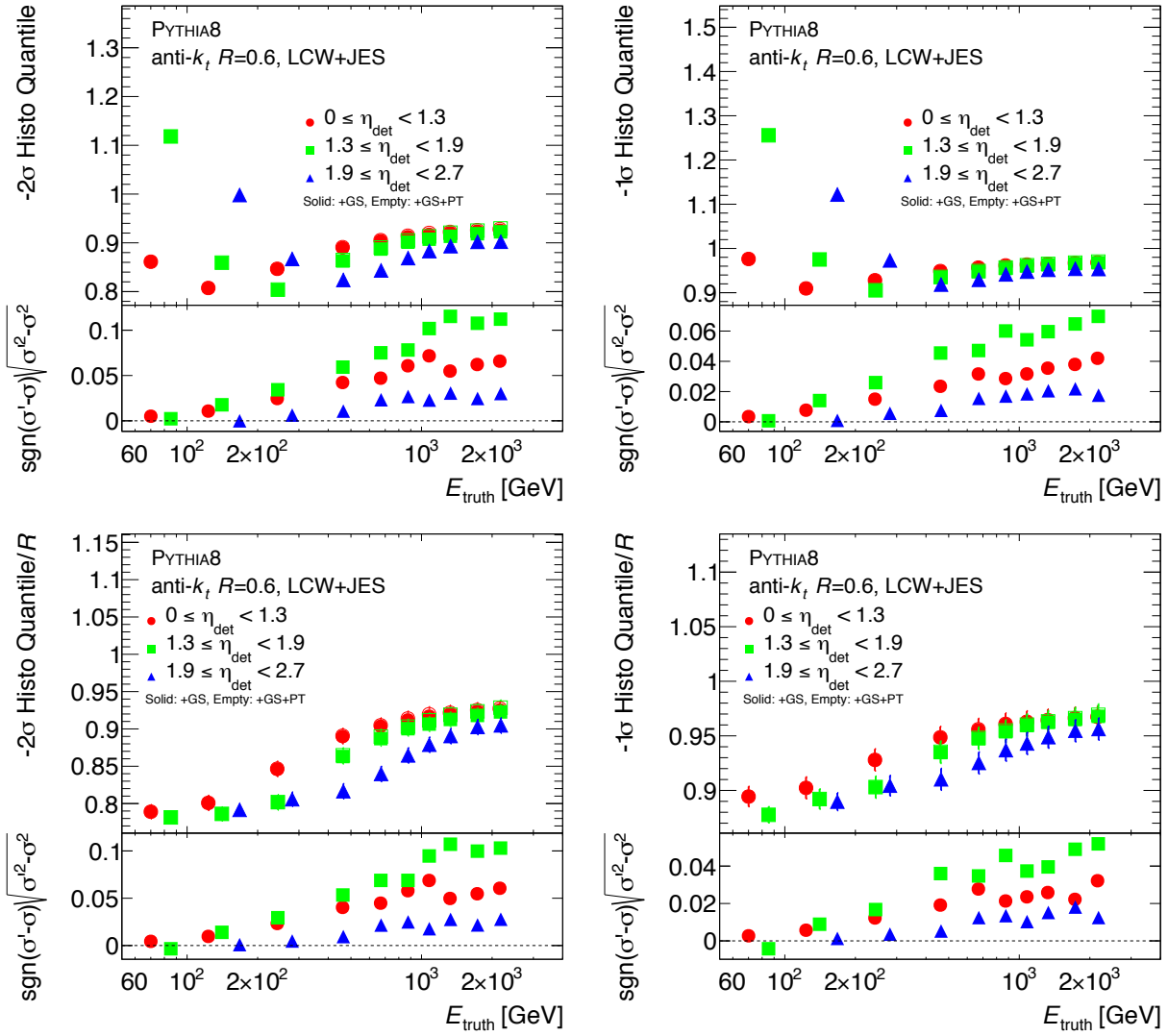


Figure D.18: Negative 1 and 2 σ response distribution quantiles (top) relative to the histogram mean response (bottom) as a function of E_{truth} in each of the three η_{det} bins, before (solid circles) and after (empty circles) the punch-through correction is applied. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

The -2σ and -1σ fit quantile positions, before and after the application of the correction, are shown in Figure D.19. As with the histogram quantiles, the fit quantiles can be, in general, seen to be pushed to higher response values by the correction. However as the fit is only done to the $\mu \pm 1.6\sigma$ core of the distribution, the shift in fit quantiles is not as substantial as the histogram quantiles, as the fit does not completely capture the behaviour of the response tails.

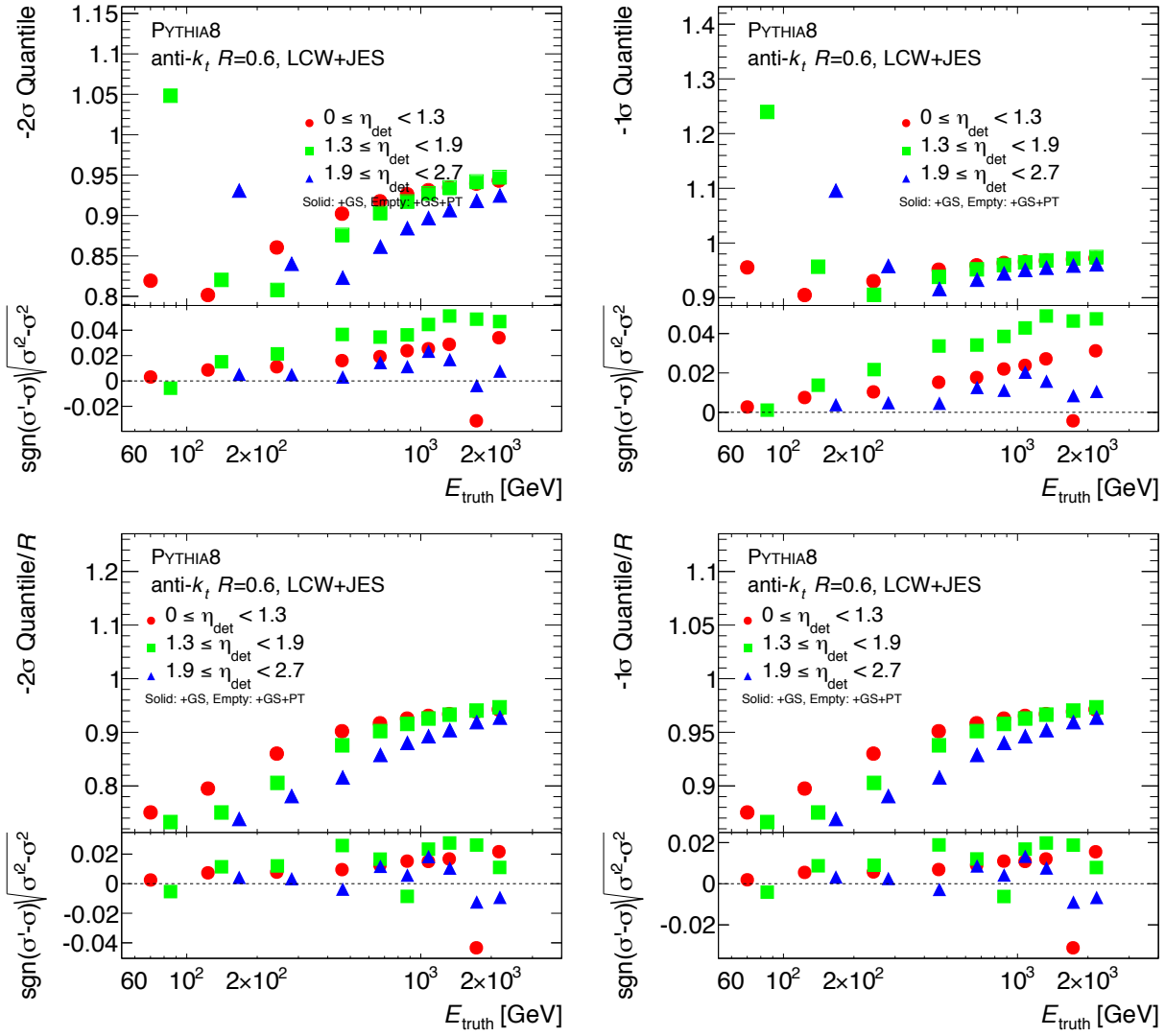


Figure D.19: Negative 1 and 2 σ response distribution quantiles extracted from fits (top) relative to the fit mean response (bottom) as a function of E_{truth} in each of the three η_{det} bins, before (solid circles) and after (empty circles) the punch-through correction is applied. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

Overall it can be seen that the correction reduces the low response tails for LCW+JES+GS-scale jets, as was also observed for EM+JES-scale jets in Section 7.1.2.

D.5.3 Impact on jet energy resolution

The jet energy resolution before and after the application of the correction as a function of $N_{segments}$ for LCW jets is shown in Figure D.20. The correction can be seen to improve the resolution by up to 10% in the $0 \leq |\eta_{det}| < 1.3$ and $1.3 \leq |\eta_{det}| < 1.9$ regions, but has little affect in the $1.9 \leq |\eta_{det}| < 2.7$ region. The dependence of the resolution on $N_{segments}$ is not completely removed in any region however. This is consistent with what was seen for EM+JES-scale jets in Figure 7.8.

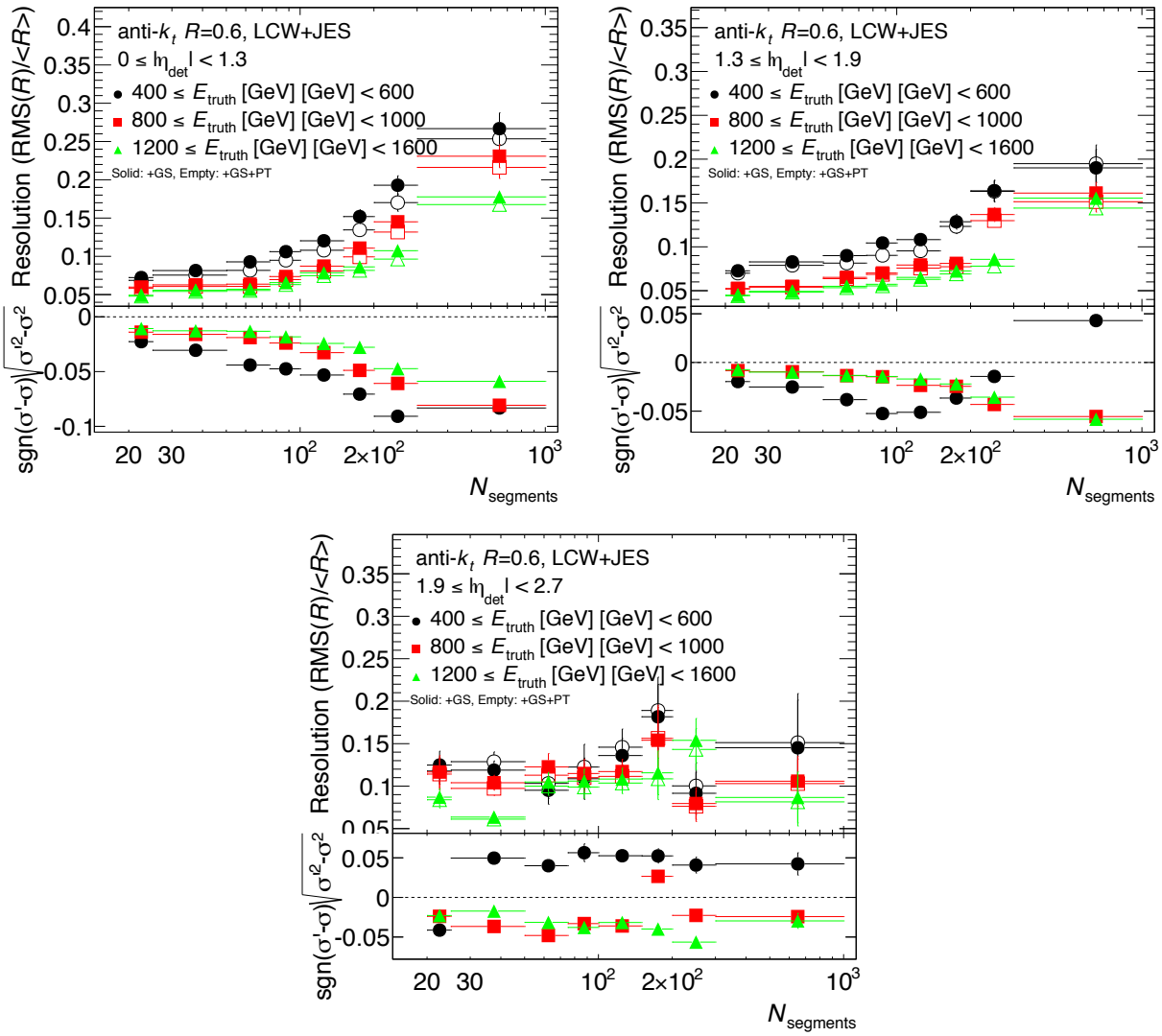


Figure D.20: The RMS($\mathcal{R}$) divided by the mean histogram response (left) as a function of $N_{segments}$ in each of the three η_{det} regions for various E_{truth} bins, before (solid circles) and after (empty circles) the application of the jet punch-through correction. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

The impact of the jet punch-through correction on the overall jet energy resolution when integrating over $N_{segments}$, with the resolution estimated from the RMS and as σ of a Gaussian fit to the jet energy response distribution, is shown in Figure D.21. Again, as the fit only captures the $\mu \pm 1.6\sigma$ core of the distribution, the impact of the correction on the low response tail is largely neglected, and thus very little improvement in the resolution extracted from the fit can be seen. However when looking at the resolution in terms of the RMS, a clear improvement of the resolution can be seen. The improvement is largest in the $1.3 \leq |\eta_{det}| < 1.9$ region, reaching up to 1%, and smallest in the $1.9 \leq |\eta_{det}| < 2.7$ region.

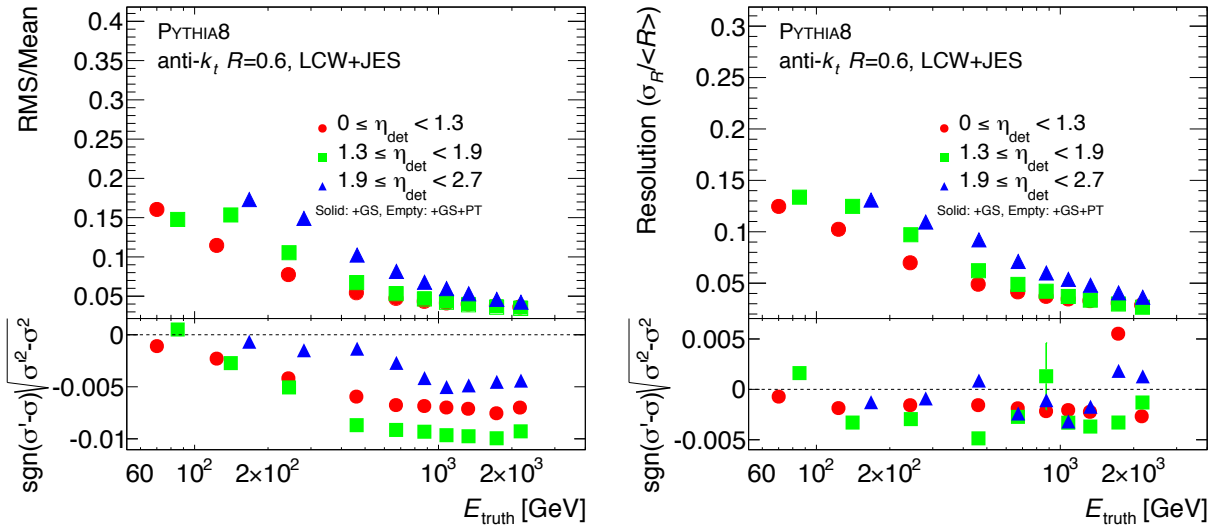


Figure D.21: The RMS($\mathcal{R}$) divided by the mean histogram response (left) and the jet energy resolution extracted from the fit (right) as a function of E_{truth} in each of the three η_{det} regions, before (solid circles) and after (empty circles) the application of the jet punch-through correction. Bottom plots show the difference in quadrature before (σ) and after (σ') the correction.

Appendix E

In-situ validation of GS corrections

This appendix details the data in-situ validation of the final 2012 GS corrections described in Section 8.1.1, and is adapted from Reference [1]. Section E.1 outlines the dijet balance method used, Section E.2 details the event selection applied to select for well balanced dijet events, the trigger strategy is then outlined in Section E.3, and finally the input variable distributions and balance in data and MC are compared in Section E.4.

Please note that as this appendix refers to studies using the final 2012 GS corrections, the EM+JES+GS and LCW+JES+GS scales referred to here include the jet punch-through correction derived in this thesis. The validation was done for anti-k_t jets, with radius R = 0.4 and R = 0.6, calibrated with the EM+JES and LCW+JES schemas. Only plots using anti-k_t R = 0.4 jets calibrated with the EM+JES scheme are shown.

E.1 Overview of dijet balance method

The GS corrections can be validated using a variation on the tag-and-probe method used to validate the jet punch-through correction, described in Chapter 6. When validating the jet punch-through correction, one can select the probe to be a jet experiencing punch-through, and the reference jet to be a fully contained jet. However for the other GS corrections, a clear choice for the probe and reference jet is not always available. As a result, choosing

which of the two back-to-back jets is the reference jet and which is the probe jet is arbitrary, and the events are always used twice, inverting the roles of reference and probe. The jet whose calorimeter response dependence on the jet structure variables is studied is referred to as the probe jet, while the other is referred to as the reference jet. The calorimeter response to jets is measured through the asymmetry variable is then defined as:

$$A(x) = \frac{p_{\mathrm{T}}^{\mathrm{probe}}(x) - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{avg}}(x)}, \tag{E.1}$$

where x is any of the variables used in the GS corrections (see Table 3.1) and $p_{\mathrm{T}}^{\mathrm{avg}}$ is the average p_{T} of the probe and reference jet[1]. Both p_T^{probe} and $p_{\mathrm{T}}^{\mathrm{ref}}$ depend on x, but the dependence is explicitly written only for the probe jet because the jet property used to build the correction is the one of the probe jet. The probe and reference jet are both calibrated with the same scheme. When computing correction number i, they are both corrected up to the $(i-1)^{\mathrm{th}}$ correction (see Sec. 3.5.4). The mean response as a function of x is given by:

$$\langle \mathcal{R}(x) \rangle = \left\langle \frac{p_{\mathrm{T}}^{\mathrm{probe}}}{p_{\mathrm{T}}^{\mathrm{ref}}} \right\rangle, \tag{E.2}$$

which correspond to a good approximation to:

$$\langle \mathcal{R}(x) \rangle = \frac{1 + \langle A(x) \rangle / 2}{1 - \langle A(x) \rangle / 2}. \tag{E.3}$$

Equations E.2 and E.3 are analogous to Equations C.1 and C.2 respectively from the jet punch-through tag-and-probe method. For this method, Equation E.3 is used to calculate the average jet response instead of Equation E.2 since $A(x)$ is Gaussian in bins of fixed $p_{\mathrm{T}}^{\mathrm{avg}}$, while the jet response distribution $\mathcal{R}(x)$, calculated as $p_{\mathrm{T}}^{\mathrm{probe}}/p_{\mathrm{T}}^{\mathrm{ref}}$ in bins of $p_{\mathrm{T}}^{\mathrm{ref}}$, is not[2].

As with the jet punch-through validation in Chapter 6, the validation of the GS correc-

[1]The average transverse momentum of the probe and the reference jet is defined as $p_{\mathrm{T}}^{\mathrm{avg}} = (p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}})/2$.

[2]This was not the case for the reconstructed response distributions when using the tag-and-probe method in Chapter 6, where both the reconstructed response and reconstructed asymmetry distributions were Gaussian.

tions involves verifying both the distributions of input jet property variables, and also the relationship between jet response and jet property variables, are well modelled.

E.2 Event selection

The dijet selection used for the GS data validation is a modified version of the event selection used for the jet punch-through correction data validation studies in Chapter 6.

Events are required to have at least one vertex with 2 or more tracks, with $p_{\mathrm{T}}^{\mathrm{track}} > 150$ MeV [1]. The longitudinal position of the vertex is required to be within 10 cm of the detector centre.

Events where the two leading jets have a $p_{\mathrm{T}}^{\mathrm{avg}} > 20$ GeV and $\Delta\phi > 2.5$ radians are used, where $\Delta\phi$ is the azimuthal angle between the two leading jets in the event and $p_{\mathrm{T}}^{\mathrm{avg}}$ correspond to the average p_{T} of the two leading jets. A cut on the Jet Vertex Fraction (see Section 3.6 ($|\mathrm{JVF}| > 0.25$) is used to reduce the effect of in-time and out-of-time pile-up for jets with $|\eta_{det}| < 2.4$ and $p_{\mathrm{T}} < 50$ GeV. If there is a third jet (J_3), additional requirements in the event are applied:

- If $|\eta_{J_3}| \geq 2.4$, $p_{\mathrm{T}}^{J_3}$ is required to be smaller than max(10 GeV, $0.4p_{\mathrm{T}}^{\mathrm{avg}}$).
- If $|\eta_{J_3}| < 2.4$, $p_{\mathrm{T}}^{J_3}$ is required to be smaller than max(12 GeV, $0.4p_{\mathrm{T}}^{\mathrm{avg}}$) and a $|\mathrm{JVF}| > 0.25$ for $p_{\mathrm{T}}^{J_3} < 50$ GeV is applied.

The same event selection criteria are used in the MC simulation and in the data. The bin boundaries are defined in Table E.1. For some variables, bins were merged in order to obtain higher statistics. Unlike the method in Chapter 6, this method requires that both jets be in the same $|\eta_{det}|$ bins to avoid the η-intercalibration contributions that this method is more sensitive to due to the extended η_{det} range.

The trigger selection applied is defined in Section E.3.

$\lvert\eta_{det}\rvert$	0.0	0.3	0.8	1.2	1.5	1.8	2.1	2.4	2.8	3.2	3.6	4.5
$p_{\mathrm{T}}^{\mathrm{avg}}$ (GeV)	20	30	40	60	80	110	160	210	260	310	400	800

Table E.1: Bin boundaries used for the GS dijet balance studies.

E.3 Trigger strategy

Table E.2: Trigger selection used anti-k_t R=0.4 jets

p_{T} range	Central Trigger ($\eta < 2.8$)	Forward Trigger ($\eta \geq 2.8$)
30 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 45 GeV	EF_j15_a4tchad	EF_fj15_a4tchad
45 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 60 GeV	EF_j25_a4tchad	EF_fj25_a4tchad
60 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 75 GeV	EF_j35_a4tchad	EF_fj35_a4tchad
75 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 90 GeV	EF_j45_a4tchad	EF_fj45_a4tchad_L2FS
90 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 125 GeV	EF_j55_a4tchad	EF_fj55_a4tchad_L2FS
125 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 160 GeV	EF_j80_a4tchad	EF_fj80_a4tchad
160 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 185 GeV	EF_j110_a4tchad	EF_fj110_a4tchad
185 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 240 GeV	EF_j145_a4tchad	EF_fj145_a4tchad
240 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 300 GeV	EF_j180_a4tchad	EF_fj180_a4tchad
300 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 480 GeV	EF_j220_a4tchad	EF_fj220_a4tchad
480 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 600 GeV	EF_j360_a4tchad	
$p_{\mathrm{T}}^{\mathrm{jet}}$ > 600 GeV	EF_j460_a4tchad	

Table E.3: Trigger selection used anti-k_t R=0.6 jets

p_{T} range	Central Trigger ($\eta < 2.8$)	Forward Trigger ($\eta \geq 2.8$)
40 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 55 GeV	EF_j15_a4tchad	EF_fj15_a4tchad
55 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 70 GeV	EF_j25_a4tchad	EF_fj25_a4tchad
70 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 85 GeV	EF_j35_a4tchad	EF_fj35_a4tchad
85 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 100 GeV	EF_j45_a4tchad	EF_fj45_a4tchad_L2FS
100 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 135 GeV	EF_j55_a4tchad	EF_fj55_a4tchad_L2FS
135 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 180 GeV	EF_j80_a4tchad	EF_fj80_a4tchad
180 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 200 GeV	EF_j110_a4tchad	EF_fj110_a4tchad
200 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 240 GeV	EF_j145_a4tchad	EF_fj145_a4tchad
240 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 300 GeV	EF_j180_a4tchad	EF_fj180_a4tchad
300 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 480 GeV	EF_j220_a4tchad	EF_fj220_a4tchad
480 GeV< $p_{\mathrm{T}}^{\mathrm{jet}}$ <= 600 GeV	EF_j360_a4tchad	
$p_{\mathrm{T}}^{\mathrm{jet}}$ > 600 GeV	EF_j460_a4tchad	

As the GS data validation studies extend into the more forward detector regions ($\eta \geq 2.8$), the trigger strategy used is different from that in Section 6.3. The trigger selection still uses single jet triggers, but the selection is designed such that the trigger efficiency is greater than 99% for a specific region of $p_{\mathrm{T}}^{\mathrm{avg}}$, instead of $p_{\mathrm{T}}^{\mathrm{jet}}$. As a result, this strategy does not use a combination of triggers for each event, but instead uses a single trigger. This trigger is either a forward or central trigger, depending on whether the jets fall into the region $\eta_{det} \geq 2.8$

or $\eta_{det} < 2.8$. A list of the trigger selection used can be found in Tables E.2 and E.3.

E.4 Results

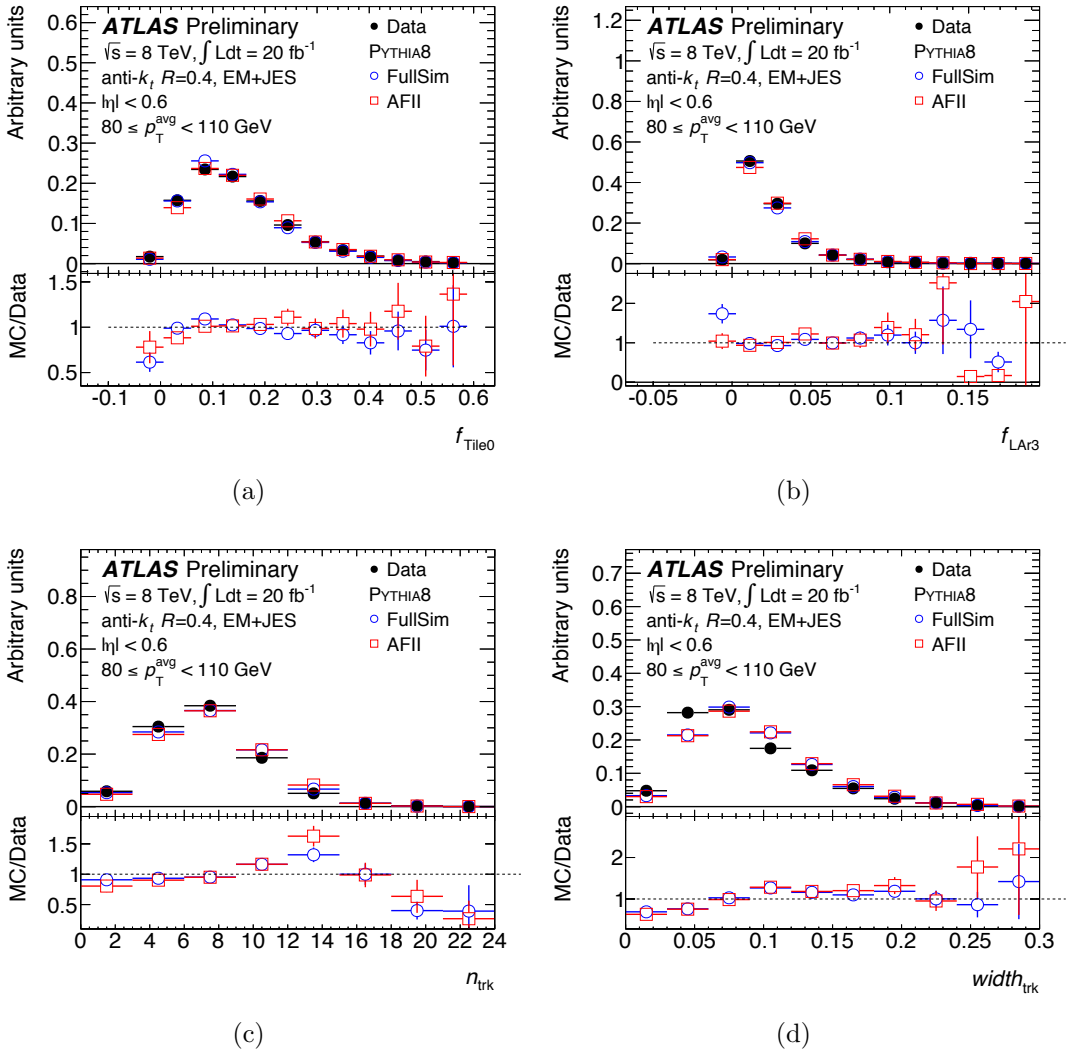
The input jet property distributions are compared between data and MC in Section E.4.1, the dijet asymmetry data/MC comparisons are shown in Section E.4.2, and comparisons between the jet punch-through uncertainties derived using the method in Chapter 6 and the method described in this appendix are shown in Section E.4.3.

All MC corrections in the final ATLAS 2012 jet energy calibration chain described in Section 8.1 were derived for both full and fast simulation samples. This includes the GS corrections, with the exception of the jet punch-through correction, for reasons mentioned in Section 2.3. Thus for the first four GS corrections, plots compare both fast (AFII) and full simulation with data.

E.4.1 Input jet property distributions

Figures E.1 and E.2 show the distributions of the f_{Tile0}, f_{LAr3}, n_{trk}, $width_{trk}$ and $N_{segments}$ variables in data and inclusive jet MC events. Distributions for the two leading jets reconstructed with the anti-k_t algorithm with $R = 0.4$ and calibrated at the EM+JES scale with $|\eta_{det}| < 0.6$ and $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV and $600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV bins are shown. The $N_{segments}$ variable is not shown for $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV due to the low number of events in data. Compared to data, the reference Pythia8 MC exhibit higher (lower) values for the f_{LAr3}, n_{trk} and $width_{trk}$ (f_{Tile0} and $N_{segments}$) variables. Similar results are found in the other $p_{\mathrm{T}}^{\mathrm{avg}}$ bins. In general, lower values of the f_{Tile0}, f_{LAr3} variables are observed for the full simulation with respect to AFII. As tracking variables are modelled in the same way in both MC simulations, differences observed are due to statistical fluctuations. These differences in the input distributions of the jet structure variables have little impact as long as the average jet response as a function of the variables is well described by the simulation, as will be

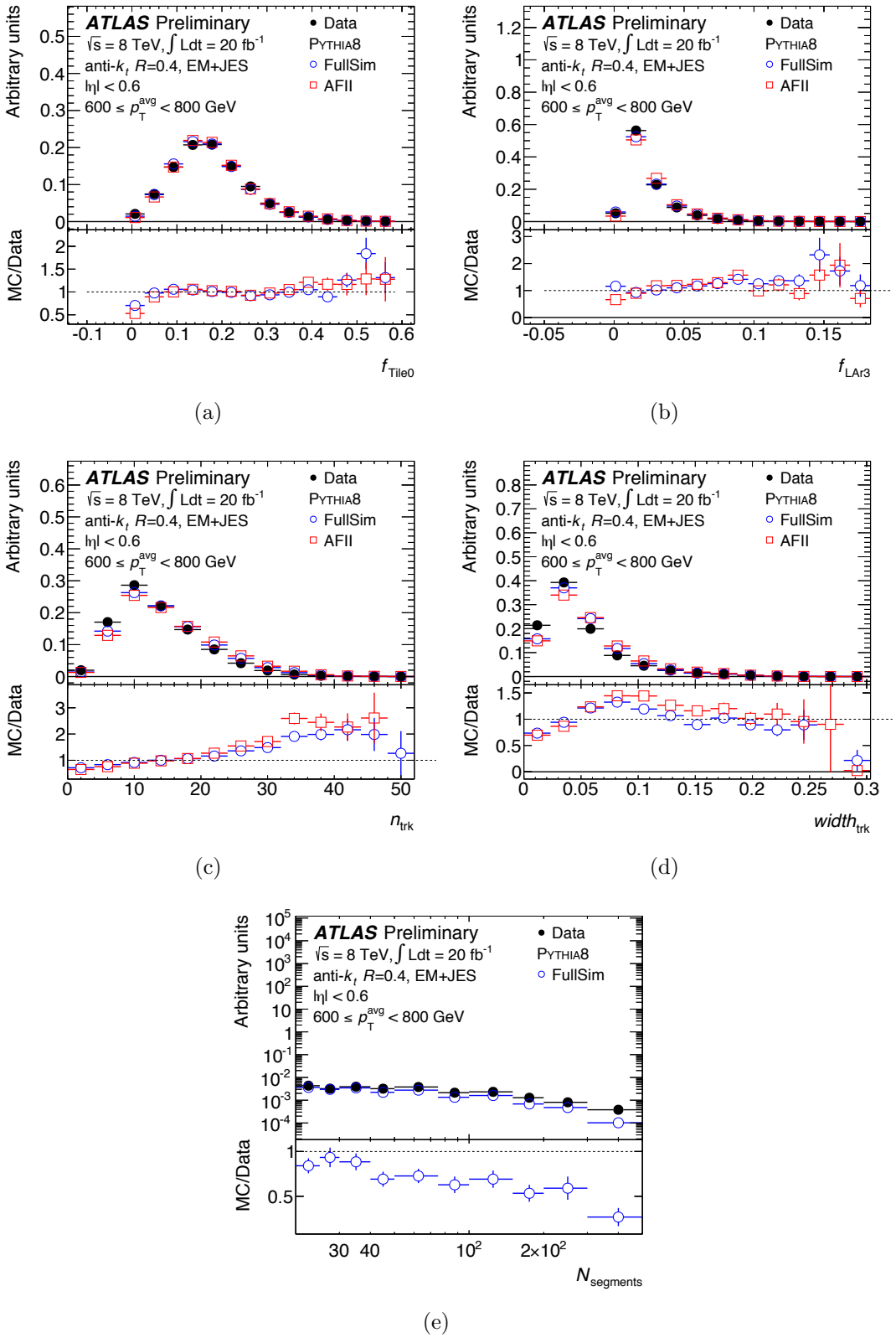
assessed in Section E.4.2.



(a) (b)

(c) (d)

Figure E.1: Distributions of the f_{Tile0} (a), f_{LAr3} (b), n_{trk} (c) and $width_{trk}$ (d) variables for jets with $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV and $|\eta_{det}| < 0.6$ in the data and the PYTHIA8 Monte Carlo simulation (full and fast simulation). Jets are reconstructed with the anti-k_t algorithm with $R = 0.4$ and calibrated with the EM+JES scheme.



(a) (b) (c) (d) (e)

Figure E.2: Distributions of the f_{Tile0} (a), f_{LAr3} (b), n_{trk} (c), $width_{trk}$ (d) and $N_{segments}$ (e) variables for jets with $600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV and $|\eta_{det}| < 0.6$ in the data and the PYTHIA8 Monte Carlo simulation (full and fast simulation, except for $N_{segments}$). Jets are reconstructed with the anti-k_t algorithm with $R = 0.4$ and calibrated with the EM+JES scheme.

E.4.2 Dijet balance comparisons

Figures E.3 and E.4 show the average reconstructed asymmetry as a function of the f_{Tile0}, f_{LAr3}, n_{trk}, $width_{trk}$ and $N_{segments}$ variables for jets at the EM+JES scale with $|\eta_{det}| < 0.6$ and $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV and $600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV respectively, in data and MC (full and AFII) simulation. The average reconstructed asymmetry as a function of the $N_{segments}$ variable is not shown for $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV due to the low number of data events.

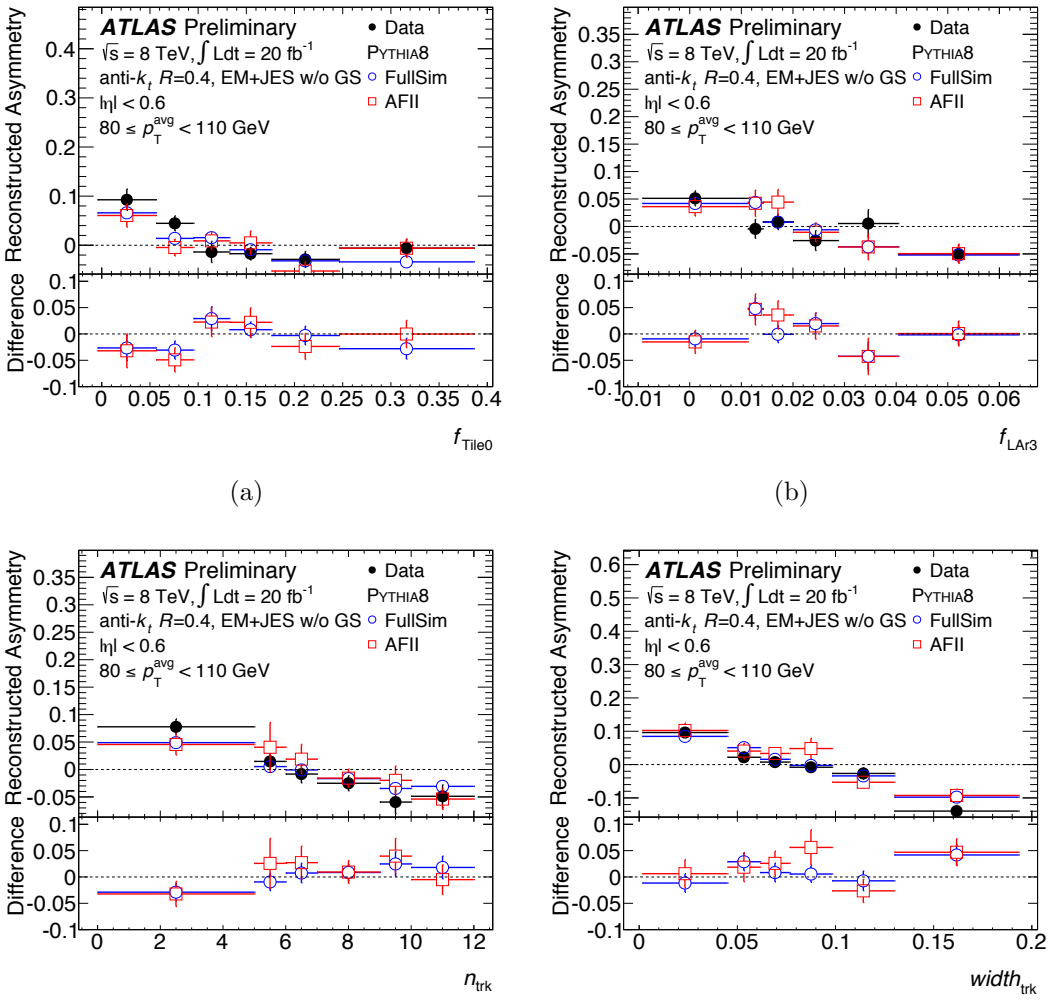
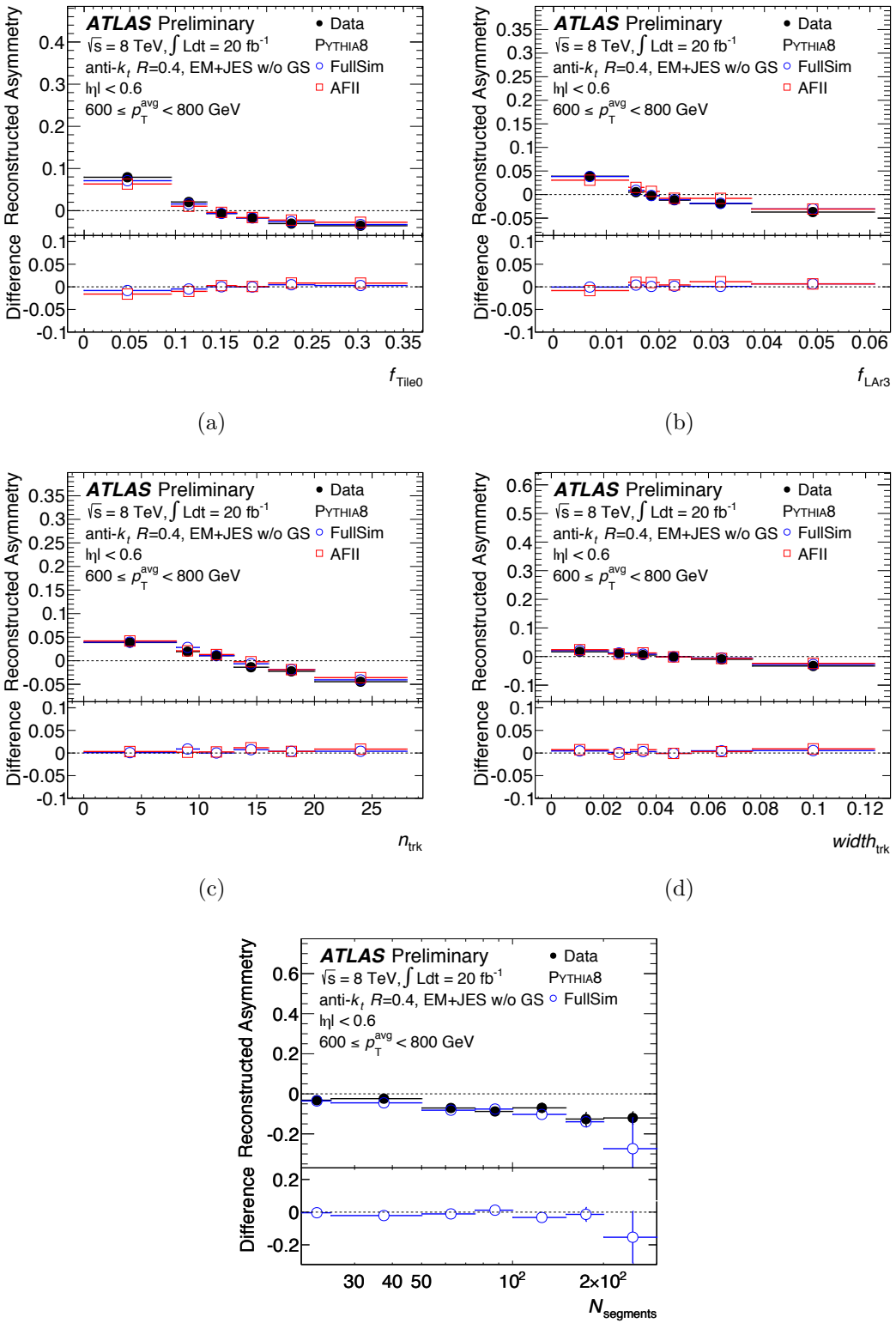


Figure E.3: Average reconstructed asymmetry as a function of the f_{Tile0} (a), f_{LAr3} (b), n_{trk} (c) and $width_{trk}$ (d) variables for jets with $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV and $|\eta_{det}| < 0.6$ in the data and the PYTHIA8 Monte Carlo simulation. Jets are reconstructed with the anti-k_t algorithm with $R = 0.4$ and calibrated with the EM+JES scheme. The lower parts of each sub-figure show the difference with respect to the data.



(a) (b) (c) (d) (e)

Figure E.4: Average reconstructed asymmetry as a function of the f_{Tile0} (a), f_{LAr3} (b), n_{trk} (c), $width_{trk}$ (d) and $N_{segments}$ (e) variables for jets with $600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV and $|\eta_{det}| < 0.6$ in the data and the PYTHIA8 Monte Carlo simulation. Jets are reconstructed with the anti-k_t algorithm with $R = 0.4$ and calibrated with the EM+JES scheme. The lower parts of each sub-figure show the difference with respect to the data.

For the layer energy fractions, the difference between data and the reference PYTHIA8 full simulation MC sample in the $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV ($600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV) bin is 4% (1%) at worst. Slightly bigger differences are observed between data and AFII for the f_{LAr3} variable. Similar results are found in the other $p_{\mathrm{T}}^{\mathrm{avg}}$ bins for the layer fractions. For the width_{trk} and n_{trk} variables, differences smaller than 5% (1%) are found in the $80 < p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV ($600 < p_{\mathrm{T}}^{\mathrm{avg}} < 800$ GeV) bin. Tracking variables are modelled in the same way in both full and AFII simulations, differences observed between both are due to statistical fluctuations.

The asymmetries measured with data and their dependence on the jet variables used to derive the GS corrections compare well to the asymmetries obtained with the MC simulation. Differences are typically small compared to the size of the proposed corrections, validating their application to data. Systematic uncertainties for the first four GS corrections are not evaluated from the asymmetry studies shown here, but are taken into account by the jet energy scale uncertainty described in Section 8.2.

Impact of truth-level imbalance on reconstructed asymmetry

The measurement of the response through the asymmetry defined in Equation E.1 assumes that the particle-jet level asymmetry is zero. This is true on average, but not when computed in bins of x. The measured asymmetry $A(x)$ is therefore a mixture of detector effects and imbalance at the truth level. In order to remove the effect of imbalance at the truth level, a new asymmetry is defined:

$$A'(x) = A(x) - A_{\mathrm{truth}}(x) \tag{E.4}$$

In this equation, $A(x)$ is given by equation E.1 and $A_{\mathrm{truth}}(x)$ is :

$$A_{\mathrm{truth}}(x) = \frac{p_{T,\mathrm{truth}}^{\mathrm{probe}}(x) - p_{T,\mathrm{truth}}^{\mathrm{ref}}}{p_{T,\mathrm{truth}}^{\mathrm{avg}}(x)} \tag{E.5}$$

where $p_{T,\mathrm{truth}}^{\mathrm{avg}}(x) = (p_{T,\mathrm{truth}}^{\mathrm{probe}}(x) + p_{T,\mathrm{truth}}^{\mathrm{ref}})/2$. $A_{\mathrm{truth}}(x)$ is the asymmetry at the particle-jet

level (or truth asymmetry) and is calculated by matching reconstructed jets to particle jets geometrically, as explained in Section 3.4. A_{truth} can of course only be determined in the MC simulation. When using $A'(x)$ instead of $A(x)$ in equation E.3, the effects of imbalance at the truth level are removed and the resulting response depends only on detector effects. Accounting for the particle-jet imbalance is particularly important for the corrections that depend on the energy in the n_{trk} and the width_{trk}.

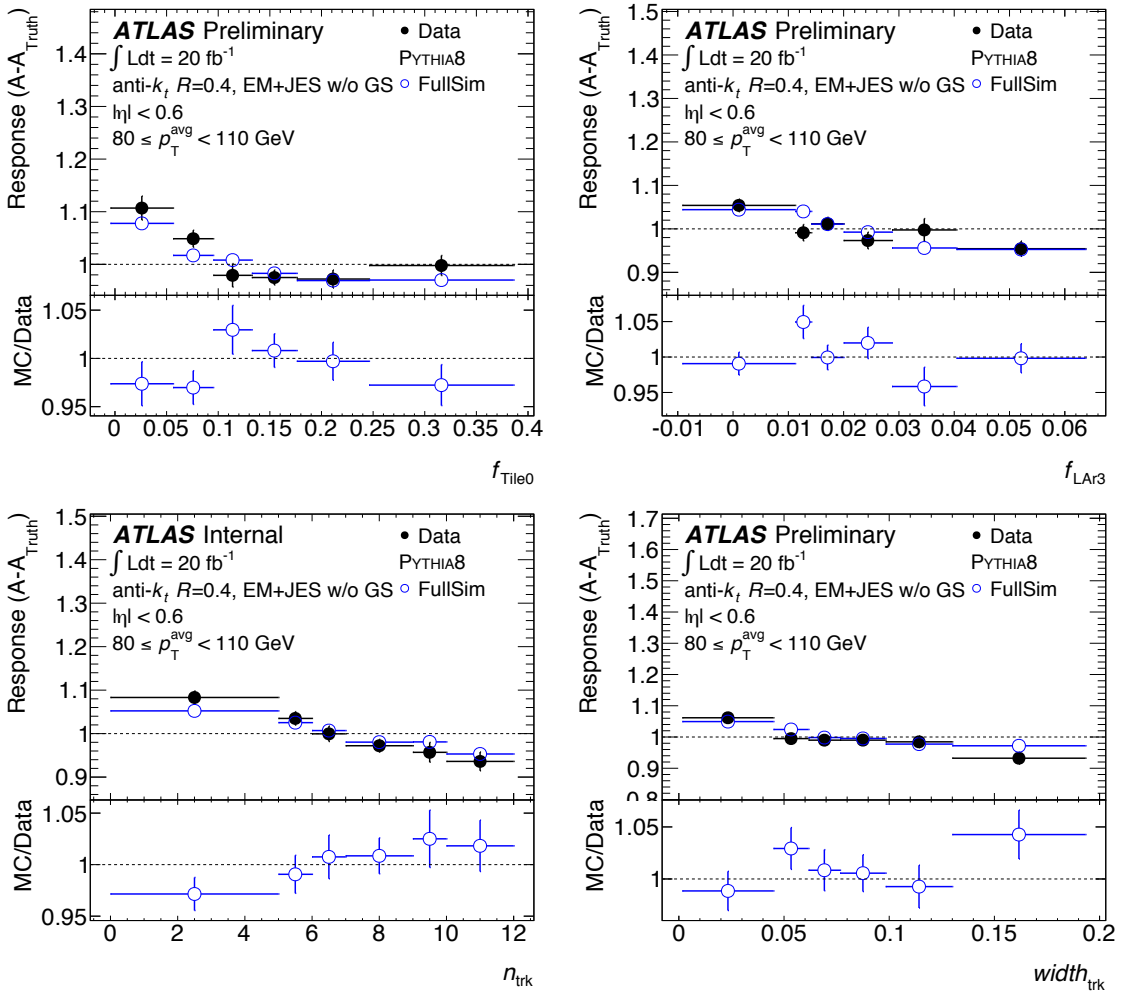


Figure E.5: Response calculated from the reconstructed asymmetry as a function of f_{Tile0}, f_{LAr3}, n_{trk} and width_{trk} for jets with $80 < p_{\text{T}} < 110$ GeV and $|\eta_{det}| < 0.6$ in the data and PYTHIA8 Monte Carlo simulation. All jets are reconstructed with anti-k_t $R = 0.4$ and calibrated with the EM+JES scheme.

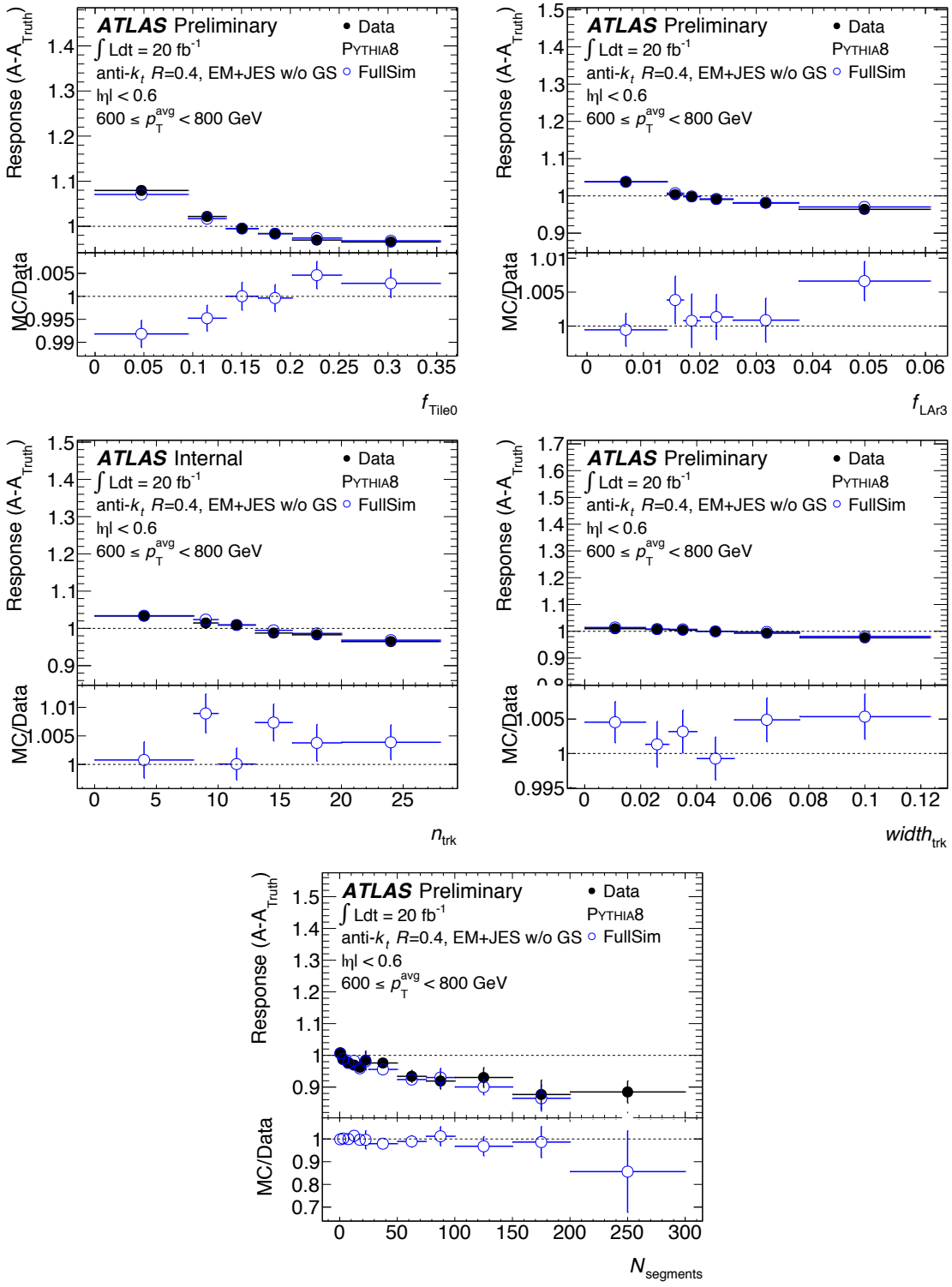


Figure E.6: Response calculated from the reconstructed asymmetry as a function of f_{Tile0}, f_{LAr3}, n_{trk}, $width_{trk}$ and $N_{segments}$ for jets with $600 < p_{\mathrm{T}} < 800$ GeV and $|\eta_{det}| < 0.6$ in the data and PYTHIA8 Monte Carlo simulation. All jets are reconstructed with anti-k_t $R = 0.4$ and calibrated with the EM+JES scheme.

Figures E.5 and E.6 show the jet response calculated from the reconstructed and truth asymmetry as a function of f_{Tile0}, f_{LAr3}, n_{trk}, $width_{trk}$ and $N_{segments}$ for jets at the EM+JES scale with $|\eta_{det}| < 0.6$ and for two p_{T} bins, $80 < p_{\mathrm{T}} < 110$ GeV and $600 < p_{\mathrm{T}} < 800$ GeV in data and MC simulation. The disagreement between data and the reference PYTHIA8 MC sample in this bin is 4% at worse. Disagreements of up to 5% are observed in other bins, particularly for the jet $width_{trk}$ and f_{Tile0}.

Comparison with truth response

The response calculated from the reconstructed and truth asymmetry, shown in Figures E.5 and E.6, can also be compared to the truth response as defined in Equation 3.4. Figure E.7 shows this comparison as a function of x. The differences are observed to be small, with the response from data and full-sim compatible with the truth response to within 1%. Similar compatibility is seen across all $|\eta_{det}|$ and p_T^{probe} regions, for all four jet collections.

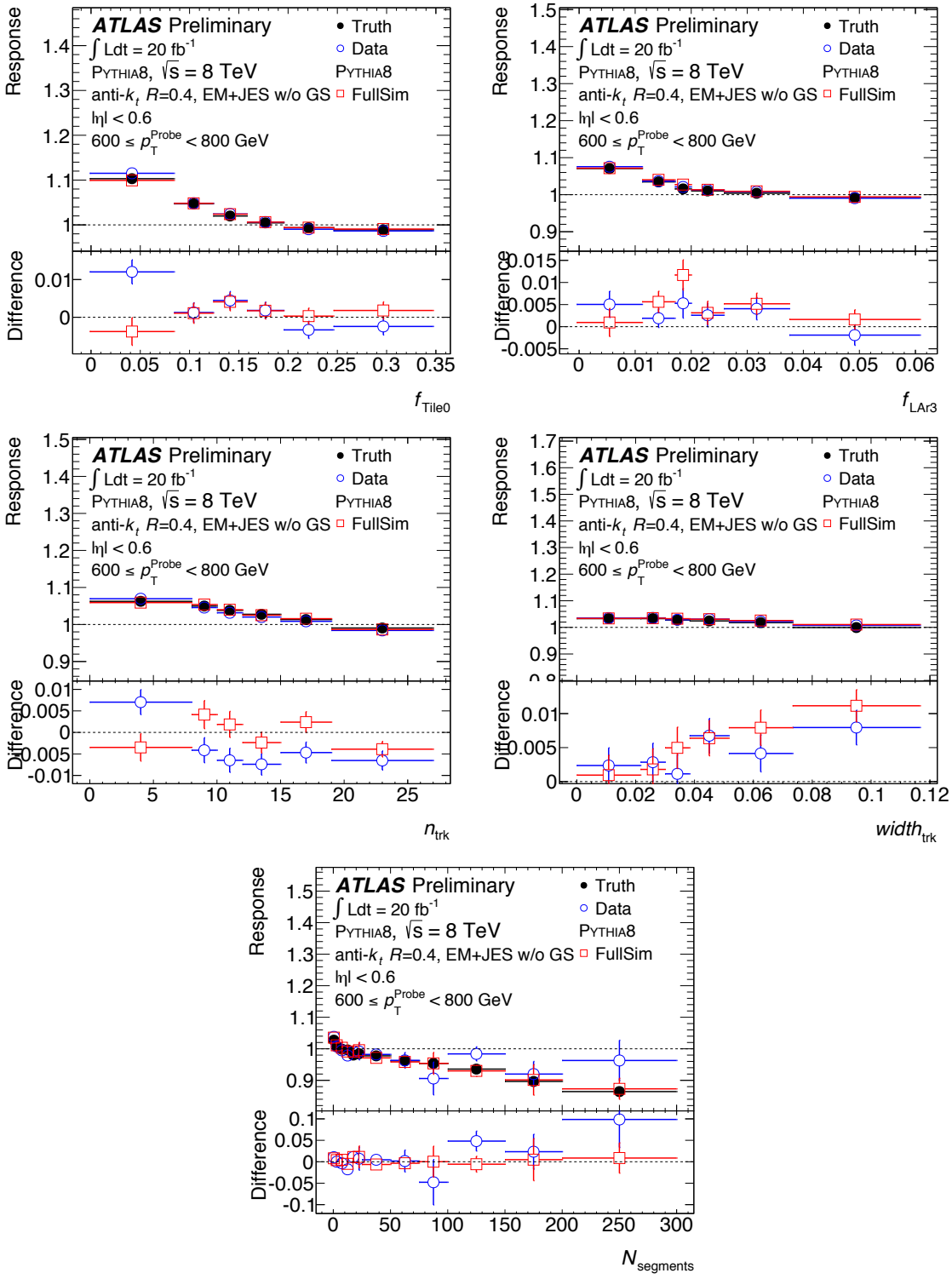


Figure E.7: Response calculated from the reconstructed asymmetry compared with truth response as a function of f_{Tile0}, f_{LAr3}, n_{trk}, $width_{trk}$ and $N_{segments}$ for jets with $600 < p_{\mathrm{T}} < 800$ GeV and $|\eta_{det}| < 0.6$ in the data and Pythia8 Monte Carlo simulation. All jets are reconstructed with anti-k_t $R = 0.4$ and calibrated with the EM+JES scheme.

E.4.3 Comparison of derived jet punch-through uncertainties

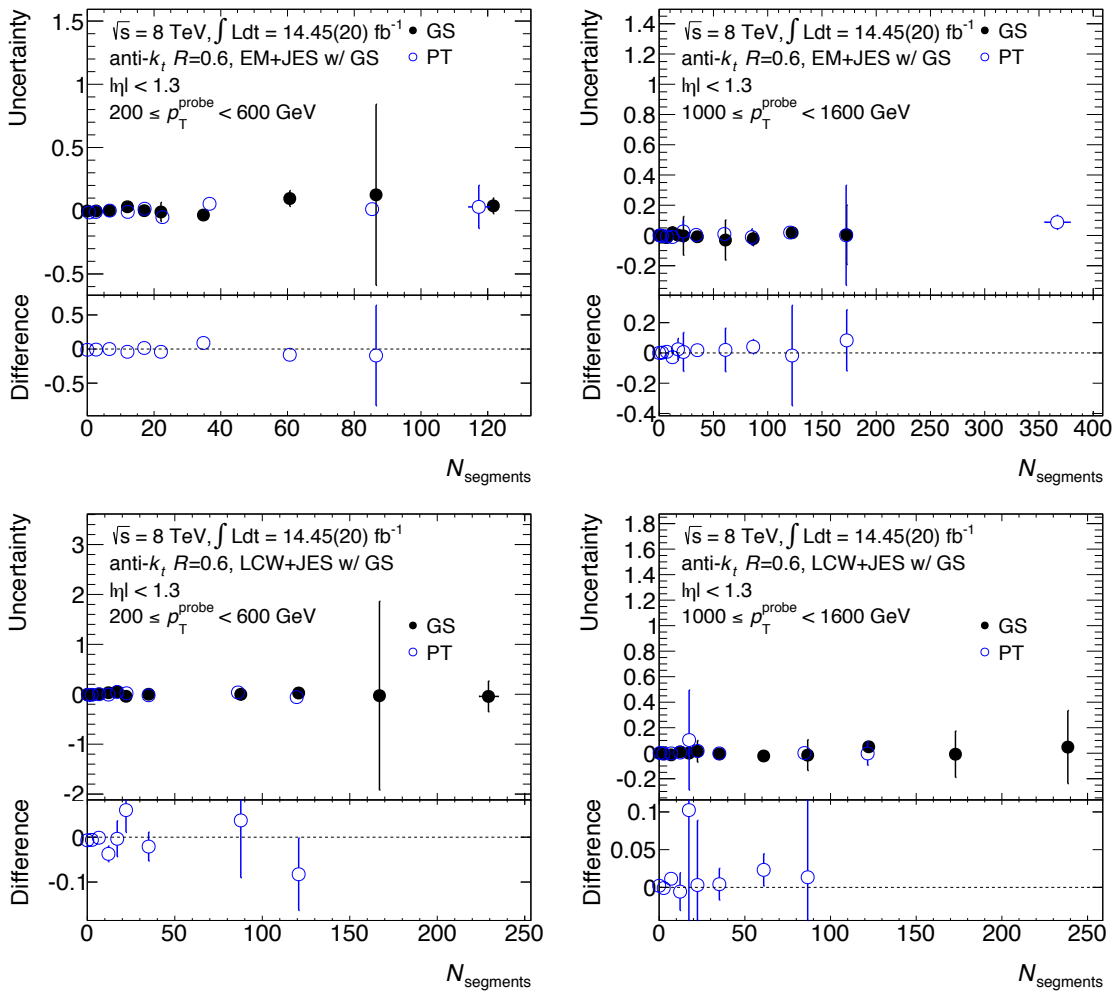


Figure E.8: Data-MC difference in the reconstructed response calculated using the GS dijet balance method and the jet punch-through (PT) dijet balance method, for jets at the EM+JES+GS (top) and LCW+JES+GS (bottom) scales, with $200 \leq \mathrm{p}_T^{probe}\,[\,\mathrm{GeV}] < 600$ (left) and $1000 \leq \mathrm{p}_T^{probe}\,[\,\mathrm{GeV}] < 1600$ (right). The bottom plot shows the difference between the two methods.

The uncertainties on the jet punch-through correction, derived using the method described in Chapter 6, can also be calculated using the dijet balance method described in this appendix. The uncertainty is the difference in the reconstructed response between data and MC. Figure E.8 compares the uncertainty derived using these two methods for EM+JES+GS and LCW+JES+GS scale, R = 0.6 jets. It is clear from this figure that the uncertainties derived from the two methods are compatible. This compatibility was observed for all other p_T^{probe} and $|\eta_{det}|$ regions.

Bibliography

[1] "Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV", Tech. Rep. ATLAS-CONF-2015-002, CERN, Geneva, Mar, 2015. https://cds.cern.ch/record/2001682.

[2] "Dijet resonance searches with the ATLAS detector at 14 TeV LHC", Tech. Rep. ATL-PHYS-PUB-2015-004, CERN, Geneva, Mar, 2015. https://cds.cern.ch/record/2002136.

[3] **ATLAS** Collaboration, G. Aad *et al.*, "Search for new phenomena in the dijet mass distribution using $p-p$ collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Phys. Rev.* **D91** no. 5, (2015) 052007, arXiv:1407.1376 [hep-ex].

[4] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[5] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[6] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory*. Westview Press, Oct., 1995. http://www.worldcat.org/isbn/0201503972.

[7] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** no. 4, (1961) 579 – 588. http://www.sciencedirect.com/science/article/pii/0029558261904692.

[8] G. ALTARELLI, "The Standard model of particle physics", arXiv:hep-ph/0510281 [hep-ph].

[9] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[10] E. Noether, "Invariante variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257.

[11] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133.

[12] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[13] D. Galbraith, "UX:Standard Model of the Standard Model", http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model.

[14] W. Heisenberg, "Über den bau der atomkerne.", *Zeitschrift für Physik* **77** no. 1-2, (1932) 1–11. http://dx.doi.org/10.1007/BF01342433.

[15] H. Yukawa, "On the Interaction of Elementary Particles", *Proc. Phys.-Math. Soc. Jpn.* no. 17, (1935) 48–57.

[16] Y. Ne'eman, "Derivation of strong interactions from a gauge invariance", *Nuclear Physics* **26** no. 2, (1961) 222 – 229. http://www.sciencedirect.com/science/article/pii/0029558261901341.

[17] M. Gell-Mann, "Symmetries of baryons and mesons", *Phys. Rev.* **125** (1962) 1067–1084.

[18] V. E. Barnes, P. L. Connolly, *et al.*, "Observation of a hyperon with strangeness minus three", *Phys. Rev. Lett.* **12** (Feb, 1964) 204–206. http://link.aps.org/doi/10.1103/PhysRevLett.12.204.

[19] M. Gell-Mann, "A schematic model of baryons and mesons", *Physics Letters* **8** no. 3, (1964) 214 – 215.
http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[20] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking; Version 2",. https://cds.cern.ch/record/570209. Version 1 is CERN preprint 8182/TH.401, Jan. 17, 1964.

[21] J. D. Bjorken, "Asymptotic sum rules at infinite momentum", *Phys. Rev.* **179** (Mar, 1969) 1547–1553. http://link.aps.org/doi/10.1103/PhysRev.179.1547.

[22] R. E. Taylor, "Deep inelastic scattering: The early years", *Rev. Mod. Phys.* **63** (Jul, 1991) 573–595. http://link.aps.org/doi/10.1103/RevModPhys.63.573.

[23] H. W. Kendall, "Deep inelastic scattering: Experiments on the proton and the observation of scaling", *Rev. Mod. Phys.* **63** (Jul, 1991) 597–614.
http://link.aps.org/doi/10.1103/RevModPhys.63.597.

[24] J. I. Friedman, "Deep inelastic scattering: Comparisons with the quark model", *Rev. Mod. Phys.* **63** (Jul, 1991) 615–627.
http://link.aps.org/doi/10.1103/RevModPhys.63.615.

[25] R. P. Feynman, "Very high-energy collisions of hadrons", *Phys. Rev. Lett.* **23** (Dec, 1969) 1415–1417. http://link.aps.org/doi/10.1103/PhysRevLett.23.1415.

[26] M. Y. Han and Y. Nambu, "Three-triplet model with double SU(3) symmetry", *Phys. Rev.* **139** (Aug, 1965) B1006–B1010.
http://link.aps.org/doi/10.1103/PhysRev.139.B1006.

[27] Y. Miyamoto, "Three kinds of triplet model", *Progress of Theoretical Physics Supplement* **E65** (1965) 187–192,
http://ptps.oxfordjournals.org/content/E65/187.full.pdf+html.
http://ptps.oxfordjournals.org/content/E65/187.short.

[28] R. E. Ruiz. http://home.fnal.gov/~rruiz/FeynLib/.

[29] G. M. Prosperi, M. Raciti, and C. Simolo, "On the running coupling constant in QCD", *Prog. Part. Nucl. Phys.* **58** (2007) 387–438, arXiv:hep-ph/0607209 [hep-ph].

[30] **ATLAS Collaboration** Collaboration, "Measurement of multi-jet cross-section ratios and determination of the strong coupling constant in proton-proton collisions at sqrt(s)=7 TeV with the ATLAS detector.", Tech. Rep. ATLAS-CONF-2013-041, CERN, Geneva, Apr, 2013. https://cds.cern.ch/record/1543225.

[31] T. Sjostrand, "Monte Carlo Generators", in *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006*, pp. 51–74. 2006. arXiv:hep-ph/0611247 [hep-ph]. http://weblib.cern.ch/abstract?CERN-LCGAPP-2006-06.

[32] J. M. Campbell, J. Huston, and W. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Rept.Prog.Phys.* **70** (2007) 89, arXiv:hep-ph/0611148 [hep-ph].

[33] T. Kinoshita, "Mass singularities of feynman amplitudes", *Journal of Mathematical Physics* **3** no. 4, (1962) 650–677. http://scitation.aip.org/content/aip/journal/jmp/3/4/10.1063/1.1724268.

[34] T. D. Lee and M. Nauenberg, "Degenerate systems and mass singularities", *Phys. Rev.* **133** (Mar, 1964) B1549–B1562. http://link.aps.org/doi/10.1103/PhysRev.133.B1549.

[35] G. Sterman and S. Weinberg, "Jets from quantum chromodynamics", *Phys. Rev. Lett.* **39** (Dec, 1977) 1436–1439. http://link.aps.org/doi/10.1103/PhysRevLett.39.1436.

[36] C. Grupen and B. Shwartz, *Particle Detectors*. Cambridge University Press, second ed., 2008. http://dx.doi.org/10.1017/CBO9780511534966. Cambridge Books Online.

[37] C. Amsler *et al.*, "Review of particle physics", *Physics Letters B* **667** no. 15, (2008) 1 – 6. http://www.sciencedirect.com/science/article/pii/S0370269308008435. Review of Particle Physics.

[38] **ATLAS Liquid Argon** Collaboration, K.-J. Grahn, "A Layer Correlation Technique for Pion Energy Calibration at the 2004 ATLAS Combined Beam Test", in *Proceedings, 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2009)*, pp. 751–757. 2009. arXiv:0911.2639 [physics.ins-det]. http://inspirehep.net/record/836891/files/arXiv:0911.2639.pdf.

[39] R. Wigmans, "Energy measurement at the tev scale", *New Journal of Physics* **10** no. 2, (2008) 025003. http://stacks.iop.org/1367-2630/10/i=2/a=025003.

[40] R. Wigmans, "Calorimeters", in *Handbook of particle detection and imaging, vol. 1 and vol.2*, C. Grupen and I. Buvat, eds., pp. 497–517. 2012.

[41] Y. A. Kulchitsky, J. A. Budagov, V. B. Vinogradov, M. Nessi, A. A. Bogush, V. V. Arkadov, and G. V. Karapetian, "Electron response and e / h ratio of iron scintillator hadron prototype calorimeter with longitudinal TILE configuration", arXiv:hep-ex/9903050 [hep-ex].

[42] Y. A. Kulchitsky, "On the e / h ratio of the electromagnetic calorimeter", *Part. Nucl. Lett.* **100** (2000) 28–34, arXiv:hep-ex/0005029 [hep-ex].

[43] Y. A. Budagov, Y. A. Kulchitskii, M. V. Kuzmin, and V. B. Vinogradov, "The e/h ratio of the ATLAS Hadronic Tile Calorimeter", Tech. Rep. ATL-TILECAL-2001-001, CERN, Geneva, Apr, 2001. https://cds.cern.ch/record/684190.

[44] **Particle Data Group** Collaboration, K. A. Olive *et al.*, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001.

[45] H.-C. Schultz-Coulon. http://www.kip.uni-heidelberg.de/~coulon/Lectures/DetectorsSoSe10/Free_PDFs/Lecture10.pdf.

[46] K. Kleinknecht, *Detectors for particle radiation*. Cambridge Univ. Press, Cambridge, 1986. http://cdsweb.cern.ch/record/108093.

[47] http://pdg.lbl.gov/2013/AtomicNuclearProperties/HTML_PAGES/026.html.

[48] Y. A. Kulchitsky, V. S. Rumyantsev, J. A. Budagov, V. B. Vinogradov, G. Karapetian, M. Nessi, and A. A. Bogush, "Effect on a hadron shower leakage on the energy response and resolution of hadron tile calorimeter", arXiv:hep-ex/9903051 [hep-ex].

[49] **RD34 Collaboration** Collaboration, F. Ariztizabal, M. Bosman, M. Cavalli-Sforza, I. Efthymiopoulos, *et al.*, "Construction and performance of an iron-scintillator hadron calorimeter with longitudinal tile configuration", *Nucl. Instrum. Methods Phys. Res., A* **349** no. CERN-PPE-94-66, (Mar, 1994) 384–397. 32 p. https://cds.cern.ch/record/262630.

[50] C. Bromberg, T. Davidek, E. Gallas, I. Hruska, J. Huston, P. Krivkova, R. Leitner, M. Lokagicek, R. Miller, S. Nmcek, J. Palacky, R. Richards, D. Schooltz, J. Slavik, K. Soustruznik, M. Suk, P. Tas, and J. Weichert, "Remote Control of Muon Walls in the Extended Barrel Module 0 Tilecal Beam Test", Tech. Rep. ATL-TILECAL-98-154. ATL-L-PN-154, CERN, Geneva, Mar, 1998. https://cds.cern.ch/record/683622.

[51] **FNAL-SDC Simulation Group** Collaboration, D. Green, A. Beretvas, K. Denisenko, N. Denisenko, J. Marraffino, A. Para, and W. Wu, "Depth requirements in ssc calorimeters", Tech. Rep. FERMILAB-FN-0570, SDC-91-00016, 1991.

[52] F. J. Sciulli, "Photon-Collecting Hadron Calorimeters", Tech. Rep. CALT-68-502, 1975.

[53] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, *et al.*, "General-purpose event generators for LHC physics", *Phys.Rept.* **504** (2011) 145–233, arXiv:1101.2599 [hep-ph].

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[55] H.-L. Lai *et al.*, "New parton distributions for collider physics", *Phys. Rev.* **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[56] **ATLAS** Collaboration, "Summary of atlas pythia 8 tunes", ATL-PHYS-PUB-2012-003.

[57] G. Folger and J. Wellisch, "String parton models in GEANT4", arXiv:0306007 [nucl-th].

[58] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the LHC", *Eur. Phys. J.* **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[59] J. Odier, S. Albrand, J. Fulachier, and F. Lambert, "Evolution of the Architecture of the ATLAS Metadata Interface (AMI)", Tech. Rep. ATL-SOFT-PROC-2015-014, CERN, Geneva, May, 2015. http://cds.cern.ch/record/2015212.

[60] **ATLAS** Collaboration, G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003.

[61] **LHCb** Collaboration, J. Alves, A. Augusto *et al.*, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005.

[62] **ALICE** Collaboration, K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002.

[63] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004.

[64] L. R. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001. 164 p. https://cds.cern.ch/record/1129806. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[65] **ATLAS** Collaboration, G. Aad *et al.*, "Luminosity Determination in *pp* Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011", Tech. Rep. ATLAS-COM-CONF-2011-130, CERN, Jul, 2011.

[66] J. Haffner, "The CERN accelerator complex. Complexe des accélérateurs du CERN",. `https://cds.cern.ch/record/1621894`. General Photo.

[67] M. Lamont, "Status of the LHC", *J. Phys.: Conf. Ser.* **455** (2013) 012001. `https://cds.cern.ch/record/1709796`.

[68] **ATLAS** Collaboration, "Luminosity Public Results", `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`, 2012.

[69] J. Pequenao, "Computer generated image of the whole ATLAS detector", Mar, 2008.

[70] J. Pequenao and P. Schaffner, "An computer generated image representing how ATLAS detects particles", Jan, 2013.

[71] J. Pequenao, "Computer generated image of the ATLAS inner detector", Mar, 2008.

[72] **ATLAS** Collaboration, G. Aad *et al.*, "Atlas pixel detector electronics and sensors", *JINST* **3** no. 07, (2008) P07007. `http://stacks.iop.org/1748-0221/3/i=07/a=P07007`.

[73] **ATLAS** Collaboration, "Atlas experiment - photos", `http://www.atlas.ch/photos/full-detector-cgi.html`, 2011.

[74] **ATLAS** Collaboration, G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003.

[75] K. Nagano, "Algorithms, performance, and development of the ATLAS High-level Trigger",. `http://cds.cern.ch/record/1632445`.

[76] D. Hall and C. Hays, *Discovery and measurement of the Higgs boson in the WW decay channel*. PhD thesis, Oxford U., Oxford, Aug, 2014. https://cds.cern.ch/record/1950648. Presented 29 Aug 2014.

[77] M. C. Tamsett, M. Begel, D. O. Damazio, F. Rubbo, C. Sandoval, and R. Subramaniam, "L1.5 Jet Trigger Performance with 2011 Data", Tech. Rep. ATL-COM-DAQ-2012-009, CERN, Geneva, Mar, 2012. https://cds.cern.ch/record/1429641.

[78] D. Gilbert, "Jet studies twiki", https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetStudies2012. ATLAS twiki.

[79] "Alignment of the ATLAS Inner Detector and its Performance in 2012", Tech. Rep. ATLAS-CONF-2014-047, CERN, Geneva, Jul, 2014. https://cds.cern.ch/record/1741021.

[80] **ATLAS** Collaboration, G. Aad *et al.*, "Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data", *Eur. Phys. J.* **C74** no. 10, (2014) 3071, arXiv:1407.5063 [hep-ex].

[81] S. Sacerdoti, G. Otero y Garzon, and R. Piegaia, "Jet momentum resolution with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV recorded in 2012.", Tech. Rep. ATL-COM-PHYS-2014-010, CERN, Geneva, Jan, 2014. https://cds.cern.ch/record/1642375.

[82] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC protonproton collision data", *Eur. Phys. J.* **C74** no. 11, (2014) 3130, arXiv:1407.3935 [hep-ex].

[83] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250–303.

[84] **ATLAS** Collaboration, G. Aad *et al.*, "The ATLAS Simulation Infrastructure", *Eur. Phys. J.* **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[85] W. Lukas, "Fast simulation for ATLAS: Atlfast-II and ISF",. http://cds.cern.ch/record/1448165.

[86] J. Pequenao, "Simulated production of a Higgs event in ATLAS. This track is an example of simulated data modeled for the ATLAS detector on the Large Hadron Collider (LHC) at CERN, which will begin taking data in 2008.", ATLAS Collection., Jan, 2008.

[87] C. Buttar, J. D'Hondt, M. Kramer, G. Salam, M. Wobisch, *et al.*, "Standard Model Handles and Candles Working Group: Tools and Jets Summary Report", arXiv:0803.0678 [hep-ph].

[88] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. D. Elvira, K. Frame, *et al.*, "Run II jet physics", arXiv:hep-ex/0005012 [hep-ex].

[89] S. V. Chekanov, "Jet algorithms: a minireview",. https://cds.cern.ch/record/592437.

[90] S. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, "Jets in hadron-hadron collisions", *Prog.Part.Nucl.Phys.* **60** (2008) 484–551, arXiv:0712.2447 [hep-ph].

[91] **ATLAS Collaboration** Collaboration, G. Aad, E. Abat, B. Abbott, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, *et al.*, *Expected performance of the ATLAS experiment: detector, trigger and physics.* CERN, Geneva, 2009. https://cds.cern.ch/record/1125884.

[92] M. Cacciari and G. P. Salam, "Dispelling the myth for the jet-finder", *Physics Letters B* **641** no. 1, (2006) 57 – 61. http://www.sciencedirect.com/science/article/pii/S0370269306010094.

[93] M. C. G. Soye and G. P. Salam. http://fastjet.fr/.

[94] G. Soyez, "The SISCone and anti-k(t) jet algorithms", arXiv:0807.0021 [hep-ph].

[95] G. P. Salam, "Towards Jetography", *Eur.Phys.J.* **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[96] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[97] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms", *JHEP* **9708** (1997) 001, arXiv:hep-ph/9707323 [hep-ph].

[98] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys.Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[99] L. Asquith, B. Brelier, J. M. Butterworth, M. Campanelli, T. Carli, G. Choudalakis, P. A. Delsart, S. De Cecco, P. O. Deviveiros, M. D'Onofrio, S. Eckweiler, E. Feng, P. Francavilla, S. Grinstein, I. La Plante, J. Huston, N. Ghodbane, D. Lopez Mateos, B. Martin, N. Makovec, S. Majewsky, M. Martinez, D. W. Miller, J. Monk, K. Perez, C. Roda, J. Robinson, A. Schwartzmann, F. Spano, K. Terashi, F. Vives, P. Weber, and S. Zenz, "Performance of Jet Algorithms in the ATLAS Detector", Tech. Rep. ATL-COM-PHYS-2009-630, CERN, Geneva, Dec, 2009. https://cds.cern.ch/record/1226299.

[100] **Particle Data Group** Collaboration, J. Beringer, J. F. Arguin, R. M. Barnett, *et al.*, "Review of particle physics*", *Phys. Rev. D* **86** (Jul, 2012) 010001. http://link.aps.org/doi/10.1103/PhysRevD.86.010001.

[101] **ATLAS** Collaboration, G. Aad *et al.*, "Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur.Phys.J.* **C75** no. 1, (2015) 17, arXiv:1406.0076 [hep-ex].

[102] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, "Calorimeter Clustering Algorithms: Description and Performance", Tech. Rep. ATL-LARG-PUB-2008-002.

ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.
https://cds.cern.ch/record/1099735.

[103] **ATLAS** Collaboration, G. Aad *et al.*, "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Eur.Phys.J.* **C73** no. 3, (2013) 2304, arXiv:1112.6426 [hep-ex].

[104] C. Cojocaru, J. Pinfold, J. Soukup, *et al.*, "Hadronic calibration of the atlas liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests", *NIM* **531** (Oct., 2004) 481–514, physics/0407009.

[105] G. Aad *et al.*, "Jet energy measurement with the atlas detector in proton-proton collisions at $\sqrt{\mathrm{s}} = 7$ TeV", *The European Physical Journal C* **73** no. 3, (2013) . http://dx.doi.org/10.1140/epjc/s10052-013-2304-2.

[106] B. Sotto-Maior Peralva, "Calibration and Performance of the ATLAS Tile Calorimeter", arXiv:1305.0550 [physics.ins-det].

[107] T. Barillari, E. Bergeaas Kuutmann, T. Carli, J. Erdmann, P. Giovannini, K. J. Grahn, C. Issever, *et al.*, "Local Hadronic Calibration", Tech. Rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun, 2008. https://cds.cern.ch/record/1112035.

[108] "Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS", Tech. Rep. ATLAS-CONF-2015-037, CERN, Geneva, Aug, 2015. https://cds.cern.ch/record/2044941.

[109] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys.Lett.* **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[110] "Pile-up subtraction and suppression for jets in ATLAS", Tech. Rep.

ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013.
https://cds.cern.ch/record/1570994.

[111] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **0804** (2008) 005, arXiv:0802.1188 [hep-ph].

[112] **ATLAS** Collaboration, "Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur. Phys. J.* **C74** no. 8, (2014) 3023, arXiv:1405.6583 [hep-ex].

[113] **ATLAS Collaboration** Collaboration, "Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data", Tech. Rep. ATLAS-CONF-2012-020, CERN, Geneva, Mar, 2012.
https://cds.cern.ch/record/1430034.

[114] **ATLAS** Collaboration, G. Aad *et al.*, "Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions", *Eur. Phys. J.* **C70** (2010) 723–753, arXiv:0912.2642 [physics.ins-det].

[115] https://twiki.cern.ch/twiki/bin/view/Sandbox/EgammaPhysCalib.

[116] "Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data", Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.

[117] J. Jovicevic, "Performance of b-jet identification in ATLAS", Tech. Rep. ATL-COM-PHYS-2014-418, CERN, Geneva, May, 2014.

[118] R. Nicolaidou, L. Chevalier, S. Hassani, J. F. Laporte, E. L. Menedeu, and A. Ouraou, "Muon identification procedure for the atlas detector at the lhc using muonboy reconstruction package and tests of its performance using cosmic rays and single beam data", *Journal of Physics: Conference Series* **219** no. 3, (2010) 032052.
http://stacks.iop.org/1742-6596/219/i=3/a=032052.

[119] S. Hassani, L. Chevalier, E. Lancon, J. Laporte, R. Nicolaidou, *et al.*, "A muon identification and combined reconstruction procedure for the ATLAS detector at the

LHC using the (MUONBOY, STACO, MuTag) reconstruction packages", *Nucl.Instrum.Meth.* **A572** (2007) 77–79.

[120] K. Assamagan, G. Azuelos, S. Bahinipati, P. Banerjee, A. B. Galtieri, J. Collot, M. Davies, A. Gaponenko, A. Haas, J. Kraus, G. Lima, W. Lockman, A. Solodkov, G. Unal, A. Watson, and Y. Zhu, "Overlay for ATLAS Simulation", Tech. Rep. ATL-SOFT-INT-2011-001, CERN, Geneva, Jul, 2011.

[121] M. Goblirsch-Kolb. Private communication.

[122] E. Ritsch, A. Salzburger, and E. Kneringer, *Fast Calorimeter Punch-Through Simulation for the ATLAS Experiment*. PhD thesis, Innsbruck U., Innsbruck, 2011. `https://cds.cern.ch/record/1388275`. Presented 28 Sep 2011.

[123] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533. [,648(1963)].

[124] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657.

[125] D. W. SCOTT, "On optimal and data-based histograms", *Biometrika* **66** no. 3, (1979) 605–610, `http://biomet.oxfordjournals.org/content/66/3/605.full.pdf+html`. `http://biomet.oxfordjournals.org/content/66/3/605.abstract`.

[126] `https://en.wikipedia.org/wiki/Bilinear_interpolation`.

[127] `https://en.wikipedia.org/wiki/Kernel_density_estimation`.

[128] E. A. Nadaraya, "On estimating regression", *Theory of Probability & Its Applications* **9** no. 1, (1964) 141–142, `http://dx.doi.org/10.1137/1109020`. `http://dx.doi.org/10.1137/1109020`.

[129] `http://www.stat.wisc.edu/~mchung/teaching/MIA/reading/diffusion.gaussian.kernel.pdf.pdf`.

[130] https://en.wikipedia.org/wiki/Similarity_measure.

[131] https://sfb649.wiwi.hu-berlin.de/fedc_homepage/xplore/ebooks/html/spm/spmhtmlnode18.html.

[132] V. Lendermann, J. Haller, M. Herbst, K. Krüger, H.-C. Schultz-Coulon, and R. Stamen, "Combining Triggers in HEP data analysis", *NIM* **604** (June, 2009) 707–718, arXiv:0901.4118 [hep-ex].

[133] "Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8\ TeV$", Tech. Rep. ATLAS-CONF-2015-017, CERN, Geneva, Apr, 2015. https://cds.cern.ch/record/2008678.

[134] **ATLAS Collaboration** Collaboration, "Response of the ATLAS calorimeters to single isolated hadrons produced in proton proton collisions at a center of mass energy of sqrts = 900 GeV", Tech. Rep. ATLAS-CONF-2010-017, CERN, Geneva, Jul, 2010. https://cds.cern.ch/record/1277648.

[135] K. Pachal, 2015. Private communication.