Title
Properties of Jets Measured with Charged Particles with the ATLAS Detector at the Large Hadron Collider

Author
Zenz, Seth Conrad

Publication Date
2011

Peer reviewed|Thesis/dissertation

Properties of Jets Measured with Charged Particles with the ATLAS Detector at the Large Hadron Collider

by

Seth Conrad Zenz

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Ian Hinchliffe, Co-Chair
Professor James Siegrist, Co-Chair
Professor Robert Jacobsen
Professor Alexei Filippenko

Fall 2011

Properties of Jets Measured with Charged Particles
with the ATLAS Detector at the Large Hadron Collider

by

Seth Conrad Zenz

Abstract

Properties of Jets Measured with Charged Particles
with the ATLAS Detector at the Large Hadron Collider
by
Seth Conrad Zenz
Doctor of Philosophy in Physics
University of California, Berkeley
Dr. Ian Hinchliffe, Co-Chair
Professor James Siegrist, Co-Chair

Jets are identified and their properties studied, in proton-proton collisions at the Large Hadron Collider at center-of-mass energy $\sqrt{s} = 7\,\text{TeV}$, using charged particles measured by the ATLAS Inner Detector. Events are selected using a minimum bias trigger, allowing jets at very low transverse momentum to be observed and the transition to high-momentum fully perturbative jets to be studied. Jets are reconstructed using the anti-k_t algorithm applied to charged particles with two radius parameter choices, 0.4 and 0.6. An inclusive charged jet transverse momentum cross section measurement from 4 GeV to 100 GeV is shown, for four ranges in rapidity extending to 1.9, and corrected to charged particle-level truth jets. The transverse momenta and longitudinal momentum fractions of charged particles within jets are measured, along with the charged particle multiplicity and the particle density as a function of radial distance from the jet axis. Comparison of the data with the theoretical models implemented in existing tunings of Monte Carlo event generators indicates reasonable overall agreement between data and Monte Carlo. These comparisons are sensitive to Monte Carlo parton showering, hadronization, and soft physics models.

To Barbara Zenz Siegel, my grandmother, who believed that adulthood only begins with the completion of one's dissertation.

Acknowledgements

Looking back at the broad sweep of the history of dissertation acknowledgements, one cannot escape the conclusion that there has been a certain amount of "mission creep." Where once the thanks were limited to those who had typeset and proofread the document itself, today the section is a sprawling multipage affair that details the important contributions made to the writer's personal development by everyone encountered all the way back to preschool. I here attempt to avoid this trend by limiting myself strictly to those who have contributed directly to the present work; this does not, alas, mean that these comments will be short.

I have been very fortunate to work with a great variety of very talented people within the Berkeley ATLAS group. My advisor, Ian Hinchliffe, has been an enthusiastic and close collaborator; he has always had clear advice on where I should take my work but has never objected to me making my own plans if I have thought them through. Marjorie Shapiro was always available to chat even about issues technical and non-technical; she and Andrei Gaponenko are the only people other than me who wrote code directly for the analysis described in this document. Jed Biesiada gave me a tremendous amount of advice on an incredible range of topics, particularly during the early phases of preparing the analysis. Max Scherzer provided me with many opportunities, over the years, to clarify my thinking about doing physics and working collaboratively. Kevin Einsweiler provided occasional but deeply valuable perspective.

More broadly, I owe thanks to all my collaborators on the ATLAS experiment. I thank the editorial board: Kiyotomo Kawagoe, Peter Maettig, Ariel Schwartzman, and Andreas Wildauer; my memories of a spring and summer made hectic by their suggestions will fade far more quickly than my pride in a final product made better by their ideas. The ATLAS Standard Model convenors, Kevin Einsweiler (again!) and Jon Butterworth, provided critical support. Zachary Marshall got Inner Detector-only simulation working for me, to say nothing of keeping simulation in general working for everyone.

Many thanks to Dr. Dr. C. A. Jenkins for referring me to the LaTeX template I used for producing this document, which was originally made available by Mark Paskin. I also thank the Graduate Division of the University of California, Berkeley for regularly modifying the layout requirements with which the template once exactingly complied, thereby giving me the opportunity to learn a number of new typesetting commands.

I also thank my Berkeley dissertation committee for their help in reviewing this document and its contents. Bob Jacobsen has given me good advice from my first semester at Berkeley to my last. Alex Filippenko taught me punctuation rules I had

never quite grasped before. Jim Siegrist bought me the beer he promised me.

I thank the Department of Energy and the National Science Foundation for funding my work, in particular through the NSF Graduate Research Fellowship Program.

Finally, I would like to thank Peter Siegel for a valuable introduction to object-oriented programming in the early 1990's, and Hope Stevens for fruitful discussions on enumerating natural phenomena using the integers during the mid-1980's. Both of these concepts proved invaluable in the completion of this work.

Contents

Chapter 1

Introduction

The Large Hadron Collider (LHC) [1] was built to produce new particles and study rare interactions at a high rate, but its first and foremost byproduct is sprays of low-energy hadrons. At its full design capacity, the LHC will cross proton bunches 31.6 million times per second at each interaction point, with an average of about 20 proton-proton collisions per beam crossing. Most of these collisions will be "soft" interactions, with relatively little energy exchanged between the protons and the outgoing particles having relatively little momentum perpendicular to the beam axis. These interactions are described in principle by Quantum Chromodynamics (QCD) [2; 3], the quantum field theory of the strong interactions. In practice, however, they are the most difficult to calculate, because the theory becomes non-perturbative at low energies. Predictions can only be made using approximations and phenomenological models. This difficulty with low-energy strong interactions appears even in processes whose core behavior is well-described by perturbation theory: outgoing high-energy quarks and gluons quickly "clothe" their strong color charge by evolving into jets of lower-energy hadrons, a process that again requires approximation and modelling.

The LHC's general-purpose experiments, ATLAS [4] and CMS [5], are equipped with multi-stage trigger systems that select against these "average" processes, for example by identifying leptons and missing energy produced in electroweak interactions. However, low-energy QCD still has a significant impact on the physics program in several areas. Because there are so many collisions in each crossing, the most interesting collisions will have many low-energy collisions whose signals in the detector overlap with the objects of interest. In order for their effects to be subtracted, these features of these *pileup* collisions must be known quantitatively. The evolution of high-energy hadronic jets, which ultimately relies on low-energy QCD radiation, must also be well-understood. This is partially to account for their contribution as pileup events, but their energy must also be calibrated so they can be studied in their own right. For example, although even very high-energy jets are relatively common at the LHC,

they can also serve as signatures of the decay of new particles [6].

The quantitative investigation of low-energy QCD is thus a foundational element of the LHC program, which will inform the studies and discoveries of the coming years. Initial low-energy QCD measurements have divided the problem between low-energy events and the study of higher-energy jet properties. In the former case, inclusive charged particle distributions are produced from events identified using a "minimum bias" trigger [7; 8], along with studies of particle properties away from the most active part of such events ("underlying event" distributions) [9; 10]. In the latter case, higher-energy jets are triggered and studied using the calorimeter system built for the purpose [11; 12; 13].

This work focuses on the additional information to be gained in the case that the two issues are not so easily factorized, by studying the emergence of low-energy jets from soft interactions. Particles are identified using the methods of the lowest-energy measurements, but grouped together into jets according to the algorithms used to study jets at higher energies. Low transverse momentum jets and their properties are measured using the ATLAS Inner Detector, the component of the ATLAS experiment that tracks charged particles, in events identified using the ATLAS Minimum Bias Trigger Scintilators. The transition from soft to high-energy jets can thus be studied. Furthermore, there is known to be tension in giving a single theoretical description that works simltaneously for "minimum bias" and "underlying event" distributions [14]; the addition of the jet-related observables measured in this work may shed light on this situation.

The objects studied in this work are referred to as *charged particle jets*. They are defined based on the acceptance of the ATLAS Inner Detector, including charged particles with momentum transverse to the beam axis (p_{T}) above 300 MeV/c. Although information is lost – specifically, the momentum carried by neutral particles and those with lower momentum – the corresponding gain is that charged particle jets can be measured very precisely. Every constituent particle in a charged particle jet is separately measured, allowing corrections with uncertainties that are dominated by well-understood detector precision rather than theoretical extrapolation.

Charged particles have previously been used to measure jets in proton-proton collisions at the CERN Intersecting Storage Rings [15], and in proton-antiproton collisions at the CERN Super Proton Synchrotron [16] and Fermilab Tevatron [17]. Fragmentation of charged particles with respect to calorimeter jets has also been measured [18; 19].

This work presents a measurement of several charged particle jet properties: (1) the cross section as a function of transverse momentum and rapidity; (2) the distribution of particles per jet; (3) the distribution of charged particle longitudinal momentum fractions with respect to the containing jet; (4) the distribution of charged particle transverse momenta with respect to the containing jet; and (5) the number

density of charged particles as a function of angle from the containing jet axis. These measurements will complement existing studies of jet properties and inclusive collisions, enabling a better understanding of strong interactions at the Large Hadron Collider. They can be used quantitatively to improve the modelling of low-energy interactions, so that the programs used for background studies as part of other LHC analyses give more accurate results.

Chapter 2 of this work describes the theory of QCD and the approximations used to understand low-energy collisions and the evolution of jets. Chapter 3 discusses the Large Hadron Collider complex and the ATLAS detector, with a particular focus on the subsystems used for this analysis. Chapter 4 presents the quantities measured and corrections used to obtain them in detail. Chapter 5 describes the systematic uncertanties in the measurement, based on uncertainties in tracking efficiency and momentum measurement, on the propogation of modelling variations through to the momentum distributions being measured, and on various assumptions made in the analysis. Finally, Chapter 6 presents and discusses the measured distributions, and Chapter 7 concludes and explores their potential impact on future measurements at the LHC.

Chapter 2

Quantum Chromodynamics in Proton-Proton Collisions

Quantum Chromodynamics (QCD) [2; 3] is the quantum field theory of strongly interacting particles. At the level of its fundamental particles, it is mathematically almost equivalent to the theory of Quantum Electrodynamics (QED) [20; 21]; both have families of fermions (quarks and leptons, respectively) and spin-1 gauge bosons (gluons and photons, respectively). The crucial difference is that the symmetry of QCD is the non-abelian $SU(3)$ group, rather than the abelian $U(1)$ of QED*, so that the gluons of QCD are self-interacting. This results in fundamentally different low-energy behavior, giving rise to many of the complexities of physics at the LHC, not to mention many of the complexities in the structure of bulk matter.

The coupling constant α_S of QCD increases with decreasing energy, becoming order unity at a scale of $\Lambda_{QCD} \sim 200$ MeV [29]. Around this energy and below, interaction rates and other quantities cannot be calculated using perturbation theory, which expands quantities in powers of α_S. Qualitatively, this corresponds to the exchange of a very large number of very low energy gluons, or simply a breakdown of the treatment of the gluon field as a collection of quanta with well-defined momentum. Therefore, in order to model the low-energy behavior of strongly interacting particles, phenomenological models and approximations must be used. As a result, parameters beyond those fundamentally present in QCD are introduced; different approximations apply over the course of a single proton-proton interaction, and there is additional freedom in determining the scales at which one model transitions into another.

*In fact, QED is now known to be the low-energy form of the non-abelian Electroweak theory [22; 23; 24], whose symmetry group is $SU(2) \times U(1)$. QED emerges from this theory when the symmetry is broken via the Higgs Mechanism [25; 26; 27]. This model, combined with QCD, is known as the Standard Model of particle interations [28]. It should be noted that only QCD interactions have a significant impact on the quantities measured in this analysis.

This chapter reviews the models of low-energy strong interactions that impact the behavior of low-energy collisions and jets at the LHC, with a particular focus on parameters that affect the modelling of the quantities measured in this work. After a discussion of the basic properties of QCD, the components of the total proton-proton cross section are described. The "soft" part of the total cross section – that is, the component involving relatively little momentum exchange between the protons – may be modelled through exchanges of hadrons and other composite states that appear in *Regge Theory*. Larger momentum exchange is modelled using QCD more directly via the *parton model*, which allows a perturbative interaction between quarks and/or gluons to be factorized from the behavior of the "remnants" of the collided protons and the soft interactions between them. Before and after the perturbative interaction, *parton evolution* is the process by which an outgoing quark or gluon radiates others. The phenomenological models that convert the resulting collection of partons into stable hadrons is known as *hadronization*. The discussion then returns to the proton remnants, which along with additional parton interactions between the protons, form the *underlying event*. The chapter then concludes with details of how these models are combined into Monte Carlo Event Generators, programs which produce simulated events whose properties are intended to match the actual distributions of particles observed at the LHC.

The simplest picture of the origin of jets is that they are a group of collimated particles formed after the radiation and hadronization of a single high-momentum quark or gluon. Realistic models, however, produce jets that contain contributions from all of the elements described in this chapter.

2.1 Quantum Chromodynamics

The Lagrangian of QCD is given by [30]

$$\mathcal{L} = \sum_q \bar{q}_a (i\gamma^\mu D_{ab\,\mu} - m_q \delta_{ab}) q_b - \frac{1}{4} F^A_{\mu\nu} F^{A\,\mu\nu}, \tag{2.1}$$

where q represents the quark fields, Greek letters are 4-vector indices, lower-case Roman letters are quark color indices (1-3), and upper-case Roman letters are gluon color indices (1-8). Repeated indices are summed over. The covariant derivative $D_{ab\,\mu}$ is given by

$$D_{ab\,\mu} = \delta_{ab}\partial_\mu - igt^C_{ab}A^C_\mu, \tag{2.2}$$

where the gluon field tensor $F^A_{\mu\nu}$ is given by

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - g f^{ABC} A^B_\mu A^C_\nu. \tag{2.3}$$

The t^A_{ab} and f^{ABC} are matrices and structure constants associated with the $SU(3)$ symmetry group. A^A_μ is the gluon field, and $g = \sqrt{4\pi\alpha_S}$. The structural similarity to QED is manifest in the Lagrangian, which in the QED case is identical except that no color indices are needed, and the f^{ABC} term is absent so that the gauge field does not couple to itself.

Quarks transform under the fundamental representation of $SU(3)$, while gluons transform under the the adjoint representation. The six quark flavor eigenstates under QCD, which are also mass eigenstates, are listed in Table 2.1.

Table 2.1: Basic quark and gluon properties in QCD. Masses from [31].

Particle	Symbol	Spin	Electric Charge	Mass
(Anti-) up quark	u ($\bar{u}$)	$\frac{1}{2}$	$+\frac{2}{3}(-\frac{2}{3})$	$1.7 - 3.3$ MeV
(Anti-) down quark	d ($\bar{d}$)		$-\frac{1}{3}$ $(+\frac{1}{3})$	$4.1 - 5.8$ MeV
(Anti-) charm quark	c ($\bar{c}$)		$+\frac{2}{3}$ $(-\frac{2}{3})$	$1.27^{+0.07}_{-0.09}$ GeV
(Anti-) strange quark	s ($\bar{s}$)		$-\frac{1}{3}$ $(+\frac{1}{3})$	101^{+29}_{-21} MeV
(Anti-) top quark	t ($\bar{t}$)		$+\frac{2}{3}$ $(-\frac{2}{3})$	$4.19^{+0.18}_{-0.06}$ GeV
(Anti-) bottom quark	b ($\bar{b}$)		$-\frac{1}{3}$ $(+\frac{1}{3})$	$173.1 \pm 0.6 \pm 1.1$ GeV
Gluon	g	1	0	0

From this Lagrangian, a set of Feynman rules can be derived, which allow the calculation of matrix elements $\mathcal{M}$ for specific perturbative processes. For example, the scattering of two gluons can be calculated at leading order in perturbation theory from the four diagrams shown in Figure 2.1. These diagrams represent a numerical value for the matrix element, as a function of incoming and outgoing colors, helicities, and momenta. Diagrams have a complex phase and hence multiple diagrams may interfere either constructively or destructively. In general, the matrix element predicts the cross section via

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 E^2_{cm}} |\mathcal{M}|^2, \tag{2.4}$$

where $d\Omega$ is the solid angle and E_{cm} is the center-of-mass energy of the incoming particles.

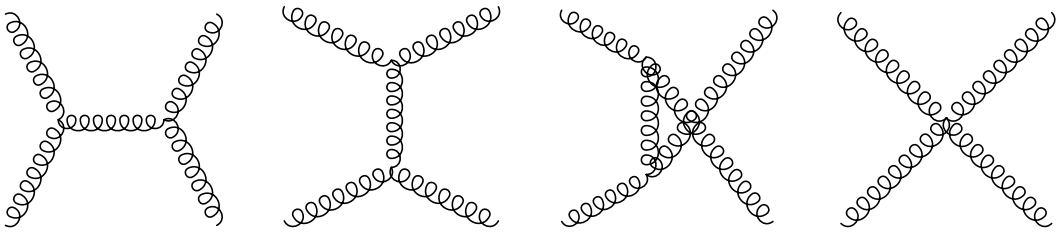


Figure 2.1: The four diagrams contributing to $gg \to gg$ at leading order.

When QCD calculations are made at higher order, corresponding to quantum corrections to the simplest ("tree-level") diagrams for a process, additional complexities arise. Loop diagrams, for example the lowest-order corrections to the gluon propagator shown in Figure 2.2, give matrix element values that diverge at high energies.

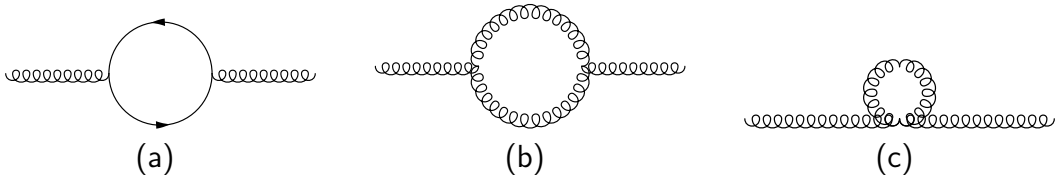


Figure 2.2: Corrections to the gluon propagator at next-to-leading order in perturbation theory. These are some of the diagrams that impact the energy scale-dependence of the coupling constant α_S.

A procedure known *renormalization* removes the divergences but introduces an arbitrary scale μ_R, at which the subtractions are performed, that does not vanish from calculations. Of course, physical observables must be independent of this parameter, which is assured by absorbing the μ_R-dependence into the physical coupling constant and quark mass. The physical coupling constant now depends on the scale Q^2 of the momentum exchanged in the process being calculated. This is given by the differential equation [29]

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = -\frac{33 - 2n_f}{12\pi} \alpha_S^2 + ..., \tag{2.5}$$

which can be solved to give

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu_R^2)}{1 + [(33 - 2n_f)/12\pi]\alpha_S(\mu_R^2)\log(Q^2/\mu_R^2)}. \tag{2.6}$$

This equation can be solved to eliminate μ_R, giving

$$\alpha_S(Q^2) = \frac{12\pi}{(33 - 2n_f)\log(Q^2/\Lambda_{QCD}^2)}, \tag{2.7}$$

where n_f is the number of kinematically accessible quark flavors. This Equation defines the Λ_{QCD}. It should be noted that the definition of Λ_{QCD} therefore depends on the number of flavors and on the order of perturbation theory at which calculations are being made, as does $\alpha_S(Q^2)$; this is acceptable because these quantities themselves are not physical observables, but it does mean that the appropriate definitions must be taken when calculating observables in order to ensure consistency.

Because there are only six known quark flavors, the factor $33 - 2n_f$ is always positive; thus, as illustrated by Equation 2.5, $\alpha_S(Q^2)$ decreases with increasing Q^2. This is the origin of asymptotic freedom, but also dictates that low-energy processes become impossible to calculate in straightforward perturbation theory. By contrast, in QED the equivalent factor is always negative, so perturbative electrodynamic calculations can be made at very low energies (although in principle they become strong at very high scales). In both cases, corrections due to fermion loops tend to make the coupling constant decrease with decreasing energy, but in QCD the impact of gluon loops is the opposite and much larger.

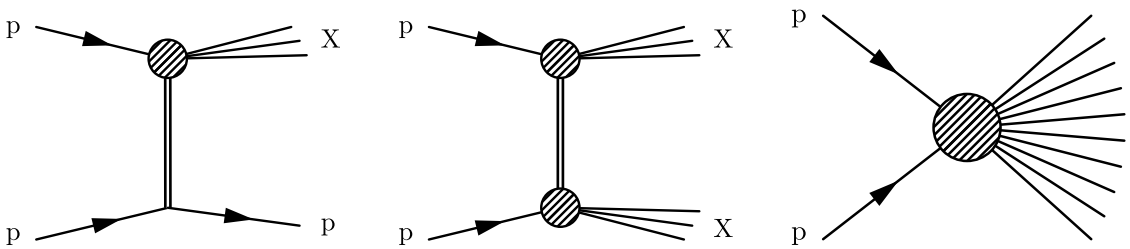
2.2 Total Proton-Proton Cross Section

The total proton-proton interaction cross section can be divided into different types of interactions, defined based on the particles exchanged in the interaction and the ultimate fate of the interacting protons themselves. It should be noted that the precise division betweens the components depends in some cases on the particular implementation of the model, rather than on model-independent physical observables.

The first division is between *elastic* and *inelastic* interactions. Elastic scattering is the straightforward scattering of two incoming protons into two outgoing protons. The remaining inelastic component may be referred to collectively as *minimum bias events*, such as might be collected using a minimum bias trigger (cf. Section 3.2.2.1). These may be further subdivided into *diffractive* and *non-diffractive* events.

Qualitatively, diffractive events are those in which no color charge is exchanged between the two protons; a colorless state called a *pomeron* is exchanged and one or both protons "break up." In the context of QCD, the pomeron may be interpreted as

colorless combinations of gluons. *Single-diffractive* collisions are those in which one proton remains intact, as illustrated in Figure 2.3(a), and *double-diffractive* collisions are those in which neither proton survives, as shown in Figure 2.3(b). Both types of events are characterized by a rapidity (y)[†] gap corresponding to the gap between the remnants of one proton and the other.



(a) Single-diffractive (b) Double-diffractive (c) Non-diffractive

Figure 2.3: A schematic of single-, double-, and non-diffractive proton-proton collisions. Double lines indicate the exchanged pomeron.

In non-diffractive collisions, depicted in general in Figure 2.3(c), objects with color charge are exchanged between the components of the proton. At higher momentum exchange between protons, non-diffractive collisions are well-modeled by factoring out a perturbative QCD interaction via the parton model, as described in the next section. At lower momentum exchange, events may be "built up" by overlapping multiple parton interactions even in the regime where QCD is non-perturbative. Although this is not theoretically well-motivated, models such as those implemented in the Pythia and Herwig++ Monte Carlo programs described in Section 2.7 are able to describe a range of experimental data through the use of tunable parameters governing the quantity and momenta of multiple interactions.

Models of soft interactions with a better theoretical justification are based on Regge Theory [32], or a combination of Regge Theory with QCD. Regge Theory is a pre-QCD model of the strong interaction that relies on model-independent symmetries of scattering processes and the analytic continuation of momenta into unphysical

[†] In this work, proton-proton collisions are analyzed in a right-handed coordinate system with the origin as the collision point. The Z-axis points along the beam direction. In the context of the LHC, it is chosen to point in the counter-clockwise direction. The X-axis points to the center of the LHC ring and the Y-axis points upwards. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ is measured from the Z-axis. The rapidity y for a track or jet is defined by $y = \frac{1}{2}\ln\frac{E+p_Z}{E-p_Z}$, where E is the object's energy and p_Z is the component of its momentum along the Z-axis. It is equivalent to the pseudorapidity $\eta = -\ln(\tan\frac{\theta}{2})$ in the massless limit. In this work, the "distance" between two objects $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$ is defined using rapidity.

regions to relate hadron properties to strong scattering cross sections. For example, the *multiperipheral model* [33] deals with the exchange of many strongly interacting pomerons and *reggeons*, which correspond to superpositions of hadron states with a particular spin. The *dual parton model* [34] adapts the techniques of the multiperipheral model to the exchange of many non-perturbative quarks and gluons, and is implemented for soft interactions in the Phojet Monte Carlo program described in Section 2.7.

2.3 The Parton Model

In the parton model [35], high-momentum interactions between two protons can be factorized into their high-momentum (short-distance) and low-momentum (long-distance) components. Because of asymptotic freedom, the high-momentum component can be computed as the interaction of one parton from each proton. The partons have momentum fraction x_1 and x_2 of the parent proton's momentum, which at the LHC is a constant P for both proton beams (3.5 TeV in the present run). The largest contribution to the cross section at the LHC is through gluon-gluon interactions, in which case the factorization can be described by the equation

$$\sigma(pp \to X + Y) = \int dx_1 dx_2 f_g(x_1) f_g(x_2) \sigma(gg \to X). \tag{2.8}$$

This equation is illustrated schematically in Figure 2.4. Here X is some particular output of the gluon-gluon scattering. Y is the *beam remnant*, a collection of particles arising from the remainder of the proton; this, along with additional soft interactions (not shown) between the protons, is described in Section 2.6. The total cross section for the interaction of two gluons of momentum $x_1 P$ and $x_2 P$, $\sigma(gg \to X)$, is illustrated by the cross-hatched blob in the diagram, The *parton distribution functions* (PDFs) $f_g(x)$ represent the probability distribution for gluons with momentum fraction x when the proton interacts. Although the equation specifies gluons as written, in its more general form it includes a sum over all possible partons, i.e., quarks and antiquarks of various flavors.

Some sample parton distribution functions are shown in Figure 2.5. The basic makeup of the proton, two up quarks and a down quark (the *valance* quarks), is reflected in the following rules for the PDFs:

$$\int [f_u(x) - f_{\bar{u}}(x)]\, dx = 2, \tag{2.9}$$

$$\int [f_d(x) - f_{\bar{d}}(x)]\, dx = 1. \tag{2.10}$$

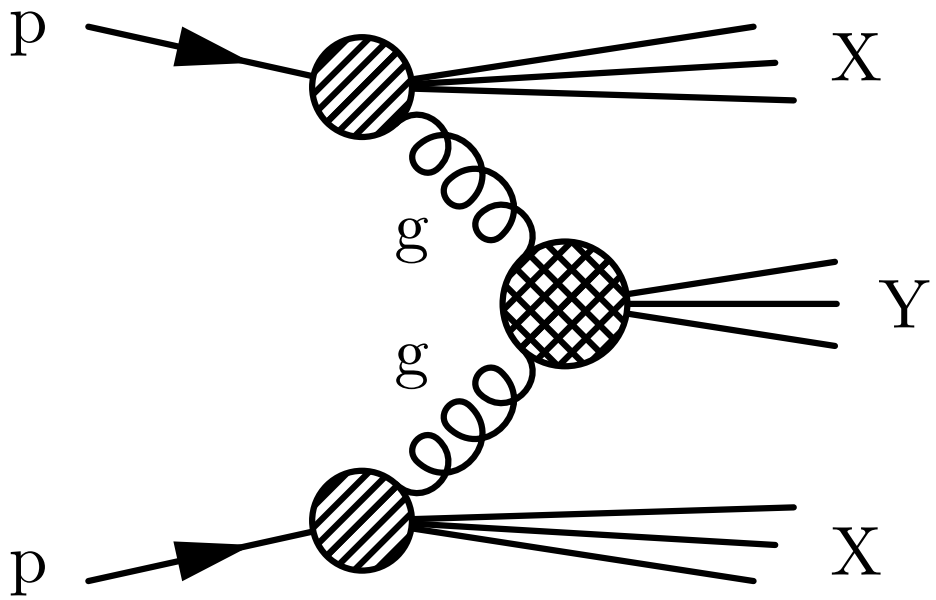


Figure 2.4: A schematic illustrating the parton model as described in Equation 2.8 and the surrounding text. Diagonally shaded blobs represent the extraction of a gluon from the proton via the PDFs, while the hatched blob represents the hard gluon-gluon scatter into some particular product Y. Beam remnants are also shown. Other aspects of high-momentum proton-proton collision modelling that are not shown include multiple parton interactions and the soft underlying event.

The *sea* quarks, including low-x up and down quarks as well as the strange quarks shown as an example, are found because of quark-antiquark pair production. Roughly half of the proton momentum is found in gluons, which dominate at low x.

The usefulness of the parton model is based on the universality of the PDFs, in the sense that they give the probability of finding partons at a given x for many processes. Thus they can be derived from one type of experiment and applied to another, in this case LHC collisions. One major source for deriving PDFs is electron-proton scattering experiments ([31], 201-214); in these experiments, there is only one unknown PDF rather than two so that results can be determined in a more straightforward manner. It can be proven that PDFs are not changed by long-range proton-proton interactions, i.e., that it is valid to use electron-proton scattering PDFs for proton-proton interactions, up to corrections of that fall with the center-of-mass energy [29].

In the "naive" formulation of the Parton Model, the cross section $\sigma(gg \to X)$ is simply the cross section calculated in parton-level QCD as described above. When

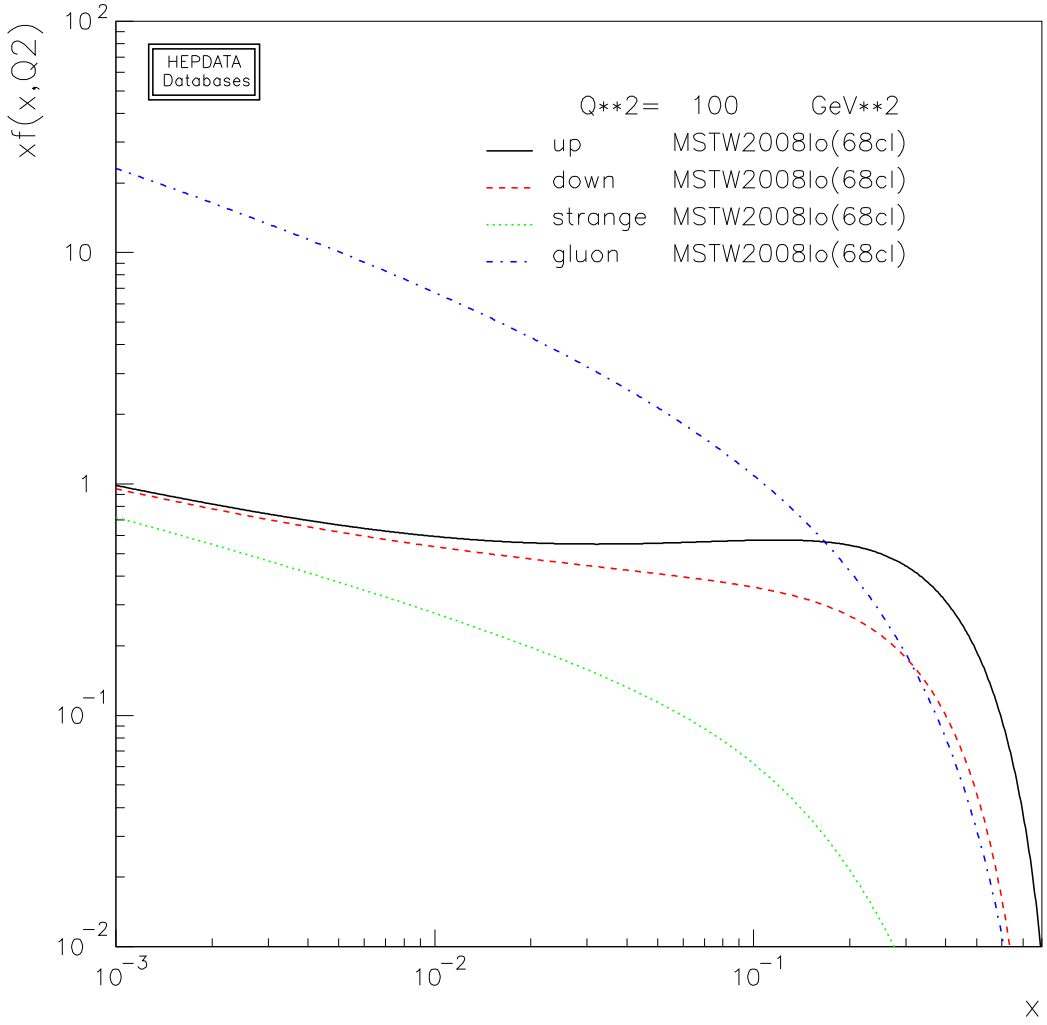


Figure 2.5: Selected parton distribution functions, illustrating some basic properties of PDFs. Gluons are dominant at low momentum fraction x, while at high x the valance quarks dominate. The up quark PDF is approximately twice that of the down quark in the latter range, because the core constituents that define a proton are two up quarks and one down quark. Generated with [36].

higher-order QCD corrections are included, they must be *factorized*; the low-energy component of the corrections includes interactions with a long timescale that should be included in adjustments to the PDFs rather than the hard scatter. Thus PDFs are dependent on the *factorization scale* μ_F as well as the renormalization scale μ_R,

both of which are typically taken to be equal to the scale of the interaction Q^2. As long as the PDFs and hard-scattering calculations use consistent values, the results should not depend greatly on the choices of μ_F and μ_R; in practice, the differences that arise from "reasonable" variations in these parameters is taken as part of the systematic uncertainty on theoretical calculations. The details of factorization, and hence the PDFs, also depend on the order of perturbation theory in which the hard scattering is calculated.

2.4 Parton Evolution

The overall picture of a single parton from each proton interacting as in Figure 2.4, and interacting at leading order as in Figure 2.1, does not provide an adequate picture of collisions with many outgoing low-energy hadrons; these final-state particles arise because both incoming and outgoing partons may radiate additional partons before hadronization. Although the production of an additional quark or gluon may be included with a complete next-to-leading-order calculation of the hard scattering process, the difficulty persists in any reasonable fixed-order calculation because the radiation of many soft and collinear partons enhances the cross section. This problem may be addressed by treating the evolution of a set of partons as a sequence of separate splittings; this is a semi-classical approximation because it neglects quantum interference between radiation from different partons and between different shower orderings. However, the approximation is most accurate in the case of soft and collinear radiation under consideration; it allows each splitting to be considered separately, as depicted in Figure 2.6. The matrix element squared for these individual splittings, generally, is of the form

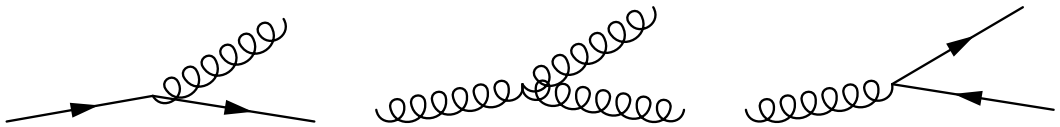
$$|\mathcal{M}|^2 \sim \int \frac{dk_{\mathrm{T}}^2}{k_{\mathrm{T}}^2} \frac{\alpha_S}{2\pi} P(z), \tag{2.11}$$

where k_{T} is the momentum of the emitted parton transverse to the incoming parton's momentum vector, z is the momentum fraction of the emitted parton, and $P(z)$ is a *splitting function* that depends only on z and the particular splitting (e.g., $g \to gg$, $g \to q\bar{q}$, $q \to qg$, etc.). The integral goes from the momentum scale μ at which the interactions become non-perturbative up to the scale of the interaction Q^2, so we have

$$|\mathcal{M}|^2 \sim \alpha_S \log(\frac{Q^2}{\mu^2}). \tag{2.12}$$

Because n splittings are treated sequentially in this approximation, they result in a factor of $\left[\alpha_S \log(\frac{Q^2}{\mu^2})\right]^n$. This factor, unlike powers of α_S occurring without a logarithm, can often be summed to all orders in α_S, either analytically or through

a random Markoff process; separating out this part of the diagram and summing it while calculating the rest at fixed order is the *leading logarithm approximation*. In the case of soft radiation from the incoming partons, these leading logarithms are incorporated into the dependence of the PDFs on Q^2 via the DGLAP equations [37; 38; 39]. These are a set of coupled differential equations for the PDFs as a function of Q in terms of the splitting functions P.



(a) $q \to qg$ (b) $g \to gg$ (c) $g \to q\bar{q}$

Figure 2.6: Diagrams illustrating the allowed types of "splitting" of quarks and gluons into each other, which are accounted for in the DGLAP equations and in the sequential modelling of parton radiation.

For the practical calculation of many-hadron final states, a Monte Carlo procedure may be implemented to produce random shower configurations on an event-by-event basis. The probability for a parton with an initial virtual mass-squared t_i[‡] to evolve to mass-squared t without radiating is given by the *Sudokov form factor*:

$$\Delta(t) = \exp\left[-\int_{t_i}^{t} \frac{dt'}{t'} \int dz \frac{\alpha_S}{2\pi} P(z)\right]. \tag{2.13}$$

The shower may then be produced as follows. For each parton that may shower, a random number generator is used to determine the value of t at which the radiation occurs, in a manner such that many such calculations would produce the appropriate probability distribution. The procedure is then repeated for both resulting partons. Branchings are added successively to all partons produced until the scale t is low enough that perturbative physics is no longer applicable and the shower is terminated. (This scale t_0 is matched with the scale of hadronization and is described in the next section.) The process is applied to both incoming partons (*initial state radiation*) and outgoing partons (*final state radiation*),

It should be noted that Equation 2.13 assumes that the shower is ordered in the virtuality t. Because there is no intrinsic order in the true quantum mechanical process, the order in which radiation occurs in this approximation is arbitrary. However, different orderings produce results with different advantages. For example, ordering

[‡] The virtual mass-squared t for a particle is given by $t = m^2 = E^2 - \vec{p}^2$; this is *not* constrained to equal the true particle mass during short-lived quantum-mechanical processes.

in the transverse momentum k_T or the angle of the radiation produces a treatment of the angular distribution of outgoing particles that better matches the expected results from a full calculation of the radiation including interference.

2.5 Hadronization

Ultimately all partons must be bound together in hadrons, a process that is entirely non-perturbative. Therefore parton showering must be cut off at some scale t_0, which is arbitrary but typically well above Λ_{QCD}. To complete the event-by-event description from this point, phenomenological models of hadronization then convert the final particles in the shower into colorless hadrons. In order to minimize the model-dependence of the showering-cutoff parameter, the hadronization model should also depend on this variable; thus, if t_0 is set high so that fewer partons are present to hadronize, the model should in some manner produce more hadrons per parton than it would with a lower t_0 [29].

Two genres of hadronization models are typically used: the *string model* [40] and the *cluster model* [41]. Both models rely on the color-connectedness of adjacent quarks and gluons, known as a planar configuration because lines of color do not "cross"; this is true to a good approximation because closed loops of color can assume any of $N_C = 3$ colors, and so are enhanced by a factor of N_C^2 over related non-planar diagrams [42].

2.5.1 String Fragmentation Model

In this model, strings of color connect the quarks from the shower to antiquarks. The strings are connected via gluons from the shower, with the gluons becoming "kinks" with localized energy. Insofar as these strings are physically realized, they correspond to flux tubes of the strong color field. The strings are decayed into hadrons by producing quark-antiquark pairs which "break" the string; these are assigned momentum fractions according to the splitting function $P(z)$ introduced in Section 2.4. Because hadrons are produced along the length of the string, this model gives an enhanced probability for hadrons in certain regions between jets in multijet events. The simplest example is three-jet events in electron-positron collisions, where strings are expected to extend from a jet associated with an outgoing quark, through a jet associated with a radiated gluon, and on to a jet associated with an outgoing antiquark. As a result, hadrons are predicted to be more plentiful in (anti)quark-gluon jet gaps and less so in the quark-antiquark jet gap, as has indeed been observed [43; 44].

The most straightforward model for producing string connections between out-

going quarks and gluons can produce an excessive energy density of strings in the central region (transverse to the beams), as color-connected partons may end up on opposite sides of the event. Models of *color reconnection* have therefore been introduced, which recompute the color reconnections after the completion of the parton shower in order to minimize the total string potential energy [45]. Conceptually, this is justified by the idea that in dense events, strings may cross and change to more energetically favorable configurations.

2.5.2 Cluster Model

The cluster model produces color-neutral clusters from quark-antiquark pairs in the parton shower, after first splitting the remaining gluons. These clusters are then interpreted as excited hadron pair resonances [46] and decayed into hadron pairs according to the available phase space. Such models must deal effectively with rare high-mass clusters, for which decay into a single hadron pair is unlikely. More complex decay schemes may be used, or the clusters may be split into appropriate masses before decay.

2.6 Beam Remnant and Underlying Event

In addition to particles produced in hard scattering and radiation from incoming and outgoing partons, there are several contributions to the particle production in a proton-proton collision. This problem is factored into the issue of beam remnants, i.e., the simulation of the remaining energy-momentum and charge of the proton after a parton is extracted, and the underlying event, a catch-all term for soft interactions between the two protons beyond the hard partons.

The beam remnant may be simulated starting with a few partons that carry the appropriate color charge. For example, if a quark has been extracted from the proton, then an additional quark pair balances its charge; this pair is given the momentum leftover from the original proton. Parton shower evolution applied to these partons, in the same manner as for initial and final state radiation. Because the system has very large momentum along the original beam direction, most of the resulting particles will be outside the acceptance of standard collider detectors.

A widely used model of the underlying event is via *multiple parton interactions*, in which it is assumed that multiple independent "semi-hard" perturbative scatterings between separate partons account for the additional activity. This model can account for the fact that the predicted jet production cross section in high-energy proton-proton collisions appears to exceed the total proton-proton interaction cross section if the cross section is extrapolated down to a jet momentum of 2 GeV. Multiple jet

production interactions occur in one proton-proton collision. In the *eikonal model* [47; 48], the number of collisions is a Poisson distribution with mean based on a proton-proton impact parameter and an overlap function reflecting the distribution of matter within the proton. Different versions of the multiple parton interaction model have a number of tunable parameters, related to the momentum cutoff for the additional scatters and the matter distribution within the proton. With appropriate parameters, this model correctly produces the increase in underlying event density based on the momentum of the hard scatter; because hard scatters simply represent the high-momentum, low-probability tail of the semi-hard scatters, they are most likely to occur in collisions with many interactions.

The *Soft Underlying Event* model [49] is a parameterization of underlying events based originally on simulations created by the UA5 experiment. It is based on a number of independent cluster decays, with free parameters corresponding to the distribution of the number of particles, cluster masses, and momentum distribution of outgoing particles.

2.7 Monte Carlo Event Generators

A Monte Carlo Event Generator is a computer program that combines choices of the models described above to form a complete description of the particles produced in a collision on an event-by-event basis. Probability distributions for event properties can thus be computed by simulating large numbers of events without an impossibly detailed non-perturbative quantum mechanical calculation of possible distributions of outgoing hadrons. A number of programs exist; some focus on particular applications, for example low-energy collisions or next-to-leading-order calculations of certain processes, while others are designed to be as general as possible.

The PYTHIA Monte Carlo [50] program computes a very large number of processes at leading order, produces minimum bias collisions and underlying events using multiple parton interactions, and uses the string hadronization model. It has a large number of tunable parameters associated with its component models, and significant work has been done to determine best values of these parameters based on a range of collider data. After tuning, PYTHIA is very succesful in describing a wide range of observables. In this work, a number of different PYTHIA 6.421 *tunes*, i.e., sets of values for model parameters, are used. Although all are designed to produce distributions consistent with existing data, they are based on different tuning strategies and on reasonable variations in how that consistency is achieved.

The Perugia family of tunes [51] uses the CTEQ5L [52] parton distributions; there is a central value ("Perugia-0") and several variants that attempt to bracket the possible changes in tune parameters (and hence event structure) that are per-

mitted while still fitting existing data. The central value ("Perugia-0") is tuned to correctly produce the Drell-Yan [53; 54] (i.e., scattering through a virtual photon or Z boson) transverse momentum spectrum, the minimum bias average momentum as a function of charged particle multiplicity, and the total particle multiplicity, based on hadron collider data at several energies. Two variants, "Perugia HARD" and "Perugia SOFT," are created changing the Initial State Radiation cutoff scale so that the perturbative contribution is different, with other parameters adjusted to bring the tuned distributions back into agreement. Perugia HARD (SOFT) has more (fewer) jets with higher (lower) momentum, and a higher (lower) charged multiplicity. Another variant, Perugia 2010, is an adjustment of Perugia-0 that improves the description of jet shapes and adjusts hadronization parameters for better consistency with data collected at the Large Electron-Positron collider (LEP), given that the original values for these parameters were based on Q^2 shower ordering. Finally, Perugia NOCR ("no color reconnection") represents an effort to reproduce the same data with no color reconnection model used; it is used in computing systematic uncertainties, but because it could not be made to fit all the original tuning data it is not included on comparison plots. All Pythia tunes described thus far use p_{T}-ordered showering; the DW tune [55] is also used in order to include the impact of Q^2 ordering.

Two Pythia 6.421 tunes produced by the ATLAS collaboration are also used, MC09 [56] and AMBT1 [57]. These use the MRST LO* PDFs [58] and are tuned based on Tevatron jet distributions, charged particle multiplicities, and momentum and charged particle densities as a function of leading jet p_{T}. The AMBT1 tune is further updated based on single track distributions in ATLAS minimum bias data [7].

This analysis also uses PYTHIA version 8.145 [59] with the 4C tune [60]; this version of the program is actually a complete rewrite of the older version, albeit with very similar models. One improvement is that, whereas in PYTHIA 6 the single- and double-diffractive events never produce central jets, this occasionally does happen in the PYTHIA 8 models; thus the jet content in inclusive minimum bias events is more realistic.

The Herwig++ Monte Carlo [46] is another general-purpose program for leading-order calculations. It uses cluster instead of string hadronization. Although earlier versions of the program used the soft underlying event model, the program now uses multiple parton interactions. A Herwig++ 2.4.2 sample is used with its default settings, along with an additional sample tuned on 7 TeV underlying event data with Herwig++ 2.5.1 (UE7) [61; 62].

Finally, the PHOJET 1.12.1.3 Monte Carlo [63; 64] is used with its default tune. This Monte Carlo specializes in minimum bias events and uses the dual parton model, allowing multiple soft and hard collisions. It relies on Pythia 6.115 [65] for the fragmentation of partons.

Chapter 3

The Large Hadron Collider and ATLAS Detector

At the CERN laboratory near Geneva, Switzerland, protons are accelerated by a chain of accelerators to progressively higher energies, finally arriving in the Large Hadron Collider. Here they are accelerated, stored, and brought in collision at up to eight points around the LHC ring, of which five are currently used by experiments. The ATLAS detector, located at Point 1 near the main CERN site, is a general purpose detector designed to conduct measurements of any new particles produced at the LHC.

The LHC and ATLAS, along with the CMS experiment, were designed and built together with the overall goal of exploring very high-energy, low-rate interactions in particle physics. The LHC is designed to produce sufficient collisions with Higgs bosons for the hypothetical particle to be discovered, if it exists, at any mass consistent with known limits. Supersymmetry [66] and other extensions of the Standard Model of particle physics can be investigated to a significant degree. This set of physics goals makes it necessary to accurately measure directly a range of high-momentum objects that emerge from collisions: electrons, muons, photons, hadronic jets (including those with displaced vertices), and missing energy from undetected particles. From these, the creation and decay of a very wide range of short-lifetime particles, both known and unknown, may be inferred. (Two other experiments at the LHC, ALICE [67] and LHCb [68], have different physics programs, focusing on the collision of heavy ions and the physics of B-hadrons respectively.)

In order to achieve these goals, the LHC must accelerate protons to very high energy – 7 TeV at design energy, currently 3.5 TeV – and collide them at a very high rate. Proton-proton collisions are chosen because protons are stable, have a higher mass than electrons (and hence can be accelerated to higher energy without excessive synchrotron radiation), and are easy to obtain and accelerate in large quantities.

Their composite structure (cf. Section 2.3) results in a range of energies being made available for high-energy processes, an advantage in the search for new particles of unknown mass. The ATLAS detector has a variety of subsystems dedicated to the identification and measurement of high-momentum objects: calorimeters to measure photons, electrons, and jets, to differentiate between them, and to capture energy over the broadest range of solid angle possible; charged particle tracking to separate photons from electrons and identify displaced vertices; and a dedicated system to identify high-energy muons.

3.1 The Large Hadron Collider Complex

The LHC is the world's largest proton-proton collider, accelerating two beams of protons to a design energy of 7 TeV per proton. Up to 2808 bunches of protons, each with 10^{11} protons, are accelerated and stored for many hours so that a large number of collisions may be produced from the same bunches. In early running, the energy per proton is limited to 3.5 TeV per beam.

The LHC was built to take advantage of existing accelerator infrastructure at CERN. It was built in the tunnel that previously housed the Large Electron-Positron collider (LEP) [69]. The LHC uses the accelerator chain previously used for the Super Proton Synchrotron (SPS), including the SPS itself; many aspects of the accelerator chain have been upgraded in order to achieve the necessary beam properties for LHC injection. The LHC accelerator chain is described in detail in [70].

A schematic of the initial stages of the LHC accelerator complex is shown in Figure 3.1. Hydrogen ions are produced and accelerated to 100 keV in a duoplasmatron source, then accelerated to 50 MeV in Linac-2, a linear accelerator. The linac produces protons in bunches, which are combined to increase intensity in the Proton Synchrotron Booster (PSB). The PSB is the first of four storage rings which use radio frequency acceleration systems, keeping bunches together at the local minimum of the electromagnetic wave; the bunches can thus be gradually accelerated by increasing the frequency. Thus the PSB accelerates protons to 1.4 GeV, the Proton Synchrotron (PS) to 25 GeV, and so on. The bunch structure for the LHC is also established in the PSB and PS. After bunches are stored in the four parallel rings of the PSB, they are injected into the PS in two separate batches, initially filling six of seven allowed positions (i.e., the 7th harmonic). The bunches are split over several steps as they are moved to the 84th harmonic, so that there are 72 bunches with 25 ns spacing followed by 12 empty bunches; this gap is necessary so that kicker magnets can be turned on to eject the beam. Before ejection and transmission further, the bunches are rotated in order to decrease their length.

After acceleration to 25 GeV, the bunches in the PS are transferred to the Super

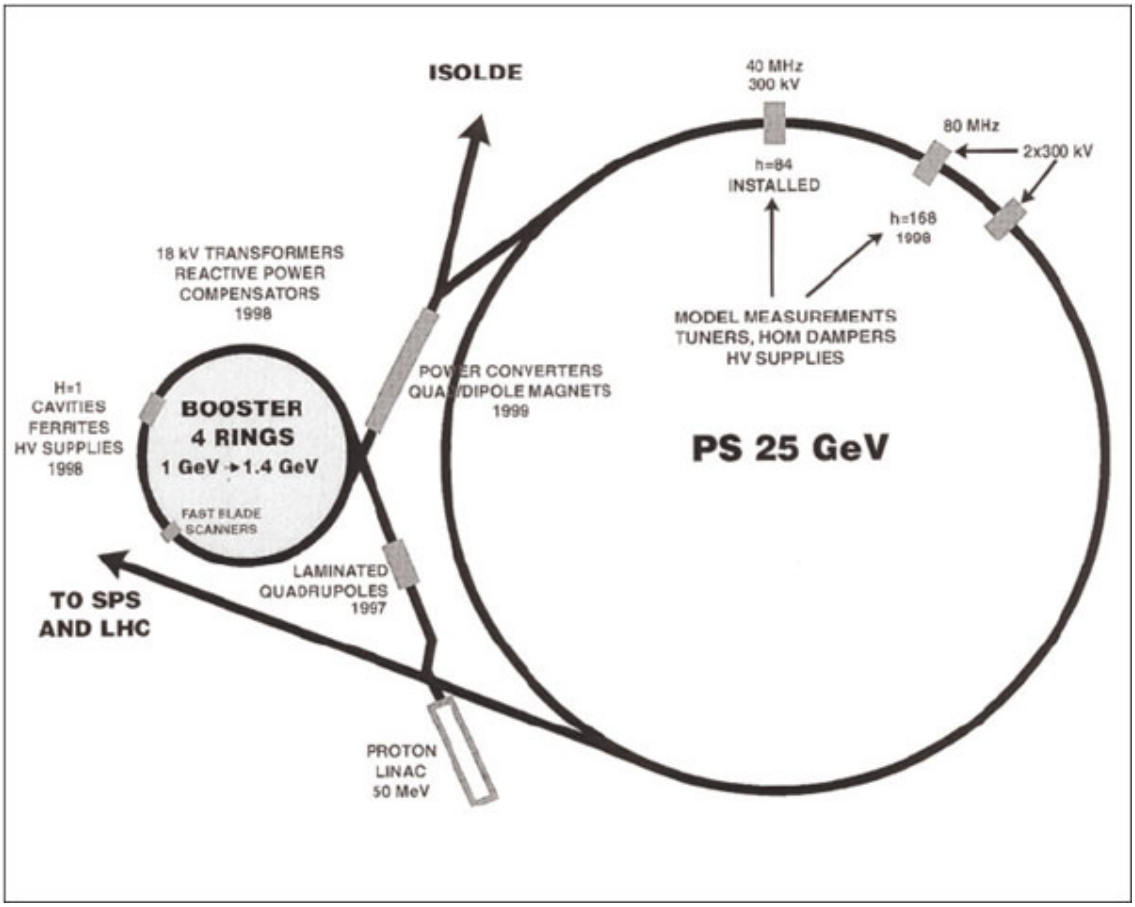


Figure 3.1: The initial stages of the Large Hadron Collider accelerator chain. Image from [71], used within terms.

Proton Synchrotron (SPS), which is shown along with the LHC in Figure 3.2. The SPS was formerly used as a proton-antiproton collider for the discovery of the W [72] and Z [73] bosons, and is still used in fixed target experiments and to produce test beams. Three or four PS batches are injected into the SPS before acceleration to 450 GeV, after which the bunches are compressed by an RF voltage increase and the tails of the bunch distribution are cleaned by copper scrapers. The full SPS batch is then injected into the LHC, giving a complex bunch structure that contains gaps reflecting the SPS injection kicker time (8 bunches after each 72-bunch PS batch) and the LHC injection kicker time (38 or 39 bunches every 3 or 4 PS batches); there is also one 119-bunch abort gap left for the kicker magnets that dump the LHC beam. This yields 2808 filled bunches in the LHC, out of 3564 possible bunch locations. It should

be noted that the above description covers the nominal LHC configuration, which has not yet been achieved. Other configurations involve fewer PS bunch splittings (e.g., for 75 ns timing), fewer SPS or PS batches, and other possibilities. The data for the present analysis was taken in runs where only a few bunches were present in the LHC machine.

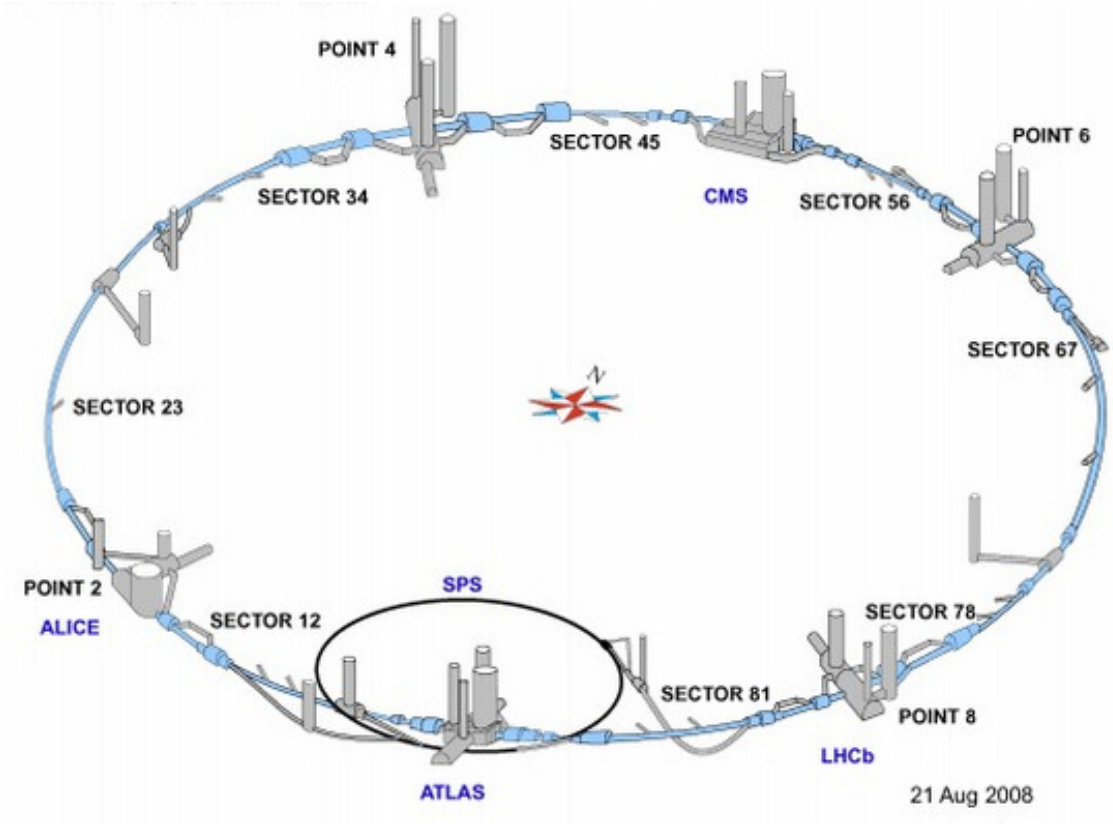


Figure 3.2: The layout of the LHC, including SPS accelerator and locations of experiments. Image from [74], used within terms.

The PS elementary cycle is 1.2 s, the PS can inject to the SPS every 3.6 s, and the SPS has a cycle of 21.6 s. This means that it takes at least 10 minutes to fill both LHC beams, after which the LHC energy is ramped to the full 7 TeV (or current 3.5 TeV) per beam.

Beyond the beam energy the other critical quantity that determines the measurements that can be made at the LHC is the *luminosity*, a variable that measures the density of proton bunches brought into collision and the rate of crossings per time. It thus determines the number of collisions per unit time N for a process with a given

cross section σ via

$$N = L\sigma. \tag{3.1}$$

The luminosity can be determined from the LHC beam parameters via

$$L = \frac{N_b^2 n_b f_{rev} \gamma F}{4\pi \epsilon_n \beta^*}. \tag{3.2}$$

Here N_b is the number of particles per bunch, n_b is the number of bunches, $f_{rev} = c/26.7\,\mathrm{km} = 11.2\,\mathrm{kHz}$ is the LHC revolution frequency, and γ is the usual relativistic γ-factor of the protons. F is a correction factor due to the LHC crossing angle at the interaction point, which is put in place to prevent "parasitic collisions" from occurring between unintended bunches, which would reduce the beam intensity and confuse detector data collection. ϵ_n is the normalized transverse beam emittance, a measure of the phase and momentum space occupied by the beam. β^* is the value of the beam envelope function beta at the interaction point, and hence a measure of the beam size and focusing distance at the collision point.

The LHC design luminosity is $10^{34}\,\mathrm{cm^{-2}s^{-1}}$, to be acheived with $N_b = 1.15 \times 10^{11}$, $\epsilon_n = 3.75\,\mu\mathrm{m}$, and $\beta^* = 0.55\,\mathrm{m}$. The data analyzed in this work reflect LHC instantaneous luminosities of at most $6.6 \times 10^{28}\,\mathrm{cm^{-2}s^{-1}}$, with only a few bunches in the machine and β^* either 2 m or 10 m.

3.2 The ATLAS Detector

The ATLAS detector is a large, general-purpose particle detector build around LHC interaction point 1, close to the main CERN site. It is designed to cover nearly the entire solid angle around the interaction point, excepting a narrow range around the LHC beam pipe: elements of ATLAS come within 1 degree of the beam axis. The angle from the beam axis is typically parameterized by the pseudorapidity η; in this variable, coverage by ATLAS extends to $|\eta| < 4.9$. ATLAS, like other such detectors, is built as a series of concentric subdetectors, which are built with cylindrical symmetry. An overall illustration of ATLAS appears in Figure 3.3.

The portion of the ATLAS detector closest to the interaction point is known as the Inner Detector (Section 3.2.1), and is devoted to the tracking of charged particles. It is surrounded by a 2T solenoidal magnet, which causes the paths of particles to curve in order to enable the measurement of their momentum. Outside of that, the calorimeter system (Section 3.2.3) absorbs and measures the energy of photons, electrons, and strongly interacting particles; its inner section (the *electromagnetic calorimeter*) has good spatial segmentation to optimize the separation of electrons, photons, and jets, while the outer *hadronic calorimeter* has sufficient depth to contain

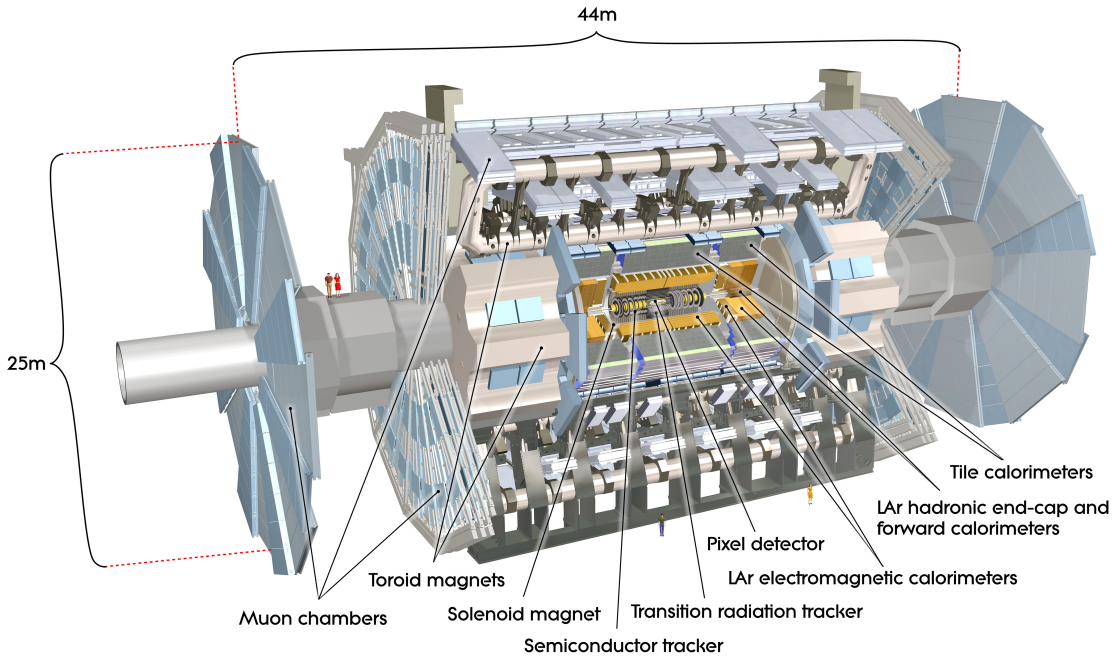


Figure 3.3: An overview of the ATLAS detector, from [4].

most hadronic jets. The outermost part of ATLAS is the muon system (Section 3.2.4), which identifies and measures muons, which pass through the rest of the detector if their momentum is above a few GeV.

3.2.1 Inner Detector

The Inner Detector, shown in Figure 3.4, occupies an overall volume 1.15 meters in radius and total length 7 meters along the Z axis, covering a pseudorapidity range of $|\eta| < 2.5$. It contains three separate technologies, with the highest precision in the innermost layers near the interaction points. The two innermost subdetectors, the Pixel Detector and Semiconductor Tracker, consist of well-separated layers of sensitive silicon, from which a charge is read out when electrons and holes are separated by a passing charged particle. The outermost Transition Radiation Tracker consists of arrays of 4 mm diameter gas-filled tubes; the gas is ionized by passing particles, with charges collected via the voltage between an outer cathode and inner anode wire. Each of these subdetector systems is separated into cylindrical *barrel* sections with active detector elements perpendicular to the radial direction, and *endcap* sections with detector elements perpendicular to the beam; this design optimizes the resolution

perpendicular to the particle path and minimizes the total material that particles pass through.

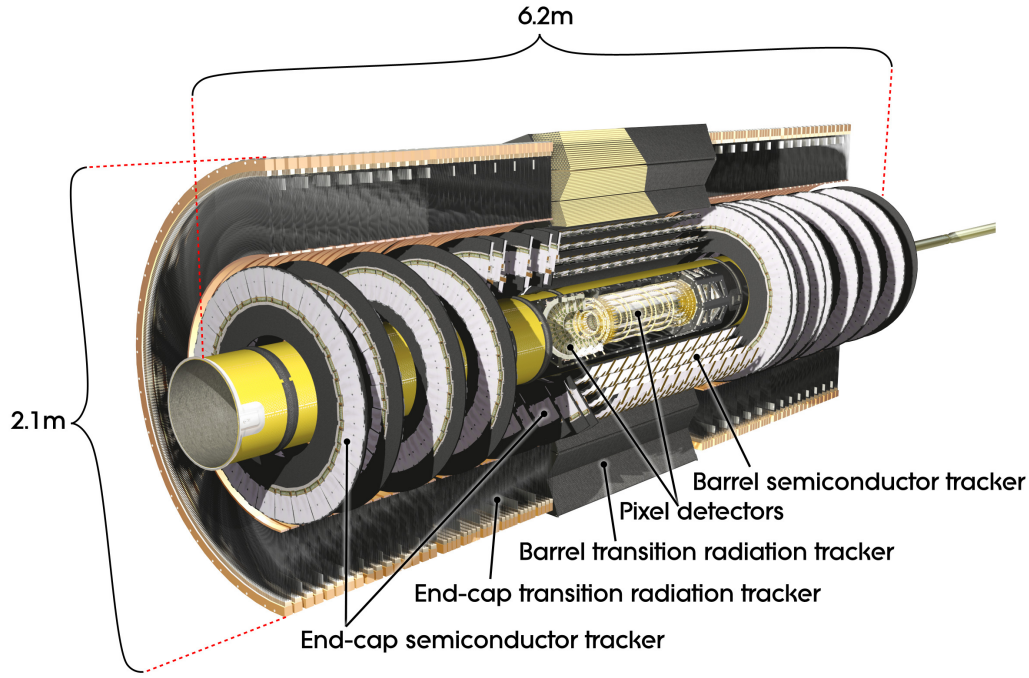


Figure 3.4: The ATLAS Inner Detector, from [4].

The track reconstruction algorithms, which determine the trajectories and momenta of charged particles by combining the pattern of energy deposits ("hits") in the Inner Detector, are discussed in Section 4.2.1.

3.2.1.1 Pixel Detector

The Pixel Detector provides precision measurements of particle paths close to the beam pipe; its precision 50×400 μm silicon pixels are designed for the identification of secondary vertices. Its central barrel region has a length of 801 mm, with three layers at a radius of 50.5, 88.5, and 122.5 mm from the center of ATLAS, so that all tracks in the region $|\eta| < 1.9$ pass through all three layers. On each side, three disks at a distance of 495 mm, 580 mm, and 650 mm are positioned in such a way that the region of three-layer coverage is extended to $|\eta| < 2.5$, as seen in Figure 3.5.

The Pixel Detector is the closest part of ATLAS to the interaction point, and will be exposed to a very high flux of particles over the lifetime of the experiment. Therefore several design features are geared toward survival in the intense radiation

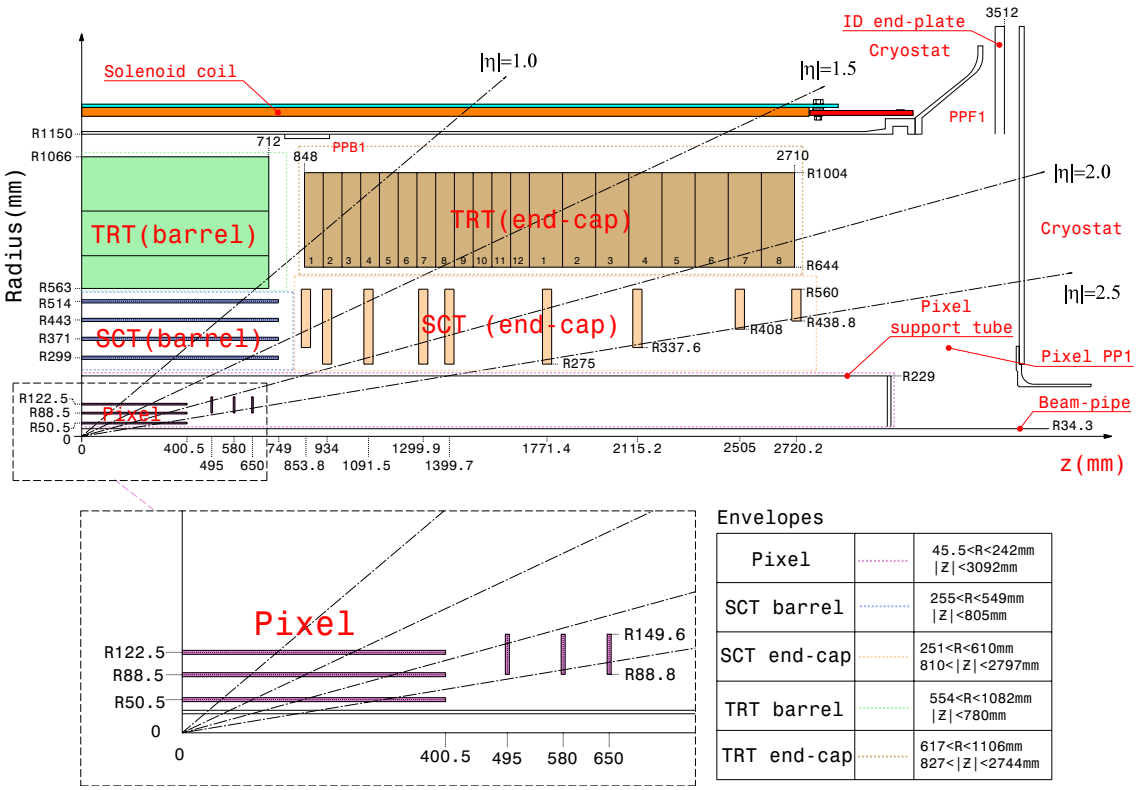


Figure 3.5: Details of the layout of the Inner Detector layers and subdetectors, from [4].

environment. The on-detector circuitry and layout are specialized to minimize radiation damage. The bulk silicon has oxygen impurities introduced to enhance resistance to damage; it is initially n-type, but becomes p-type as radiation damage changes the effective doping concentration. The sensor structure is designed to ensure functionality through this type inversion. The pixel sensors are n^+-doped, and there are p-type implants on the opposite side of the bulk; thus during initial operation, the depletion region grows from the back side but extends to the pixels. After type inversion, the depletion region grows from the pixel side so that readout remains possible. The leakage current increases with radiation damage, but can be reduced by operating the detector at roughly $-7\,^{\circ}\mathrm{C}$; further damage from annealing is also reduced at this temperature. Despite these measures, the innermost layer of the pixel detector (the *B-layer*) will have to be replaced after a few years of high-luminosity running.

3.2.1.2 Semiconductor Tracker

The Semiconductor Tracker (SCT) is designed to provide precision silicon tracking similar to that of the Pixel Detector, but because it is at a greater distance from the Interaction Point, the requirements placed on its measurement precision and radiation response are less stringent. It contains silicon strips of width 80 μm in the precision ϕ direction, but 12 cm in extent in the perpendicular direction (Z in the barrel, r in the endcaps). It nevertheless provides precision measurements in the perpendicular direction, because each of its layers have a double layer of strips. The strips on one side of the layer are offset by 40 μrad with respect to each other, so that when a charged particle passes through, information on which strips on each side have charge deposited can be combined. The SCT consists of four double barrel layers at radii from 29.9 cm to 56.3 cm, and each of its endcaps have a total of nine layers with Z distance from 85.4 to 272.0 cm; the geometry may be seen in more detail in Figure 3.5.

3.2.1.3 Transition Radiation Tracker

The Transition Radiation Tracker covers $|\eta| < 2.0$ with polyimide drift tubes 4 mm in diameter, such that a typical track passes through at least 36 straws (except in the transition region between barrel and endcap). In the barrel, tubes are 144 cm long and run along the z direction; in the endcaps, they are 37 cm in length and run radially. The central anode wire, 31 μm in diameter, is made of tungsten plated with gold. The cathode, on the inside of the tube itself, is made of aluminum protected by a layer of graphite-polyimide. The anode is held at ground voltage, while the cathode is operated at -1530V. When the standard gas mixture is ionized by a charged particle passing through the tube, this gives a gain of 2.5×10^4 and an electron collection time of 48 ns. The standard gas mixture is 70% Xe, 27% CO_2, and 3% O_2.

The TRT straws are interleaved with polypropeline fibers or foils that cause electrons with momentum above 2 GeV to produce significant numbers of transition radiation photons. The Xe absorbs these low-energy photons, producing a significantly amplified signal; the front-end electronics have a high-threshold discriminator to detect these signals, allowing electrons to be identified as tracks with a significant number of high-threshold hits.

In the context of the present work, the TRT is used primarily to provide additional high-radius measurements on tracks already found by the silicon detectors, thereby refining measurements of the track parameters. In particular, the curvature measurement for very straight tracks is enhanced, thereby improving momentum resolution for high-momentum tracks and reducing mismeasurements (cf. Section 5.2).

3.2.2 Trigger System

The ATLAS trigger system has three levels of processing, each progressively taking more time and making more refined decisions of which events should be recorded for further analysis. The Level-1 trigger system is built on specialized on-detector hardware, which makes a decision on each event within 2.5 μs in order for all subdetectors to read out the data from that event; its maximum rate is up to 75 kHz, about three orders of magnitude below the LHC bunch-crossing rate at full capacity. The Level-2 trigger system is based on off-detector processing using a special configuration of the ATLAS reconstruction software that focuses on Regions of Interest defined by the location of the L1 Trigger; it selects events at a rate of roughly 3.5 kHz. The Event Filter is based on full reconstruction of events, and selects roughly 200 events per second for permanent storage.

As the present work makes an inclusive measurement using early, low-rate ATLAS data, the full power of this trigger system is not used. Events are selected based on the Level 1 signal from the Minimum Bias Trigger Scintillators (MBTS) [75], which are designed to collect as inclusive a signature is possible. Studies of this trigger selection use events selected randomly from all events with LHC beam crossings, as measured using the Beam Pickup Timing devices (BPTX).

3.2.2.1 Minimum Bias Trigger Scintillators

The ATLAS MBTS system is composed of segmented polystyrene scintillator counters mounted on the liquid argon calorimeter endcap cryostat on each side of the detector, 3.65 m from the interaction point. They cover the pseudrapidity region $2.09 < |\eta| < 3.84$. They were installed in order to identify bunch crossings with inelastic proton-proton collisions with minimal further requirements during low-luminosity running, and trigger on almost any significant particle activity. On each side, there are two segments in the radial direction and 8 in ϕ, for a total of 32 channels. The MBTS system is not designed for substantial radiation exposure, and is expected to lose efficiency as the LHC luminosity increases during 2011.

The MBTS trigger is the primary trigger used for this analysis. Events with even the low-momentum jets studied have many more particles than the average pp collision, so the trigger is extremely efficient, as discussed in Section 4.3.1.

3.2.2.2 Beam Pickup Timing Devices

The BPTX stations are part of the instrumentation of the LHC itself, situated along the beam pipe 175 m from the ATLAS collision point in each direction. On each side, they consist of four electrostatic button pick-up detectors, which sum the signal

induced by the passage of each charged LHC bunch. They are operated by ATLAS, and used for both the overal timing and for "zero bias" triggering.

3.2.3 Calorimetry

The calorimeter system is divided into a number of subdetectors, and also into two conceptual regions: an inner electromagnetic calorimeter, and an outer hadronic calorimeter. The former is finely segmented and designed to differentiate between photons, electrons, and strongly interacting particles through the shape of the particle cascades they produce when they interact with the material. The electromagnetic calorimeter measures energy deposits, but does not completely contain the typical high-energy hadronic jet. This is the task of the hadronic calorimeter, which is less finely segmented and precise but has sufficient depth to completely absorb showers from even the highest-energy jets.

A schematic of the calorimeter system is shown in Figure 3.6. The inner parts of the system are contained within one central and two endcap cryostats, allowing them to use liquid argon as their active readout material. The central cryostat contains the central electromagnetic calorimeter, which covers $|\eta| < 1.475$ and uses lead as its absorber. The endcap cryostats contain the similarly-designed electromagnetic endcap calorimeter (EMEC), covering $1.375 < |\eta| < 3.2$, beyond which is the hadronic endcap (HEC), which covers $1.5 < |\eta| < 3.2$ and uses copper as its absorbing material. Inside the EMEC and HEC, at radii occupied by the Inner Detector at smaller $|z|$, is the forward calorimeter (FCal), which as one electromagnetic module with copper absorber followed by two hadronic modules using tungsten absorber. The Tile Calorimeter system surrounds the cryostats and provides additional depth at central rapidity, $|\eta| < 1.7$, with central and extended barrel sections. It uses lead as its absorber and doped polystyrene scintillators to measure the deposited energy.

Although the calorimeter system is the primary ATLAS System for measuring jets, it is not used in the present analysis (except for a limited set of cross-checks described in Section 4.5). This is because it is optimised for good resolution in measuring high-momentum jets, and not for measurement of individual low-momentum particles or low-momentum jets. Calorimeter-based measurements of jet shapes and jet constituents have been made [11], but the present analysis is designed to be complementary and to have independent systematic uncertainties. In principle, an algorithm that combined tracking information for charged particles with calorimeter-based measurements for individual neutral particles (primarily pions) could be considered as an extension to this analysis. Such an algorithm has been implemented by the CMS experiment, where it enhances jet energy resolution at all energies; the gains for ATLAS would be considerably less.

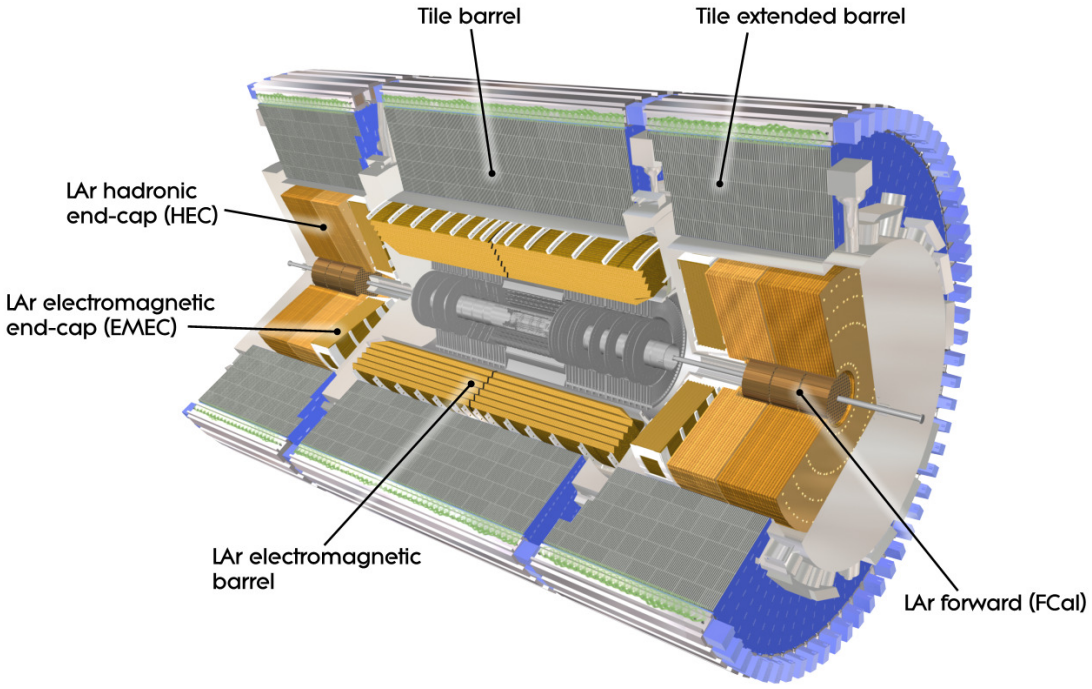


Figure 3.6: The ATLAS calorimeter systems, from [4].

3.2.4 Muon System

The ATLAS muon system is the outermost part of the ATLAS detector, which identifies muons by relying on the fact that high-momentum muons pass through the calorimeter system. A measurement of their momentum is also made, which is very important for extremely high-momentum muons whose tracks in the Inner Detector appear essentially straight; the muon system resolution is approximately 10% for 1 TeV muons, and acceptable momentum resolution and good charge identification are available even at 3 TeV. This momentum measurement is made possible by the large air core toroid magnet systems: eight barrel toroid coils and eight endcap coils produce a magnetic field that provides significant bending power $\int Bdl$ over the region $|\eta| < 2.7$: up to 7.5 Tm and over 1 Tm everywhere except in the transition region between barrel and encap magnets ($1.4 < |\eta| < 1.6$).

Four technologies are used for muon measurements: the Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) for precision tracking, and the Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) for triggering and transverse coordinate measurement. There are several layers of MDT chambers; each

chamber consists of three to eight layers of drift tubes conceptually similar to those found in the TRT, although larger. The MDTs have a resolution of about 80 μm per tube, or 35 μm per chamber, in the precision direction. (Because the toroidal magnetic field curves tracks forward or backward, the precision direction is z in the barrel and R in the endcaps.) The MDTs cover all of $|\eta| < 2.7$, except for the innermost layer at $|\eta| > 2.0$, where they are replaced by the CSCs. The CSCs are designed to better cope with the higher muon flux closer to the IP and LHC beam; these are multiwire proportional chambers with strips designed for measurements in both R and ϕ, which have a faster response time and better radiation hardness than the MDTs. RPCs are responsible for triggering and transverse measurements in the barrel region ($|\eta| < 1.05$); these chambers contain parallel plates without wires, spaced by about 2 mm and with a voltage difference of 10 kV; the plates are divided into strips measuring z and ϕ with a width of 23-35 mm. At higher η, up to 2.4, these functions are performed by the multiwire porportional chamber TGCs, whose wires are separated by 1.8 mm and operate at 2900 V. Both RPCs and TGCs provide fast measurements that can be used for Level 1 Triggering based on a rough estimate of muon momentum.

Chapter 4

Data Analysis

In this work, distributions of observables for charged particle jets are measured by reconstructing the paths of charged particles (*tracks*) in the ATLAS detector, applying a jet-finding algorithm to group these particles together, and then correcting these distributions for the effects of track- and event-finding efficiencies and measurement resolutions. The distributions to be measured are defined in Section 4.1. The procedure for reconstructing tracks and identifying their points of origin is described in Section 4.2. The selection of collision events to be used is described in Section 4.3, and the selection of tracks in Section 4.4. The properties of selected tracks and validation of tracking simulation appear in Section 4.5. The algorithm used to produce jets from tracks is discussed in Section 4.6. Finally, the procedure for correcting distributions to the level of true charged particle jets, based on simulated events, is described in detail in Section 4.7.

4.1 Measured Variables

Five quantities are measured for charged particle jets, which together encapsulate global properties such as the rate at which they occur and their total momentum, as well as the number, momentum, and spatial distribution of their constituent charged particles. The first is the inclusive jet cross section, given differentially as a function of jet transverse momentum (p_{T}) and rapidity (y):

$$\frac{d^2\sigma_{jet}}{dp_{\mathrm{T}jet}dy_{jet}}. \tag{4.1}$$

(Note that, throughout this work, the subscript *jet* will be used to clarify that a variable is a property of a jet. Likewise, *ch* will be used for a charged particle, and *trk* for a measured track. These subscripts may be omitted when the object being

discussed is clear from context.)

The second quantity measured is the probability for a jet to have a given number of tracks N_{jet}^{ch}:

$$\frac{1}{N_{jet}}\frac{dN_{jet}}{dN_{jet}^{ch}}. \tag{4.2}$$

The third is the distribution of charged particle z values over jets in a given p_{T} range:

$$\frac{1}{N_{jet}}\frac{dN_{ch}}{dz}, \tag{4.3}$$

where the variable z is the momentum fraction of the particle along the jet axis, given by.

$$z = \frac{\vec{p}_{ch}\cdot\vec{p}_{jet}}{|\vec{p}_{jet}|^2}, \tag{4.4}$$

where $\vec{p}_{ch}$ is the momentum of the charged particle and $\vec{p}_{jet}$ is the momentum of the jet that contains it. The z distribution presented here differs from the usual definition (in [18], for example), which would include neutral particles and low-momentum charged particles in the total jet momentum.

Fourth, the charged particle momentum distribution transverse to the jet axis is measured:

$$\frac{1}{N_{jet}}\frac{dN_{ch}}{dp_{\mathrm{T}}^{rel}}. \tag{4.5}$$

Because QCD predicts that the scale of jet transverse momenta is Λ_{QCD}, this distribution is measured in GeV rather than being a fraction with respect to the jet momentum. The variable p_{T}^{rel} is the momentum of charged particles in a jet transverse to that jet's axis:

$$p_{\mathrm{T}}^{rel} = \frac{|\vec{p}_{ch}\times\vec{p}_{jet}|}{|\vec{p}_{jet}|}. \tag{4.6}$$

Because QCD predicts that the scale of jet transverse momenta is Λ_{QCD}, this distribution is measured in absolute units rather than being a fraction with respect to the jet momentum.

Finally, the density of charged particles in ϕ-y space, $\rho_{ch}(r)$, is measured as a function of the radial distance r of charged particles from the axis of the jet that

contains them, where r is given by

$$r = \Delta R(ch, jet) = \sqrt{(\phi_{ch} - \phi_{jet})^2 + (y_{ch} - y_{jet})^2}. \tag{4.7}$$

Thus for an annulus of inner radius r_1 and outer radius r_2, containing a total of $N_{ch}^{r_1<r<r_2}$ charged particle entries, $\rho_{ch}(r)$ is given by

$$\rho_{ch}(r) = \frac{1}{N_{jet}} \frac{N_{ch}^{r_1<r<r_2}}{\pi(r_2^2 - r_1^2)}. \tag{4.8}$$

Note that this is a particle number density, rather than the related energy density variable used for calorimeter-based jet shape measurements.

The multiplicity and per-particle distributions are separately corrected and produced for five jet momentum ranges: $4-6\,\text{GeV}$, $6-10\,\text{GeV}$, $10-15\,\text{GeV}$, $15-24\,\text{GeV}$, and $24-40\,\text{GeV}$. All distributions are corrected as a function of rapidity in four ranges: $0-0.5$, $0.5, 1.0$, $1.0-1.5$, and $1.5-1.9$.

4.2 Track Reconstruction

Charged particle tracks in the ATLAS detector are identified by reconstructing the curved paths they produce as they pass through the Inner Detector. In the approximation of a uniform magnetic field and no energy loss in the Inner Detector, a charged particle follows a helix that may be parameterized with the following variables.

- q/p_{T}, the charge divided by the component of the particle momentum transverse to the beam, which can be derived from the track curvature ρ and magnetic field strength B via $\rho = Bq/p_{\text{T}}$.
- d_0, the *transverse impact parameter*, defined as the distance of closest approach of the particle, in the transverse plane, to some fixed origin point.
- Z_0, the *longitudinal impact parameter*, defined as the Z position of the point of closest approach used in the d_0 definition.
- ϕ_0, the azimuthal angle of the particle trajectory at the distance of closest approach.
- η, the pseudorapidity of the particle trajectory.

The "fixed origin point" is ultimately the *primary vertex* (PV), the point in space where the two protons originally collided. An event may have multiple primary vertices if several proton pairs from the same bunches collide. In order to identify the primary vertices in an event, tracks must be reconstructed and selected, and in this case the *beam spot* (BS) is used as the origin point; this is the average location of proton-proton collisions. The beam spot itself must also be determined by observ-

ing the distribution of primary vertices over many events. Thus track-, PV-, and BS-finding are an iterative process; track-finding algoirthms are described in Section 4.2.1, and PV and BS reconstruction are discussed in discussed in Section 4.2.2.

4.2.1 Track Reconstruction Algorithms

The *track reconstruction* algorithms [76] are modularized and highly customizable. The present analysis uses default settings and cuts that match initial charged particle analyses [7].

Track reconstruction begins with the raw signals from the Inner Detector, the information on deposits of charge in the Pixel Detector and SCT; raw data from the TRT is incorporated at a later step. First, within each pixel or SCT layer, deposits in adjacent pixels or SCT strips are combined into *clusters*. The raw silicon cluster information is then processed to produce *space points*, abstract, detector-independent, three-dimensional objects to which the track algorithms can be applied. For the Pixel Detector, the space point location is taken from the charge-weighted location of the constituent hits. In the SCT detector, after adjacent strips are combined into clusters, the position in the long strip dimension is determined by combining adjacent clusters on double layers; the 40 μrad angle between the two strips gives a well-defined position.

Space points are then combined into *track seeds*, which are constructed from three space points on separate layers assuming a uniform magnetic field and no particle scattering. These seeds are then used to define *roads*, to which pixel and SCT clusters (not space points) are attached. Hits are attached sequentially, with track fitting updated after each added hit using a Kalman filter [77]; during this step, a model of the ATLAS magnetic field and detector geometry is used to incorporate scattering into the trajectory calculations.

Once the process of adding hits to seeds is complete, an additional step called ambiguity solving [78] is applied. This step removes *fake* tracks assembled out of chance combinations of hits, as well as reducing the number of hits that are shared between multiple tracks. Tracks are given scores based on the quality of the fit, and the number of hits, and the number of *holes* (missing hits in places where one would be expected along the track). Hits assigned to multiple tracks are iteratively assigned to the highest-scoring track, then tracks are refit and rescored as necessary before the next hit is processed. Remaining tracks that pass basic quality cuts have roads extended into the TRT, TRT hits added, and are refit for a final time.

4.2.2 Primary Vertex and Beam Spot Reconstruction

After track-finding is run, tracks meeting the following criteria are used in primary vertex reconstruction [79]:

- transverse momentum $p_{\mathrm{T}} > 150\,\mathrm{MeV}$;
- transverse impact parameter $|d_0| < 4$ mm;
- estimated uncertainty on the transversee impact parameter $\sigma(d_0) < 5$ mm;
- estimated uncertainty on the longitudinal impact parameter $\sigma(Z_0) < 10$ mm;
- at least 4 hits in the SCT detector; and
- at least 6 total hits in the pixel and SCT detectors.

The impact parameters are calculated with respect to the beam spot if it has already been reconstructed; otherwise they are calculated with respect to the center of the detector.

The seed for the initial primary vertex is identified based on the maximum in the distribution of the tracks' Z_0 distributions. The vertex position is then determined using *adaptive vertex fitting* [80], a χ^2-based algorithm which iteratively reduces the weighting for outlying tracks. If the beam spot has been found, it is also used as a constraint for the fit. Tracks whose χ^2 value for consistency with the calculated vertex is greater than 49 are excluded from the vertex and instead used to seed a new vertex. The procedure is repeated until all tracks are used or no additional vertex fits succeed.

Beam spot finding [81] is done by means of an unbinned maximum likelihood fit applied to all primary vertices in a subset of the data. It is computed separately for each period of roughly 10 minutes.

4.3 Event Selection

This measurement uses a sample of early ATLAS data, with total integrated luminosity of $799\,\mu\mathrm{b}^{-1}$ and a peak luminosity of $6.6 \times 10^{28}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ ($0.066\,\mu\mathrm{b}^{-1}\mathrm{s}^{-1}$). Luminosities are calibrated using the Van-der-Meer method of scanning the beams against each other in the horizontal and vertical planes [75]. The systematic uncertainty is 3.4% which is dominated by the uncertainty on the beam currents.

Events from colliding proton bunches are selected if the MBTS trigger recorded one or more counters above threshold on either side. They are further required to have a primary vertex [82] reconstructed using beam-spot information [83]. Events with additional reconstructed primary vertices are rejected. The number of average collisions per bunch crossing μ depends on luminosity; most of the data are collected with $\mu \lesssim 0.01$, with a maximum value of 0.14. A total of 42,719,645 selected events

satisfy these criteria.

To determine the differential cross section from the number of corrected jets per p_{T} and y bin, the following formula is applicable:

$$\frac{d^2\sigma_{jet}}{dp_{\mathrm{T}jet}dy_{jet}} = \frac{N_{corr}(p_{\mathrm{T}}, y)}{\epsilon_{evt}\mathcal{L}\Delta p_{\mathrm{T}}\Delta y}, \tag{4.9}$$

where $N_{corr}(p_{\mathrm{T}}, y)$ is the number of corrected jets per p_{T} and y bin (as computed via the procedure described in Section 4.7), ϵ_{evt} is the event selection efficiency, $\mathcal{L}$ is the total luminosity, and Δp_{T} and Δy are the bin widths. The event selection efficiency may be further factorized into

$$\epsilon_{evt} = \epsilon_{trig}\epsilon_{vtx}\epsilon_{vtx<2}, \tag{4.10}$$

where ϵ_{trig} is the trigger efficiency, ϵ_{vtx} the primary vertex reconstruction efficiency, and $\epsilon_{vtx<2}$ the efficiency of the additional primary vertex rejection cut, all given the presence of a track jet used in this analysis.

4.3.1 Trigger Efficiency

The efficiency of the trigger given the presence of one or more track jets was determined using a random subset of the data triggered using only the BPTX. Events were selected that passed the track trigger and and contained a track jet reconstructed offline passing the cuts in Section 4.4. Of 150,981 (234,856) such events with $R = 0.4$ (0.6), all passed the MBTS trigger as well, demonstrating that the MBTS trigger efficiency has a negligible difference from 100% for events containing jets of interest to this analysis.

4.3.2 Vertexing Efficiency

The efficiency for vertex finding given the presence of one or more jets is determined from data by removing the track selection cuts that use impact parameter with respect to the primary vertex described in Section 4.4. Other cuts are kept as described, and a requirement is added that the transverse impact parameter with respect to the beam spot satisfy $|d_0| < 4.0$ mm. Jets are then reconstructed from tracks satisfying this new selection, and the fraction of events containing jets that also have a primary vertex reconstructed is determined. In the full dataset, out of 4,685,266 (7,301,911) events containing jets with $R = 0.4$ (0.6), a primary vertex is found in 4,685,167 (7,301,809) events, demonstrating that the vertex-finding efficiency is essentially 100% for events containing jets used in this analysis.

4.3.3 Rejection of Events with Additional Vertices

The efficiency of having exactly one vertex may be further factorized into

$$\epsilon_{vtx<2} = \epsilon_{pileupfree}\epsilon_{nosplit}, \tag{4.11}$$

where $\epsilon_{pileupfree}$ is the fraction of collisions retained after rejection of real additional vertices, and $\epsilon_{nosplit}$ is the efficiency of reconstructing only a single vertex in events with only one collision (i.e., avoiding fake vertices). The former is a function of the average number of collisions per crossing μ, while the latter is a function of the activity in the primary collision, parameterized by the number of charged particles in the event.

The fraction of collisions that occur without additional primary vertices, $\epsilon_{pileupfree}$, is computed from data for each luminosity block* based the value of μ in that block, and accounts for the fact that multiple events that would otherwise have passed the event selection are lost. It amounts to roughly a 3% effect, averaged over the entire dataset, which can be considered a correction to the luminosity. For each luminosity block, it is given by the ratio between the number of collisions in events with only one reconstructed vertex and the total number of collisions:

$$\epsilon_{pileupfree} = \frac{N_{coll,vtx<2}}{N_{coll}} = \frac{1}{N_{coll}} \sum_{N_{vtx}=1}^{\infty} N_{vtx} P(N_{vtx}, \mu_{LB})(1 - \epsilon_{additional})^{N_{vtx}-1}, \tag{4.12}$$

where

$$N_{coll} = \sum_{N_{vtx}=1}^{\infty} N_{vtx} P(N_{vtx}, \mu_{LB}). \tag{4.13}$$

Here μ_{LB} is the average number of collisions in that luminosity block, $P(N_{vtx}, \mu_{LB})$ is the probability of N_{vtx} vertices given by the Poisson distribution with mean μ_{LB}, and $\epsilon_{additional}$ is the efficiency of reconstructing additional vertices beyond the first. (The efficiency of reconstructing the first vertex, in events of interest, is essentially 100%, as described in Section 4.3.2.) Because the probability of each collision is independent, one can take $\epsilon_{additional}$ to be the ratio of σ_{vis} to σ_{tot}. These are the visible and total inelastic cross sections respectively, as defined in [75]; σ_{vis} has been corrected due to improved luminosity calculations, from 58.7 mb to 56.6 mb. The ratio $\sigma_{vis}/\sigma_{tot}$ can be varied by a substantial amount without changing this correction significantly; hence there is no systematic uncertainty assigned based on the unknown total cross section.

*A luminosity block is a period of 1-2 minutes during a run in which detector conditions remain roughly constant.

There is a further 5.7% correction due to the probability of additional collisions occurring at nearly the same Z is the primary collision, thereby rendering the vertex unresolvable, as illustrated in Figure 4.1. (Of course, such events are then included in the sample, but due to the small average number of collisions per event and the low activity of the average visible collision, this contribution is negligible.)

The efficiency $\epsilon_{pileupfree}$ varies from roughly 88% in luminosity blocks in which $\mu \sim 0.15$ up to nearly 100% in luminosity blocks with $\mu \lesssim 0.01$. The number of events per unit luminosity declines with μ in a manner consistent with this calculation; $N_{evt}/\epsilon_{pileupfree}\mathcal{L}$ is independent of μ, except for a 1% disagreement between runs with $\mu \lesssim 0.04$ and the last few runs with $\mu \gtrsim 0.1$, as is shown in Figure 4.2. This disagreement is consistent with differences between luminosity computation methods in different runs, and is well within the overall luminosity uncertainty. Over all luminosity blocks used, the adjustment to the total luminosity associated with this efficiency is

$$\mathcal{L}\epsilon_{pileupfree} = \sum_{LB} \mathcal{L}(LB)\epsilon_{pileupfree}(LB) = 773\ \mu\mathrm{b}^{-1}, \tag{4.14}$$

where the sum runs over all luminosity blocks the analysis.

The fraction of collisions with no pileup, but which are removed due to the erroneous reconstruction of multiple vertices, is determined from simulation as a function of selected track multiplicity in the event. The value of $\epsilon_{nosplit}$ is 0.25% for 10 or fewer selected tracks, rising to 1% around 30 tracks and 4% at about 100 tracks, in events with one truth jet. This is corrected for by weighting the data events according to the number of selected tracks in the event. Because the determination is made entirely from simulation, the full value of the correction ($1.8\% - 2.2\%$) is taken as a systematic uncertainty, as described in Section 5.3.

4.4 Track Selection

For this analysis, each reconstructed track is required to have

- transverse momentum $p_{\mathrm{T}} > 300$ MeV;
- pseudorapidity $|\eta| < 2.5$;
- transverse impact parameter with respect to the primary vertex $|d_0| < 1.5$ mm (0.2 mm) for tracks with p_{T} less than (greater than) 10 GeV;
- longitudinal impact parameter with respect to the primary vertex Z_0 satisfying $|Z_0 \sin\theta| < 1.5$ mm;
- if a signal (or *hit*) is expected in the Pixel Detector B-layer (i.e. if the extrapolated track passes through a section of that layer with functioning instrumen-

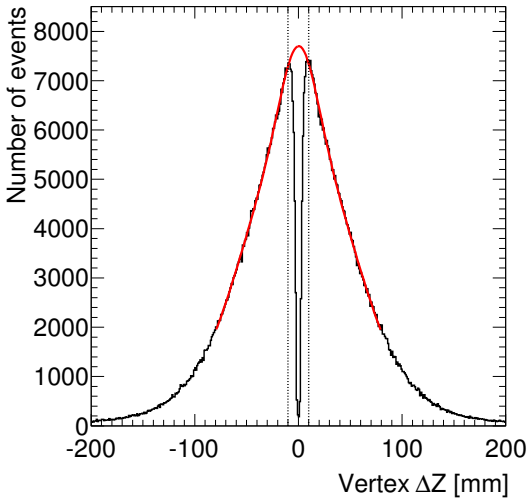


Figure 4.1: The difference in Z coordinate of the two vertices in events where exactly two primary vertices are identified. The distribution is fit to a double Gaussian in the range $|\Delta Z| < 80$ mm, excluding the "dip" region with $|\Delta Z| < 10$mm which arises because of the resolution of the primary vertex reconstruction algorithm. Vertical dotted lines illustrated the excluded region; the solid red line illustrates the fit. The contribution of Z-overlaps to $\epsilon_{additional}$, the efficiency of reconstructing vertices beyond the first, is computed using the difference between the fit and the distribution in the excluded region: it is found to be about 5.7%.

tation), then such a hit is required, with one pixel hit in any layer required otherwise;

- at least 6 SCT hits.

The momentum requirement was selected to ensure that tracks traverse, at minimum, all layers of the pixel and SCT detectors. The pixel and SCT hit requirements are necessary to ensure sufficient information for a good track fit. The requirement of a B-layer hit, if expected, is to select primary tracks rather than those resulting from long-lived particle decays and interaction with the detector material. The impact parameter requirements are designed to reject tracks from secondary vertices. At $p_{\mathrm{T}} > 10$ GeV, the $|d_0|$ resolution is improved to the point that the tightened cut is highly efficient for tracks from the primary vertex in simulated events; the tightened cut is designed to remove misreconstructed tracks, particularly those with artificially high momentum, as discussed in Section 5.2.

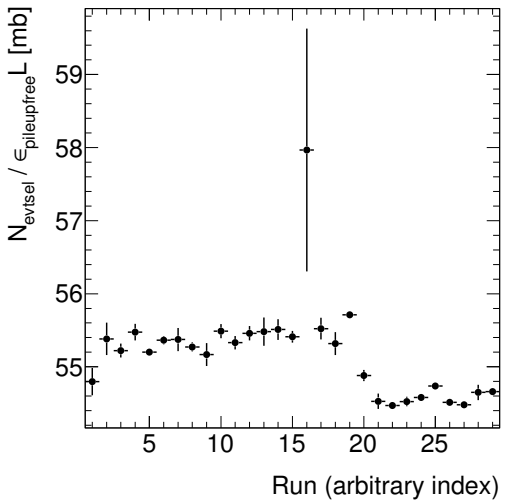
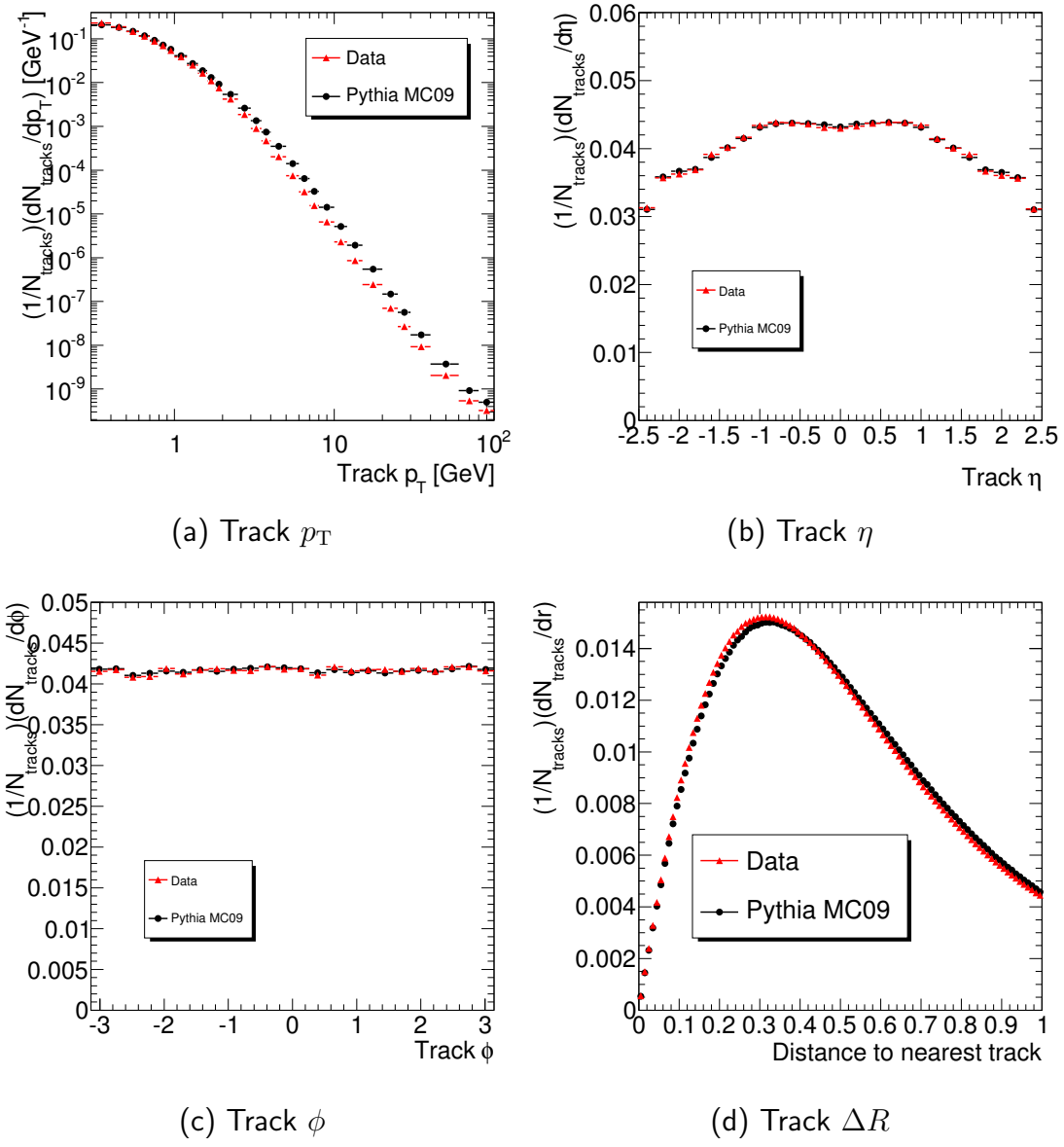


Figure 4.2: The number of events per unit luminosity, after correcting for the efficiency, is flat over runs, except for a 1% discontinuity as discussed in the text.

4.5 Tracking Efficiencies and Distributions

In order to have confidence in the correction procedure, which is based on simulated data, it is important to verify that the properties of tracks and track selection agree between the data and simulation. The p_{T}, η, and ϕ distributions of all reconstructed tracks, compared with the ATLAS MC09 tune of Pythia, are shown in Figure 4.3, along with the ΔR to the nearest track. All track selection cuts are applied except the tighter of the d_0 cut at high momentum. The p_{T} spectra do not agree, reflecting physics differences between data and MC; this is accounted for in the systematic uncertainties based on range of tunes (Section 5.1.3).

The efficiency of the high-p_{T} d_0 cut is shown in Figure 4.4. The efficiency as a function of p_{T} is extended down to 5 GeV (where the cut is not otherwise applied for the analysis), to illustrate that the cut efficiency has plateaued by 10 GeV. The efficiency then decreases at higher momenta. This is most likely caused by mismeasured high-momentum tracks, which result from lower-momentum particles that scatter in such a way that they are reconstructed at high momentum; this conclusion is supported by the decrease because as momentum increases, fewer and fewer of the reconstructed tracks are really high-p_{T} particles. The systematic uncertainty for this effect, discussed in Section 5.2, also allows for the possibility that the d_0 resolution at high-p_{T} is mismodelled. The efficiency as a function of η exhibits the greatest deviations in the most forward parts of the Inner Detector, especially for $|\eta| > 2.3$.



(a) Track p_{T}

(b) Track η

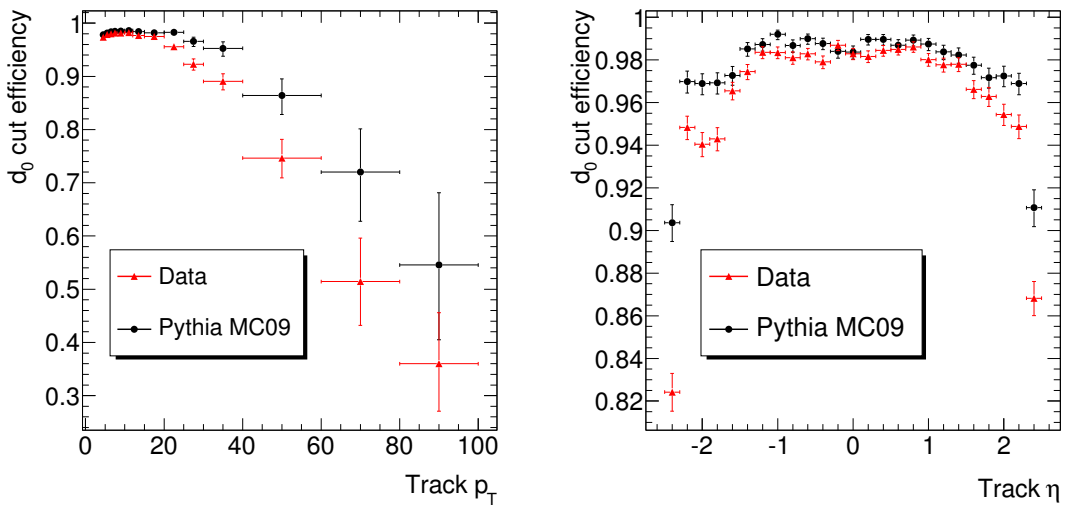
(c) Track ϕ

(d) Track ΔR

Figure 4.3: Raw distributions of reconstructed tracks selected in this analysis, compared with simulation, excluding the high-momentum d_0 cut. ΔR is the distance in $\eta - \phi$ space to the nearest selected track.

However, after the requirement that the tracks have $p_{\mathrm{T}} > 10$ GeV, all tracks appearing in the analysis have $|\eta| < 2.1$; this is because tracks at high η will "pull" the jet

that contains them out of the rapidity range used ($|y_{jet}| < 1.9$).

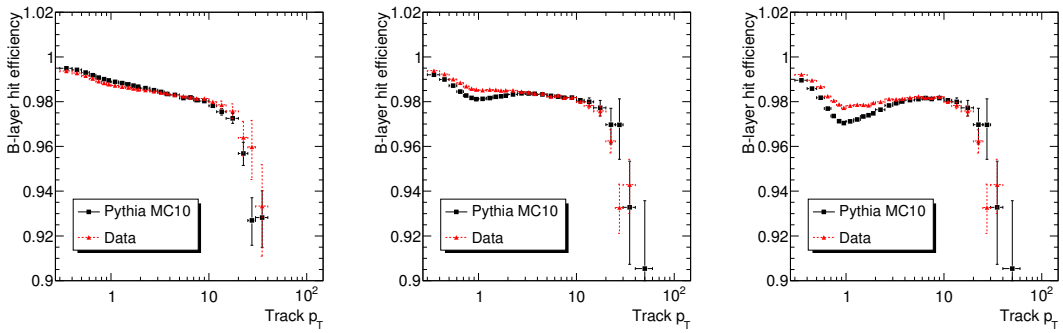


(a) p_{T} (b) η ($p_{\mathrm{T}} > 10$ GeV)

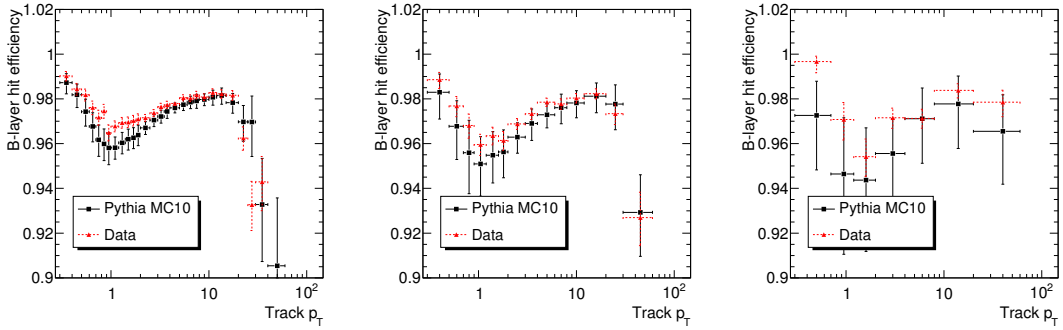
Figure 4.4: Efficiency of the high momentum d_0 cut, compared with simulation, as a function of p_{T} and η.

In order to investigate the tracking performance in the high-density environment of jets, the B-layer hit efficiency is studied. This is the layer with the highest density of track hits, and also plays a special role in the track selection. The denominator of the efficiency calculation uses tracks which pass all selection requirements except the pixel hit requirement, and which have a B-layer hit expected. The efficiency for identifying a B-layer hit for all such tracks is shown as a function of track p_{T} in Figure 4.5(a). In Figures 4.5(b)-4.5(f), the efficiency is shown for the subset of these tracks which are within $\Delta R < 0.2$ of the nearest track jet; these plots are separated based on the jet momentum. As the nearest jet momentum increases, the B-layer hit efficiency decreases markedly for tracks of p_{T} less than a few GeV. The B-layer hit efficiency rises again at very low momentum, because tracks are less likely to be found in the first place without a B-layer hit.

Although the tracking is clearly impacted inside higher-momentum jets, the impact is accounted for fairly well by the detector simulation and will thus be properly accounted for by the corrections described in Section 4. The deviations between data and simulation are significantly smaller than the tracking efficiency uncertainties described in Section 5.1.1. Therefore no additional systematic uncertainty is applied. (N.B. The local minimum in plots (b)-(f) around a track p_{T} in fact reflects two



(a) All tracks (b) Jet $p_{\mathrm{T}} > 4$ GeV (c) Jet $p_{\mathrm{T}} > 10$ GeV

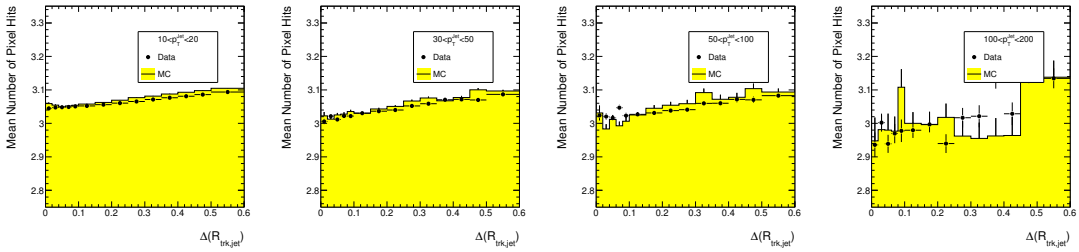


(d) Jet $p_{\mathrm{T}} > 20$ GeV (e) Jet $p_{\mathrm{T}} > 30$ GeV (f) Jet $p_{\mathrm{T}} > 50$ GeV

Figure 4.5: B-layer hit efficiency, as a function of track p_{T}, for (a) all tracks, and (b)-(f) tracks within $\Delta R < 0.2$ of the jet axis of track jets with p_{T} above the momentum indicated, for data and Monte Carlo.

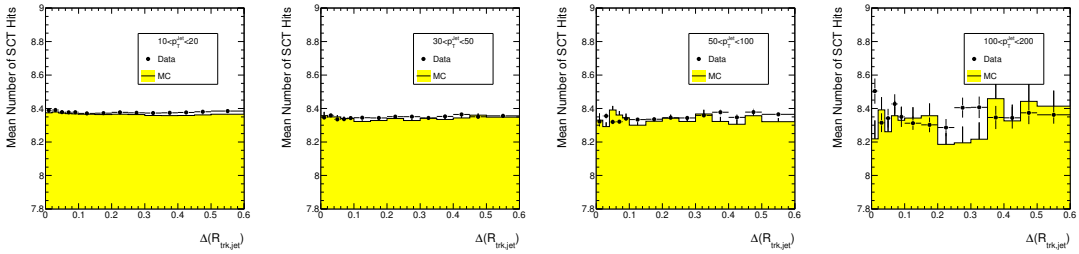
competing effects: B-layer hit efficiency is rising with track momentum, but at very low momenta tracks are not accurately extrapolated to the primary vertex without a B-layer hit. Thus in the latter case, tracks fail the d_0 cut and enter neither the numerator nor the denominator.)

Additional plots with hit properties of tracks, as a function of their distance to the nearest calorimeter jet, appear in Figure 4.6 to Figure 4.8. Track cuts are the same as in the rest of the analysis, except that no B-layer hit is required, and no check is made whether one is expected; instead, the requirement is that the track have a pixel hit in any layer. Agreement is generally good between data and the MC09 simulated events.



(a) 10 - 20 GeV (b) 30 - 50 GeV (c) 50 - 100 GeV (d) 100 - 200 GeV

Figure 4.6: Mean number of Pixel Detector hits per track with $p_T > 300$ MeV, as a function of ΔR from the nearest calorimeter jet with $|y_{jet}| < 1.9$ and $p_{\mathrm{T},jet}$ as indicated. Track cuts are the same as in the rest of the analysis, except that no B-layer hit is required, and no check is made whether one is expected; instead, the requirement is that the track have a pixel hit in any layer.

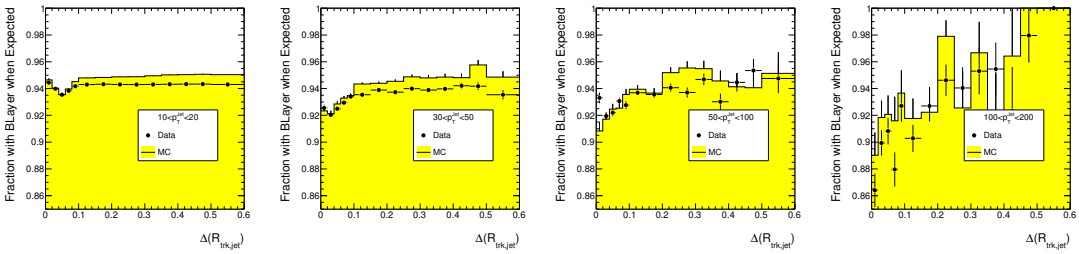


(a) 10 - 20 GeV (b) 30 - 50 GeV (c) 50 - 100 GeV (d) 100 - 200 GeV

Figure 4.7: Mean number of SCT hits per track with $p_T > 300$ MeV, as a function of ΔR from the nearest calorimeter jet with $|y_{jet}| < 1.9$ and $p_{\mathrm{T},jet}$ as indicated. Track cuts are the same as in the rest of the analysis, except that no B-layer hit is required, and no check is made whether one is expected; instead, the requirement is that the track have a pixel hit in any layer.

4.6 Jet Definition

Although the basic picture presented by the parton model (Section 2.3) identifies jets as originating from a single outgoing quark or gluon, the particles arising from the underlying event contribution (Section 2.6) and other soft physics make contributions to the topology of the event that cannot be distinguished on an event-by-event basis. Jets are therefore produced using an algorithm that combines charged particles regardless of their "source," which can be applied consistently to reconstructed tracks.



(a) 10 - 20 GeV (b) 30 - 50 GeV (c) 50 - 100 GeV (d) 100 - 200 GeV

Figure 4.8: Fraction of tracks with $p_T > 300$ MeV that have a B-layer hit, as a function of ΔR from the nearest calorimeter jet with $|y_{jet}| < 1.9$ and $p_{\mathrm{T},jet}$ as indicated. Track cuts are the same as in the rest of the analysis, except that no B-layer hit is required, and no check is made whether one is expected; instead, the requirement is that the track have a pixel hit in any layer.

The anti-k_t algorithm [84] combines objects (tracks, jets, etc.) into jets by iteratively combining pairs of objects. It does this by forming a list of all "distances" d_{ij} between each pair of objects i and j, along with the distances d_{iB} from each object to the beam, defined by

$$d_{ij} = \min(p_{\mathrm{T}i}^{-2}, p_{\mathrm{T}j}^{-2}) \frac{[\Delta R(i,j)]^2}{R^2}, \tag{4.15}$$

$$d_{iB} = p_{\mathrm{T}i}^{-2} \tag{4.16}$$

.

where $\Delta R(i,j)$ is the usual distance in ϕ-y space between objects i and j and R is the "radius parameter" for the algorithm. If the smallest value on this list of distances is a d_{ij}, then the objects i and j are combined into a single object (by adding their four-momenta) and the list is recreated. If the smallest value is a d_{iB}, then this object is designated a final jet and removed from the list. The process terminates when only final jets remain.

In the anti-k_t algorithm, objects that are most collinear and have the highest momentum are combined first. This may be contrasted with the k_t algorithm [85; 86], in which the negative exponents on the transverse momenta in Equation 4.15 are replaced by positive ones so that low-momentum collinear objects are combined first. This procedure comes closest to "reversing" the parton branching by which high-momentum jets are formed, but it produces jets of large irregular spatial extent for which experimental background corrections are more difficult. Anti-k_t produces

"cone-like" jets, in the sense that the region in which a potential constituent is added to a jet is roughly circular in ϕ-y space. Anti-k_t, like k_t, has a distinct advantage over older cone algorithms in that it is *infrared-safe*, i.e., two jets cannot be merged together by the addition of an arbitrarily soft additional particle.

Measurements in this work are made both with radius parameter $R = 0.4$ and $R = 0.6$. In order to produce 4-vectors for reconstructed tracks for the application of the algorithm, the pion mass is assumed for all tracks. After the algorithm is applied, track jets with $p_{\mathrm{T}} > 4$ GeV and $|y| < 1.9$ are accepted, including those with only one constituent track. The rapidity cut is chosen to ensure that jets are fully contained within the tracking acceptance $|\eta| < 2.5$ in the $R = 0.6$ case; for the narrower $R = 0.4$ jets, the same rapidity cut is used for consistency.

Truth-level charged particle jets, which are the objects ultimately measured after the completion of the correction procedure, are derived from MC events. Similarly to reconstructed track jets, they also have the anti-k_t algorithm applied to primary charged particles with $p_{\mathrm{T}} > 300$ MeV. The momentum cut is chosen in order to reflect the choice of cut track level, in order to minimize model-dependent corrections and extrapolation. Each particle's true mass is used in the application of the algorithm. Primary charged particles are defined as charged particles with a mean lifetime $\tau > 0.3 \times 10^{-10}$ s, including muons, which are produced in the primary collision or from subsequent decays of particles with a shorter lifetime. Thus the charged decay products of K_S particles are not included. Charged particle jets are required to meet the p_{T} and y requirements given for track jets above.

4.7 Unfolding Procedure

The corrections used in this analysis, from detector-level track jets to truth-level charged particle jets, are derived from a Monte Carlo (MC) simulated sample using the Pythia 6.421 event generator program [50] with the ATLAS AMBT1 tune [7], which is described in Section 2.7. In order to derive corrections, which account for tracking efficiency and momentum resolution, these events are then passed through the ATLAS detector simulation [87], based on GEANT 4 [88].

Reconstructed jets are binned in p_{T}, constituent multiplicity, and rapidity, i.e., a 3-dimensional distribution is produced for the purpose of applying corrections. Similarly, for the z corrections, each track in each jet is binned in z, the parent jet momentum, and parent jet rapidity; the p_{T}^{rel} and r distributions likewise are binned with parent jet momentum and rapidity as two additional variables. For the jet-level quantities as well as track-level z, p_{T}^{rel}, and r, corrections are applied simultaneously in the three binned variables using the Bayesian Iterative Unfolding [89] algorithm implemented in the RooUnfold [90] program. As this correction method is central to

the analysis, its general properties are reviewed in Section 4.7.1. Its implementation in the present analysis is discussed in Section 4.7.2, the properties of the simulated distributions used for corrections are discussed in Section 4.7.3, and the validation of the method with data-sized simulated samples is described in Section 4.7.4.

4.7.1 Bayesian Iterative Unfolding Algorithm

The Bayesian Iterative Unfolding algorithm is a method for correcting an experimentally measured distribution to the true values of that distribution that would have been measured by a perfect detector. It assumes that both distributions are binned; the binning is identical in the present case, although this is not necessary. It accounts for efficiency, which is defined as the probability that a true entry will appear in any measured bin (as opposed to simply being lost). It also accounts for transfers between bins; that is, cases where an entry appears in a different bin than its original true bin. In contrast to bin-by-bin unfolding, this algorithm does not perform differently for stable events than for those with bin-to-bin transfers. Nor does it assume any sort of conceptual ordering between bins; therefore, multi-dimensional unfolding is mathematically equivalent to one-dimensional unfolding with a large number of bins.

In general, there are three sets of numbers used by the algorithm. There are N_M measured bins, labeled $M_1, M_2, ..., M_i, ..., M_{N_M}$; the number of events in these bins $n(M_i)$ is measured in the experiment being run. There are N_T true bins $T_1, T_2, ..., T_j, ..., T_{N_T}$; these are to be determined, but a prior value is also used by the algorithm, as will be discussed below. Finally there is the $N_M \times N_T$ response matrix R, for which R_{ij} is the probability of an entry in true bin i being reconstructed in measured bin j. Thus we have

$$n(M_j) = R_{ij} n(T_i), \tag{4.17}$$

where n denotes the number of events in the given bin.

The method does not account for unmatched reconstructed objects, which must be corrected for before it is applied, so we have the probability of a measured entry coming from *any* true entry is unity:

$$\sum_i R_{ij} = 1. \tag{4.18}$$

On the other hand, a true entry need not be reconstructed anywhere; if the efficiency for it being found is ϵ_i, then

$$\sum_j R_{ij} = \epsilon_i. \tag{4.19}$$

The response matrix must be derived from simulated events, which of course do not necessarily reflect the true efficiencies or resolutions of the ATLAS detector. If the response matrix entries are systematically incorrect, then the algorithm does not produce the correct T_i values from the measured R_i. The systematic uncertainties on the measurement that are assigned to account for this are discussed in Chapter 5.

The core calculation in the Bayesian Iterative Unfolding algorithm is based on Bayes's Theorem, which in this context is

$$P(T_i|M_j) = \frac{P(M_j|T_i)P(T_i)}{P(M_j)} = \frac{P(M_j|T_i)P(T_i)}{\sum_k P(M_j|T_k)P(T_k)}. \tag{4.20}$$

In this equation, repeated indices are not summed unless explicitly indicated. $P(T_i|M_j)$ is the probability, given an entry measured in bin j, of the true value being in bin i. $P(M_j|T_i)$ uses analogous notation and therefore equals R_{ij}. $P(T_i)$ is the probability distribution for the true bins, and $P(M_j)$ the probability distribution for the measured bins. As written, the formula is therefore trivial, because it depends on the quantity we are trying to compute. However, if we take the distribution $P(T_i)$ to be a prior (initial) probability, we can then calculate iteratively via

$$n_1(T_i) = \frac{1}{\epsilon_i}\sum_j n(M_j)P(T_i|M_j) = \frac{1}{\epsilon_i}\sum_j n(M_j)\frac{R_{ij}n_0(T_i)}{\sum_k R_{kj}n_0(T_k)}, \tag{4.21}$$

where $n_0(T_i)$ is the initial distribution from MC, and $n_1(T_i)$ is the newly calculated unfolded distribution. If the result of this computation is self-consistent, then we have shown that, given our measured distribution and our response matrix, our initial probability distribution was the correct one. In practice, this will not be the case. However, we can take the new distribution $n_1(T_i)$ and repeat the procedure to calculate $n_2(T_i)$, and so on until reasonable consistency is achieved.

4.7.2 Unfolding Algorithm Implementation

In the present analysis, unfolding is performed over four different three-dimensional distributions. Distributions in $p_{\mathrm{T}jet}$ and N_{jet}^{ch} are derived from a distribution with one entry per jet, binned in ($p_{\mathrm{T}jet}$,N_{jet}^{ch}, y_{jet}). Distributions in z are derived from a distribution with one entry per charged particle in a jet, binned in (z,$p_{\mathrm{T}jet}$,y_{jet}); distributions in p_{T}^{rel} and r_{ch} are derived from similar distributions. Thus for all distributions, the bin indices i and j from the previous section run over bins that cover particular ranges in three variables, and the response matrices are six-dimensional.

A reconstructed track jet is defined to be matched to a truth-level charged particle jet if it is within $\Delta R < 0.2$ (0.3) for jet radius 0.4 (0.6). Because this is half the jet radius, ambiguities in the matching are impossible. (The analysis was also performed

with a match cut up to 0.75 times the jet radius; this resulted in higher efficiencies by up to 2% in some bins, but had a negligible impact on the final results.) In the case of the z response matrix, if a truth jet is unmatched then all its tracks are counted as lost to inefficiency. If it is matched, then each of its constituent charged particles are matched to the closest track within $\Delta R < 0.04$; if there is no such track, then that particular particle is lost to inefficiency. The response matrices for p_{T}^{rel} and r are handled similarly.

As noted above, before the Bayesian Iterative Unfolding algorithm can be applied, a correction must be made for unmatched reconstructed jets. This correction is determined from the AMBT1 sample by the ratio between the reconstructed distributions and the matched reconstructed distributions. A systematic uncertainty is computed for this correction, and described in Section 5.4.

Three iterations of the algorithm are used; validation of that choice is discussed in Section 4.7.4. Statistical uncertainties on the unfolded variables are determined by producing 100 pseudoexperiment distributions, which vary the contents of each input bin randomly, then calculating the covariance matrix from the variation in the output of the unfolding.

4.7.3 Response Matrix Properties

The response matrix is best visualized with some variables projected out or fixed, to reduce the dimensionality of the 3 to 3 variable mapping. The efficiency of matching a jet with $|y| < 0.5$ and any number of constituents is shown as a function of p_{T} in Figure 4.9(a), and the efficiency of matching a jet with $|y| < 0.5$ and $10 < p_{\mathrm{T}}/\mathrm{GeV} < 15$ as a function of multiplicity is shown in Figure 4.9(b). Efficiency is defined by

$$\text{Efficiency} = \frac{\text{Truth jets in the bin with } \textit{any} \text{ matching reconstructed jet}}{\text{Truth jets in the bin}}. \qquad (4.22)$$

The distribution of reconstructed jet p_{T}, as a function of truth jet p_{T}, for $|y| < 0.5$ and any number constituents, is shown in Figure 4.10. It may be seen that there is a narrow distribution around the correct value, corresponding to smearing of track p_{T}, along with a large low-momentum tail corresponding to tracks lost to inefficiency. This is further illustrated in Figure 4.11, which shows the probability distribution of reconstructed jet p_{T} and multiplicity, for $|y| < 0.5$, $10 < p_{\mathrm{T}}/\mathrm{GeV} < 11$, and 4 tracks at truth level.

Purity and stability measure the number of bin-by-bin transfers in the response

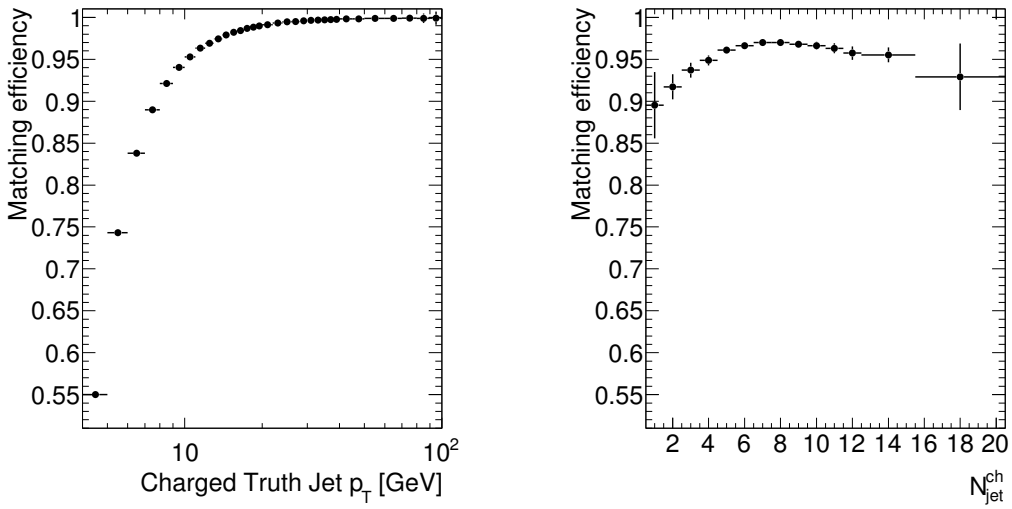


Figure 4.9: The efficiency for matching a track jet to a truth jet with $|y| < 0.5$ (a) as a function of p_{T}, (b) as a function of multiplicity (for $10 < p_{\mathrm{T}}/\ \mathrm{GeV} < 15$).

matrix. They are defined by

$$\text{Purity} = \frac{\text{Reconstructed jets in the bin with a matching truth jet in the same bin}}{\text{Reconstructed jets in the bin}}, \tag{4.23}$$

and

$$\text{Stability} = \frac{\text{Truth jets in the bin with a matching reconstructed jet in the same bin}}{\text{Truth jets in the bin}}. \tag{4.24}$$

The purity and stability of the response matrix are shown in Figure 4.12. Truth jets in the same p_{T} bin as their reconstructed counterpart, but with a different number of charged particles than the number of tracks and the reconstructed jet, are counted here as pure or stable. It is worth noting that a transfer in momentum bin usually also implies a transfer in charged particle multiplicity bin, as is illustrated by Figure 4.11.

As an additional consistency check, one can visualize the response matrix through the ratio between truth jets and reconstructed jets as a function of jet momentum and charged particles per jet. This is a simple ratio that does not depend on whether

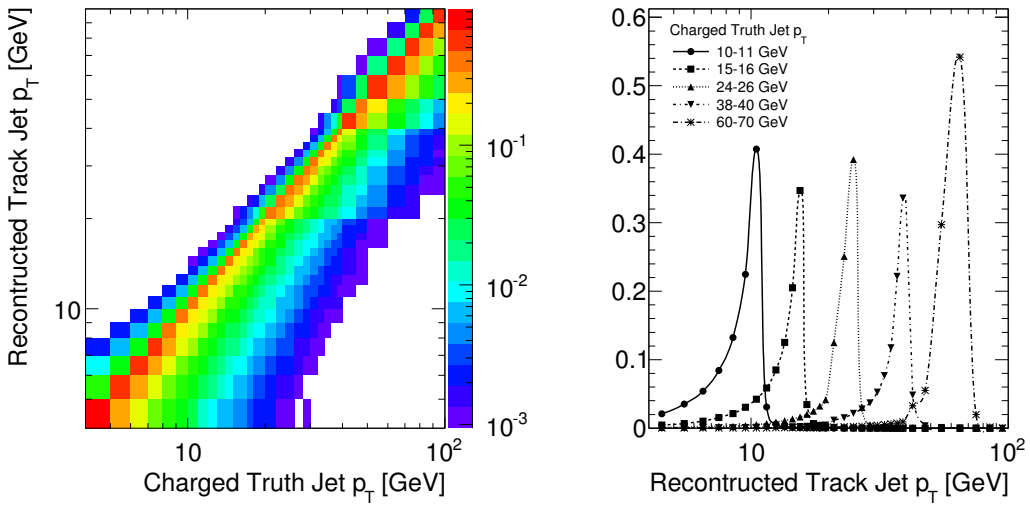


Figure 4.10: Distribution of matching track jets p_{T} as a function of a truth jet p_{T}, shown as (a) a 2-dimensional matrix, and (b) 1-dimensional track jet p_{T} distributions for selected truth jet p_{T} bins. It should be noted that lines in (b) are included only to guide the eye, and do not represent fits and have not been normalized to bin widths. (Thus for example the 60-70 GeV bin has the best probability to have the jet reconstructed in the same bin because the bin is widest.)

the jets migrated between bins; it corresponds to what a bin-by-bin correction *would* be if that were the correction being made. For (p_{T}, N) and $|y| < 0.5$, the ratio in the MC sample and the ratio in the data are compared in Figure 4.13. For the bins with most of the data, these ratios are of order one and are consistent with each other.

4.7.4 Unfolding Validation and Choice of Iteration Parameter

In order to validate the behavior of the unfolding and validate the choice of three iterations, 196 subsamples of the training MC sample are used; these subsamples have roughly the same number of jets as the data, although the low-momentum events must be reweighed. (These separately generated and simulated subsamples are distinct from the internal pseudoexperiments done to determine the statistical uncertainties output by the unfolding algorithm.) These subsamples are then each unfolded, and for each bin, the distribution over the subsamples for the pulls and

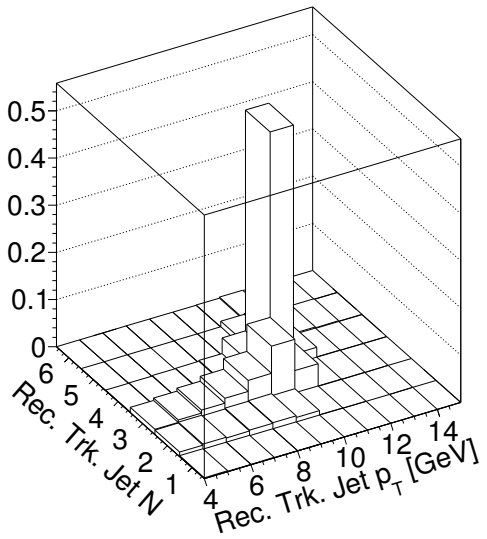


Figure 4.11: Distribution of matching track jet p_{T} and N^{ch}, given a truth jet with $10 < p_{\mathrm{T}}/\mathrm{GeV} < 11$ and 4 constituent tracks. This illustrates that the low tails in Figure 4.10 are associated primarily with jets where one or more tracks are not reconstructed.

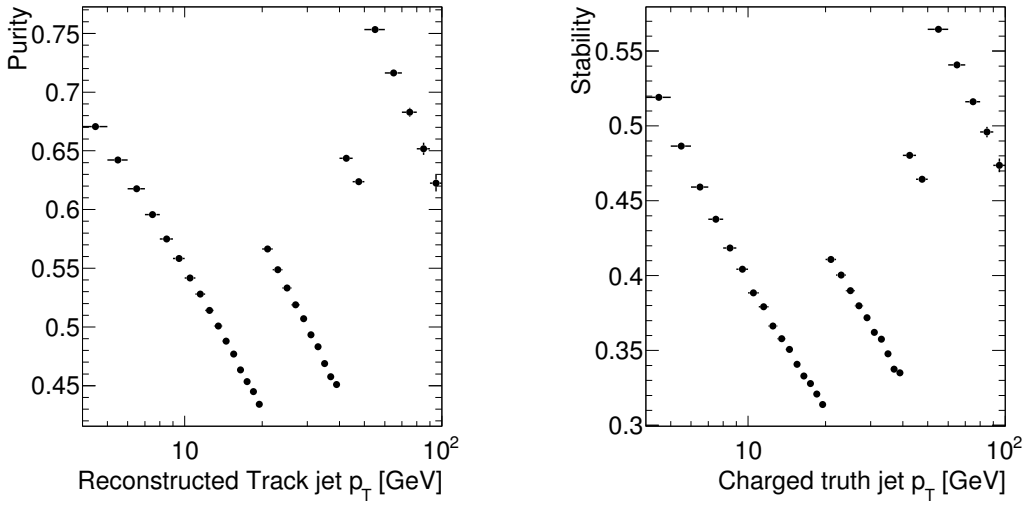
residuals are studied:

$$\text{Residual} = \frac{N_{corr} - N_{truth}}{N_{truth}}, \tag{4.25}$$

$$\text{Pull} = \frac{N_{corr} - N_{truth}}{\sigma_{corr}}, \tag{4.26}$$

where N_{corr} is the corrected number of events in a given bin, N_{truth} is the true number of events in that bin, and σ_{corr} is the statistical error for that bin as reported by the unfolding procedure. Residual and pull distributions for a typical momentum bin are shown in Figure 4.14.

The mean of the pull and residual distributions would be consistent with zero if there were no bias introduced by the unfolding. This is not always the case. For example, Figure 4.15(a) illustrates a 2% deviation at low momentum, which falls with increasing momentum but eventually rises again to 6% at the very highest momenta. The deviation in the residual mean is taken as a systematic uncertainty, described further in Section 5.5. The root mean square (RMS) of the pull distribution should be consistent with one if the statistical uncertainty is correctly computed. As is shown in Figure 4.15(b), the distributions improve with the number of iterations used; after



(a) Purity (b) Stability

Figure 4.12: The (a) purity and (b) stability of the response matrix for anti-k_t jets with $R = 0.6$, as defined in the text. Note that charged truth jets with the same momentum range as their matching reconstructed track jet are defined here to be stable, and the reconstructed track jet to be pure, even if they have a different number of charged particles. Jumps in these quantities occur where the bin width changes.

three, the average pull RMS is close to one, although some bins are not consistent with one. In order to account for this, the statistical errors in the final results are scaled by the RMS values found in the closure tests. This scaling is at most roughly 1.5 except for a few outlying bins in the multiplicity and p_{T}^{rel} distributions that have very low statistics.

This procedure is repeated with the same samples, but using the Phojet truth spectrum as the prior distribution in computing the output. As the response matrix itself is unchanged, an additional bias in this case would correspond to a systematic uncertainty associated with initial "guess" being different from the true answer. However, there is no significant additional bias after 3 iterations, as is shown in Figure 4.16(a). (It can be seen that there *is* significant additional bias after only 1 or 2 iterations.) This demonstrates that no uncertainty needs to be assigned and that 3 iterations are sufficient to "converge" independent of reasonable initial assumptions. The pull RMS distributions also remain similar, as shown in Figure 4.16(b).

Similar distributions are studied for 20 data-sized subsamples of the Perugia2010 MC. In this case, bias in the residual mean should be consistent with the systematic

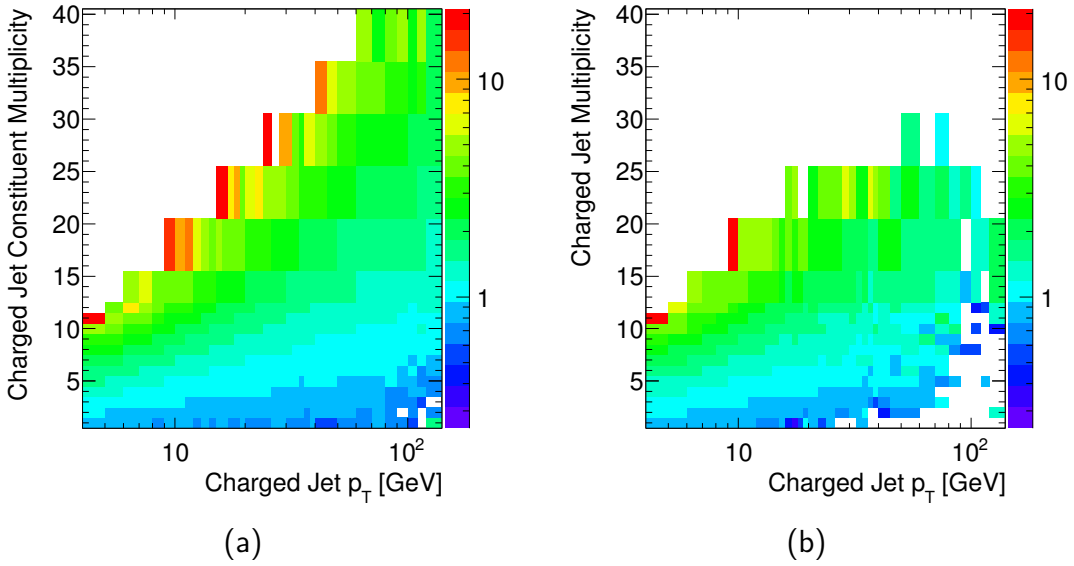
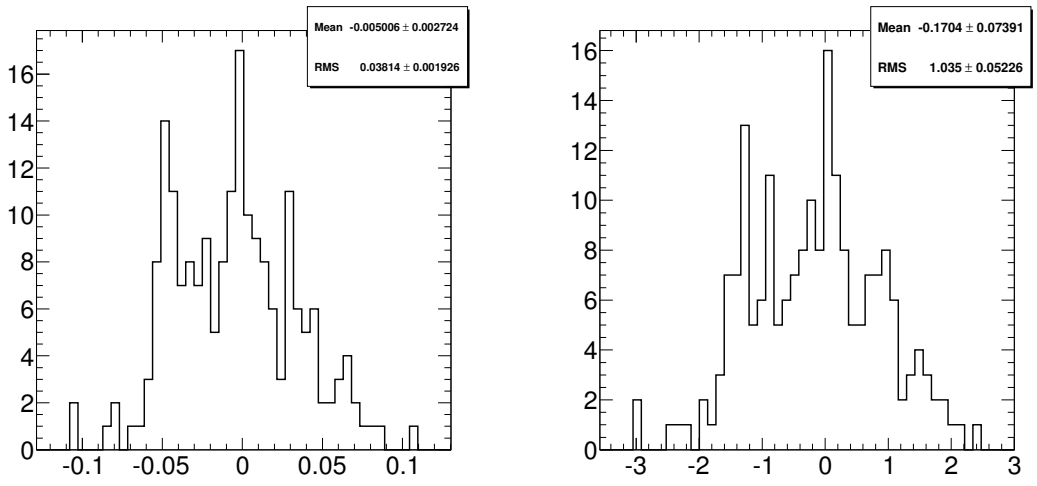


Figure 4.13: (a) The ratio of truth to reconstructed jets as determined from MC, and (b) the ratio of unfolded result to reconstructed data, both as a function of charged jet p_{T} and charged particles per jet.

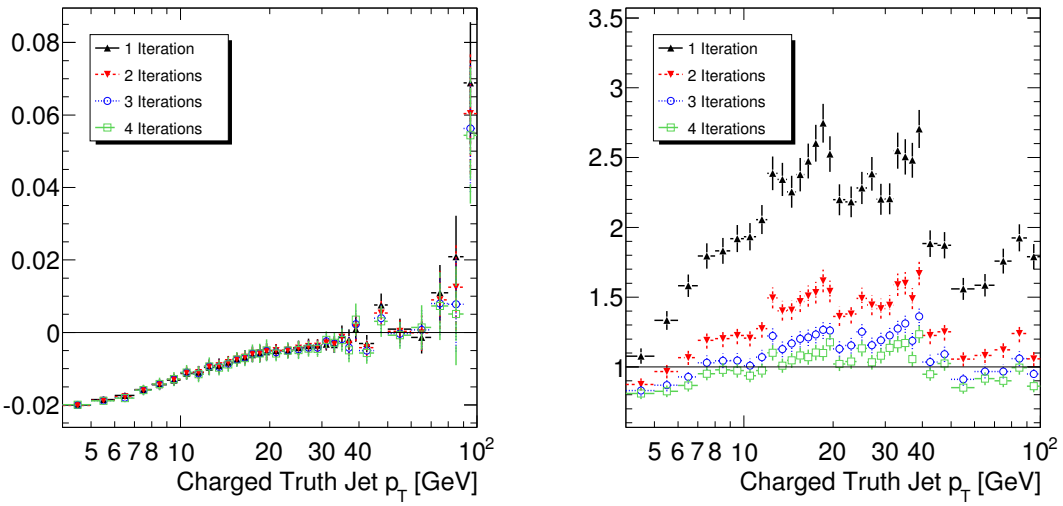
uncertainty due to MC tuning described in Section 5.1.3, added in quadrature to the unfolding biases described above. The pull widths should remain consistent with one despite the additional bias. Indeed, both of these consistency checks are passed, as shown in Figure 4.17.

Some of the closure test RMS distributions in p_{T}^{rel} or charged particle multiplicity have significant deviations from 1 even after 3 or 4 iterations. These distributions, being "doubly differential" in the sense of being made for subranges of jet momentum, suffer more from non-Gaussian bin migrations; the residual mean and pull RMS are shown in Figure 4.18 for one of the worst such distributions, the p_{T}^{rel} distribution for $4\,\mathrm{GeV} < p_{\mathrm{T}jet} < 6\,\mathrm{GeV}$, which has scale factors significantly larger than the typical upper bound of 1.5 in the highest-momentum (very low statistics) bins. As noted above, the final statistical errors are scaled to account for this.



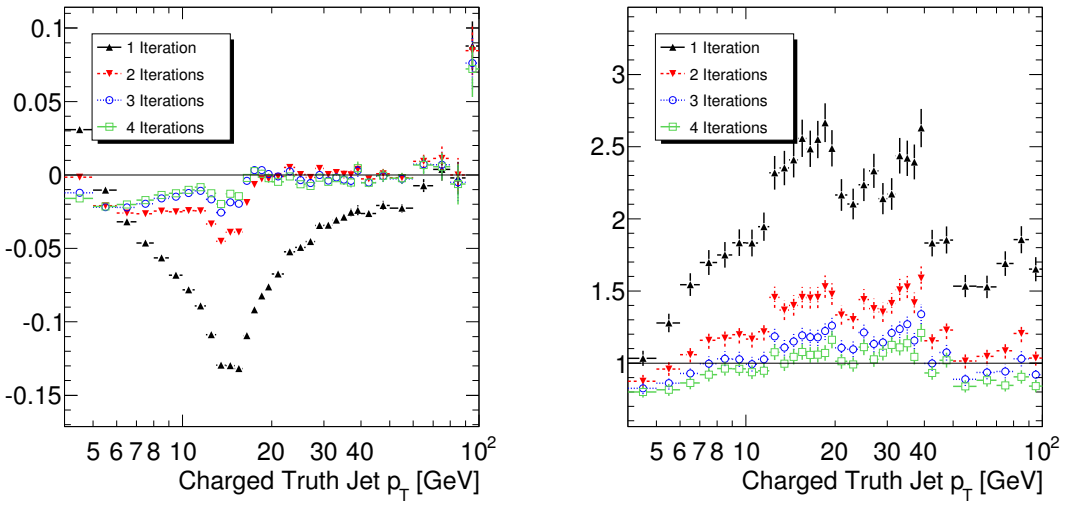
(a) Residual (b) Pull

Figure 4.14: The distribution of (a) residuals, and (b) pulls over 196 data-sized subsamples as described in the text. This is for the 40-45 GeV bin for $R = 0.6$ jets, after 3 iterations. The mean of the left plot appears as a point in Figure 4.15(a), and the RMS of the right plot as a point in Figure 4.15(b).



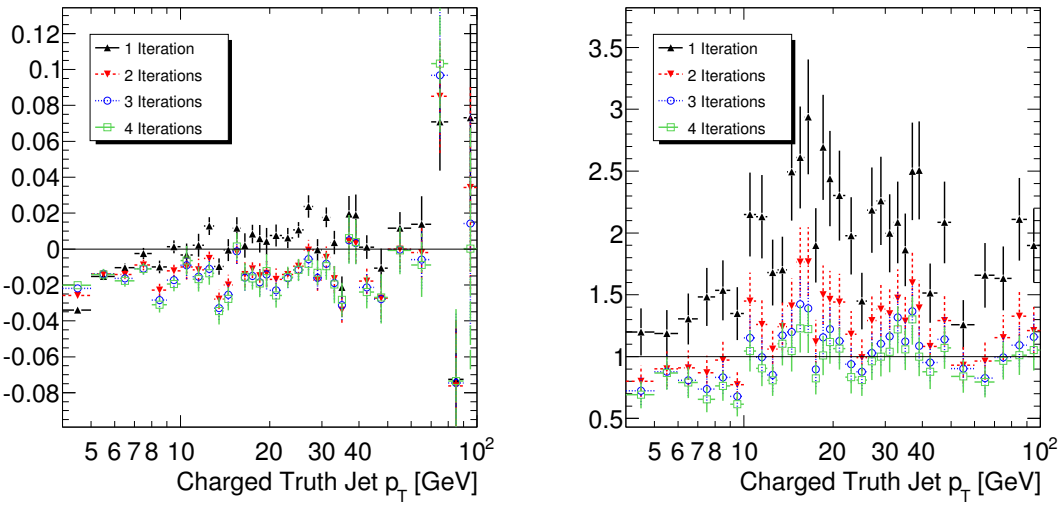
(a) Residual Mean (b) Pull RMS

Figure 4.15: For the p_{T} distribution with $0 < |y| < 0.5$, (a) the mean of the residuals and (b) the RMS of the pulls, over 196 data-sized fully simulated samples.



(a) Residual Mean (b) Pull RMS

Figure 4.16: For the p_{T} distribution with $0 < |y| < 0.5$, (a) the mean of the residuals and (b) the RMS of the pulls, over 196 data-sized fully simulated samples. The response matrix and data-sized samples both use the AMBT1 tuning, but the "initial guess" in the unfolding algorithm is set to the truth spectrum from Phojet. In figure (a), the process of convergence from the incorrect initial prior distribution to the final value is illustrated.



(a) Residual Mean (b) Pull RMS

Figure 4.17: For the p_{T} distribution with $0 < |y| < 0.5$ (a) the mean of the residuals and (b) the RMS of the pulls, over 20 data-sized fully simulated samples. The response matrix is constructed the AMBT1 tuning, but the samples use the Perugia2010 tuning.

(a) Residual Mean (b) Pull RMS

Figure 4.18: For the p_{T}^{rel} distribution with $|y_{jet}| < 1.9$ and $4\,\mathrm{GeV} < p_{\mathrm{T}jet} < 6\,\mathrm{GeV}$, (a) the mean of the residuals and (b) the RMS of the pulls, over 196 data-sized fully simulated samples. As discussed in the text, this distribution has amongst the largest scale pull RMS values, particularly in the highest momentum bins.

Chapter 5

Systematic Uncertainties

> As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know.
>
> Donald Rumsfeld

Systematic uncertainties quantify the degree to which the corrected measurement of a physical observable may differ from the true value because the assumptions used in making the corrections are incorrect. They are distinguished from statistical uncertainties because such assumptions are wrong by a consistent amount rather than a random amount. Statistical uncertainties vary randomly from one bin to the next, because by chance there will simply be more objects in some bins than the average and fewer in others; thus, statistical uncertainties become smaller as more data are recorded. By contrast, if an assumption is systematically wrong, the result will be that every bin disagrees with the true value in a correlated way.

To take a concrete example from this work, the uncertainty on the track-finding efficiency is the dominant systematic uncertainty for most bins in the charged particle jet cross section and multiplicity distributions. This is computed by creating data-like samples which include the largest "reasonable" deviation in the track finding efficiency, then applying the correction procedure to these samples; the resulting change in the corrected distributions is taken as the systematic uncertainty. If tracking efficiency is, for example, decreased from the baseline used in corrections, then jets are more likely to be lost when their core tracks are not reconstructed; jets will also be reconstructed with a smaller fraction of their true momentum and a lower multiplicity. As a result, the cross section would be reduced for every momentum bin, and the multiplicity distribution would be biased toward lower values. By contrast,

the particle property distributions are only modestly biased by changes in efficiency, because they are normalized to the number of jets; these distributions have dominant systematics from other sources.

The systematic uncertainties that contribute to the overall uncertainty in this analysis are described below. The calculated uncertainty values for representative bins appear in Appendix 5.A.

5.1 Uncertainties Computed using Smeared Samples

Systematic uncertainties due to tracking efficiency, tracking resolution, and variations in MC simulation are computed through use of simulated events in which aspects of detector performance or the physics simulation is varied. For each type of variation, an ensemble of 50 or more subsamples with the same number of events as the data is required.

Because full simulation of a sufficient number of events is impractical, a "toy" smearing mechanism is used. Using the baseline MC sample, the tracking efficiency and the resolution in $1/p_T$ are determined in bins of η, $1/p_T$, and ΔR from the nearest charged particle. For each generator tune, the events in the samples then have their particles vetoed or smeared according to the determined efficiency and resolution. In order to bring event generation to a manageable scale, events with low leading jet momentum are produced in reduced numbers and given weights; at high jet momentum, events are always used with a weight of one.

Truth particles are smeared, with the smearing possibly modified, and then the jet algorithm is applied and the usual distributions produced. Many data-sized samples are produced in this manner, and they are then unfolded. The baseline is an ensemble of 196 data-sized AMBT1 samples. Uncertainties due to track-finding efficiency are found by reusing the same truth particles but using different efficiencies to smear them (Section 5.1.1). Likewise, uncertainties in tracking resolution are investigated by applying an additional Gaussian smearing (Section 5.1.2). Uncertainties associated with varying MC parameters are determined by producing ensembles of data-sized events using other tunes, applying the same smearing as to the baseline sample, and comparing the derived corrections. In all cases, the largest (smallest) increase is taken as an upper-bound (lower-bound) systematic uncertainty.

5.1.1 Tracking Efficiency

The systematic uncertainty associated with uncertainties on the tracking efficiency is determined by varying the properties of the smearing. The same set of 196 data-sized AMBT1 samples is used to produce three smeared samples: one using the

baseline, one with the efficiency increased, and one with the efficiency decreased. The amount of the change is obtained from effciency uncertainties computed for the ATLAS experiment's inclusive charged particle measurement [7], which are measured as a function of momentum and rapidity; they vary from roughly 2% for particles with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ and $|\eta| < 1.3$, up to 7% for $300\,\mathrm{MeV} < p_{\mathrm{T}} < 350\,\mathrm{MeV}$ and $2.3 < |\eta| < 2.5$. When the efficiency used for the smearing is increased (decreased), the correction factors become smaller (larger), so that the computed cross section is reduced (increased). Thus for each bin, the positive and negative uncertainties due to tracking efficiency are given by:

$$U_{Trk.\,eff.,+} = \frac{N_{smeared\,down}}{N_{smeared,\,base}} - 1, \tag{5.1}$$

$$U_{Trk.\,eff.,-} = 1 - \frac{N_{smeared\,up}}{N_{smeared,\,base}}, \tag{5.2}$$

where $N_{smeared,\,base}$ is the number of events found in that bin after unfolding the baseline smeared sample, while $N_{smeared\,down}$ ($N_{smeared\,up}$) is the number of events found after unfolding the sample smeared with lowered (raised) efficiency.

5.1.2 Tracking Resolution

High-momentum tracks are known from muon studies [91] to have a wider Inner Detector momentum resolution than appears in MC; this is determined by comparing a muon's momentum as measured in the Inner Detector with the muon system measurement. It is assumed that this discrepancy is relevant for all particles. The difference between the MC baseline and the measurement from data is taken as the basis for a systematic uncertainty due to resolution. An additional "wide resolution" jet collection is built for the primary AMBT1 ensemble of data-sized samples. The tracks with $p_{\mathrm{T}} > 5$ GeV in this collection have an additional Gaussian smearing to bring the resolution into agreement with that measured from data. (Below this momentum, the additional smearing is too small to have a significant impact.) The difference in average correction between this ensemble and the baseline is taken (symmetrically) as the uncertainty due to resolution smearing, so the uncertainty is given by

$$U_{Trk.\,Res.,\pm} = \left| \frac{N_{smeared,\,wide\,res.}}{N_{smeared,\,base}} - 1 \right|, \tag{5.3}$$

where $N_{smeared,\,wide\,res.}$ is the number of events found in a given bin after unfolding the sample with widened resolution. It should be noted that symmetrizing this uncertainty is somewhat artificial, corresponding to *narrowing* the resolution as compared with data. This is a conservative way to obtain a two-sided uncertainty given the

available information, and the impact on the measurement is small.

5.1.3 Monte Carlo Tuning

The uncertainties associated with differing MC tunings are explored through smeared data-sized subsamples using all event generators and tunings described in Section 2.7. The ratio of unfolded-to-reconstructed values in each bin is compared to that of the central tuning; the largest increase is taken as the upper bound systematic uncertainty, while the largest decrease is taken as the lower bound:

$$U_{MC,+} = \max_{MC} \left\{ \left(\frac{N_{smeared}}{N_{true}} \right)_{MC} \Big/ \left(\frac{N_{smeared}}{N_{true}} \right)_{base} \right\} - 1, \tag{5.4}$$

$$U_{MC,-} = 1 - \min_{MC} \left\{ \left(\frac{N_{smeared}}{N_{true}} \right)_{MC} \Big/ \left(\frac{N_{smeared}}{N_{true}} \right)_{base} \right\}, \tag{5.5}$$

where N_{true} denotes the number of true events in a given bin, $N_{smeared}$ denotes the number of events in the bin after smearing, and the subscript MC refers to each Monte Carlo sample, while *base* refers to the baseline AMBT1 sample. Unlike the foregoing samples, this is formulated as a double ratio because it is the sample used that changes, rather than the properties of the smearing; thus the number of true events that should be present changes from sample to sample, whereas it previously did not. As in other cases, the computation is based on averaging many data-sized subsamples – 196 for AMBT1, 20 for Perugia2010, 50 for others, as dictated by the available number of events – in order to reduce statistical fluctuations.

Although the response matrix primarily models "detector effects," there are several ways that truth-level details can impact the corrections. Differing jet or track momentum distributions change the population *within* bins; so for example if one MC sample has an average jet p_{T} of 5.2 GeV in the 5-6 GeV bin, and another has an average of 5.3, the response associated with the 5-6 GeV bin will reflect a higher weight on higher-momentum jets in the latter. Differences in the amount or distribution of activity in the underlying event can significantly change the fragmentation distribution, especially in the low-momentum bins, because these tracks are not correlated with the jet direction and so may appear at large radii. Differing strangeness fractions can change the distribution of long-lifetime tracks that decay and produce kinks in the tracker, leading to momentum mismeasurements and/or loss of tracks due to failed hit requirements.

The variation over tunes also accounts for PDF uncertainties, because MC properties are tuned with a particular PDF and several different PDFs are used. Reweighting events according to a set of PDF errors is not performed, because the samples used are all minimum bias samples, which use leading order PDFs with very large uncertain-

ties at low x. Because MC properties are tuned to data assuming a particular PDF set, such a variation would greatly overestimate the variation in event properties.

5.2 High-Momentum Mismeasured Tracks

Despite the tightening of the d_0 cut for high-momentum tracks, in simulation a significant fraction of selected tracks at very high momentum do not have a matching truth particle with momentum within 50% (Figure 5.1(a)). The impact of these significantly mismeasured tracks on the measurement is accounted for the unfolding procedure provided that the fraction is similar in data; however, the data have a larger fraction of tracks failing the d_0 cut, as shown in Figure 5.1(b). A systematic uncertainty on the number of mismeasured tracks in data is computed and propagated to assess its impact on jets. The upper bound of this uncertainty assumes that all such extra tracks are well-measured, but did not pass the d_0 cut due to poorer-than-modelled resolution; the lower bound is based on the assumption that the increase in rejected tracks corresponds to a proportional increase in accepted mismeasured tracks. Therefore the uncertainties for tracks are given by

$$u_{\text{High-p}_\text{T},+} = I_{Data} - I_{MC}, \tag{5.6}$$

$$u_{\text{High-p}_\text{T},-} = \frac{I_{Data}}{I_{MC}} M_{MC}, \tag{5.7}$$

where $u_{\text{High-p}_\text{T},\pm}$ is the upper/lower uncertainties, I is the inefficiency of the d_0 cut, and M is the fraction of mismeasured tracks. These uncertainties are shown as a function of track momentum in Table 5.1. Similar uncertainties, albeit slightly smaller, are obtained by replacing the high-momentum d_0 cut with a high-momentum TRT extension requirement.

These uncertainties are then propagated to each measured jet bin, with the scale factor determined by the fraction of jets in a given bin with leading track in each momentum range:

$$U_{\text{High-p}_\text{T},\pm}(p_{\text{T}jet}, N^{ch}_{jet}) = \sum_{p_{\text{T},lead\,track}} f\,(p_{\text{T}jet}, N^{ch}_{jet}, p_{\text{T},lead\,track}) u_{\text{High-p}_\text{T},\pm}(p_{\text{T},lead\,track}) \tag{5.8}$$

,

where $U_{\text{High-p}_\text{T},\pm}(p_{\text{T}jet}, N^{ch}_{jet})$ is the upper/lower systematic uncertainty on the number of reconstructed jets in that $(p_{\text{T}jet}, N^{ch}_{jet})$ bin, and $f(p_{\text{T}jet}, N^{ch}_{jet}, p_{\text{T},lead\,track})$ is the frac-

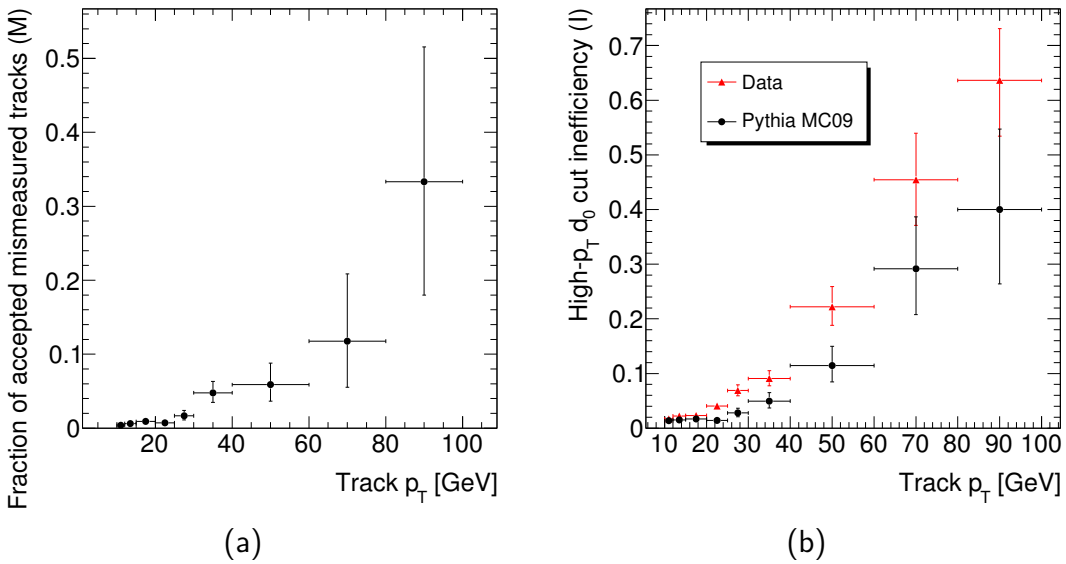


Figure 5.1: Information used to compute the high-momentum mismeasured track uncertainty. (a) Fraction of accepted tracks in MC that do not match to any charged truth particle whose momentum is within 50%. (b) Fraction of tracks failing the high-momentum d_0 cut in data (red triangles) and MC09 simulation (black circles).

tion of tracks in that bin with lead track momentum $p_{\mathrm{T},lead\,track}$ in the given range. Thus the correction is large at high jet momentum and low number of particles per jet, because these jets have the highest leading track momenta, and hence the greatest impact from high-momentum mismeasured tracks.

This uncertainty on the reconstructed bins is then used to scale up and down the data, with the unfolding applied in each case and compared to the unfolding central value. The resulting differences in each measured bin give the final uncertainty due to high-momentum mismeasured tracks.

This correction is also applied to the z, p_{T}^{rel}, and r distributions. Each bin is scaled in proportion to the fraction of tracks in that bin that come from a jet whose leading track is in a given range, in proportion to the uncertainty on tracks in that range. The corrections are only significant for high-z bins in high-p_{T} jets. The corrections are then propagated through the unfolding as in the jet-based distributions, and the result compared to the central value.

Table 5.1: Fractional uncertainties on the number of high-momentum tracks, computed from the text. These uncertainties are then propagated to jets based on the distribution of the leading track momenta within each (p_{T},N^{ch}) jet bin.

Track p_{T} [GeV]	Upper uncertainty ($u_{\text{High-pT},+}$)	lower uncertainty ($u_{\text{High-pT},-}$)
10-12	+0.3%	-0.5%
12-15	+0.7%	-0.9%
15-20	+0.7%	-1.3%
20-25	+2.7%	-2.1%
25-30	+4.1%	-4.1%
30-40	+4.1%	-8.7%
40-60	+10.8%	-11.4%
60-80	+16.3%	-18.3%
80-100	+23.6%	-53.0%

5.3 Misreconstruction of Multiple Vertices

The data are corrected as a function of the number of selected tracks in the event, as discussed in Section 4.3.3. A weight of $1/\epsilon_{nosplit}$ is applied on an event-by-event basis; the correction is less than 0.25% for 10 or fewer selected tracks, rising to 1% around 30 tracks and continuing steadly to 4% at about 100 tracks. As the correction is derived entirely from MC, the full value of the correction is taken as a (symmetrized) uncertainty. After unfolding, the value of this correction is relatively flat as a function of jet momentum, varying from about 1.8% at 4-5 GeV to 2.2% at 90-100 GeV.

5.4 Unmatched Reconstructed Tracks

As described in Section 4.7.3, a scaling correction is applied to the data to account for unmatched reconstructed jets, because these jets would otherwise not appear in the response matrix and not be accounted for in the unfolding. In order to assign an uncertainty to this correction, the unfolding is repeated with a correction determined from the Perugia 2010 sample rather than AMBT1. The fractional difference in the unfolding output between this and the baseline configuration is taken to be the

uncertainty on the correction; the uncertainty is symmetrized.

$$U_{Unmatched,\pm} = \left| \frac{N_{data,\,Perugia\,2010}}{N_{data,\,AMBT1}} - 1 \right|, \tag{5.9}$$

where $N_{data,\,Perugia\,2010}$ is the number of events in a given bin when the data is unfolded using the Perugia 2010 correction, while $N_{data,\,AMBT1}$ is the number of events obtained when unfolding the data with the baseline AMBT1 correction. Perugia 2010 is the only sample that can be used for this purpose, because it is the only fully simulated sample other than the baseline AMBT1 that has sufficient events in the high-momentum bins.

5.5 Unfolding Uncertainties

As discussed in Section 4.7.4, the closure tests do not achieve perfect consistency between the true value and the result after unfolding for all bins. The deviation is determined by averaging over the 196 data-sized AMBT1 samples, with the unfolding applied to fully simulated events; it is typically a few percent or less. The uncertainty on the final result is taken to be the size of the disagreement for each bin, and is symmetrized. Thus the uncertainty for each bin is thus given by

$$U_{Closure,\pm} = \left| \frac{N_{unfolded}}{N_{true}} - 1 \right|, \tag{5.10}$$

where $N_{unfolded}$ is the average number of events after full simulation and unfolding, while N_{true} is the average number of true events in the bin.

Appendix 5.A Uncertainties for Selected Bins

Uncertainties for representative bins and distributions are given in the tables on the following pages. Cross section and p_{T}^{rel} uncertainties appear in Table 5.2; multiplicity uncertainties appear in Table 5.3; and z and $\rho_{ch}(r)$ uncertainties appear in Table 5.4.

Table 5.2: Summary of systematic uncertainties for selected bins in selected cross section and p_{T}^{rel} distributions, for R = 0.6. An overall luminosity uncertainty of 3.4% is not shown, and applies to the cross section only.

Distribution	Bin [GeV]	Track eff.	Track res.	Monte Carlo	High-p_{T} Tracks	Unmatched Jets	Split Vertex	Closure	Total
$\frac{d^2\sigma_{jet}}{dp_{\mathrm{T}jet}dy_{jet}}$	4 - 5	$^{+3.2\%}_{-3.3\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+0.07\%}_{-1.2\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.21\%}_{-0.21\%}$	$^{+1.8\%}_{-1.8\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+4.2\%}_{-4.4\%}$
	20 - 22	$^{+6.6\%}_{-6.3\%}$	$^{+0.34\%}_{-0.34\%}$	$^{+2.2\%}_{-0.69\%}$	$^{+0.10\%}_{-0.14\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+0.54\%}_{-0.54\%}$	$^{+7.3\%}_{-6.7\%}$
$\lvert y_{jet}\rvert$<0.5	40 - 45	$^{+6.8\%}_{-6.9\%}$	$^{+0.39\%}_{-0.39\%}$	$^{+1.4\%}_{-2.3\%}$	$^{+1.1\%}_{-1.4\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+2.1\%}_{-2.1\%}$	$^{+0.50\%}_{-0.50\%}$	$^{+7.3\%}_{-7.7\%}$
	80 - 90	$^{+7.1\%}_{-6.9\%}$	$^{+0.34\%}_{-0.34\%}$	$^{+7.1\%}_{-7.5\%}$	$^{+5.5\%}_{-9.1\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.3\%}_{-2.3\%}$	$^{+0.78\%}_{-0.78\%}$	$^{+12\%}_{-14\%}$
$\frac{d^2\sigma_{jet}}{dp_{\mathrm{T}jet}dy_{jet}}$	4 - 5	$^{+5.4\%}_{-5.3\%}$	$^{+0.02\%}_{-0.02\%}$	$^{+3.0\%}_{-2.1\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.19\%}_{-0.19\%}$	$^{+1.8\%}_{-1.8\%}$	$^{+5.2\%}_{-5.2\%}$	$^{+8.3\%}_{-7.9\%}$
	20 - 22	$^{+10\%}_{-9.5\%}$	$^{+0.02\%}_{-0.02\%}$	$^{+2.6\%}_{-3.2\%}$	$^{+0.10\%}_{-0.15\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+0.96\%}_{-0.96\%}$	$^{+11\%}_{-10\%}$
1.5<$\lvert y_{jet}\rvert$<1.9	40 - 45	$^{+11\%}_{-10\%}$	$^{+0.30\%}_{-0.30\%}$	$^{+2.9\%}_{-1.5\%}$	$^{+1.1\%}_{-1.3\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+2.1\%}_{-2.1\%}$	$^{+0.95\%}_{-0.95\%}$	$^{+11\%}_{-11\%}$
	80 - 90	$^{+12\%}_{-11\%}$	$^{+1.3\%}_{-1.3\%}$	$^{+7.3\%}_{-12\%}$	$^{+7.3\%}_{-8.2\%}$	$^{+9.3\%}_{-9.3\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+6.9\%}_{-6.9\%}$	$^{+20\%}_{-21\%}$
$\frac{1}{N_{jet}}\frac{dN_{ch}}{dp_{\mathrm{T}}^{rel}}$	0.0 - 0.05	$^{+0.44\%}_{-0.32\%}$	$^{+0.18\%}_{-0.18\%}$	$^{+16\%}_{-4.8\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.7\%}_{-2.7\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.17\%}_{-0.17\%}$	$^{+16\%}_{-5.6\%}$
$\lvert y_{jet}\rvert$<1.9	0.5 - 0.55	$^{+0.30\%}_{-0.20\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+3.0\%}_{-1.7\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+3.2\%}_{-3.2\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+4.4\%}_{-3.6\%}$
4 GeV<$p_{\mathrm{T}jet}$<6 GeV	0.85 - 0.9	$^{+1.6\%}_{-0.03\%}$	$^{+1.2\%}_{-1.2\%}$	$^{+10\%}_{-6.5\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.9\%}_{-2.9\%}$	$^{+0.02\%}_{-0.02\%}$	$^{+2.1\%}_{-2.1\%}$	$^{+11\%}_{-7.5\%}$
$\frac{1}{N_{jet}}\frac{dN_{ch}}{dp_{\mathrm{T}}^{rel}}$	0.0 - 0.05	$^{+1.2\%}_{-1.1\%}$	$^{+0.10\%}_{-0.10\%}$	$^{+3.4\%}_{-3.0\%}$	$^{+0.21\%}_{-0.17\%}$	$^{+0.58\%}_{-0.58\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+3.6\%}_{-3.2\%}$
	0.5 - 0.55	$^{+1.2\%}_{-1.2\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+3.3\%}_{-2.4\%}$	$^{+0.38\%}_{-0.29\%}$	$^{+0.40\%}_{-0.40\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+3.5\%}_{-2.7\%}$
$\lvert y_{jet}\rvert$<1.9	0.85 - 0.9	$^{+0.95\%}_{-1.0\%}$	$^{+0.06\%}_{-0.06\%}$	$^{+3.3\%}_{-2.6\%}$	$^{+0.32\%}_{-0.25\%}$	$^{+0.57\%}_{-0.57\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+3.5\%}_{-2.9\%}$
24 GeV<$p_{\mathrm{T}jet}$<40 GeV	3.0 - 3.5	$^{+2.6\%}_{-2.4\%}$	$^{+0.56\%}_{-0.56\%}$	$^{+8.7\%}_{-1.1\%}$	$^{+0.05\%}_{-0.04\%}$	$^{+0.27\%}_{-0.27\%}$	$^{+0.14\%}_{-0.14\%}$	$^{+0.48\%}_{-0.48\%}$	$^{+9.1\%}_{-2.7\%}$

Table 5.3: Summary of systematic uncertainties for selected bins in selected multiplicity bins, for $R = 0.6$.

Distribution	Bin	Track eff.	Track res.	Monte Carlo	High-p_{T} Tracks	Unmatched Jets	Split Vertex	Closure	Total
$\frac{1}{N_{jet}}\frac{dN_{jet}}{dN^{ch}_{jet}}$	1	$^{+6.8\%}_{-6.3\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+7.9\%}_{-10\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.15\%}_{-0.15\%}$	$^{+0.21\%}_{-0.21\%}$	$^{+0.25\%}_{-0.25\%}$	$^{+10\%}_{-12\%}$
	3	$^{+2.8\%}_{-2.6\%}$	$^{+0.14\%}_{-0.14\%}$	$^{+3.3\%}_{-0.27\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.16\%}_{-0.16\%}$	$^{+0.08\%}_{-0.08\%}$	$^{+0.26\%}_{-0.26\%}$	$^{+4.3\%}_{-2.6\%}$
$\lvert y_{jet}\rvert$<1.9	5	$^{+1.2\%}_{-1.3\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.23\%}_{-2.4\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.04\%}_{-0.04\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+1.2\%}_{-2.7\%}$
$4\,GeV<p_{\mathrm{T}jet}<6\,GeV$	9	$^{+11\%}_{-10\%}$	$^{+0.36\%}_{-0.36\%}$	$^{+11\%}_{-8.4\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+0.24\%}_{-0.24\%}$	$^{+1.4\%}_{-1.4\%}$	$^{+15\%}_{-13\%}$
$\frac{1}{N_{jet}}\frac{dN_{jet}}{dN^{ch}_{jet}}$	1	$^{+19\%}_{-11\%}$	$^{+3.2\%}_{-3.2\%}$	$^{+50\%}_{-8.5\%}$	$^{+3.7\%}_{-5.5\%}$	$^{+0.13\%}_{-0.13\%}$	$^{+0.35\%}_{-0.35\%}$	$^{+35\%}_{-35\%}$	$^{+64\%}_{-38\%}$
	3	$^{+10\%}_{-9.9\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+3.1\%}_{-3.3\%}$	$^{+1.7\%}_{-1.9\%}$	$^{+0.14\%}_{-0.14\%}$	$^{+0.28\%}_{-0.28\%}$	$^{+0.77\%}_{-0.77\%}$	$^{+11\%}_{-11\%}$
$\lvert y_{jet}\rvert$<1.9	5	$^{+6.8\%}_{-6.5\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+3.3\%}_{-1.4\%}$	$^{+0.64\%}_{-0.75\%}$	$^{+0.06\%}_{-0.06\%}$	$^{+0.21\%}_{-0.21\%}$	$^{+0.26\%}_{-0.26\%}$	$^{+7.6\%}_{-6.7\%}$
$24\,GeV<p_{\mathrm{T}jet}<40\,GeV$	9	$^{+1.6\%}_{-1.7\%}$	$^{+0.14\%}_{-0.14\%}$	$^{+2.9\%}_{-0.21\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+0.02\%}_{-0.02\%}$	$^{+0.08\%}_{-0.08\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+3.3\%}_{-1.7\%}$

Table 5.4: Summary of systematic uncertainties for selected bins in selected z and $\rho_{ch}(r)$ distributions, for $R = 0.6$.

Distribution	Bin	Track eff.	Track res.	Monte Carlo	High-p_{T} Tracks	Unmatched Jets	Split Vertex	Closure	Total
$\frac{1}{N_{jet}}\frac{dN_{ch}}{dz}$	0.1-0.125	$^{+1.7\%}_{-1.7\%}$	$^{+0.11\%}_{-0.11\%}$	$^{+1.5\%}_{-8.9\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+3.0\%}_{-9.3\%}$
	0.5-0.525	$^{+1.2\%}_{-1.3\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+3.2\%}_{-0.54\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.0\%}_{-2.0\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.18\%}_{-0.18\%}$	$^{+3.9\%}_{-2.4\%}$
$\lvert y_{jet}\rvert$<1.9	0.85-0.9	$^{+4.5\%}_{-4.7\%}$	$^{+0.20\%}_{-0.20\%}$	$^{+3.7\%}_{-4.0\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.25\%}_{-0.25\%}$	$^{+0.12\%}_{-0.12\%}$	$^{+0.33\%}_{-0.33\%}$	$^{+5.9\%}_{-6.1\%}$
$4\,GeV<p_{\mathrm{T}jet}<6\,GeV$	1.0	$^{+7.0\%}_{-6.5\%}$	$^{+0.08\%}_{-0.08\%}$	$^{+4.8\%}_{-16\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.23\%}_{-0.23\%}$	$^{+0.28\%}_{-0.28\%}$	$^{+8.5\%}_{-17\%}$
$\frac{1}{N_{jet}}\frac{dN_{ch}}{dz}$	0.1-0.125	$^{+1.2\%}_{-1.3\%}$	$^{+0.14\%}_{-0.14\%}$	$^{+3.0\%}_{-5.8\%}$	$^{+0.42\%}_{-0.32\%}$	$^{+0.01\%}_{-0.01\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+3.3\%}_{-5.9\%}$
	0.5-0.525	$^{+3.9\%}_{-3.6\%}$	$^{+0.29\%}_{-0.29\%}$	$^{+2.7\%}_{-0.86\%}$	$^{+0.40\%}_{-0.65\%}$	$^{+0.64\%}_{-0.64\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+0.24\%}_{-0.24\%}$	$^{+4.8\%}_{-3.9\%}$
$\lvert y_{jet}\rvert$<1.9	0.85-0.9	$^{+9.0\%}_{-9.3\%}$	$^{+1.7\%}_{-1.7\%}$	$^{+1.8\%}_{-6.6\%}$	$^{+3.3\%}_{-3.7\%}$	$^{+0.21\%}_{-0.21\%}$	$^{+0.20\%}_{-0.20\%}$	$^{+1.3\%}_{-1.3\%}$	$^{+10\%}_{-12\%}$
$24\,GeV<p_{\mathrm{T}jet}<40\,GeV$	0.95-1.0	$^{+13\%}_{-13\%}$	$^{+1.3\%}_{-1.3\%}$	$^{+6.8\%}_{-3.5\%}$	$^{+3.8\%}_{-5.5\%}$	$^{+0.81\%}_{-0.81\%}$	$^{+0.30\%}_{-0.30\%}$	$^{+3.1\%}_{-3.1\%}$	$^{+16\%}_{-15\%}$
$\rho_{ch}(r)$	0.0 - 0.01	$^{+7.1\%}_{-6.8\%}$	$^{+0.12\%}_{-0.12\%}$	$^{+20\%}_{-46\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.30\%}_{-0.30\%}$	$^{+0.22\%}_{-0.22\%}$	$^{+0.36\%}_{-0.36\%}$	$^{+21\%}_{-46\%}$
	0.09 - 0.1	$^{+0.00\%}_{-0.06\%}$	$^{+0.22\%}_{-0.22\%}$	$^{+2.6\%}_{-10\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.4\%}_{-2.4\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.12\%}_{-0.12\%}$	$^{+3.5\%}_{-10\%}$
$\lvert y_{jet}\rvert$<1.9	0.28-0.3	$^{+0.90\%}_{-1.1\%}$	$^{+0.11\%}_{-0.11\%}$	$^{+2.5\%}_{-10\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+2.4\%}_{-2.4\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+0.03\%}_{-0.03\%}$	$^{+3.6\%}_{-11\%}$
$4\,GeV<p_{\mathrm{T}jet}<6\,GeV$	0.55 - 0.6	$^{+2.4\%}_{-2.8\%}$	$^{+0.06\%}_{-0.06\%}$	$^{+0.94\%}_{-3.1\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+3.8\%}_{-3.8\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+5.4\%}_{-5.4\%}$	$^{+7.1\%}_{-7.8\%}$
$\rho_{ch}(r)$	0.0 - 0.01	$^{+3.3\%}_{-3.4\%}$	$^{+0.08\%}_{-0.08\%}$	$^{+4.2\%}_{-5.1\%}$	$^{+0.48\%}_{-0.59\%}$	$^{+0.20\%}_{-0.20\%}$	$^{+0.09\%}_{-0.09\%}$	$^{+0.20\%}_{-0.20\%}$	$^{+5.4\%}_{-6.2\%}$
	0.09 - 0.1	$^{+0.56\%}_{-0.60\%}$	$^{+0.04\%}_{-0.04\%}$	$^{+2.9\%}_{-4.6\%}$	$^{+0.33\%}_{-0.26\%}$	$^{+0.21\%}_{-0.21\%}$	$^{+0.00\%}_{-0.00\%}$	$^{+0.05\%}_{-0.05\%}$	$^{+3.0\%}_{-4.7\%}$
$\lvert y_{jet}\rvert$<1.9	0.28-0.3	$^{+2.0\%}_{-2.1\%}$	$^{+0.07\%}_{-0.07\%}$	$^{+3.4\%}_{-8.0\%}$	$^{+0.45\%}_{-0.35\%}$	$^{+0.86\%}_{-0.86\%}$	$^{+0.10\%}_{-0.10\%}$	$^{+0.04\%}_{-0.04\%}$	$^{+4.1\%}_{-8.4\%}$
$24\,GeV<p_{\mathrm{T}jet}<40\,GeV$	0.55 - 0.6	$^{+2.5\%}_{-2.3\%}$	$^{+0.02\%}_{-0.02\%}$	$^{+3.8\%}_{-6.7\%}$	$^{+0.47\%}_{-0.36\%}$	$^{+1.3\%}_{-1.3\%}$	$^{+0.15\%}_{-0.15\%}$	$^{+0.64\%}_{-0.64\%}$	$^{+4.8\%}_{-7.2\%}$

Chapter 6

Results and Discussion

A selection of the distributions measured in this analysis appear in Figures 6.1–6.10; other rapidity ranges may be found in Appendix A. They are compared the MC distributions described in Section 2.7.

6.1 Charged particle jet cross section

cross sections as a function of jet p_{T} are shown in Figure 6.1. The simulated cross sections shown for comparison are scaled to the data, using the scale factor S defined by:

$$S = \sigma^{total}_{data}/\sigma^{total}_{MC}, \tag{6.1}$$

where

$$\sigma^{total} = \int_{4\,\mathrm{GeV}}^{100\,\mathrm{GeV}} dp_{\mathrm{T}jet} \int_{-1.9}^{1.9} dy_{jet} \frac{d^2\sigma_{jet}}{dp_{\mathrm{T}jet} dy_{jet}}. \tag{6.2}$$

The scale factors for the various MC's are given in Table 6.1.

In all cases except Herwig++ 2.4.2, the scale factors S for $R = 0.6$ are larger than those for $R = 0.4$. Since the larger radius parameter results in the inclusion of more particles not directly associated with perturbative scattering, this implies that the models underestimate the contribution of the underlying event required to reproduce the data. Phojet has the best agreement between the scale factors at $R = 0.4$ and $R = 0.6$. The Pythia Perugia tunes also agree well for the two radii, and are most consistent with one.

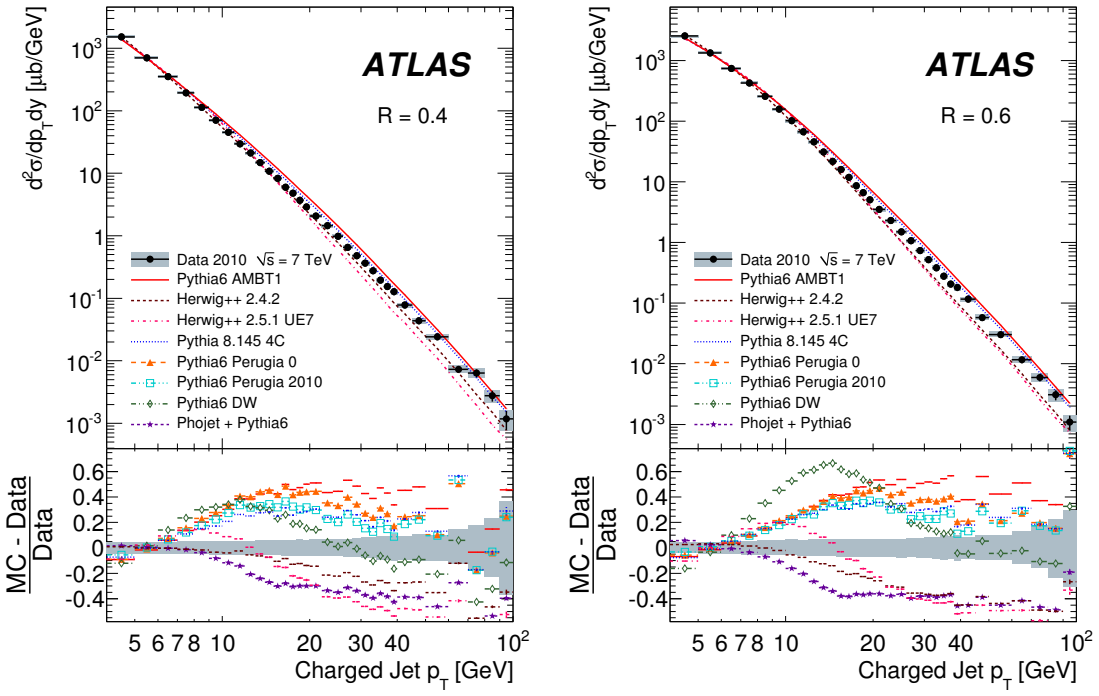
The jet cross section distributions (Fig. 6.1) fall by 6 orders of magnitude between jet momenta of 4 and 100 GeV. The MC models considered agree broadly with this trend, but do not agree well in detail. By construction of the normalization factor

Table 6.1: Scale factors S for Monte Carlo cross section normalization as defined in Equation 6.1. The total Monte Carlo cross sections are normalized to the total for the data, over the full momentum and rapidity ranges investigated in this analysis, $4\,\mathrm{GeV} < p_{\mathrm{T}jet} < 100\,\mathrm{GeV}$ and $|y_{jet}| < 1.9$. The statistical uncertainties on these ratios are less than 0.1%. The systematic uncertainties are $^{+4.9\%}_{-4.8\%}$ ($\pm 5.0\%$) for $R = 0.4$ (0.6); these uncertainties are entirely correlated within columns, and largely correlated between columns.

Tune	S ($R = 0.4$)	S ($R = 0.6$)
AMBT1	0.838	0.896
Perugia-0	0.981	1.087
Perugia HARD	0.936	1.058
Perugia SOFT	0.968	1.036
Perugia 2010	0.976	1.044
DW	0.894	1.045
Herwig++ 2.4.2	0.753	0.612
Herwig++ 2.5.1 UE7	0.425	0.458
Pythia 8	0.777	0.815
Phojet	0.643	0.668

S, all distributions agree with the data in the lowest momentum bins; most also give qualitative agreement for the shape at the lowest momenta. The MC distributions diverge from the data in the 10–20 GeV range, with some having a harder and others a softer momentum dependence,. At higher p_{T}, many of the models' momentum dependence agrees well with the data. If one identifies the higher-momentum region as dominated by perturbative modeling and the low-momentum region as dominated by soft physics, this indicates that perturbative modeling of charged particle jets is in fair agreement for most of the tunes. It is the transition from soft physics to the perturbative region that is not successfully modeled.

The Pythia models give a harder shape for the momentum spectrum than the data below a jet p_{T} of about 20 GeV, after which they exhibit roughly the same momentum dependence or become slightly softer. By contrast, the Phojet and Herwig++ models produce spectra that are softer than the data in the 10–20 GeV range but have relatively good shape agreement outside, although the Herwig++ 2.5.1 UE7 tune has the additional feature of producing too hard a spectrum at momenta below 10 GeV. The DW tune, has a cross section that rises much too rapidly in the transition region,

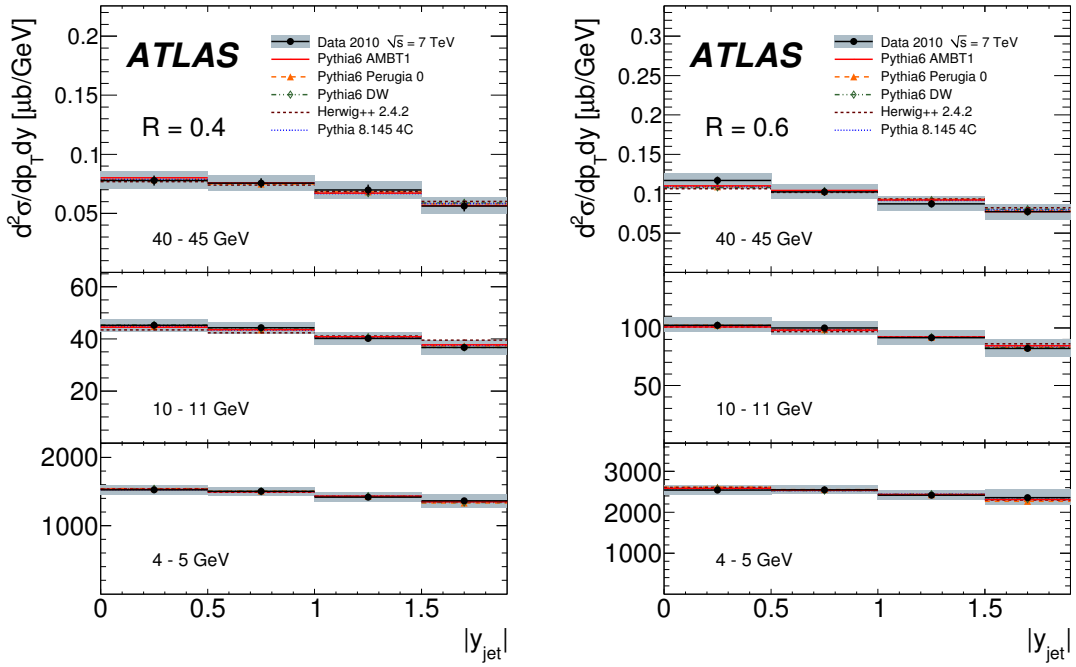


(a) $R = 0.4$ (b) $R = 0.6$

Figure 6.1: The cross section for anti-k_t charged particle jets as a function of p_{T}, with $|y| < 0.5$ and radius parameter R as indicated. The shaded area is the total uncertainty for the corrected data distribution, excluding the overall 3.4% luminosity uncertainty. The data are compared to a range of theoretical results from Monte Carlo event generators, and normalized to the data over the full momentum and rapidity range measured, using the scale factor S as defined in the text. The bottom inserts show the fractional difference between these distributions and the data. The distributions for the Perugia HARD (Perugia SOFT) tune, not shown, agree qualitatively with the Perugia2010 (Perugia 0) tune.

and falls much too rapidly at high momentum; this effect is especially pronounced for $R = 0.6$. This suggests that the Q^2-ordered showering used by the DW tune is less successful in modeling the jet momentum spectrum.

Cross sections as a function of rapidity are shown in Figure 6.2. The MC distributions are normalized to the data in each momentum bin separately rather than to the scale factor S. The rapidity dependence of the cross section shows generally good agreement between data and MC. The cross section decreases only slightly with increasing rapidity at low momenta, but by a somewhat higher amount at higher



(a) $R = 0.4$ (b) $R = 0.6$

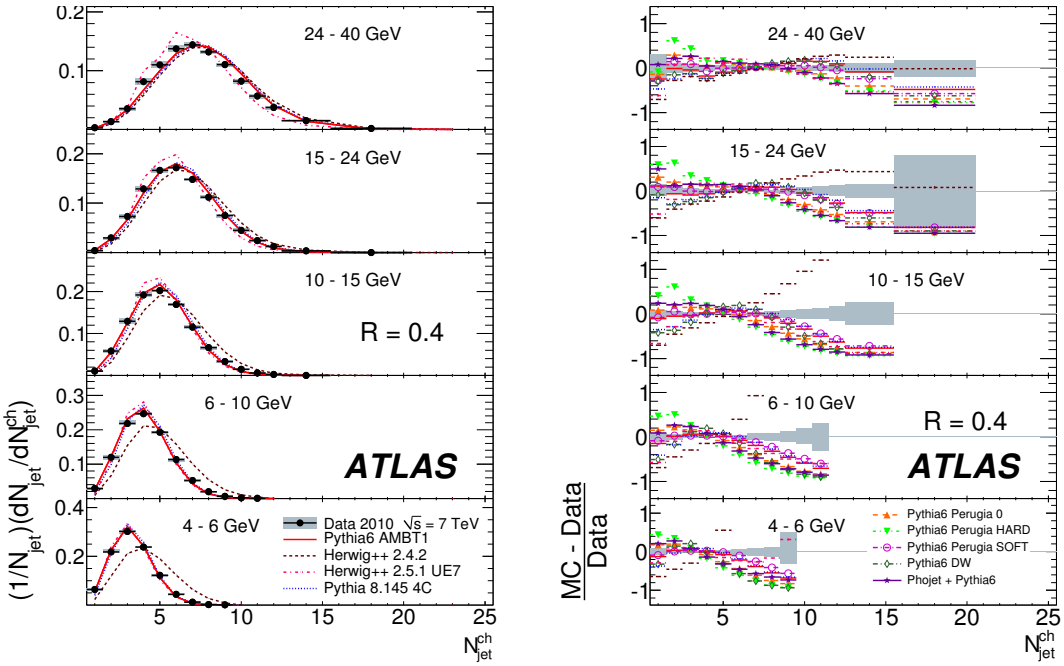
Figure 6.2: The cross section for anti-k_t charged particle jets as a function of rapidity, for selected momentum bins and radius parameter R as indicated. The shaded area is the total uncertainty for the corrected data distribution, excluding the overall 3.4% luminosity uncertainty. The data are compared to a range of theoretical results from Monte Carlo event generators, which are normalized to the data separately for each momentum range. The distributions for all Perugia samples agree qualitatively, so only Perugia 0 is shown. The two Herwig++ tunes agree, so only Herwig++ 2.4.2 is shown.

momenta.

6.2 Charged particle kinematics and multiplicity in jets

The multiplicity of charged tracks per jet, for several momentum ranges over the full range $|y_{jet}| < 1.9$, is shown in Figure 6.3 (6.4) for R = 0.4 (0.6). The charged particle z, p_{T}^{rel}, and $\rho_{ch}(r)$ distributions are shown for same central rapidity range in Figures 6.5 (6.6), 6.7 (6.8), and 6.9 (6.10) respectively for R = 0.4 (0.6).

No tune describes well all of the kinematic distributions and multiplicities of charged particles within jets. For z, the ATLAS AMBT1 and Perugia SOFT give

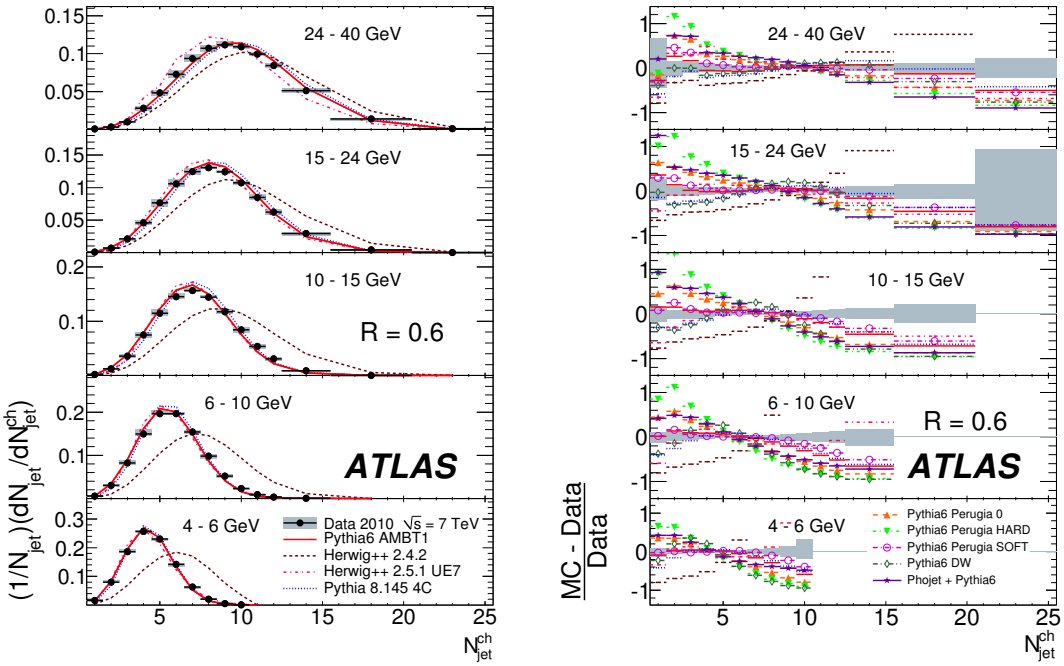


(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure 6.3: Multiplicity of particles per charged particle jet, over the full measured rapidity range $|y| < 1.9$, with anti-k_t radius parameter $R = 0.4$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data. The distributions for the Perugia 2010 tune, not shown, agree qualitatively with the Perugia 0 tune.

good descriptions of the data. For $\rho_{ch}(r)$, AMBT1 gives a good description. For p_{T}^{rel} and multiplicity, no tune correctly describes all data. For all distributions, Herwig++ 2.4.2 shows strong disagreement with the data, characterized by an excess of low-momentum particles, which is especially pronounced for the larger jet-finding parameter R and at large particle r (as defined in Equation 4.7). Herwig++ 2.5.1 UE 7000 represents a significant improvement, so Herwig++ 2.4.2 will not be discussed further.

All models give good agreement for the average charged particle multiplicity per jet. The AMBT1 and Perugia SOFT tunes agree well with the multiplicity distributions (Figures 6.3 and 6.4) for the vast majority of jets, and Perugia 0, Perugia 2010, and Phojet give fair agreement for $R = 0.4$, although the high-multiplicity tail in data is greatly underestimated by all models. Herwig++ 2.5.1 UE7 and Perugia

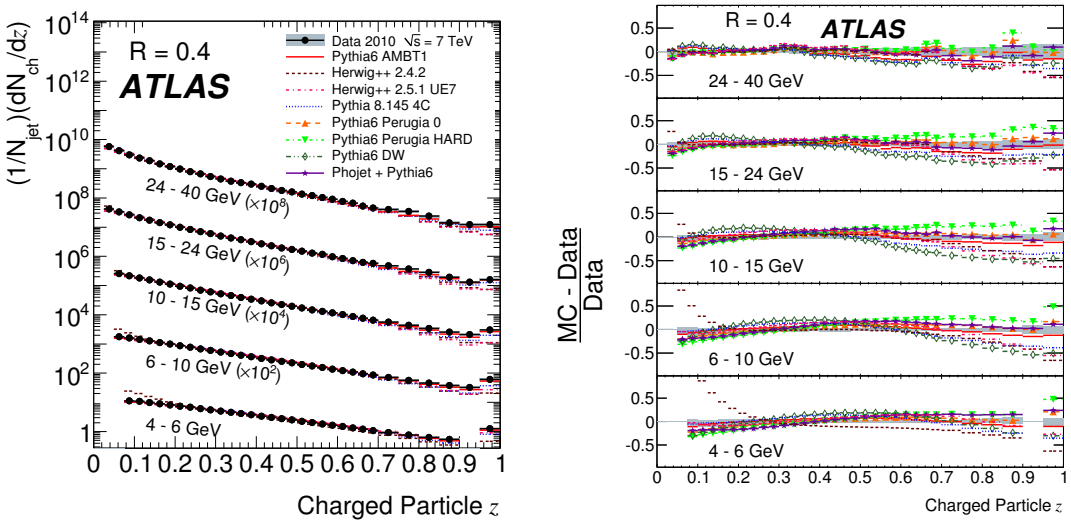


(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure 6.4: Multiplicity of particles per charged particle jet, over the full measured rapidity range $|y| < 1.9$, with anti-k_t radius parameter $R = 0.6$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data. The distributions for the Perugia 2010 tune, not shown, agree qualitatively with the Perugia 0 tune.

HARD have a significant excess of low-multiplicity jets, while Pythia 8.145 4C and Pythia DW exhibit a deficit.

The AMBT1 and Perugia SOFT tunes give good agreement with the measured longitudinal momentum fraction z (Figures 6.5 and 6.6); Perugia-0 also agrees well for $R = 0.4$. The other MC's (except Herwig++ 2.4.2) agree within 30% at low z, but diverge more significantly at high z. Perugia HARD has the most significant excess of high-z particles, with excesses also present for Phojet and Perugia 0. The excess is particularly large at lower jet momenta and $R = 0.6$, suggesting that the soft physics model is characterized by fewer particles with higher momentum. By contrast, Pythia 8.145 4C, Pythia DW, and Herwig++ 2.5.1 UE 7000 have too few high-z particles, with variations again larger for $R = 0.6$. Pythia DW exhibits an excess at mid-z at low jet momenta, which is seen at progressively lower z values as the jet momentum



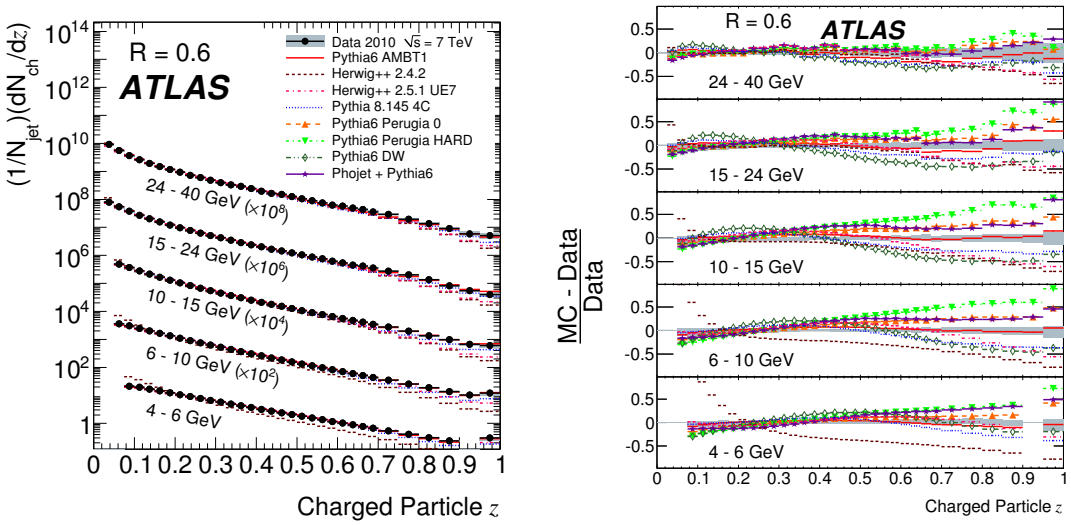
(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure 6.5: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter $R = 0.4$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data. The distributions for the Perugia 2010 (Perugia SOFT) tune, not shown, agree qualitatively with the Perugia 0 (AMBT1) tune.

increases, implying an excess of particles with a momentum of roughly 2 GeV that is not associated with jet structure.

The transverse momentum p_{T}^{rel} (Figures 6.7 and 6.8) is in fair agreement ($\sim$20%) at low-to-moderate values for all MC generators except Herwig++ 2.4.2. At the lowest jet momenta and highest measurable p_{T}^{rel}, Phojet and Herwig++ 2.5.1 UE7 have an excess of particles, while Pythia DW has a deficit. At higher jet momenta, the data have more high-p_{T}^{rel} particles than any tune, with Perugia 2010 and Perugia HARD giving the closest description and Perugia SOFT the furthest. Perugia 2010 and Perugia HARD agree better than do the other Perugia tunes.

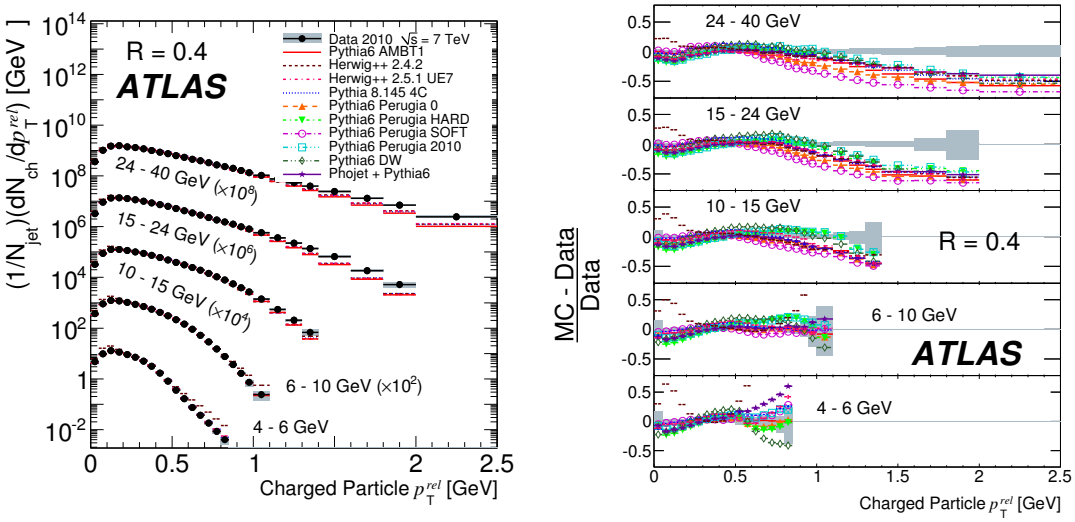
The AMBT1, Pythia 8.145 4C, and Herwig++ 2.5.1 UE7 tunes provide a good description of the charged particle number density $\rho_{ch}(r)$ (Figures 6.9 and 6.10) at all radii. Phojet and the Perugia tunes (especially SOFT) have an excess of particles very close to the jet axis, which is most pronounced at high jet momentum and for $R = 0.6$; Perugia 2010 agrees better in this region than do the other Perugia tunes. At high r, Phojet and all Pythia tunes except AMBT1 and Perugia SOFT have too



(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure 6.6: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter $R = 0.6$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data. The distributions for the Perugia 2010 (Perugia SOFT) tune, not shown, agree qualitatively with the Perugia 0 (AMBT1) tune.

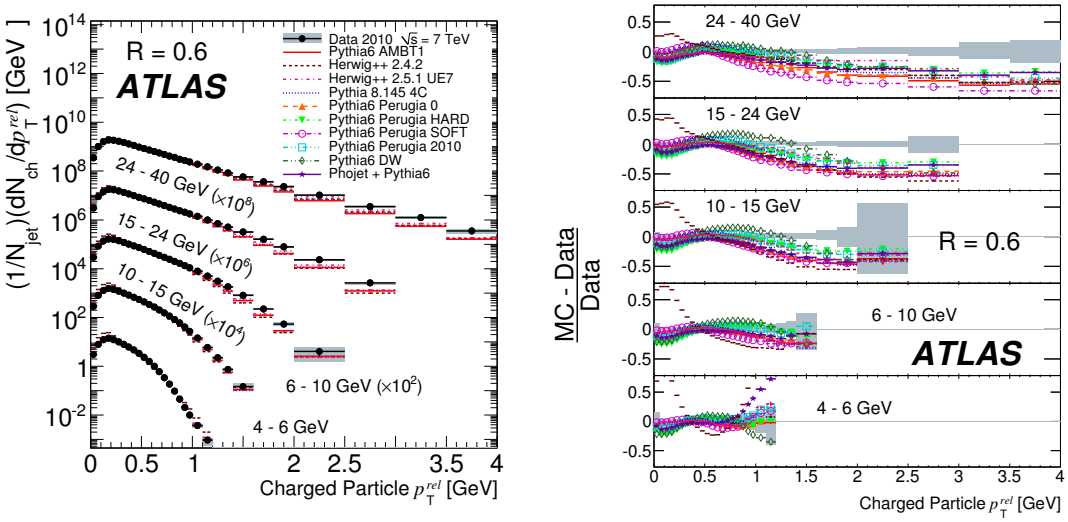
few particles. However, the disagreement is less pronounced than is seen at high p_{T}^{rel}, implying that high-radius soft particles from the underlying event are better-described than high-radius hard radiation.



(a) $R = 0.4$ Distribution

(b) $R = 0.4$ Data-MC Difference

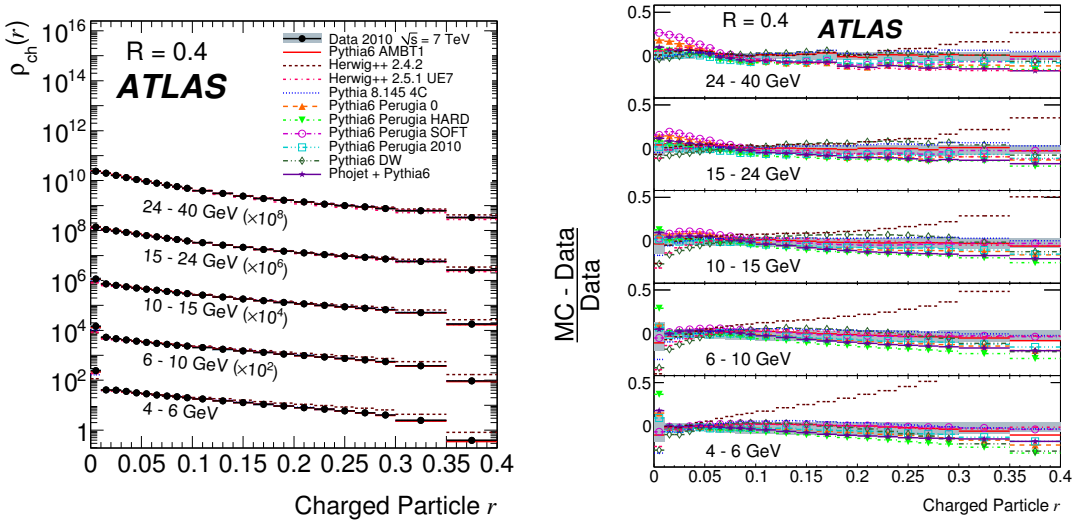
Figure 6.7: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter $R = 0.4$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



(a) $R = 0.6$ Distribution

(b) $R = 0.6$ Data-MC Difference

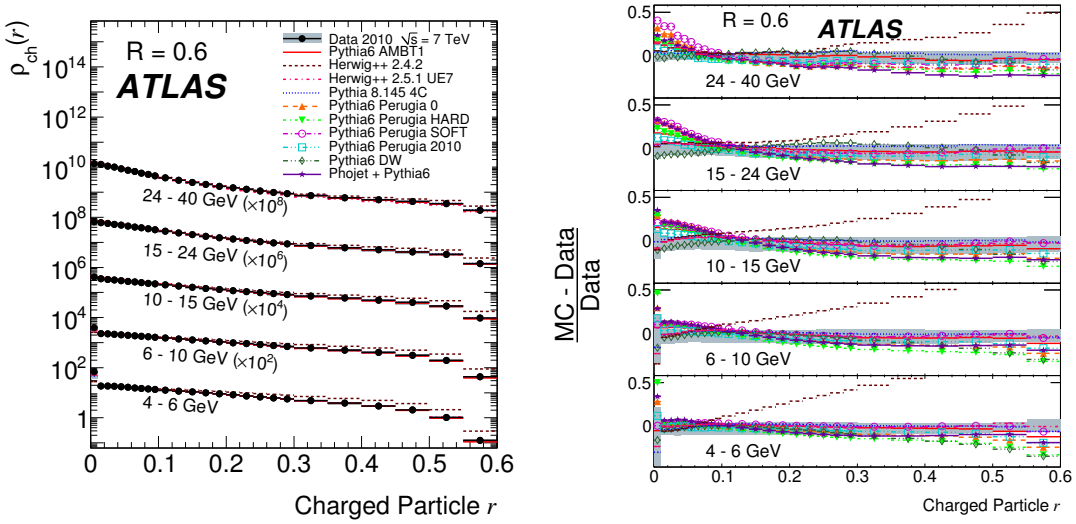
Figure 6.8: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter $R = 0.6$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



(a) $R = 0.4$ Distribution

(b) $R = 0.4$ Data-MC Difference

Figure 6.9: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter $R = 0.4$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure 6.10: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter $R = 0.6$, in the rapidity range $|y| < 1.9$. figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.

Chapter 7

Conclusions

A measurement is presented of the charged particle jet cross section as a function of transverse momentum and rapidity, along with the transverse momentum, longitudinal momentum fraction, and number density as a function of radius for charged particles within these jets, with the ATLAS detector using early 7 TeV LHC collisions. The results for all distributions are in qualitative agreement with most of the theoretical models considered, but differences are quite large for some distributions. No tune presented here agrees with all quantities measured within their uncertainties, suggesting that future MC tunes may be improved.

Difficulty in modeling the transition between soft and perturbative physics is indicated by disagreements between data and all MC distributions in the 10-20 GeV range in the dependence of the charged jet cross section on jet momentum. Dependence of the cross section on rapidity is consistent with predictions. Particles with large transverse momentum p_{T}^{rel} with respect to the jet that contains them are produced more copiously than any model predicts, as are jets with large charged particle multiplicity. The longitudinal momentum fraction z is best described by the Pythia 6.421 AMBT1 tune. The charged particle number density $\rho_{ch}(r)$ is well-described by the Pythia 6.421 AMBT1, Pythia 8.145 4C, and Herwig++ 2.5.1 UE7 tunes. With the exception of the Herwig++ 2.4.2 default tune, which greatly disagrees with these measurements, all models appear to underestimate the contribution of the underlying event required to model the data.

The study of jets with tracks allows for precise measurements of low-momentum jets and their properties, thus complementing calorimeter-based jet measurements and allowing the study of the transition from soft collisions to jet production in the perturbative regime of QCD. It also provides additional observables for consideration in the tuning of MC event generators, which complement existing studies such as "minimum bias" and underlying event measurements. These results may thus be used to derive better MC tunes and models, which will in turn lead to a better

understanding of QCD at the Large Hadron Collider. These improvements will feed back into models of pileup and other backgrounds, thereby improving future searches and discoveries at the LHC.

Bibliography

[1] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[2] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories," *Phys. Rev. Lett.*, vol. 30, pp. 1343–1346, 1973.

[3] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?," *Phys. Rev. Lett.*, vol. 30, pp. 1346–1349, 1973.

[4] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[5] R. Adolphi *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[6] G. Aad *et al.*, "Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC," *Phys.Rev.Lett.*, vol. 105, p. 161801, 2010.

[7] G. Aad *et al.*, "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC," *New J.Phys.*, 2010.

[8] V. Khachatryan *et al.*, "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV," *Phys. Rev. Lett.*, vol. 105, p. 022002, 2010.

[9] G. Aad *et al.*, "Measurements of underlying-event properties using neutral and charged particles in *pp* collisions at 900 GeV and 7 TeV with the ATLAS detector at the LHC," *Eur. Phys. J.*, vol. C71, p. 1636, 2011.

[10] S. Chatrchyan *et al.*, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s}$= 7 TeV and Comparison with $\sqrt{s} = 0.9$ TeV," 2011.

[11] G. Aad *et al.*, "Study of Jet Shapes in Inclusive Jet Production in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector," *Phys. Rev.*, vol. D83, p. 052003, 2011.

[12] G. Aad *et al.*, "Measurement of Dijet Azimuthal Decorrelations in pp Collisions at $\sqrt{s}$=7 TeV," *Phys. Rev. Lett.*, vol. 106, p. 172002, 2011.

[13] S. Chatrchyan *et al.*, "Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s}$=7 TeV," *Phys. Lett.*, vol. B700, pp. 187–206, 2011.

[14] ATLAS Collaboration, "New ATLAS event generator tunes to 2010 data," Tech. Rep. ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011.

[15] A. G. Clark *et al.*, "Large transverse momentum jets in high-energy proton-proton collisions," *Nucl. Phys.*, vol. B160, p. 397, 1979.

[16] G. Arnison *et al.*, "First observation of correlations between high transverse momentum charged particles in events from the CERN proton-antiproton collider," *Phys. Lett.*, vol. B118, pp. 173–177, 1982.

[17] A. A. Affolder *et al.*, "Charged jet evolution and the underlying event in proton - anti-proton collisions at 1.8-TeV," *Phys. Rev.*, vol. D65, p. 092002, 2002.

[18] G. Arnison *et al.*, "Jet fragmentation into charged particles at the CERN proton-antiproton collider," *Phys. Lett.*, vol. B132, p. 223, 1983.

[19] D. E. Acosta *et al.*, "Momentum distribution of charged particles in jets in dijet events in p anti-p collisions at $\sqrt{s}$ = 1.8-TeV and comparisons to perturbative QCD predictions," *Phys. Rev.*, vol. D68, p. 012003, 2003.

[20] F. J. Dyson, "The radiation theories of tomonaga, schwinger, and feynman," *Phys. Rev.*, vol. 75, pp. 486–502, Feb 1949.

[21] F. J. Dyson, "The s matrix in quantum electrodynamics," *Phys. Rev.*, vol. 75, pp. 1736–1755, Jun 1949.

[22] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, pp. 579–588, 1961.

[23] A. Salam and J. C. Ward, "Electromagnetic and weak interactions," *Phys. Lett.*, vol. 13, pp. 168–171, 1964.

[24] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967.

[25] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–322, 1964.

[26] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.

[27] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, 1964.

[28] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry," *Phys. Rev.*, vol. D2, pp. 1285–1292, 1970.

[29] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge University Press.

[30] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory," Reading, USA: Addison-Wesley (1995).

[31] K. Nakamura *et al.*, "Review of particle physics," *J. Phys.*, vol. G37, p. 075021, 2010.

[32] T. Regge *Il Nuovo Cimento*, vol. 14, p. 951, 1959.

[33] G. F. Chew and A. Pignotti, "Multiperipheral Bootstrap Model," *Phys. Rev.*, vol. 176, pp. 2112–2119, 1968.

[34] A. Capella, U. Sukhatme, C.-I. Tan, and J. Tran Thanh Van, "Dual parton model," *Phys. Rept.*, vol. 236, pp. 225–329, 1994.

[35] R. P. Feynman, "Very high-energy collisions of hadrons," *Phys. Rev. Lett.*, vol. 23, pp. 1415–1417, 1969.

[36] "Online PDF plotting and calculation." `http://hepdata.cedar.ac.uk/pdf/pdf3.html`.

[37] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972.

[38] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.*, vol. B126, p. 298, 1977.

[39] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics," *Sov. Phys. JETP*, vol. 46, pp. 641–653, 1977.

[40] X. Artru and G. Mennessier, "String model and multiproduction," *Nucl. Phys.*, vol. B70, pp. 93–115, 1974.

[41] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference," *Nucl. Phys.*, vol. B238, p. 492, 1984.

[42] G. 't Hooft, "A Planar Diagram Theory for Strong Interactions," *Nucl. Phys.*, vol. B72, p. 461, 1974.

[43] D. P. Barber *et al.*, "Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies," *Phys. Rev. Lett.*, vol. 43, p. 830, 1979.

[44] C. Berger *et al.*, "Evidence for Gluon Bremsstrahlung in e+ e- Annihilations at High-Energies," *Phys. Lett.*, vol. B86, p. 418, 1979.

[45] M. Sandhoff and P. Z. Skands, "Colour annealing - a toy model of colour reconnections," Tech. Rep. FERMILAB-CONF-05-518-T. Presented at Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 2-20 May 2005.

[46] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J.*, vol. C58, pp. 639–707, 2008.

[47] L. Durand and P. Hong, "QCD and Rising Total Cross-Sections," *Phys. Rev. Lett.*, vol. 58, pp. 303–306, 1987.

[48] L. Durand and H. Pi, "Semihard QCD and High-Energy p p and anti-p p Scattering," *Phys. Rev.*, vol. D40, p. 1436, 1989.

[49] G. J. Alner *et al.*, "The UA5 High-Energy anti-p p Simulation Program," *Nucl. Phys.*, vol. B291, p. 445, 1987.

[50] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[51] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev.*, vol. D82, p. 074018, 2010.

[52] H. L. Lai *et al.*, "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions," *Eur. Phys. J.*, vol. C12, pp. 375–392, 2000.

[53] A. A. Affolder *et al.*, "The Transverse momentum and total cross-section of e+ e- pairs in the Z-boson region from p anti-p collisions at $\sqrt{s}$ = 1.8 TeV," *Phys. Rev. Lett.*, vol. 84, pp. 845–850, 2000.

[54] V. M. Abazov *et al.*, "Measurement of the shape of the boson transverse momentum distribution in p anti-p $\rightarrow$ Z / γ^* $\rightarrow$ e+ e- + X events produced at $\sqrt{s}$ = 1.96 TeV," *Phys. Rev. Lett.*, vol. 100, p. 102002, 2008.

[55] P. Z. Skands, "Peter's Pythia Plots." `http://home.fnal.gov/~skands/leshouches-plots/LHPlotsFrontpage.html`.

[56] ATLAS Collaboration, "ATLAS Monte Carlo tunes for MC09," Tech. Rep. ATL-PHYS-PUB-2010-002, CERN, Geneva, Mar 2010.

[57] ATLAS Collaboration, "Charged particle multiplicities in pp interactions at $\sqrt{s}$ = 0.9 and 7 TeV in a diffractive limited phase-space measured with the ATLAS detector at the LHC and new PYTHIA6 tune," Tech. Rep. ATLAS-CONF-2010-031, CERN, Geneva, 2010.

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

[59] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[60] R. Corke and T. Sjostrand, "Interleaved Parton Showers and Tuning Prospects," *JHEP*, vol. 03, p. 032, 2011.

[61] S. Gieseke, D. Grellscheid, K. Hamilton, A. Papaefstathiou, S. Platzer, *et al.*, "Herwig++ 2.5 Release Note," 2011.

[62] S. Gieseke, D. Grellscheid, K. Hamilton, A. Papaefstathiou, S. Platzer, *et al.*, "Minimum Bias (MB) and Underlying Event (UE), 900GeV and 7TeV." `http://projects.hepforge.org/herwig/trac/wiki/MB_UE_tunes`.

[63] R. Engel and J. Ranft, "Hadronic photon-photon interactions at high-energies," *Phys. Rev.*, vol. D54, pp. 4244–4262, 1996.

[64] R. Engel and J. Ranft, "Color singlet exchange between jets and the PHOJET Monte Carlo," *Nucl. Phys. Proc. Suppl.*, vol. 75A, pp. 272–274, 1999.

[65] T. Sjostrand *et al.*, "High-energy physics event generation with PYTHIA 6.1," *Comput. Phys. Commun.*, vol. 135, pp. 238–259, 2001.

[66] P. Fayet, "Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions," *Phys. Lett.*, vol. B69, p. 489, 1977.

[67] K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[68] A. A. Alves *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[69] The LEP Injector Study Group, "LEP Design Report: VOL. 2. The LEP Main Ring," Tech. Rep. CERN-LEP-84-01, CERN, Geneva, 1984.

[70] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, "LHC Design Report. 3. The LHC injector chain," Tech. Rep. CERN- 2004-003-V-3, CERN, Geneva, 2004.

[71] "Canada makes a TRIUMFant contribution." CERN Courier, Geneva, `http://cerncourier.com/cws/article/cern/35872/2`, April 1998.

[72] G. Arnison *et al.*, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV," *Phys.Lett.*, vol. B122, pp. 103–116, 1983.

[73] G. Arnison *et al.*, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider," *Phys.Lett.*, vol. B126, pp. 398–410, 1983.

[74] "CERN website, Hardware Commissioning Coordination." `http://hcc.web.cern.ch/HCC/nonconform.php`.

[75] G. Aad *et al.*, "Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector at the LHC," 2011.

[76] T. Cornelissen *et al.*, "The new ATLAS track reconstruction (NEWT)," *J. Phys. Conf. Ser.*, vol. 119, p. 032014, 2008.

[77] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl. Instrum. Meth.*, vol. A262, pp. 444–450, 1987.

[78] D. Wicke, "A new algorithm for solving tracking ambiguities," Tech. Rep. LC-TOOL-1999-007-TESLA, 1999.

[79] "Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment," Tech. Rep. ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010.

[80] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting," *J. Phys.*, vol. G34, p. N343, 2007.

[81] "Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC," Tech. Rep. ATLAS-CONF-2010-027, CERN, Geneva, May 2010.

[82] G. Piacquadio, K. Prokofiev, and A. Wildauer, "Primary vertex reconstruction in the ATLAS experiment at LHC," *J. Phys. Conf. Ser.*, vol. 119, p. 032033, 2008.

[83] ATLAS Collaboration, "Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC," Tech. Rep. ATLAS-CONF-2010-027, CERN, Geneva, 2010.

[84] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[85] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev.*, vol. D48, pp. 3160–3166, 1993.

[86] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions," *Nucl. Phys.*, vol. B406, pp. 187–224, 1993.

[87] G. Aad *et al.*, "The ATLAS Simulation Infrastructure," *Eur.Phys.J.*, vol. C70, pp. 823–874, 2010.

[88] S. Agostinelli *et al.*, "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[89] G. D'Agostini, "A Multidimensional unfolding method based on Bayes' theorem," *Nucl. Instrum. Meth.*, vol. A362, pp. 487–498, 1995.

[90] T. Ayde, "RooUnfold webpage." `http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html`.

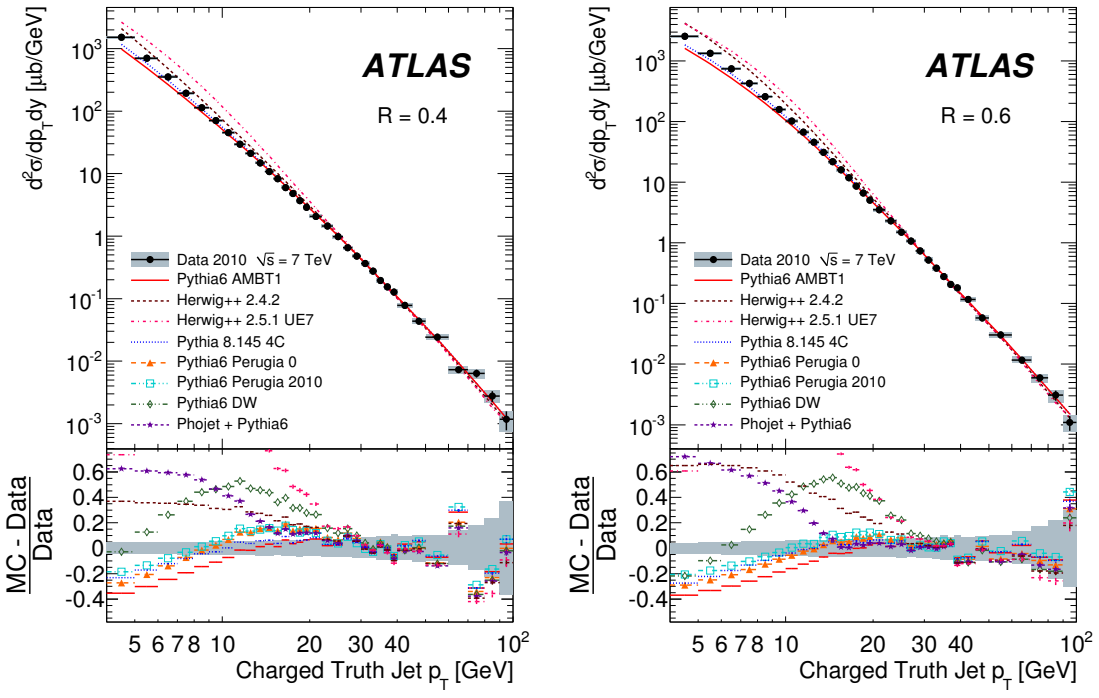
[91] "Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010," Tech. Rep. ATLAS-CONF-2011-046, CERN, Geneva, Mar 2011.

[92] "The Durham HepData Project." `http://hepdata.cedar.ac.uk`.

Appendix A

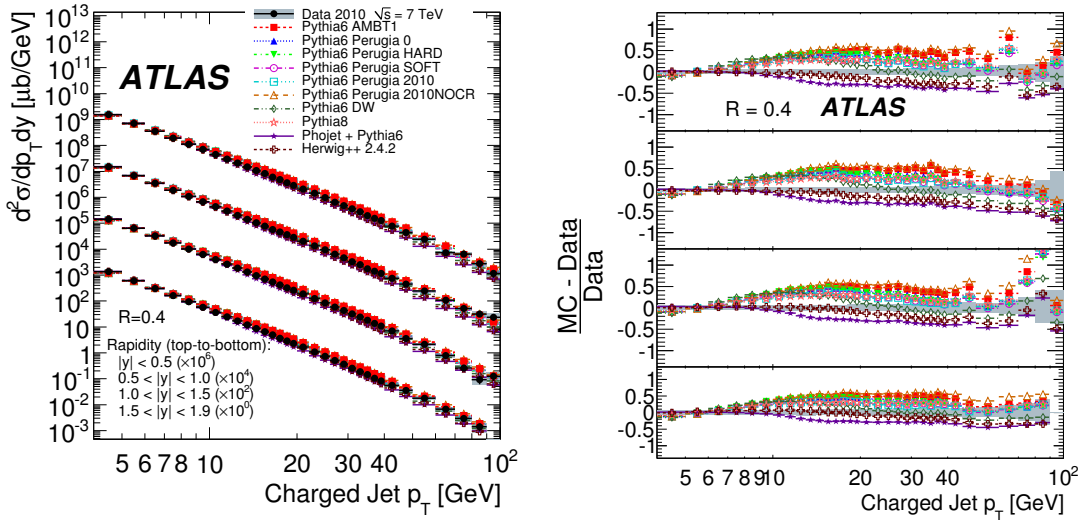
Plots of Additional Results

In this Appendix, additional plots analagous to those appearing in Chapter 6 are presented, but for all individual rapidity ranges. They do not suggest additional qualitative conclusions, but the measurements for all bins will be made available [92] for the purposes of Monte Carlo tuning.



(a) R = 0.4 (b) R = 0.6

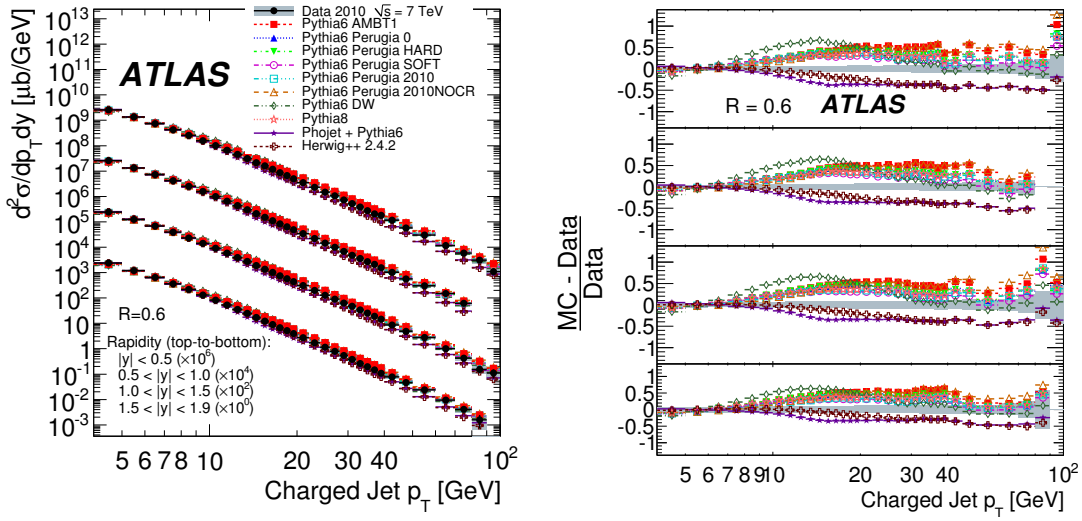
Figure A.1: The cross section for anti-k_t charged particle jets as a function of p_{T}, with $|y| < 0.5$ and radius parameter R as indicated. The shaded area is the total uncertainty for the corrected data distribution, excluding the overall 3.4% luminosity uncertainty. The data are compared to a range of theoretical results from Monte Carlo event generators, which are normalized to the data over the full momentum and rapidity range measured, using the a scale factor analagous to the scale factor S from the text but normalizing only to the p_{T} range 30–100 GeV. The bottom inserts show the fractional difference between these distributions and the data. The distributions for the Perugia HARD (Perugia SOFT) tune, not shown, agree qualitatively with the Perugia2010 (Perugia 0) tune.



(a) $R = 0.4$ Cross Section

(b) $R = 0.4$ Fractional difference

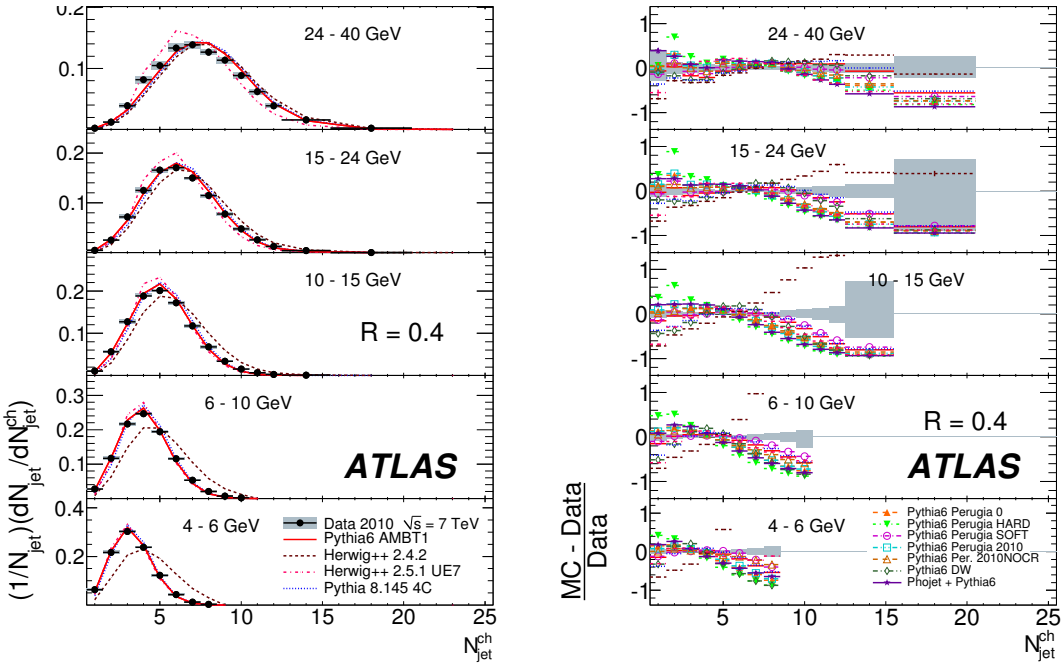
Figure A.2: The cross section for anti-k_t charged particle jets as a function of p_{T}, with rapidity range and radius parameter $R = 0.4$. The shaded area is the total uncertainty for the corrected data distribution, excluding the overall 3.4% luminosity uncertainty. The data are compared to a range of theoretical results from Monte Carlo event generators, and normalized to the data over the full momentum and rapidity range measured, using the scale factor S as defined in the text. The section at right shows the fractional difference between these distributions and the data.



(a) $R = 0.6$ Cross Section

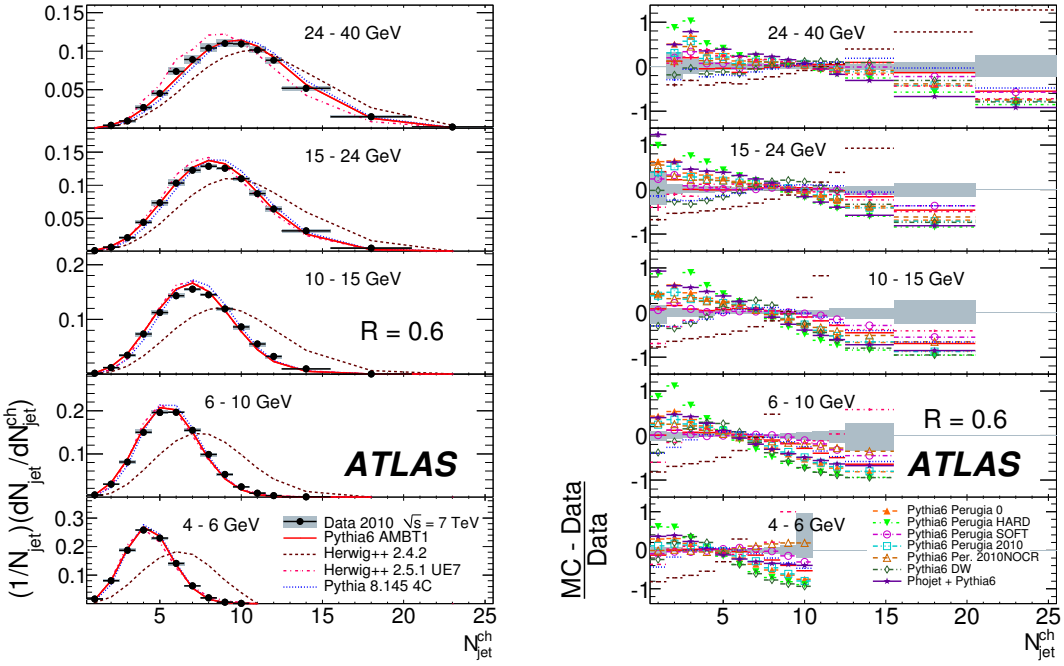
(b) $R = 0.6$ Fractional difference

Figure A.3: The cross section for anti-k_t charged particle jets as a function of p_{T}, with rapidity range and radius parameter $R = 0.6$. The shaded area is the total uncertainty for the corrected data distribution, excluding the overall 3.4% luminosity uncertainty. The data are compared to a range of theoretical results from Monte Carlo event generators, and normalized to the data over the full momentum and rapidity range measured, using the scale factor S as defined in the text. The section at right shows the fractional difference between these distributions and the data.



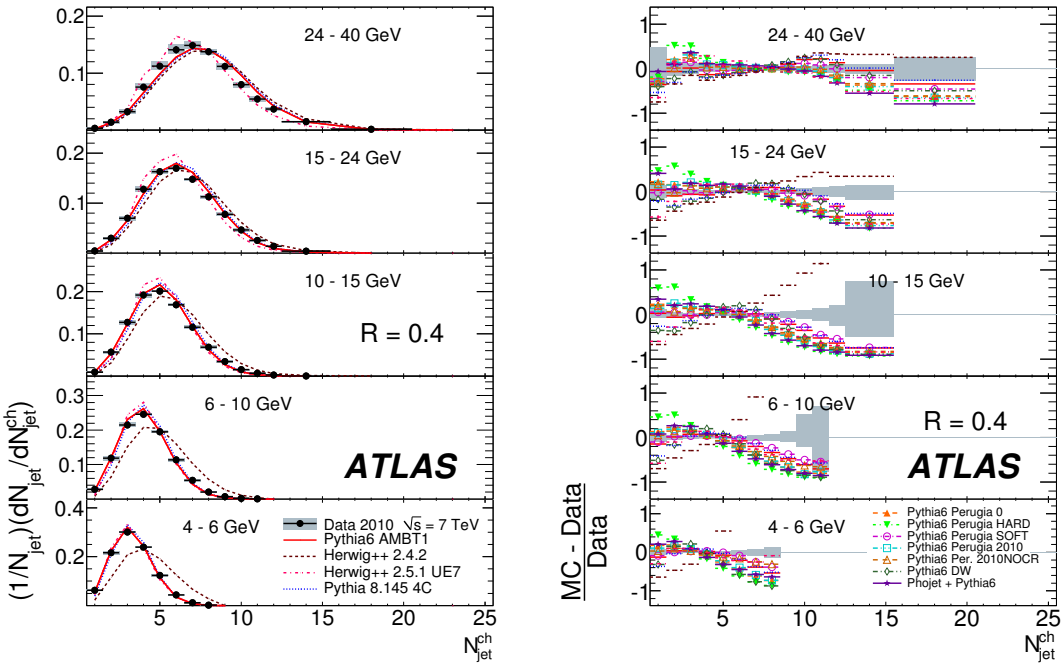
(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure A.4: Multiplicity of particles per charged particle jet, over the rapidity range $|y| < 0.5$, with anti-k_t radius parameter $R = 0.4$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



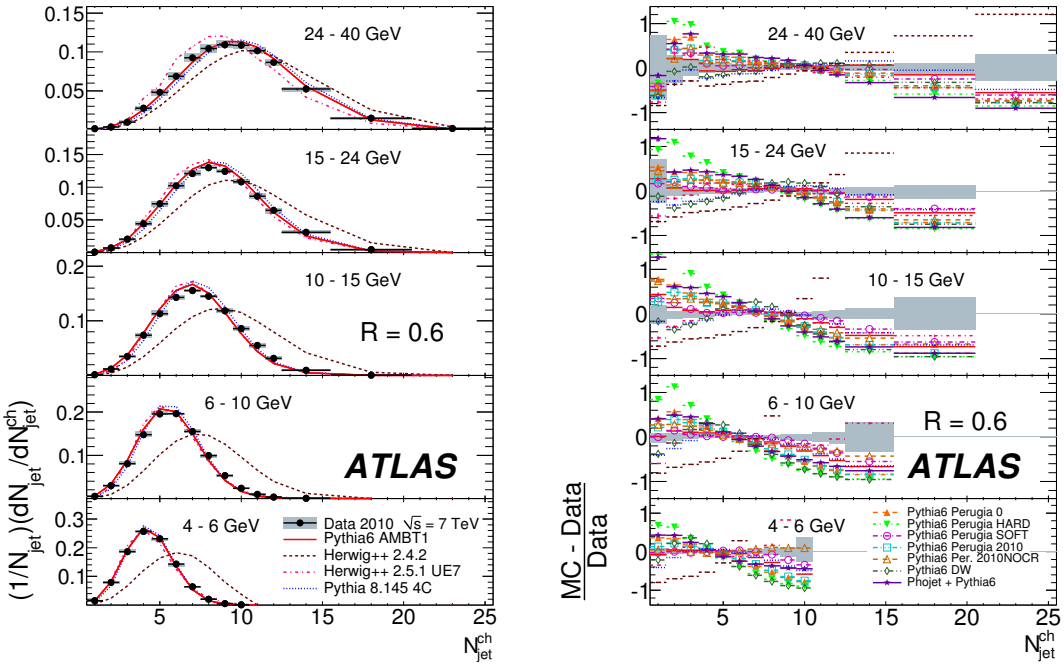
(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure A.5: Multiplicity of particles per charged particle jet, over the rapidity range $|y| < 0.5$, with anti-k_t radius parameter $R = 0.6$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



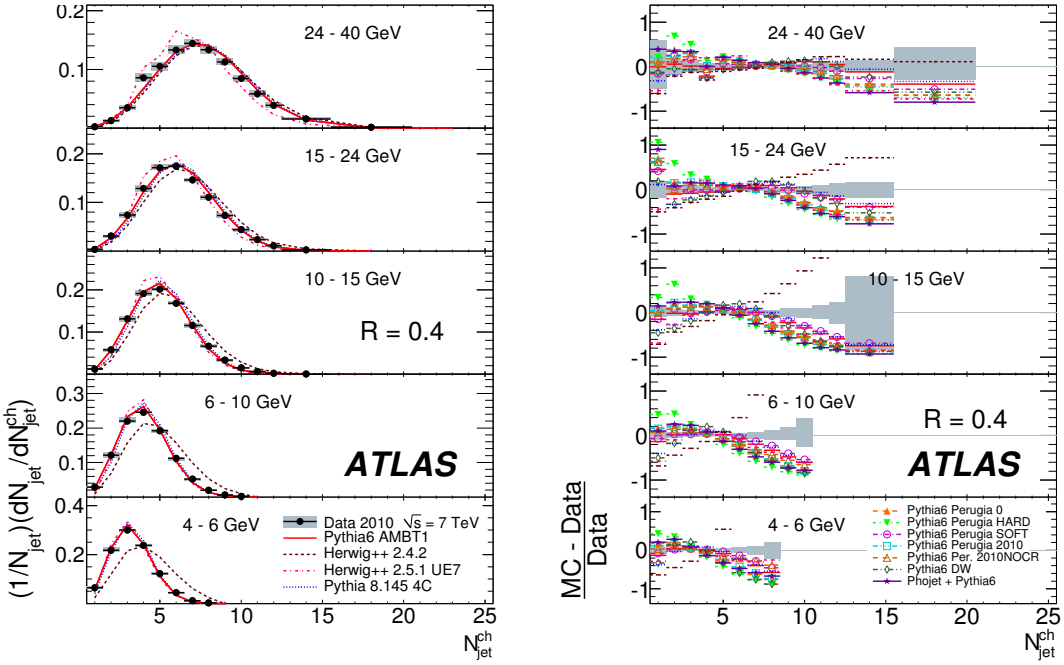
(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure A.6: Multiplicity of particles per charged particle jet, over the rapidity range $0.5 < |y| < 1.0$, with anti-k_t radius parameter $R = 0.4$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



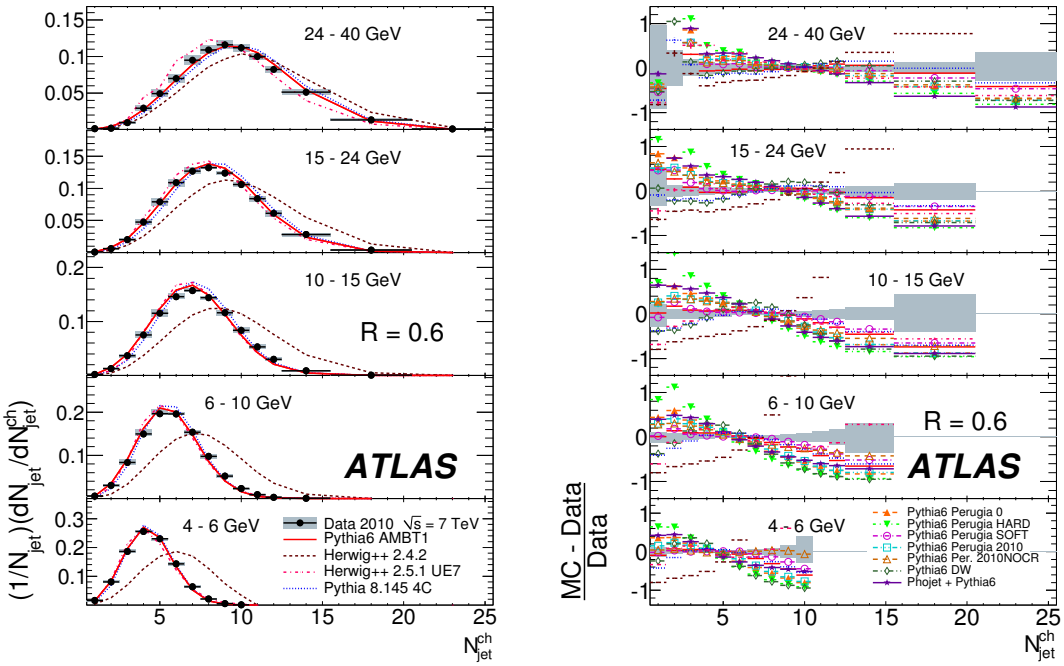
(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure A.7: Multiplicity of particles per charged particle jet, over the rapidity range $0.5 < |y| < 1.0$, with anti-k_t radius parameter $R = 0.6$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



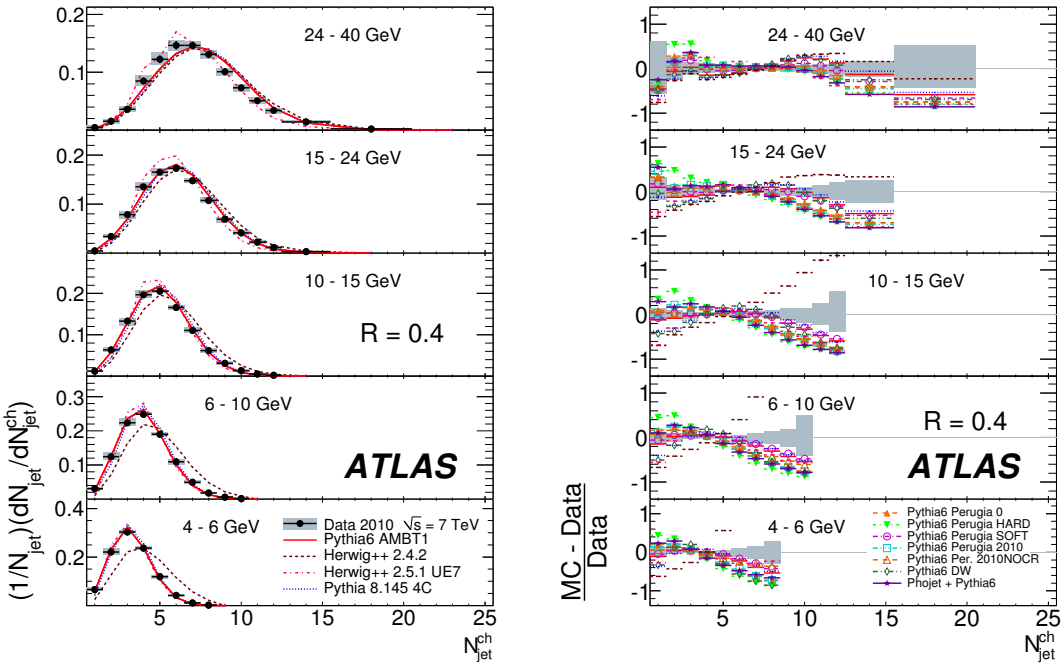
(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure A.8: Multiplicity of particles per charged particle jet, over the rapidity range $1.0 < |y| < 1.5$, with anti-k_t radius parameter $R = 0.4$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



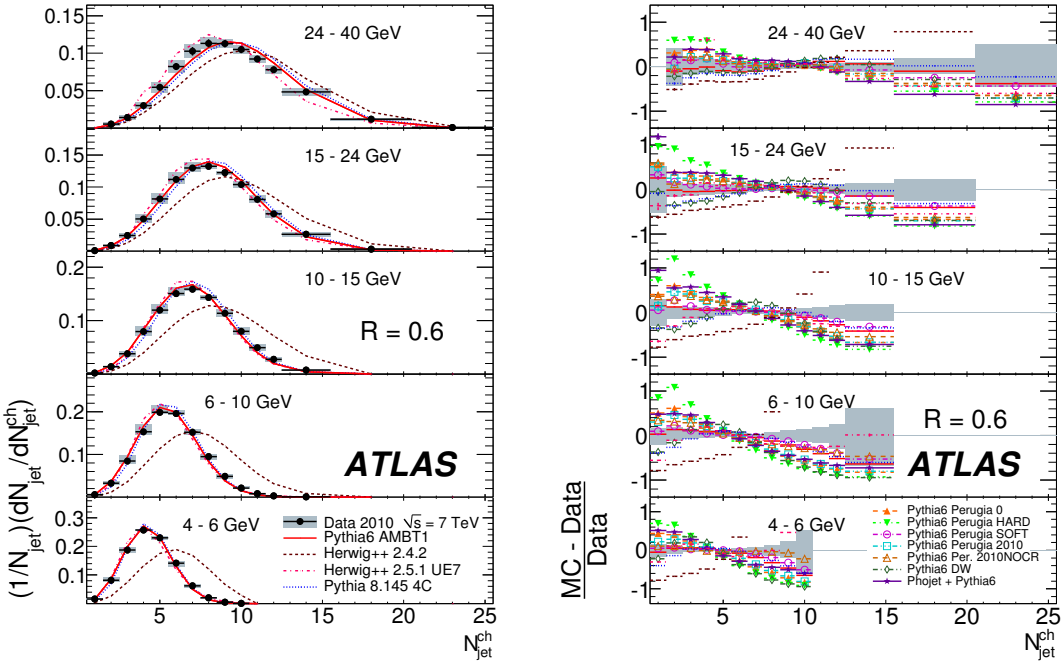
(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

Figure A.9: Multiplicity of particles per charged particle jet, over the rapidity range $1.0 < |y| < 1.5$, with anti-k_t radius parameter $R = 0.6$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



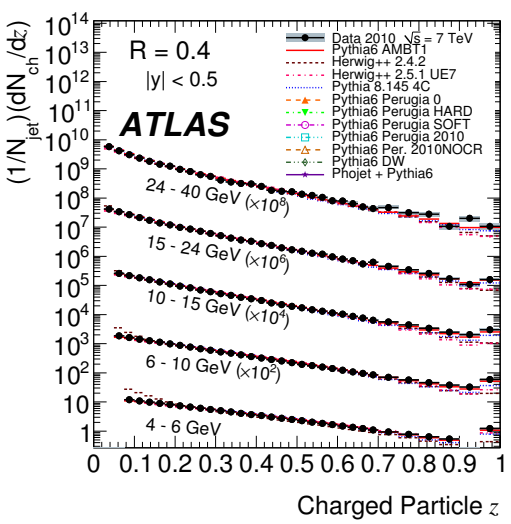
(a) $R = 0.4$ Distribution (b) $R = 0.4$ Data-MC Difference

Figure A.10: Multiplicity of particles per charged particle jet, over the rapidity range $1.5 < |y| < 1.9$, with anti-k_t radius parameter $R = 0.4$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.

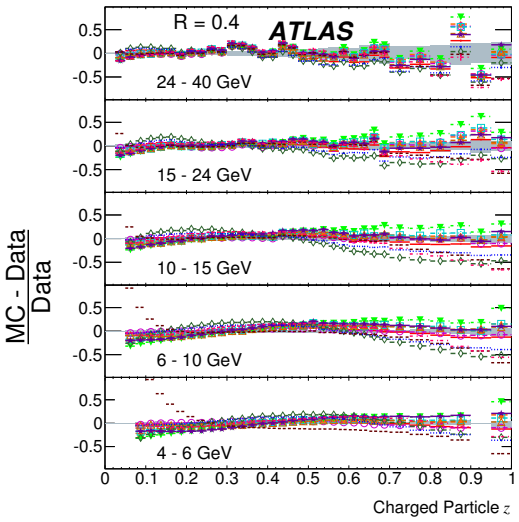


(a) $R = 0.6$ Distribution (b) $R = 0.6$ Data-MC Difference

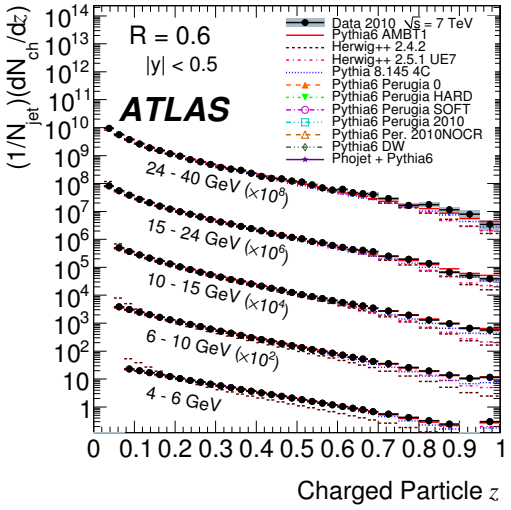
Figure A.11: Multiplicity of particles per charged particle jet, over the rapidity range $1.5 < |y| < 1.9$, with anti-k_t radius parameter $R = 0.6$. Figure (a) shows the distributions for five momentum ranges, and Figure (b) shows the fractional difference between a range of Monte Carlo event generator predictions and the data.



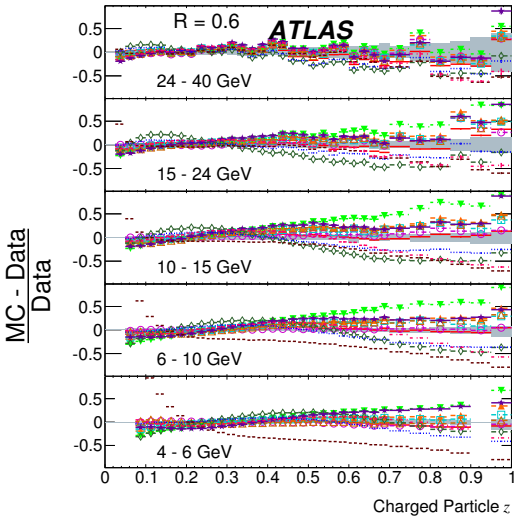
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

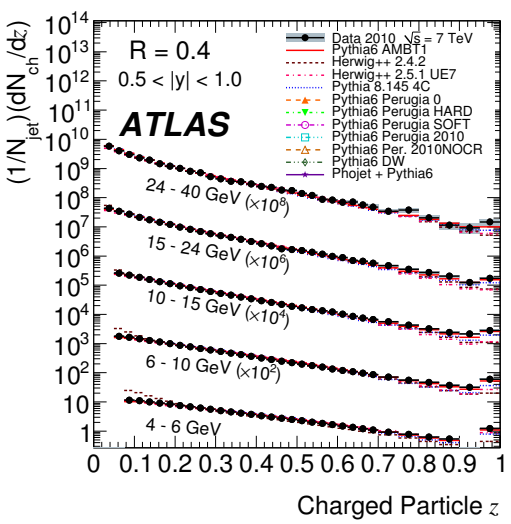


(c) $R = 0.6$ Distribution

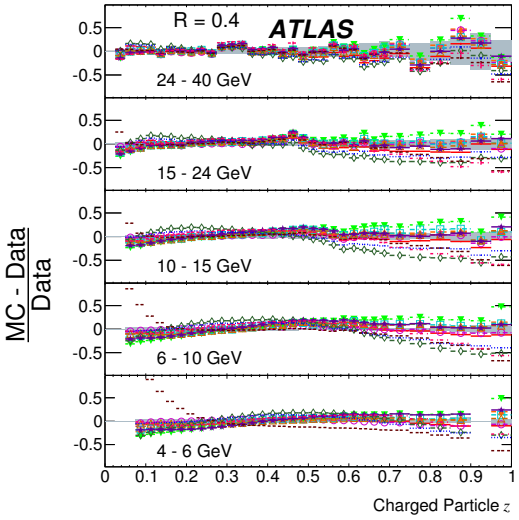


(d) $R = 0.6$ Data-MC Difference

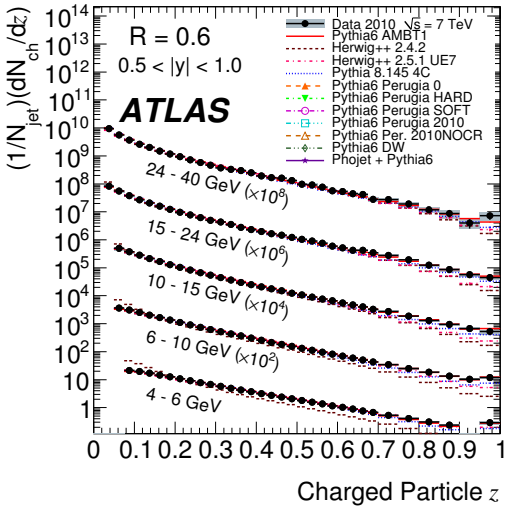
Figure A.12: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter R as indicated, in the rapidity range $|y| < 0.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



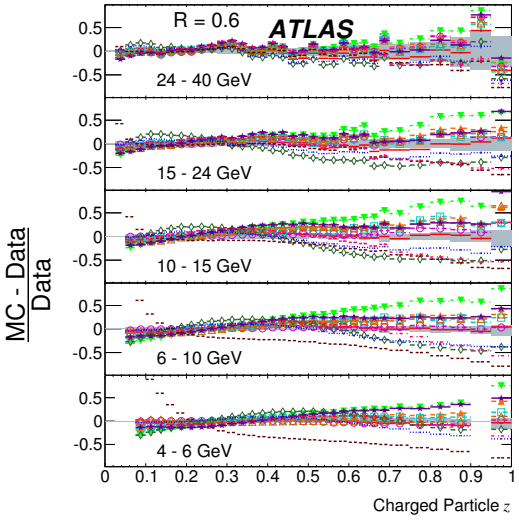
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

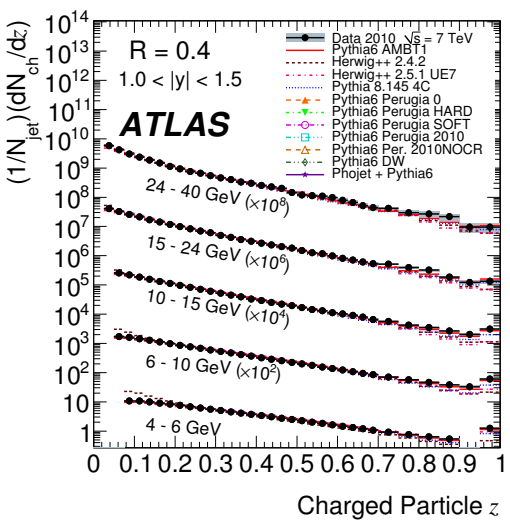


(c) $R = 0.6$ Distribution

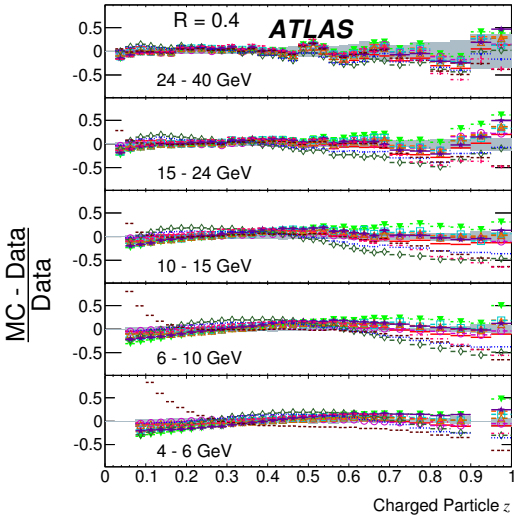


(d) $R = 0.6$ Data-MC Difference

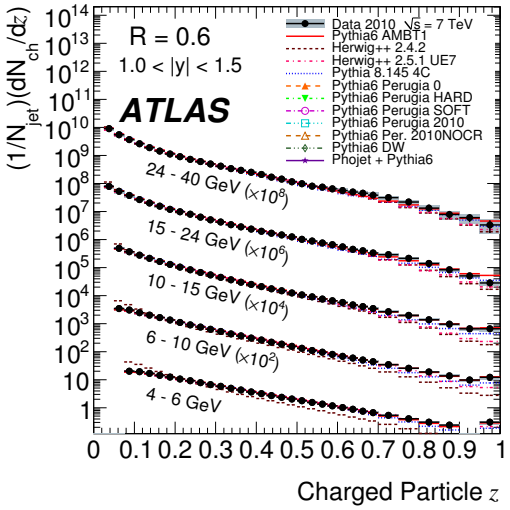
Figure A.13: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter R as indicated, in the rapidity range $0.5 < |y| < 1.0$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



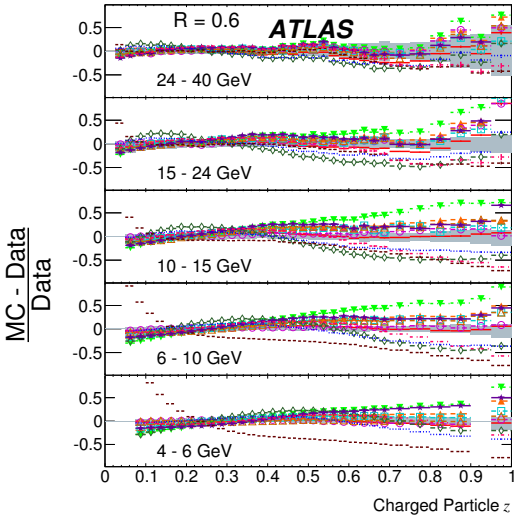
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

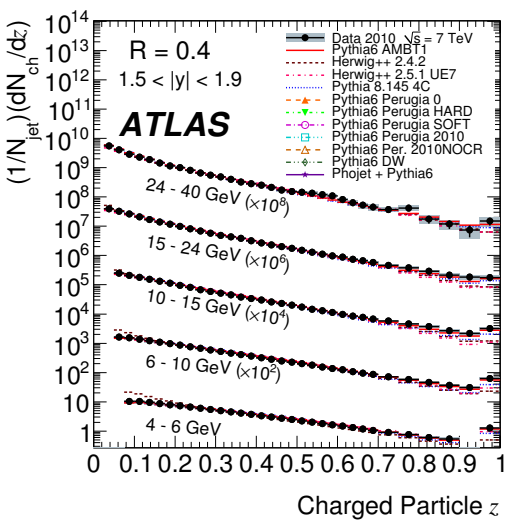


(c) $R = 0.6$ Distribution

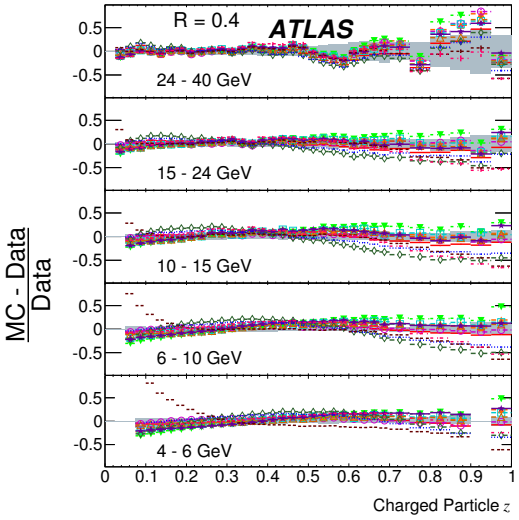


(d) $R = 0.6$ Data-MC Difference

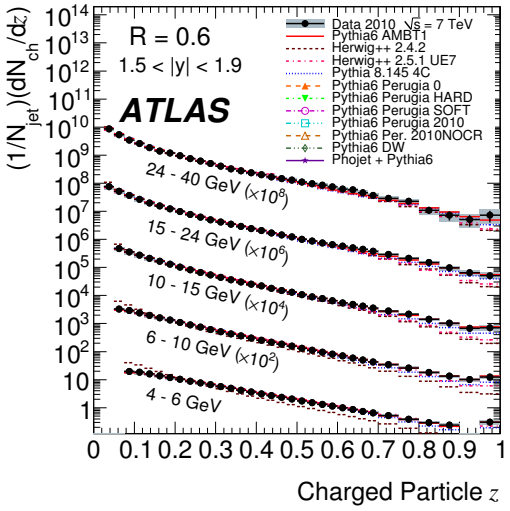
Figure A.14: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.0 < |y| < 1.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



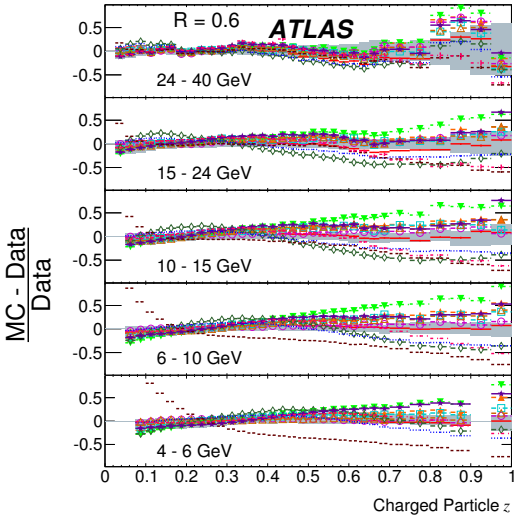
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

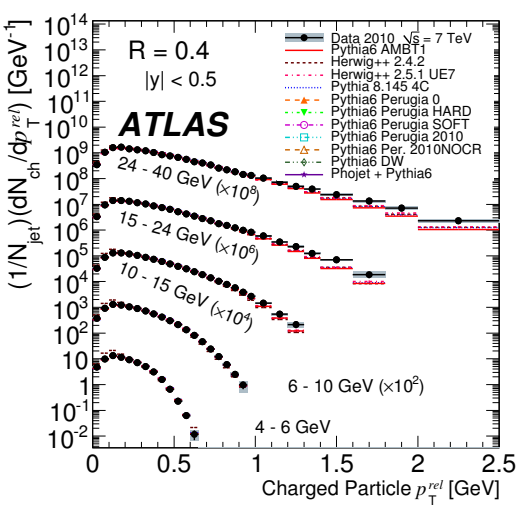


(c) $R = 0.6$ Distribution

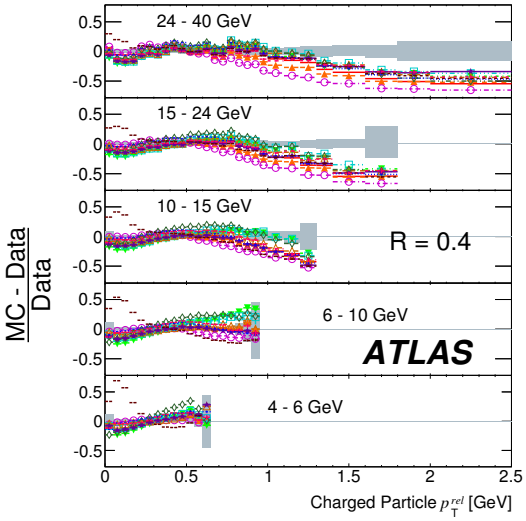


(d) $R = 0.6$ Data-MC Difference

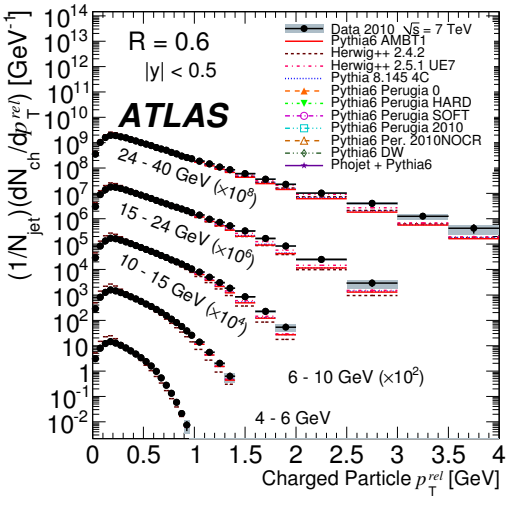
Figure A.15: The distribution of the fragmentation variable z for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.5 < |y| < 1.9$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



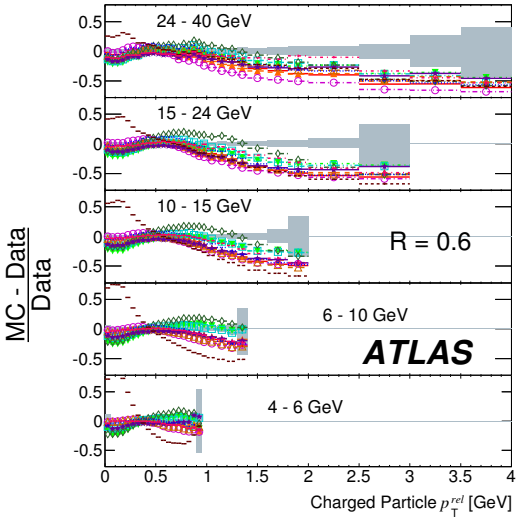
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

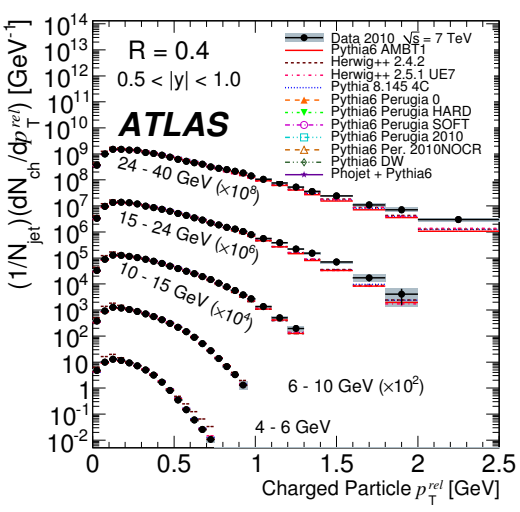


(c) $R = 0.6$ Distribution

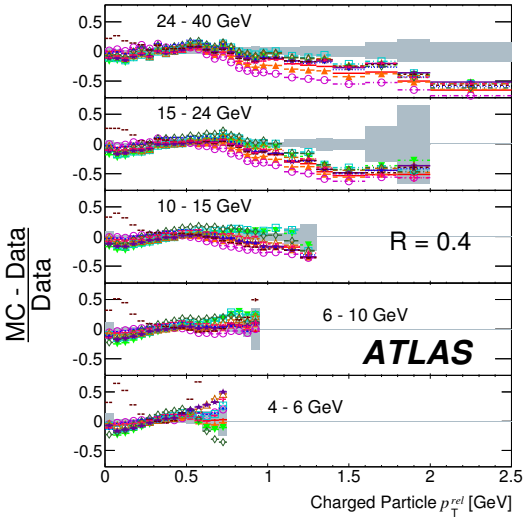


(d) $R = 0.6$ Data-MC Difference

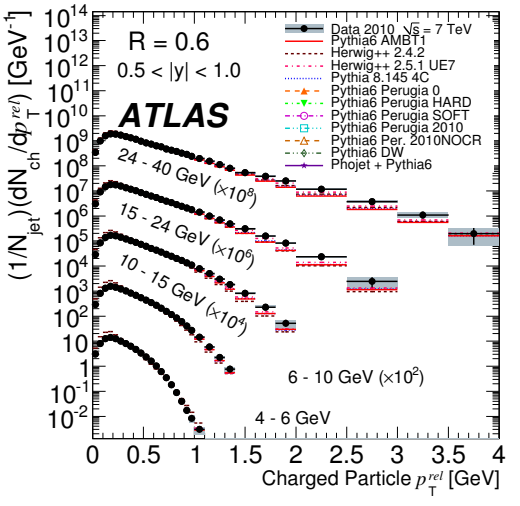
Figure A.16: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter R as indicated, in the rapidity range $|y| < 0.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



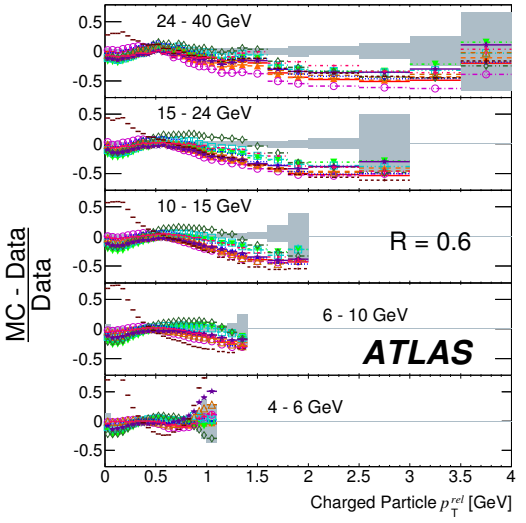
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

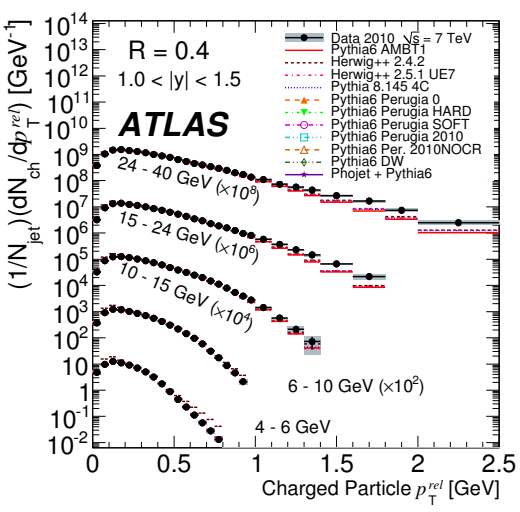


(c) $R = 0.6$ Distribution

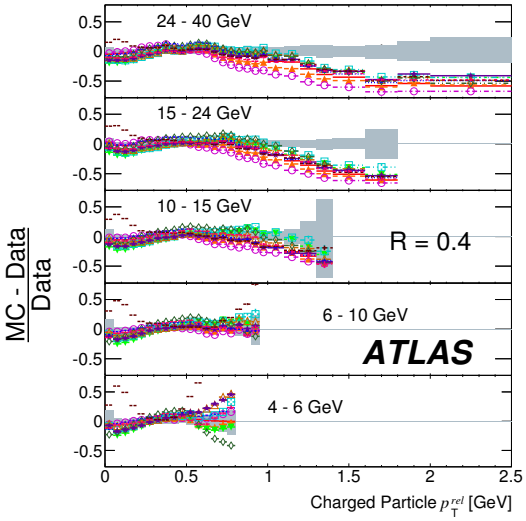


(d) $R = 0.6$ Data-MC Difference

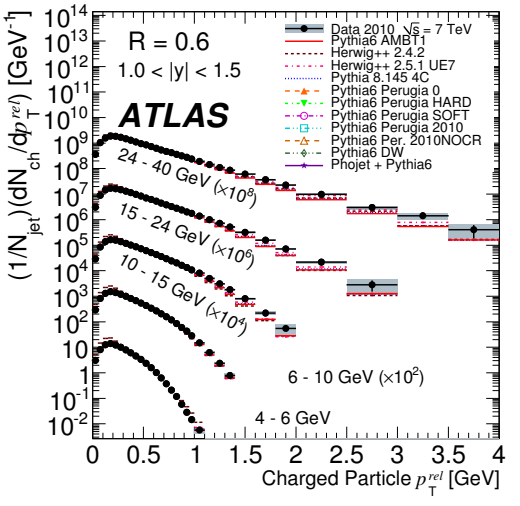
Figure A.17: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter R as indicated, in the rapidity range $0.5 < |y| < 1.0$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



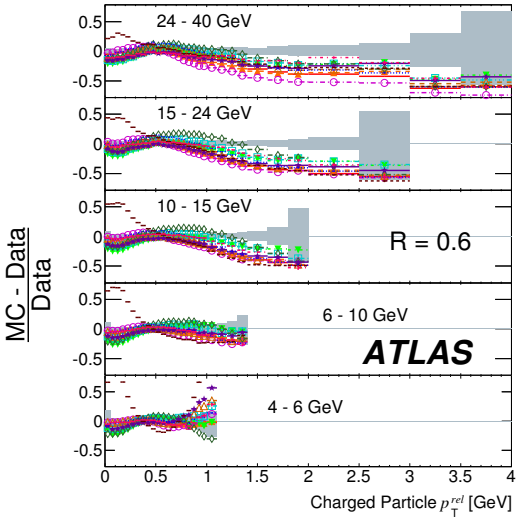
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

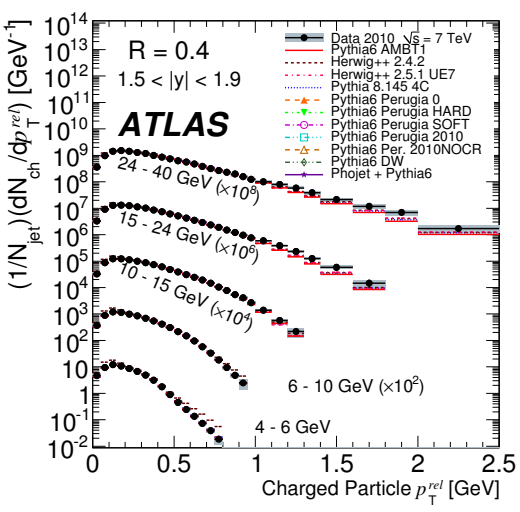


(c) $R = 0.6$ Distribution

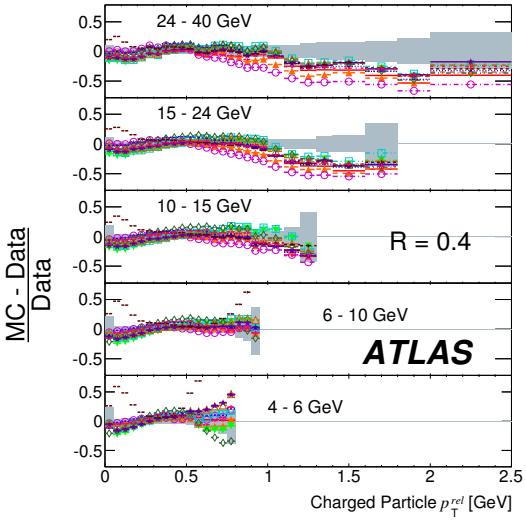


(d) $R = 0.6$ Data-MC Difference

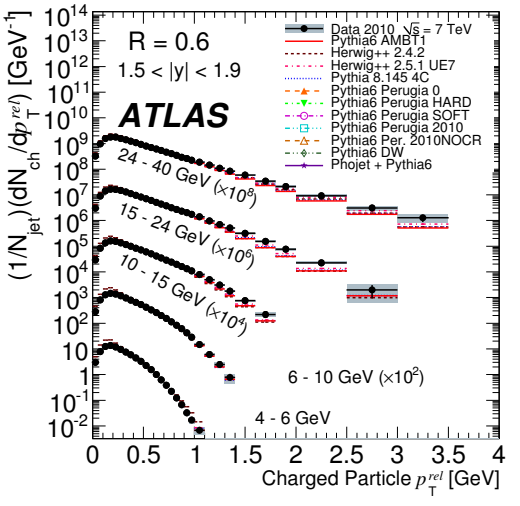
Figure A.18: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.0 < |y| < 1.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



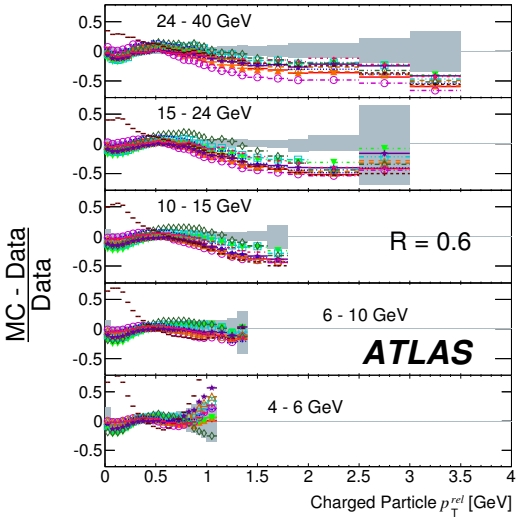
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

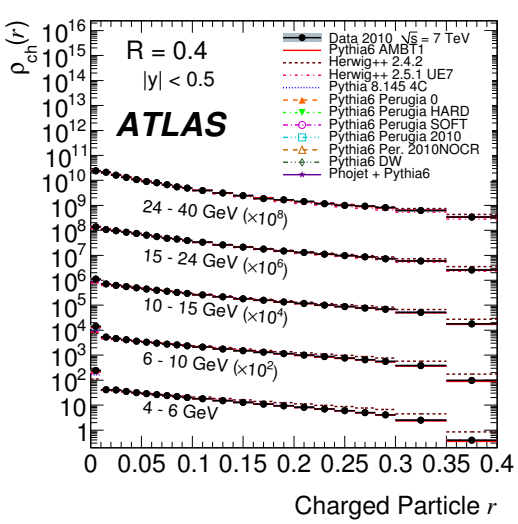


(c) $R = 0.6$ Distribution

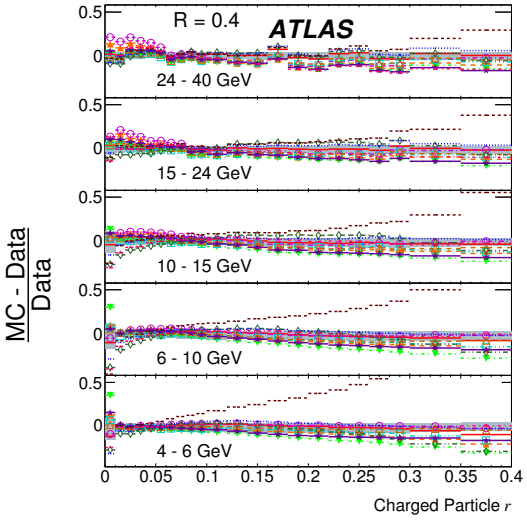


(d) $R = 0.6$ Data-MC Difference

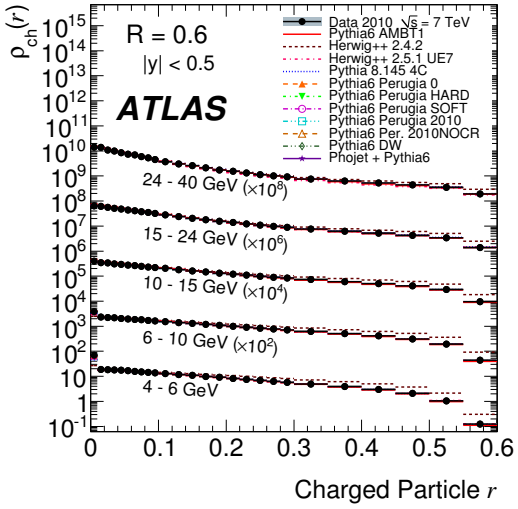
Figure A.19: The distribution of the fragmentation variable p_{T}^{rel} for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.5 < |y| < 1.9$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



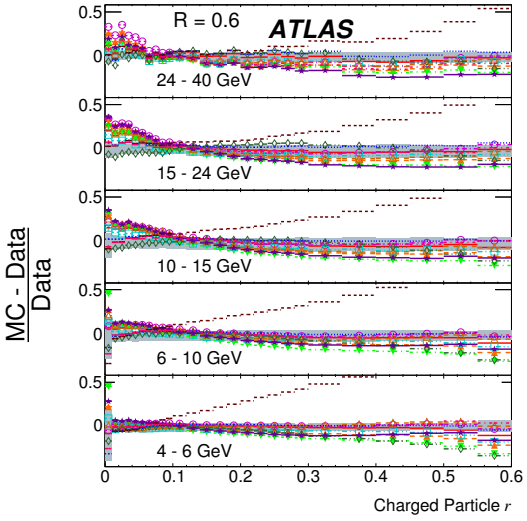
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

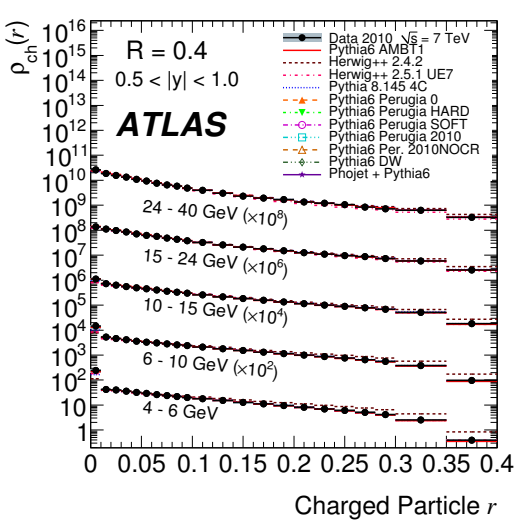


(c) $R = 0.6$ Distribution

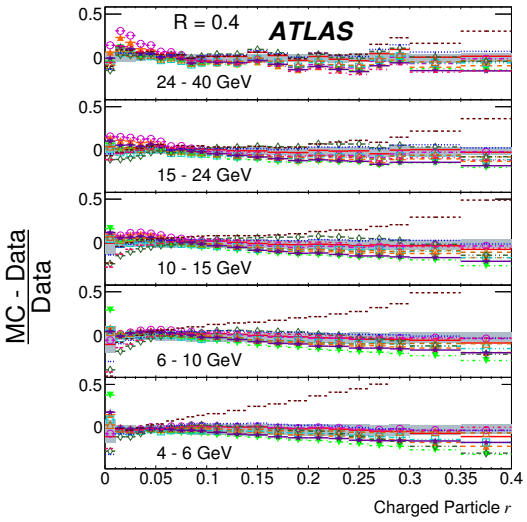


(d) $R = 0.6$ Data-MC Difference

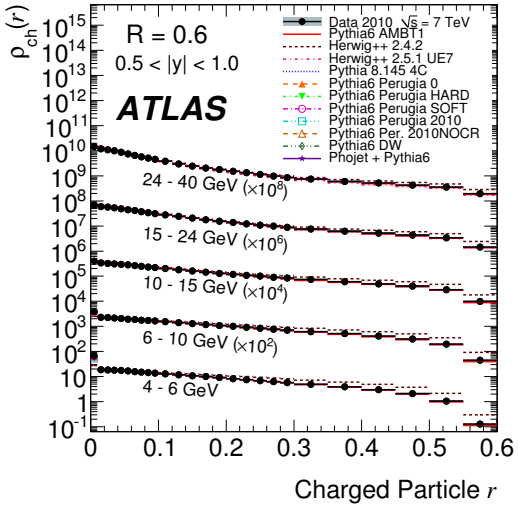
Figure A.20: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter R as indicated, in the rapidity range $|y| < 0.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



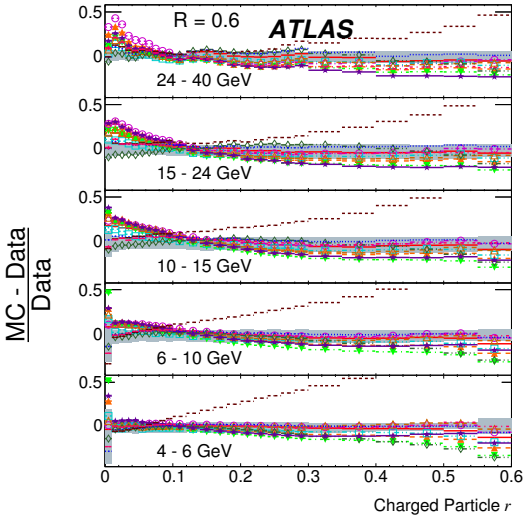
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

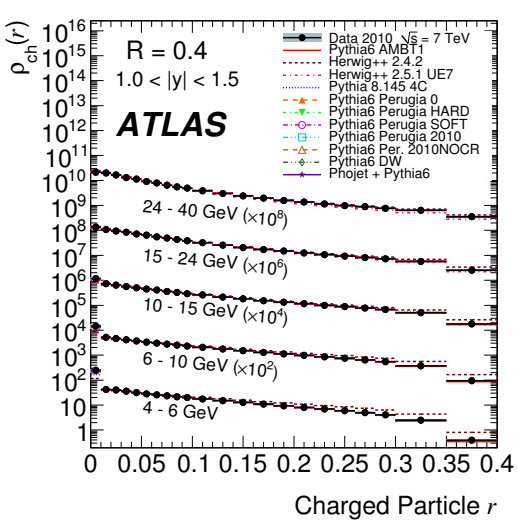


(c) $R = 0.6$ Distribution

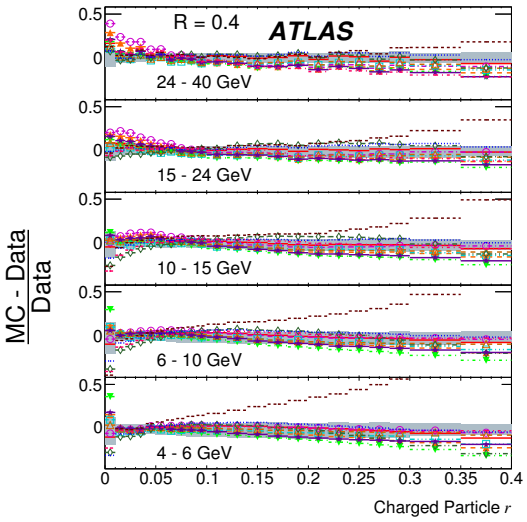


(d) $R = 0.6$ Data-MC Difference

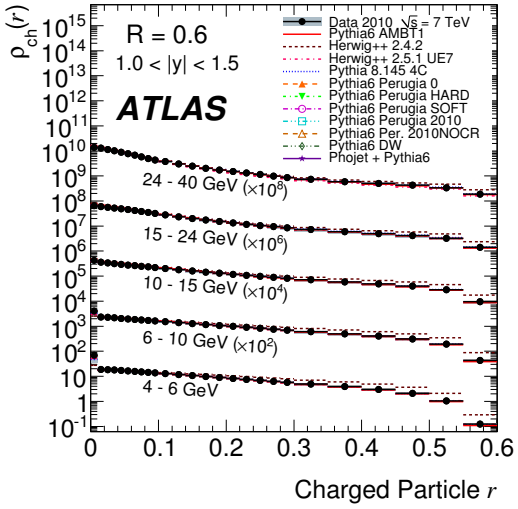
Figure A.21: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter R as indicated, in the rapidity range $0.5 < |y| < 1.0$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



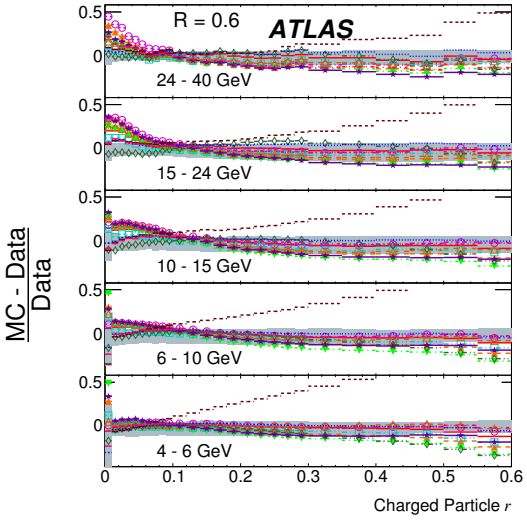
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference

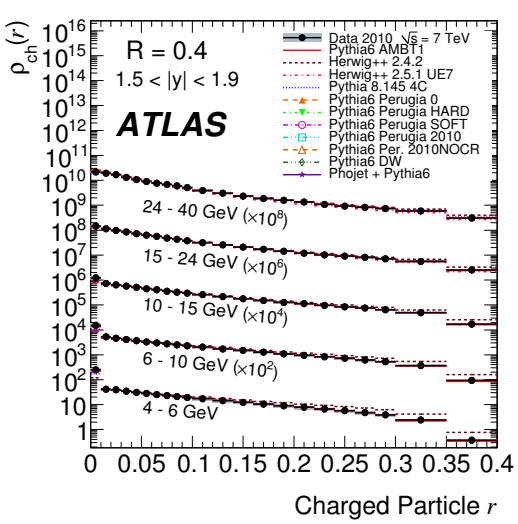


(c) $R = 0.6$ Distribution

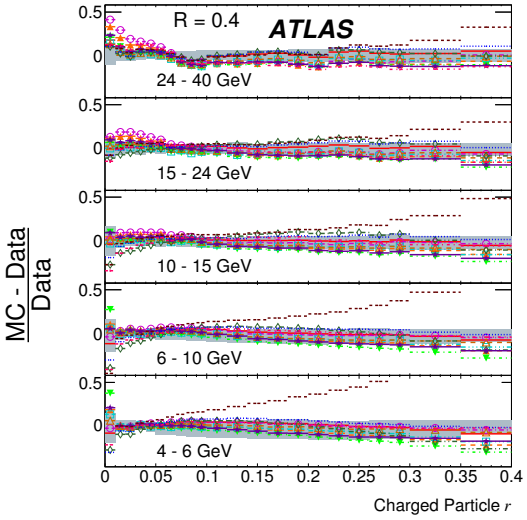


(d) $R = 0.6$ Data-MC Difference

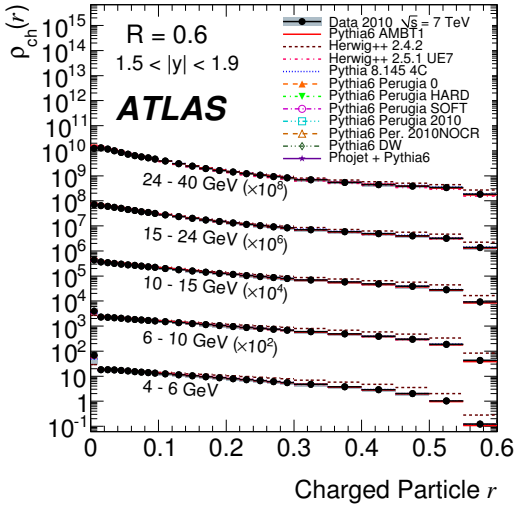
Figure A.22: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.0 < |y| < 1.5$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.



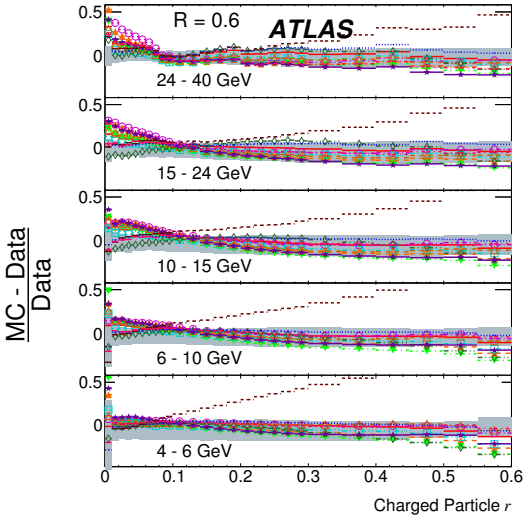
(a) $R = 0.4$ Distribution



(b) $R = 0.4$ Data-MC Difference



(c) $R = 0.6$ Distribution



(d) $R = 0.6$ Data-MC Difference

Figure A.23: The distribution of the fragmentation variable r for anti-k_t jets with radius parameter R as indicated, in the rapidity range $1.5 < |y| < 1.9$. Figures (a) and (c) show the distributions for five momentum ranges with $R = 0.4$ and 0.6, respectively, and Figures (b) and (d) show the fractional difference between a range of Monte Carlo event generator predictions and the data.