Measurements of Z boson plus jet production cross section using $\sqrt{s} = 8$ TeV data and studies of jet quark-gluon decomposition

Dissertation

zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Nataliia Kondrashova

Hamburg
2016

Gutachter/in der Dissertation:	PD Dr. Alexandre Glazov Prof. Dr. Johannes Haller
Gutachter/in der Disputation:	Prof. Dr. Daniela Pfannkuche Prof. Dr. Elisabetta Gallo PD Dr. Krisztian Peters
Datum der Disputation:	21.03.2016
Vorsitzender des Prüfungsausschusses:	Prof. Dr. Daniela Pfannkuche
Vorsitzender des Promotionsausschusses:	Prof. Dr. Jan Louis
Dekan des Fachbereichs Physik:	Prof. Dr. Heinrich Graener

Abstract

A measurement of the differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production and a study of the jet quark-gluon decomposition are presented. The data of 21.3 fb^{-1} collected with the ATLAS detector at the Large Hadron Collider in 2012 at the centre-of-mass energy $\sqrt{s} = 8$ TeV are used. The double-differential $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross-section is measured as a function of the absolute rapidity and the transverse momentum of jets. The jet quark-gluon decomposition study is performed in bins of the transverse momentum and the absolute rapidity of the highest-p_T jet.

The possibility to distinguish between quark-initiated and gluon-initiated jets is especially important for beyond Standard Mode searches, where a lot of signal processes have quarks in the final states, while background processes in Quantum Chromodynamic have mostly gluons. The performance of the discrimination between these two types of jets using different jet properties is studied using data-driven techniques with purified quark-like and gluon-like jet samples.

The $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production provides an important test of perturbative Quantum Chromodynamics and is an important background for many Standard Model processes and beyond Standard Model searches. In addition, the measurement of the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross section as a function of the absolute rapidity and the transverse momentum of inclusive jets provides constraints on the uncertainties on the parton distribution functions. The rapidity of jets provides the information on the fraction of the initial proton's momentum carried by the interacting partons, which provides the sensitivity to the parton distribution functions, while the transverse momentum of jets allows to probe different transfer momentum scales.

The measured cross-section is compared to the predictions from Monte Carlo generators based on leading order matrix elements and supplemented by parton showers, where the predictions are normalised to the inclusive $pp \to Z/\gamma^* \to e^+e^-$ cross-section calculated at the next-to-next-to-leading order. The predictions are in a good agreement with the measured cross-section within the uncertainty on the measurement, except particular absolute rapidity and transverse momentum regions.

Zusammenfassung

Diese Arbeit behandelt die Messung des differenziellen Wirkungsquerschnitts der $pp \to Z/\gamma^*(\to e^+e^-)$ + jet Produktion und eine Studie der Quark-Gluon Zusammensetzung von Jets. Es werden Daten mit einer integrierten Luminosität von 21.3 fb^{-1} genutzt, die 2012 vom ATLAS Detektor bei einer Schwerpunktsenergie von $\sqrt{s} = 8$ TeV gesammelten wurden. Der doppelt-differenzielle $pp \to Z/\gamma^*(\to e^+e^-)$ + jet Wirkungsquerschnitt wird als Funktion der absoluten Rapidität und des transversalen Impulses der inklusiven Jets gemessen. Die Studie der Quark-Gluon-Zusammensetzung der Jets wird in Einheiten des transversalen Impulses und der absoluten Rapidität der Jets mit dem höchsten p_T durchgeführt.

Die Möglichkeit zwischen Quark-iniziierten und Gluon-iniziierten Jets unterscheiden zu können ist speziell für Suchen jenseits des Standardmodells wichtig, in denen viele Signalprozesse einen Quark-Endzustand besitzen, während die quantumchromodynamischen Hintergrundprozesse meistens einen Gluon-Endzustand aufweisen. Es wird die Qualität der Abgrenzung zwischen diesen beiden Jet-Typen unter Verwendung unterschiedlicher Jet-Eigenschaften studiert, wobei datenbasierte Techniken mit bereinigten Quark-ähnlichen und Gluon-ähnlichen Proben verwendet werden.

Die $pp \to Z/\gamma^*(\to e^+e^-)$ + jet Produktion liefert eine wichtige Testmöglichkeit für die pertubative Quantumchromodynamik und ist wichtiger Hintergrund für viele Standardmodell-Prozesse sowie Suchen jenseits des Standardmodells. Außerdem kann die Messung des $pp \to Z/\gamma^*(\to e^+e^-)$ + jet Wirkungsquerschnitts als Funktion der absoluten Rapidität und des transversalen Impulses die Unsicherheiten der Partondichtefunktion reduzieren. Die Rapidität der produzierten Jets liefert Informationen über den Impulsbruchteil des interagierenden Partons vom initialen Proton welcher sensitiv auf die Partondichtefunktion ist, während der Transversalimpulsder Jets das Testen verschiedener Skalen ermöglicht.

Der gemessene Wirkungsquerschnitt wird mit Vorhersagen von Monte-Carlo-Generatoren verglichen, welche auf Matrixelementen führender Ordnung basieren und mit Partonschauern verbessert sind. Die Vohersagen wurden auf den inklusiven $pp \to Z/\gamma^* \to e^+e^-$ Wirkungsquerschnitt normiert, der in next-to-next-to-leading oder berechnet wurde. Die Vorhersagen sind, bis auf bestimmte Bereiche der absoluten Rapidität und des transversalen Impulses, in guter Übereinstimmung mit dem gemessenen Wirkungsquerschnitt innerhalb der Unsicherheiten der Messung.

Hiermit erklaäre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, 18/12/2015

Contents

Contents i

1 Thesis Organization 1

2 Theory Introduction 3
- 2.1 The Standard Model of particle physics . . . 3
- 2.2 Proton-proton collisions . . . 5
  - 2.2.1 Renormalisation and running of the strong coupling . . . 6
  - 2.2.2 Parton distribution functions . . . 7
- 2.3 Jets . . . 9
  - 2.3.1 Jet algorithms . . . 9
  - 2.3.2 Jet properties . . . 10
- 2.4 Z boson in proton-proton collisions . . . 12
  - 2.4.1 Jets in Drell-Yan process . . . 13
- 2.5 Monte Carlo simulation . . . 14
  - 2.5.1 Hard subprocess . . . 15
  - 2.5.2 Parton shower . . . 15
  - 2.5.3 Hadronisation . . . 15
  - 2.5.4 Underlying event . . . 16
  - 2.5.5 Monte Carlo generators for ATLAS . . . 16

3 Experimental Setup 19
- 3.1 The Large Hadron Collider . . . 19
  - 3.1.1 Overview . . . 19
  - 3.1.2 Luminosity . . . 20
- 3.2 The ATLAS detector . . . 21
  - 3.2.1 Overview . . . 21
  - 3.2.2 The magnet system . . . 23
  - 3.2.3 The inner detector . . . 24
  - 3.2.4 The calorimeter system . . . 26
    - 3.2.4.1 The electromagnetic calorimeters . . . 27
    - 3.2.4.2 The hadronic calorimeters . . . 28
  - 3.2.5 The muon spectrometer . . . 30
  - 3.2.6 The trigger system . . . 32

4 Data and Monte Carlo Samples **34**
4.1 Data taking . . . 34
4.2 Monte Carlo samples . . . 35

5 Event Reconstruction and Selection **38**
5.1 Tracks and vertices . . . 38
5.2 Electron reconstruction and identification . . . 39
5.2.1 Electron reconstruction . . . 39
5.2.2 Electron identification . . . 41
5.3 Jet reconstruction . . . 41
5.3.1 Calorimeter jet inputs . . . 42
5.3.2 Jet cleaning . . . 44
5.3.3 Jet vertex fraction . . . 45
5.4 Monte Carlo reconstruction . . . 45
5.5 $Z/\gamma^*(\rightarrow e^+e^-)$ + jet event selection . . . 46

6 Data and Monte Carlo Corrections **49**
6.1 Pile-up reweighting . . . 49
6.2 Electron corrections . . . 50
6.2.1 Reconstruction, identification and trigger scale factors . . . 50
6.2.2 Electron energy scale and resolution . . . 51
6.3 Jet calibration . . . 52

7 Background Extraction **56**
7.1 Multijet background estimation . . . 57

8 Control Distributions **61**

9 Z + jet Events for the Quark-Gluon Tagging Purposes **65**
9.1 Introduction . . . 65
9.2 Data samples . . . 66
9.3 Discriminating variables . . . 66
9.4 Quark-Gluon templates extraction . . . 67
9.4.1 Two-samples extraction . . . 68
9.4.2 Three-samples extraction . . . 68
9.4.3 Variable ranking . . . 70
9.5 Uncertainties in the Quark-Gluon Jets Discrimination . . . 73
9.6 Validation of Extracted Quark/Gluon Templates . . . 81
9.7 Conclusions . . . 81

10 Unfolding in the Cross-Section Measurement **82**
10.1 Bayes' iterative method . . . 83
10.2 Matching efficiency, purity & stability . . . 84
10.3 Optimisation of number of iterations . . . 88
10.4 Closure tests . . . 90

11 Uncertainties in the Cross-Section Measurement **91**
11.1 Statistical uncertainties 91
11.2 Systematic uncertainties 92
11.2.1 Electron uncertainties 94
11.2.2 Jet uncertainties 94
11.2.3 Uncertainty from the data-driven background subtraction . . 98
11.2.4 Uncertainty from the Monte Carlo based background subtraction 98
11.2.5 Uncertainty in the unfolding procedure 99
11.2.6 Summary 99

12 Results of the Cross-Section Measurement **103**

13 Summary **111**

Acknowledgements **121**

Chapter 1

Thesis Organization

This thesis presents the measurement of the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet double-differential cross-section and the study of the jet quark-gluon decomposition using data collected with the ATLAS detector at the LHC in 2012 at a center-of-mass energy $\sqrt{s} = 8$ TeV. The layout of the thesis is organised as follows:

Chapter 2, Theory Introduction, contains the brief theoretical introduction to the Standard Model, the theory of proton-proton collision, jets and Z boson production in proton-proton collision. The motivation for the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross-section measurement is discussed. The description of the Monte Carlo simulation steps and the overview of the Monte Carlo generators used in ATLAS are also provided.

Chapter 3, Experimental Setup, describes the LHC accelerator complex and the ATLAS detector together with all subdetector systems.

Chapter 4, Data and Monte Carlo Samples, provides an overview of the data collected with the ATLAS detector in 2012 and Monte Carlo samples used for this measurement.

Chapter 5, Event Reconstruction and Selection, describes the reconstruction of the physics objects which are used in this analysis and discuss the criteria used to select $Z/\gamma^*(\to e^+e^-)$ + jet events.

Chapter 6, Data and Monte Carlo Corrections, outlines the corrections applied to measured and simulated events in order to account for detector effects and to improve the agreement between data and simulation.

Chapter 7, Background Extraction, describes the techniques used for the subtraction of the background contributions in the selected signal sample.

Chapter 8, Control Distributions, shows the level of agreement between data and Monte Carlo predictions for different kinematic distributions of electrons, Z bosons and jets.

Chapter 9, Z+jets Events for Quark Gluon Tagging Purposes, presents the study of the jet quark-gluon decomposition using $Z/\gamma^*(\rightarrow e^+e^-) + \text{jet}$ sample. The data-driven method used for the discrimination between quark-initiated and gluon-initiated jets is described. The resulting quark-gluon jet tagging performance together with the related systematic uncertainties are shown.

Chapter 10, Unfolding in the Cross-Section Measurement, explains the unfolding technique used for the correction of the measured cross-section for detector effects.

Chapter 11, Uncertainties in the Cross-Section Measurement, describes the sources of the uncertainty in this measurement and the procedure of the uncertainties propagation to the final result.

Chapter 12, Results of the Cross-Section Measurement, presents the results of the double-differential cross-section measurements as well as the comparison to the predictions from Monte Carlo generators based on the leading order matrix elements supplemented by the parton showers. The systematic uncertainties in the final result are described.

Chapter 2

Theory Introduction

2.1 The Standard Model of particle physics

The Standard Model (SM) [1–3] describes the universe in terms of fundamental particles and interactions between them. There are two categories of elementary particles: fermions with half-integer spin, which are matter particles, and bosons with integer spin, which carry interaction forces. The summary of SM elementary particles and their properties is shown in Figure 2.1.

There are a few boson fields in the SM, namely twelve gauge bosons with spin equal to 1 and one scalar Higgs boson with spin 0. The photon is a mediator of the electromagnetic interaction, $W^{\pm}$ and Z bosons mediate the weak interaction, and the gluons, which carry colour charge, are responsible for the strong interaction. All bosons except $W^{\pm}$, Z and Higgs bosons are massless. Electrically neutral bosons, H, γ, Z and gluons coincide with their antiparticles.

The fermions are divided into two groups by the different ways in which they react to the fundamental forces: quarks and leptons. There are six flavours of quarks: up, down, charm, strange, top and bottom. Each quark carries colour charge, electric charge Q and weak isospin I and, hence, can interact through the strong force, as well as through the electromagnetic and weak forces. There are three charged leptons: the electron e, the muon μ and the tau τ, and three neutral leptons called neutrinos, ν_e, ν_μ, ν_τ. All leptons are colorless and do not interact through the strong force, while neutrinos are weakly interacting only. All fermions are grouped into three generations, two quarks and two leptons in each of them.

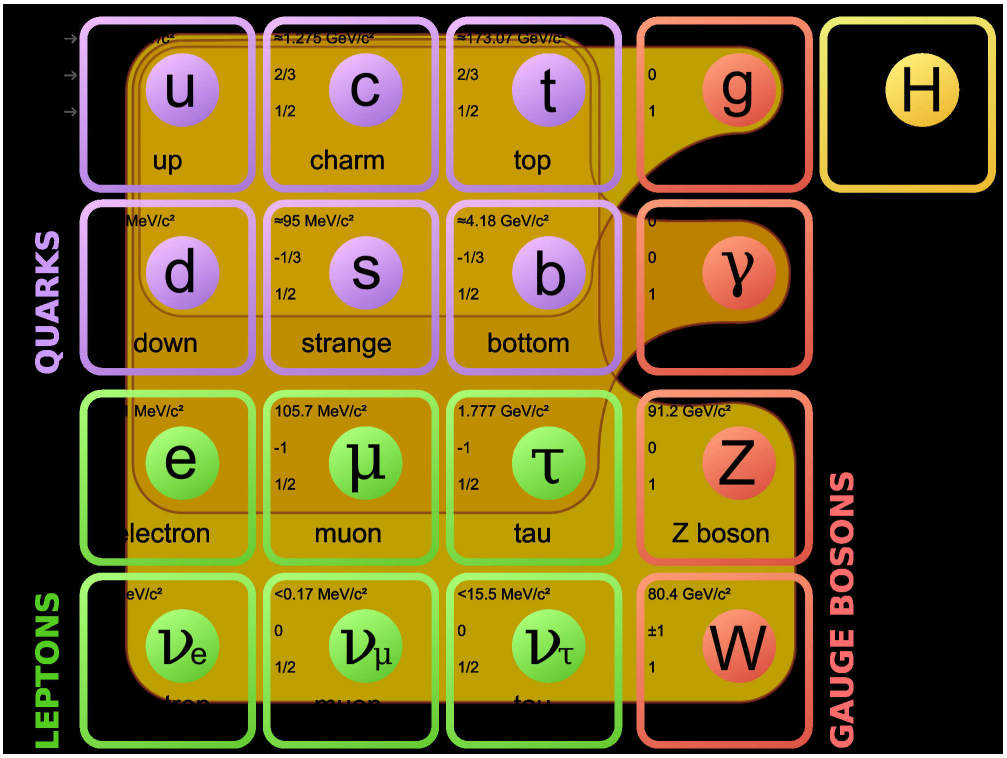


FIGURE 2.1: Standard Model of elementary particles. Taken from [4].

The lightest and most stable particles are in the first generation, while heavier and less stable particles are in the second and the third. All left-handed fermions have isospin $I = 1/2$, and within each generation particles in each quark or lepton pair are defined to have a different third component of the isospin vector, I_3, one with $I_3 = 1/2$, and the second one with $I_3 = -1/2$. Each fermion particle has a corresponding antiparticle.

Due to the phenomenon known as color confinement, quarks cannot be directly observed or found in isolation. They combine to composite particles called hadrons, held together by the strong force. The hadrons are divided into two groups: baryons, made of three quarks or antiquarks, and mesons, composed of one quark and one antiquark. The hadrons have integer electric charge and no colour charge. Except of so-called "valence" quarks, which define quantum properties of hadrons, they contain also virtual quark-antiquark pairs, known as "sea" quarks, which are the result of gluons splittings.

The SM is constructed within the Quantum Field Theory (QFT) based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The gauge group of the strong interaction $SU(3)_C$, where C refers to the colour quantum number and 3 to the number of the possible colour states, is described by a theory called Quantum

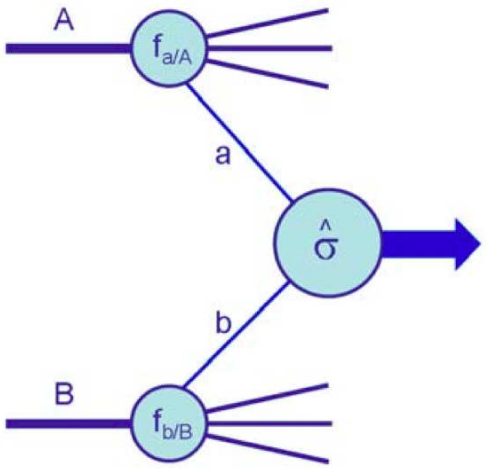


FIGURE 2.2: Hard scattering process in proton-proton collision, taken from [6].

Chromodynamics (QCD). The gauge group of electromagnetic and weak interactions $SU(2)_L \times U(1)_Y$, where L shows that only left-handed fermions are included in the SM, and Y denotes the weak hypercharge and is related to the isospin and the electric charge as $Y = 2(Q - I_3)$, is described by Quantum Electrodynamics (QED).

2.2 Proton-proton collisions

The proton is composed of three valence quarks, two up and one down (uud), which are held together by the strong force, mediated by gluons. In addition, the proton contains also "sea" quarks. Each of the partons inside the proton carries a fraction x of the proton momentum.

Since the protons are not elementary particles, the high energy proton-proton collisions cannot be completely described by the perturbative QCD theory. However, factorisation theorem [5] allows to separate the long-distance effects, which are not perturbatively calculable, from the short distance cross-section for the hard scattering of partons, which can be predicted with a good precision using the perturbation theory. The scale at where this separation is done is called factorisation scale, denoted by μ_F.

The diagram of a generic hadron-hadron hard scattering process $AB \to X$ is shown in Figure 2.2. The hadronic cross-section for this process, following

factorisation theorem, can be obtained by weighting the partonic cross-section $\hat{\sigma}_{ab\to X}$ with the parton distribution functions (PDFs) $f_{i/p}(x, \mu_F^2)$, the probability density of a parton i to carry the momentum fraction x of the proton at the scale μ_F. Therefore, the cross-section of this process can be calculated using the perturbation theory as follows [6]:

$$\sigma_{AB\to X} = \int dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \times \left[\hat{\sigma}_0 + \alpha_S(\mu_R^2)\hat{\sigma}_1 + \alpha_S^2(\mu_R^2)\hat{\sigma}_1 + ...\right]_{ab\to X}, \tag{2.1}$$

where μ_R is the renormalisation scale for the QCD running strong coupling α_S.

The calculated cross-section for the hard scattering process should be additionally corrected for the low scale non-perturbative effects in proton-proton collisions. The short description of these effects and their modelling are briefly summarised in Section 2.5.

2.2.1 Renormalisation and running of the strong coupling

The perturbative component $\hat{\sigma}_{ab\to X}$ of the hard scattering process in proton-proton collisions can be calculated analytically by integrating the contributions from all order Feynman diagrams for a given process over the phase space. However, since higher order diagrams include internal loops, the integrals over the phase space become ultraviolet (UV) divergent for integration-momenta going to infinity. These UV divergences can be absorbed by the renormalisation procedure [7], which introduces an arbitrary renormalisation scale μ_R, and the UV divergences are absorbed in the definition of the coupling strength. Consequently, for calculations performed at a finite perturbative order, the coupling parameter $\alpha_S(\mu_R)$ and any physical quantity which is expressed as a series of the coupling parameter become functions of μ_R [8] .

The dependence of the strong coupling on the scale is given by the renormalisation group equation (RGE):

$$\mu_R^2 \frac{\partial \alpha_S(\mu_R)}{\partial \log \mu_R^2} = \beta(\alpha_S(\mu_R)) = -\beta_0 \alpha_S^2(\mu_R) - \beta_1 \alpha_S^3(\mu_R) - ..., \tag{2.2}$$

where $\beta(\alpha_S(\mu_R))$ is calculated perturbatively in QCD. The first coefficient is:

$$\beta_0 = \frac{33 - 2f}{12\pi}, \tag{2.3}$$

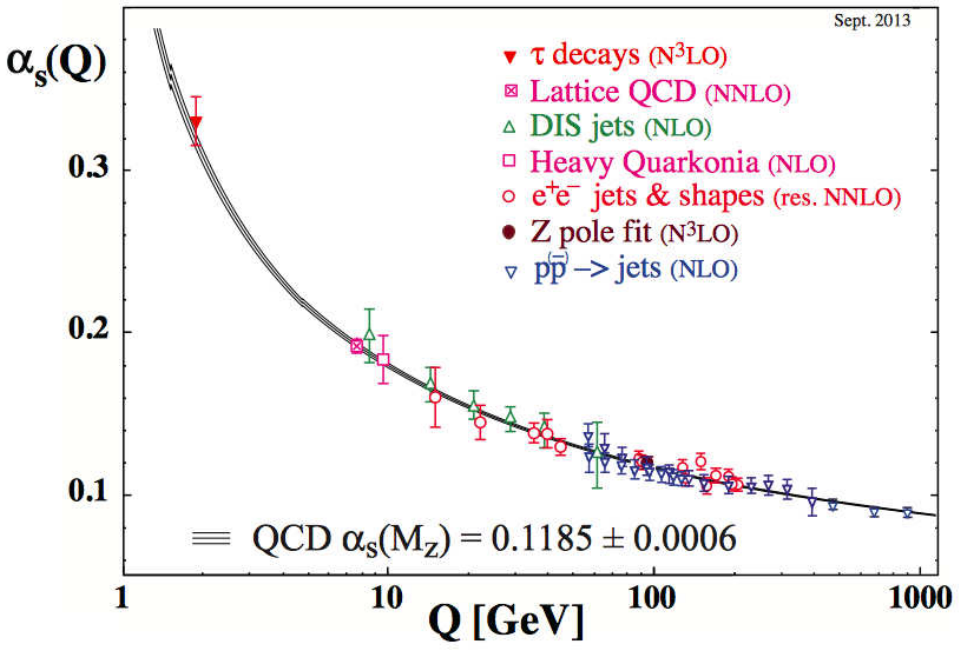


FIGURE 2.3: The status of the running of α_S in 2013, taken from [9].

where f is the number of flavours and colour factor 3 is used. The energy dependence of the renormalised coupling $\alpha_S(\mu_R)$ can be precisely determined by integrating Eq. 2.2.

The running of α_S is observed by measuring processes sensitive to the strong coupling at different scales, as shown in Figure 2.3.

2.2.2 Parton distribution functions

The theoretical calculation of the production cross sections at hadron colliders requires the knowledge of the distribution of the parton momentum fraction x in the proton. The parton density function (PDF) $f_i(x, \mu_F^2)$ gives the probability density of finding in the proton a parton of flavour i (quarks or gluon) carrying a fraction x of its momentum, with μ_F being the factorisation scale, which is typically chosen as the momentum transfer in the hard interaction, Q^2.

The parton distributions used in the hard-scattering calculations are the solutions of the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations:

$$\frac{\partial q_i(x,\mu_F^2)}{\partial log \mu_F^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}, \left\{ P_{q_i q_j}(z,\alpha_S) q_j(\frac{x}{z},\mu_F^2) + P_{q_i g}(z,\alpha_S) g(\frac{x}{z},\mu_F^2) \right\}, \quad (2.4)$$

$$\frac{\partial g(x,\mu_F^2)}{\partial log \mu_F^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left\{P_{gq_j}(z,\alpha_S)q_j(\frac{x}{z},\mu_F^2) + P_{gg}(z,\alpha_S)g(\frac{x}{z},\mu_F^2)\right\} \tag{2.5}$$

where P_{qq}, P_{qg}, P_{gq} and P_{gg} are the splitting functions. The splitting function P_{ab} is the probability for a parton b with a momentum fraction z to emit a parton a with a momentum fraction x. The splitting functions have perturbative expansions:

$$P_{ab}(x,\alpha_S) = P_{ab}^{(0)}(x) + \frac{\alpha_S}{2\pi}P_{ab}^{(1)}(x) + ... \tag{2.6}$$

The DQLAP equations determine the momentum scale Q^2 dependence of the PDFs. The x dependence has to be obtained from the global fits to Deep Inelastic Scattering (DIS), the Drell-Yan process and jets production data. Modern tools are able to perform fits in the leading-order (LO), the next-to-leading order (NLO) or in the next-to-next-to-leading order (NNLO) in the strong coupling α_S expansion. Measurements of the DIS in the lepton-hadron scattering and of the dilepton production in the hadron-hadron collisions are the main source of information about the quark distribution functions $f_{q/p}(x,Q^2)$, while the gluon distribution function $f_{g/p}(x,Q^2)$ is probed in events with jets in the final state [6].

Since quark distributions are determined from DIS and DY data sets with the large statistics and systematic uncertainties of the order of few percent, the sum of the quark distributions is known to a similar accuracy. However, the individual quark flavour distributions have larger uncertainty. This can be very important in predicting observables which depend on a specific quark flavour. The uncertainty in the gluon distribution function is in particular large in a high x region, where direct measurements are necessary.

Several PDF fitting collaborations provide regular updates to the parton distributions. Figure 2.4 shows the parton distribution functions determined at transfer momentum scales $Q^2 = 10\text{GeV}^2$ and $Q^2 = 10^4\text{GeV}^2$ using MMHT 2014 NNLO parametrisation [10]. At low Q the valence quarks (u, d) carry about one third of the proton momentum, while the sea quarks ($\bar{u}$, $\bar{d}$, c, $\bar{c}$) and the gluon dominate at low x. The contribution from the sea quarks and the gluon becomes larger with increasing the transfer momentum scale.

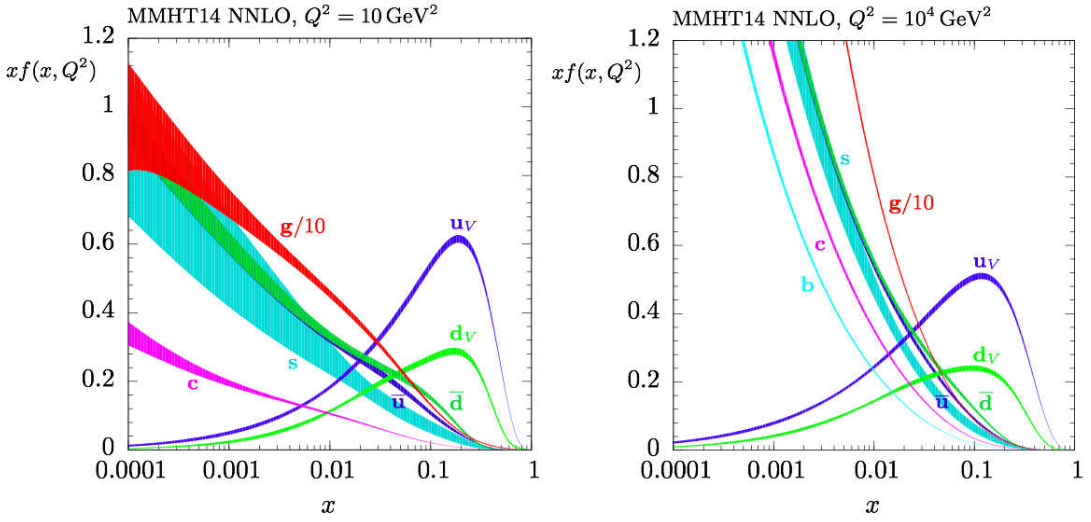


FIGURE 2.4: MMHT 2014 parton distribution functions [10] as a function of x for momentum transfer scales (left) $Q^2 = 10$ GeV^2 and (right) $Q^2 = 10^4$ GeV^2.

2.3 Jets

The production of quarks and gluons via strong interaction is the high-rate process at the LHC. These partons eventually are measured as ensembles of well collimated charged and neutral hadrons, called jets. Jets are reconstructed using jet algorithms, and, depending on the parton which initiates the hadronisation, they can have different properties.

2.3.1 Jet algorithms

The jet definition strongly depends on the algorithm which was used for the jet reconstruction. There are several main requirements for such algorithms. The most important requirement is infrared and collinear safety (IRC), which means that a presence of the soft partons or collinear splitting of the parton originating the jet does not change the output jet. The algorithm is also required to be applicable on parton, hadron and detector levels with identical output, to have minimal sensitivity to the hadronisation, underlying event and pile-up.

Jet algorithms are divided into two classes, depending on the approach of defining a jet.

• **Fixed cone algorithms** – the jet is defined as a cone with a fixed radius in $y - \phi$ space. This approach is based on the collinear nature of the parton showers;

the decay products and emissions of the initial state parton will tend to form a cone of particles in $(y - \phi)$ plane. The idea is to maximise the energy in a fixed cone radius. The difficulty for most of implementations of such algorithms is a sensitivity to soft and collinear particles.

• **Sequential recombination algorithms** are specially designed to be IRC safe. The approach is such that they attempt to follow the parton showering starting from the hardest particles.

Recombination algorithms can be generalised in the following way. The clustering starts with introducing a distance d_{iB} for each jet constituent between this constituent (i) and the beam (B) and a distance between each pair of constituents i and j, d_{ij}, and proceeds with identifying the smallest of these distances. If it is d_{ij} algorithm recombines constituents i and j, and if it is d_{iB} i is considered to be a jet and is removed from the list of constituents. The distances are recalculated then and the procedure is repeated until no undefined constituents are left.

The definition of the distances is the following:

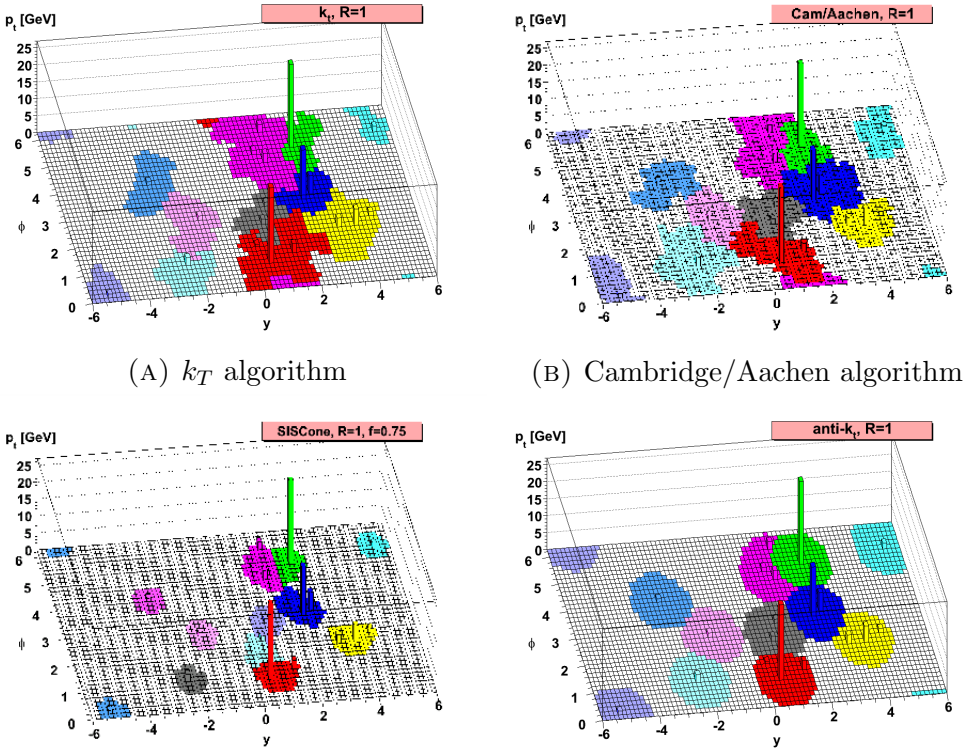
$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{2.7}$$

$$d_{iB} = k_{ti}^{2p}, \tag{2.8}$$

where $\Delta = sqrt(y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_T, y_i and ϕ_i are the transverse momentum, rapidity and azimuth of the particle i, respectively, and R is the radius parameter. Parameter p sets the power of the transverse momentum versus geometrical scales. $p = 1$ corresponds to the k_T algorithm [11], $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm [12] and $p = -1$ – to the anti-k_T algorithm [13], which is the default jet algorithm for ATLAS. The reason of such choice is a regular, cone-like geometry of output jets thanks to the clusterisation of the soft particles at the end of the sequence, while k_T and Cambridge/Aachen first cluster soft and collinear branching, and only at the end – high-p_T particles, leading to the irregular shapes (see Figure 2.5).

2.3.2 Jet properties

The properties of the jet depend on the initiating parton. Due to different probabilities for a parton splitting, gluon jets tend to be broader and fragment into more



(A) k_T algorithm

(B) Cambridge/Aachen algorithm

(C) SISCone algorithm [14]

(D) anti-k_T algorithm

FIGURE 2.5: Comparison of different jet algorithms behaviors, taken from [13].

particles. These probabilities are governed by so-called colour factors, C_F, C_A and T_F. As shown in Figure 2.6, C_F determines the coupling of a gluon to a quark, C_A the gluon self-coupling and T_F the splitting of a gluon into a quark-antiquark pair. The latter process is kinematically suppressed. The QCD predicts C_F to be 4/3, C_A to be 3, and T_F to be 1/2, which means that we can naively expect C_A/C_F=9/4 times larger number of particles in gluon jets in comparison to quark jets. However, the measured difference in jet constituent multiplicities is not as large as 9/4 due to significant non-perturbative effects and to multiplicity losses in the gluon jet due to particles emitted at large angles. The ratio of multiplicities is expected to reach the asymptotic value of 9/4 at large values of the jet transverse momentum.

In addition to colour-based differences between gluon- and quark-initiated jets, they also are expected to have different electric charge-based properties. Due to the charge conservation law, the sum of charged tracks in gluon jets is expected to be closer to 0 than in quark jets.

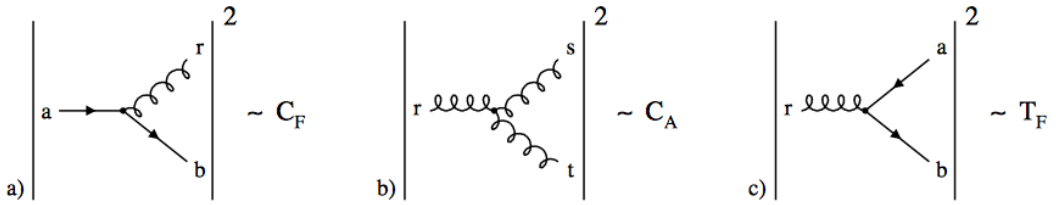


FIGURE 2.6: Parton splitting processes for quarks and gluons, taken from [15].

Z decay modes	Fraction
e^+e^-	(3.363 ± 0.004) %
$\mu^+\mu^-$	(3.366 ± 0.007) %
$\tau^+\tau^-$	(3.370 ± 0.008) %
$\nu\bar{\nu}$	(20.00 ± 0.006) %
$q\bar{q}$	(69.9 ± 0.006) %

TABLE 2.1: Main branching fractions of the Z boson, taken from [17].

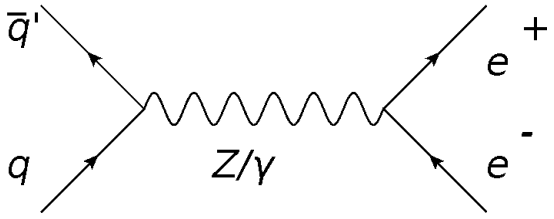


FIGURE 2.7: Z boson production in proton-proton collision with Z decay mode $Z \to e^+e^-$.

2.4 Z boson in proton-proton collisions

The Z boson was discovered at CERN in 1983. Together with the photon, it is a neutral mediator of the electroweak force, and it can decay to a quark-antiquark or lepton-antilepton pair. The main decay modes of Z boson together with the corresponding branching fractions are listed in Table 2.1. Though in about 70 % of cases Z boson decays to a quark-antiquark pair, it is not feasible to measure this process in proton-proton collisions due to the overwhelming hadronic background. Instead, since the experimental signature of leptons can be very accurately measured experimentally, the $Z \to l^+l^-$ processes are used for precision tests of QCD and electroweak interactions [16].

At hadron colliders Z bosons are mainly produced via the quark-antiquark annihilation mechanism, as shown in Figure 2.7. The production of Z boson decaying to leptons is commonly referred to as the Drell-Yan process, since in 1970 Drell and Yan suggested the description of the lepton-antilepton pairs production

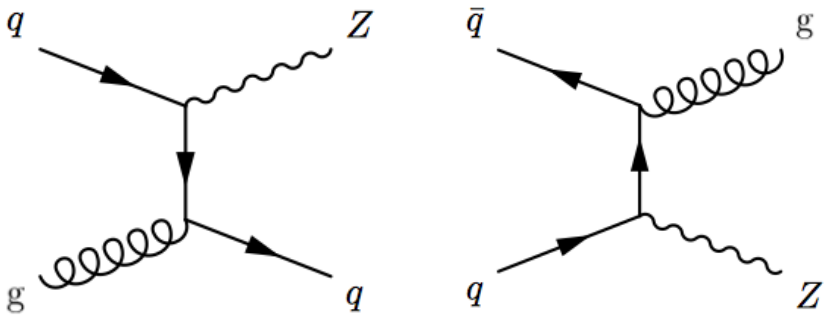


FIGURE 2.8: Z + jet production at LO in proton-proton collision.

in the high energy hadron-hadron collisions as: $q\bar{q} \to Z/\gamma^* \to l^+l^-$. Both photon and Z boson can mediate interaction in this process, but they dominate in different dilepton invariant mass M_{ll} regions. The Drell-Yan dilepton production via Z boson exchange has larger cross-section around the Z mass peak region, while photon exchange dominates at lower and higher M_{ll}.

2.4.1 Jets in Drell-Yan process

The production of jets together with the Z boson is an interesting process by itself and is also an important background for many SM processes and BSM searches. Feynman diagrams for LO Z + jet production in proton-proton collisions are shown in Figure 2.8. Due to the larger values of the gluon PDF $f_{g/p}(x, Q^2)$ in comparison to those of sea (anti)quarks (see Figure 2.4), the process with the gluon in the initial state has larger probability than the $q\bar{q} \to Zg$ production mechanism (Figure 2.8, right). Therefore, in about 80 % of cases the $gq \to Zq$ channel (Figure 2.8, left) is the dominant mechanism of the Z + jet production. Consequently, the final state observables in the Z + jet production are expected to be sensitive to the initial gluon density.

In the centre-of-mass system of proton-proton collisions, the momenta of two incoming partons, $\overrightarrow{p_1}$ and $\overrightarrow{p_2}$ are:

$$\overrightarrow{p_1} = x_1\sqrt{s}/2(1,0,0,-1), \quad \overrightarrow{p_2} = x_2\sqrt{s}/2(1,0,0,+1), \tag{2.9}$$

where x_1 and x_2 are corresponding fractions of the proton momenta carried by the first and the second parton, and $\sqrt{s}$ is the centre-of-mass energy of proton-proton

collision. The momentum transfer scale for the hard parton-parton interaction is:

$$Q^2 = (\overrightarrow{p_1} + \overrightarrow{p_2})^2 \simeq x_1 x_2 s. \tag{2.10}$$

The rapidity of the final state, y, is related to the momentum fractions carried by the initial state partons, following the energy and momentum conservation laws, and using Eq. 2.9, as:

$$y = \frac{1}{2} ln \frac{E - p_z}{E + p_z} = \frac{1}{2} ln \frac{x_1}{x_2}. \tag{2.11}$$

Using Eq. 2.10 and 2.11, the relation connecting the proton momentum fractions carried by initial state partons, the momentum transfer scale Q^2 and the rapidity of the final state:

$$x_{1,2} = \frac{Q}{\sqrt{s}} e^{\pm y}. \tag{2.12}$$

For $Z/\gamma^*(\rightarrow l^+l^-)$ + jet process the scale Q is related to the dilepton mass and the transverse momentum of associated jets, p_T^{jet}. Consequently, the $x_{1,2}$ values, which can be probed using Z/γ^* + jet events, depend on the transverse momentum of the jets and the rapidity of the final state system. The measurement of the $pp \rightarrow Z/\gamma^*(\rightarrow l^+l^-)$ + jet production cross-section as a function of the jet rapidity and the transverse momentum is, therefore, sensitive to the initial parton momentum fractions and to the QCD dynamics of the final state.

2.5 Monte Carlo simulation

There are several steps by which event generators build the structure of hadron-hadron collision [18]: a primary hard subprocesses, parton showers associated with initial and final particles in the subprocess, non-perturbative interactions that convert the showers into hadrons, additional interactions of other partons, beyond the leading ones in the hard subprocess, that give rise to the underlying event, and the decays of unstable particles that do not escape from the detector. These steps are briefly described below.

2.5.1 Hard subprocess

Hard scattering subprocesses can be calculated analytically at a fixed order of perturbative QCD using factorisation theorem, as described in Section 2.2. The calculation is performed for a given PDF set and for a certain choice of the factorisation and renormalisation scales. Current event generators allow to get the cross section at LO or NLO accuracy of α_S.

2.5.2 Parton shower

The coloured quarks and gluons involved in the hard scattering radiate gluons, which themselves are coloured particles and can trigger a new radiation, which leads to the showers. The parton shower can be simulated as step-by-step evolution in momentum transfer scales. It starts from the high scales associated with the hard subprocess and works downwards to the low scales of order 1 GeV, where perturbation theory breaks down and the confinement of the parton systems into hadrons starts.

In order to avoid double counting of partons between matrix elements and parton showers, they should be properly combined by dedicated matching procedure. There are two groups of combination approaches widely used. The matching approach means that high-order corrections to an inclusive process are integrated with the parton shower. Another approach is the merging strategy, which is usually defined in terms of a jet resolution scale, where any parton produced above this scale is generated with a corresponding matrix element, and any parton below this scale is generated by the shower.

2.5.3 Hadronisation

The recombination of the partons after the parton shower into the colour-neutral hadronic final state is called hadronisation. The two most used hadronisation models are the string model and the cluster model [19], which are schematically shown in Figure 2.9. The string model uses string dynamics to describe colour flux stretched between the initial $q\bar{q}$. It assumes that the string produces a linear confinement potential. The string breaks up into hadrons via $q\bar{q}$ pairs production in its intense colour field. Gluons which are produced in the parton shower give rise

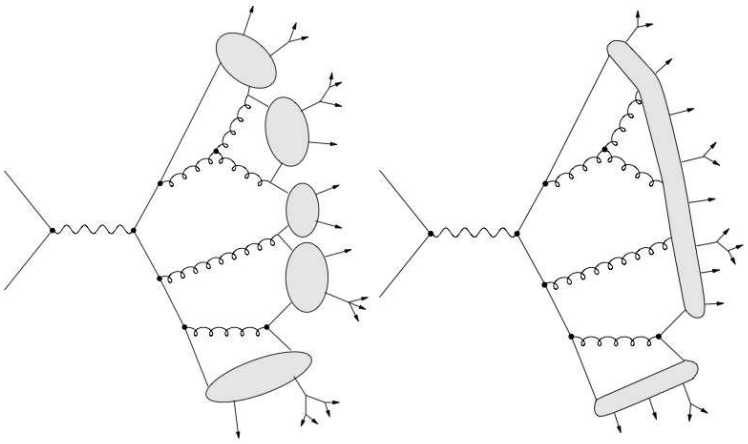


FIGURE 2.9: String and cluster hadronisation models, taken from [19]

to "kinks" on the string. The cluster model is based on the colour preconfinement. It starts from the gluon splitting to a $q\bar{q}$ pair, and quarks and antiquarks from different gluons are combined into colour-singlet clusters which typically decay into two hadrons.

2.5.4 Underlying event

The incoming hadrons are complex bound states of partons, and it is possible that in a collision more than one pair of partons interact. These additional interactions can produce partons which contribute to the final state observables. This part of event structure is described as the underlying event. These additional parton interactions happen at low energy transfer scales and involve colour and flavour connection to the hard scattering, and, therefore, cannot be described perturbatively. Since the dominant QCD cross-sections decrease rapidly with the transverse momentum of the hard scattering $\hat{p_T}$, the underlying event is modeled as minimum bias events with low $\hat{p_T} > \hat{p_T^{min}}$, where the parameter $\hat{p_T^{min}}$ is tuned to experimental data.

2.5.5 Monte Carlo generators for ATLAS

A large variety of Monte Catlo event generators are available to simulate high energy hadron-hadron collisions, studied at LHC. The following generators are used for this analysis:

• **Sherpa** [20] (Simulation of High Energy Reactions of PArticles) is a general-purpose event generator, which can can be used for the simulation of the physics of lepton-lepton, lepton-hadron, and hadron-hadron collisions as well as photon induced processes. The hard scattering processes are described by a tree-level leading-order matrix-element generator, and an additional QCD radiation off the initial and final states is described with a parton shower model. The matching between multiparton matrix elements and QCD parton showers is performed using the CKKW approach [21]. For the hadronisation process the cluster model is employed, and underlying events are described with a multiple interactions model.

• **Powheg** [22, 23] is a hard event generator for hadronic collisions. It is accurate at the next-to-leading order in QCD, and it can be interfaced to showers from Monte Carlo programs like Herwig and Pythia, in such a way that both the leading logarithmic accuracy of the shower and the NLO accuracy are maintained in the output.

• **Alpgen** [24] event generator is dedicated to the study of multiparton hard processes in hadronic collisions, based on the merging of multijets matrix elements with the shower development. The samples with different parton multiplicity are combined in an inclusive multiparton sample by weighting them by corresponding matrix element cross-sections. The MLM matching scheme [25] is used to identify overlapping contributions.

• **Herwig** [26] is a general-purpose Monte Carlo event generator, which includes the simulation of hard lepton-lepton, lepton-hadron and hadron-hadron scattering and soft hadron-hadron collisions. It uses the parton-shower approach for initial- and final-state QCD radiation, including colour coherence effects and azimuthal correlations both within and between jets. The first-order matrix elements are matched with the parton showers. A cluster model based on non-perturbative gluon splitting is used for a jet hadronisation, as well as for soft and underlying hadronic events.

• **Jimmy** [27] is a library of routines which should be linked with the Herwig Monte Carlo event generator. It is designed to generate multiple parton scattering events in hadron-hadron, photon-photon or photon-hadron events.

• **Pythia** is a program for the description of collisions at high energies between elementary particles such as e^+, e^-, p and $\bar{p}$ in various combinations. It contains theory and models for a number of physics aspects, including hard and soft

interactions, parton distributions, initial- and final-state parton showers, multiparton interactions, fragmentation and decay. In this analysis two different versions are used, Pythia 6 [28] and Pythia 8 [29].

• **Tauola** [30] program is designed for simulation of leptonic and semileptonic decays of τ leptons.

• **Photos** [31] package is developed for QED radiative corrections.

Chapter 3

Experimental Setup

3.1 The Large Hadron Collider

3.1.1 Overview

The Large Hadron Collider (LHC) is the largest and the most powerful particle accelerator in the world. It was built by the European Organization for Nuclear Research (CERN), in the tunnel which previously hosted the Large Electron Positron (LEP) collider, from 1998 to 2008. The LHC consists of a 27-kilometers ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way.

The LHC was designed to provide proton-proton collisions at a centre of mass energy of 14 TeV. As can be seen in Figure 3.1, the protons are pre-accelerated in several stages before being injected in the LHC ring. First a linear accelerator, LINAC 2, accelerates the protons to the energy of 50 MeV and injects them into the Proton Synchrotron Booster (PBS), where they are accelerated to 1.4 GeV. Subsequently they are injected into the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), where their energy is increased to 26 GeV and 450 GeV, respectively. And finally, they are accelerated in the LHC up to their operating energy, which was 3.5 TeV in 2010 and 2011, 4 TeV in 2012, 1.38 TeV during short period in 2013 and 6.5 TeV in 2015. Here the two beams, containing $\sim 10^{11}$ protons per bunch, are collided. The collisions took phase every 50 ns during Run 1 in 2009 – 2013, and every 25 ns in Run 2, started in 2015.

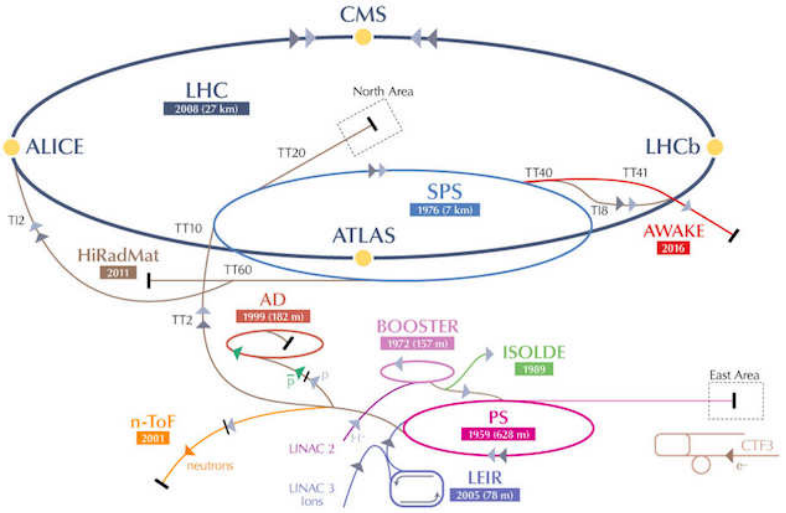


FIGURE 3.1: CERN Accelerator Complex and Experiments. Taken from [32]

The beams are collided at four interaction points of the ring, corresponding to the locations of the four LHC experiments. Two large multipurpose detectors, A Toroidal LHC ApparatuS (ATLAS) and a Compact Muon Solenoid (CMS) are designed to cover a wide range of high-energy physics goals. The design of another two mid-size detectors, A Large Ion Collider Experiment (ALICE) and a Large Hadron Collider beauty (LHCb), was optimised to study heavy ion collisions and b-quark physics, respectively.

3.1.2 Luminosity

The number of events N for a certain process is related to the corresponding cross section σ as:

$$\mathcal{N} = \mathcal{L} \cdot \sigma, \tag{3.1}$$

where $\mathcal{L}$ is the integrated luminosity, that can be written in terms of the accelerator parameters as [33]:

$$\mathcal{L} = \frac{n_b f_r n_1 n_2}{2\pi \sum_x \sum_y}, \tag{3.2}$$

where n_1 and n_2 are the numbers of protons per bunch in beam 1 and beam 2 respectively, $\sum_x$ and $\sum_y$ characterise the horizontal and vertical beam profile width. The revolution frequency is given as f_r, and the number of bunch pairs colliding per revolution as n_b.

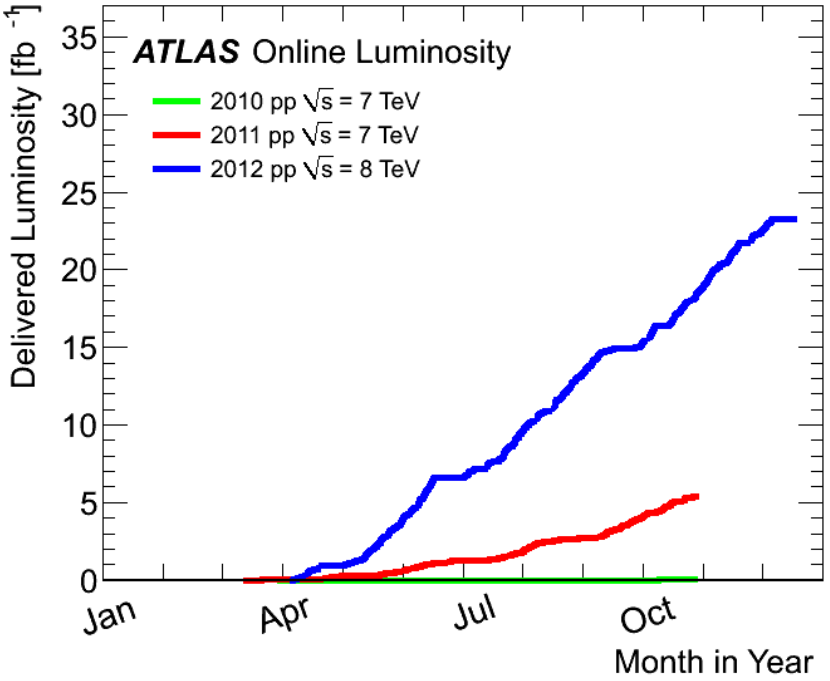


FIGURE 3.2: The delivered luminosity versus time for 2010, 2011, 2012 (p-p data only). Picture taken from [34]

The luminosity delivered to the ATLAS experiment for pp collisions in 2010, 2011 and 2012 is shown in Figure 4.1. In 2010 and 2011 at centre-of-mass energy $\sqrt{s} = 7$ TeV the ATLAS detector recorded 45 pb^{-1} and 5.08 fb^{-1} respectively, and in 2012 at $\sqrt{s} = 8$ TeV the ATLAS detector recorded 21.3 fb^{-1} [34].

3.2 The ATLAS detector

3.2.1 Overview

The ATLAS detector layout is shown in Figure 3.3. The detector is placed in one of the LHC interaction points, which defines the center of the ATLAS coordinate system. The ATLAS detector is forward-backward symmetric with respect to the interaction point. The z-axis of ATLAS coordinate system is defined to be along the beam pipe, and x-y plane is transverse to the beam direction. The positive x-axis points from the interaction point to the center of the LHC ring, the positive y-axis points upward to the surface of the earth. The azimuthal angle ϕ is measured from the x-axis around the z-axis and the polar angle θ is measured from the positive z-axis. The polar angle is often reported in terms of pseudorapidity,

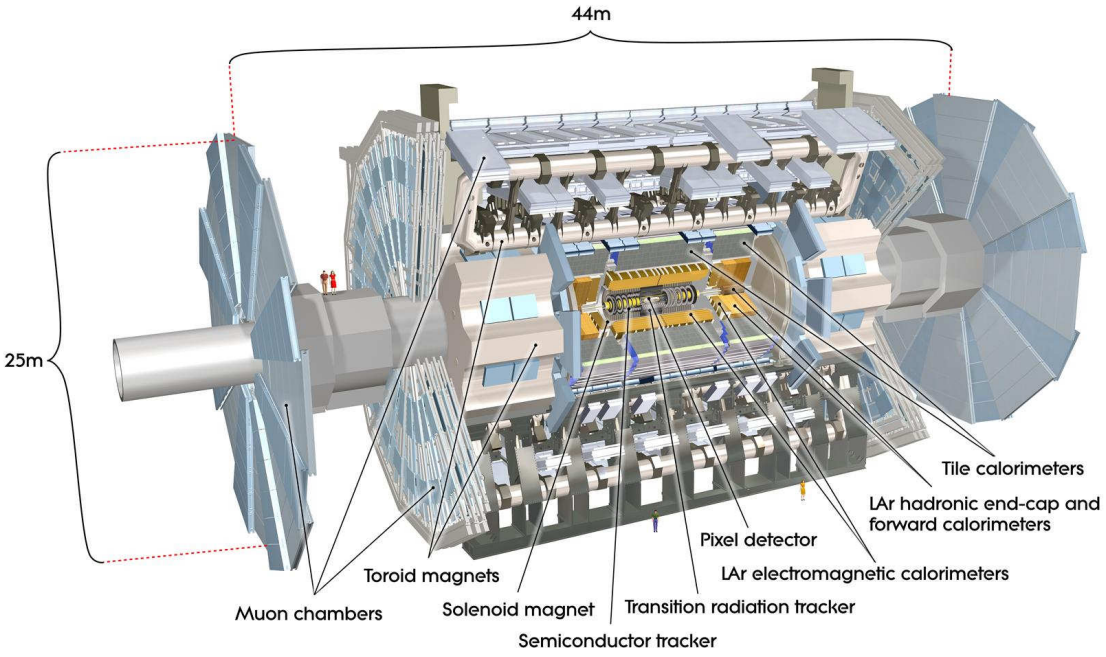


FIGURE 3.3: Schematic view of the ATLAS detector. Taken from [35].

which is defined as:

$$\eta = -\ln(\tan(\theta/2)). \tag{3.3}$$

In case of massive objects such as jets, the rapidity is used:

$$y = \frac{1}{2}\ln\frac{(E+p_z)}{(E-p_z)}. \tag{3.4}$$

The distance ΔR in the pseudorapidity-azimuthal angle space is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.5}$$

The ATLAS detector consists of four subsystems, which are designed to detect and measure different types of particles. The magnet system, described in Section 3.2.2, defines the design of the rest of the detector parts. It includes one thin solenoid surrounding the Inner Detector, and three large toroids around the calorimeter system, described in Section 3.2.3 and Section 3.2.4, respectively. The muon spectrometer, described in Section 3.2.5, is located in the toroidal magnetic field.

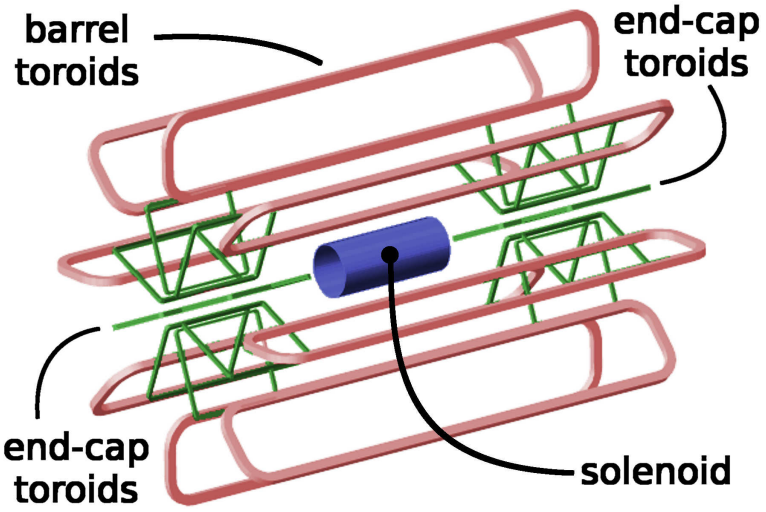


FIGURE 3.4: Spacial arrangement of the ATLAS magnet system. Picture taken from [36].

3.2.2 The magnet system

The ATLAS magnet system [35, 37] provides the bending power for the momentum measurement of charged particles. It consists of one solenoid providing the magnetic field for the inner detector, and three toroids, one barrel and two end-caps, which generate the magnetic field for the muon spectrometer. The spacial arrangement of the ATLAS magnet system is shown in Figure 3.4.

• **The central solenoid** is aligned on the beam axis and provides an axial magnetic field of 2 T for the Inner Detector. The inner and outer diameters of the solenoid are 2.46 m and 2.56 m respectively, and its axial length is 5.8 m. The steel of the ATLAS hadronic calorimeter is used as a return yoke for the magnetic flux.

• **The barrel toroid** is the largest component of the ATLAS magnet system, with an axial length 25.3 m, and the inner and outer diameters 9.4 m and 20.1 m respectively. The magnet generates a toroidal magnetic field which has almost perpendicular direction with respect to tracks of particles. The toroid is composed of eight coils assembled radially and symmetrically around the beam axis, each encased in individual racetrack-shaped, stainless-steel vacuum vessel. The provided bending power is 2-6 Tm in the pseudorapidity range $|\eta| < 1.3$.

• **The end-cap toroids** provide the bending power about 4-8 Tm in the forward region of the ATLAS detector within the pseudorapidity range $1.6 < |\eta| <$

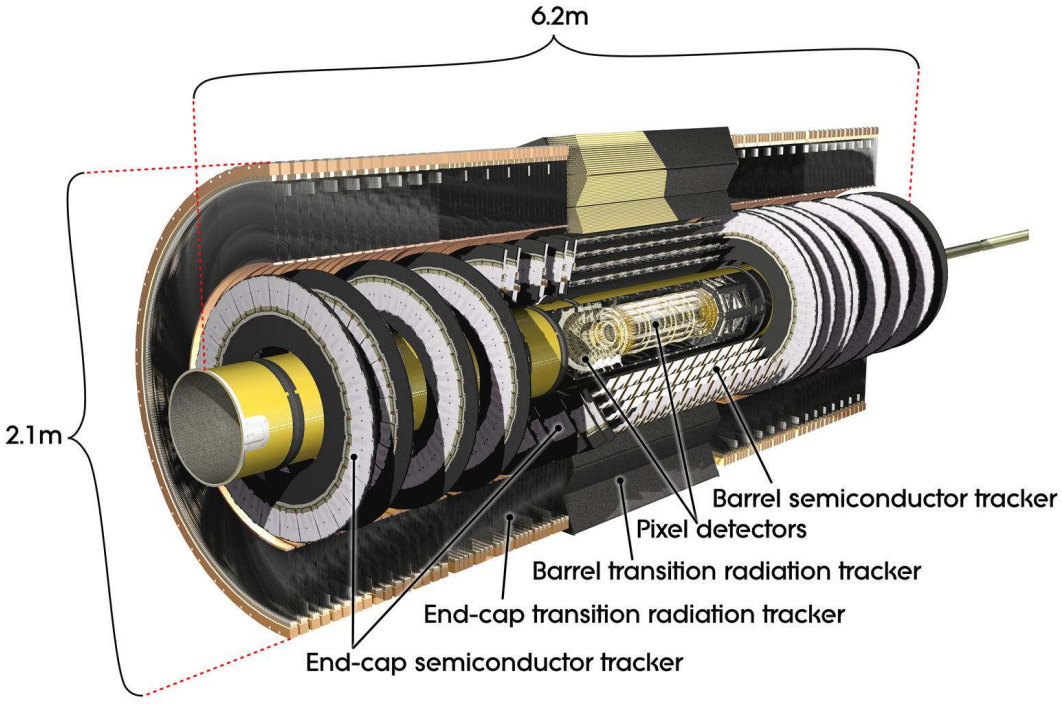


FIGURE 3.5: Cut-away view of the ATLAS Inner Detector, picture taken from [35].

2.7. The axial length of each magnet is 5 m, and inner and outer diameters 1.65 m and 10.7 m respectively. Each end-cap toroid consists of eight racetrack-shaped coils, which are assembled similarly to barrel toroid coils and housed as a single unit inside one large cryostat.

3.2.3 The inner detector

The ATLAS Inner Detector (ID) [35, 38], shown in Figure 3.5, is designed to reconstruct with high efficiency vertices and tracks of charged particles in the event. Its acceptance covers the pseudorapidity region of $|\eta| < 2.5$. The ID consists of three independent but complementary sub-detectors: the pixel and the silicon microstrip (SCT) trackers and the transition radiation tracker (TRT). The design is such that a particle originating from the interaction point crosses at least three pixel layers, four SCT strip layers and around 36 TRT straws, giving on average 44 measurements in total to reconstruct the trajectory and the momentum of a charged particle.

Mechanically, the ID consists of three units: a barrel part and two identical end-cap parts. Barrel detector layers are arranged on concentric cylinders around the beam axis within the pseudorapidity region $|\eta| < 1$, and the end-cap detectors are mounted on perpendicular to the beam axis disks.

• **The pixel detector** is designed to provide a high-granularity, high-precision measurements as close to the interaction point as possible in order to measure secondary vertices and contribute to primary vertex position measurement and the impact parameter resolution determination. The system consists of three barrel layers and three disks on each side segmented in $R-\phi$ and z with the same pixel size in $R-\phi \times z$ of $50 \times 400 \mu m^2$, providing three precision measurements over the full acceptance range. The position resolution parameters are 10 μm in the $R-\phi$ plane and 115 μm along z. The total number of readout channels in the pixel detector is approximately 80.4 million.

• **The silicon microstrip (SCT) detector** consists of four strip layers in the barrel and nine disks in each end-cap, providing at least four precision space point measurements per track in the intermediate radial range and contributing to the measurement of momentum, impact parameter and vertex position, as well as providing good pattern recognition by the use of high granularity. In the barrel region small-angle (40 mrad) stereo strips are used to measure both coordinates, with one set of strips in each layer parallel to the beam direction. In the end-cap detectors each disk has a set of radial strips and a set of stereo strips at an angle of 40 mrad. The mean pitch of the strips is around 80 μm. The position resolution parameters are 17 μm $(R-\phi)$ and 580 μm (z). The total number of readout channels in the SCT is approximately 60.3 million.

• **The transition radiation tracker (TRT)** provides the possibility of track-following as particles enter the EM calorimeter up to $|\eta| = 2.0$. This detector provides only $R-\phi$ information with a position resolution of 130 μm. Though the TRT has lower precision per point in comparison with silicon tracker, the larger number of measurements (around 36) and the higher radius allow to contribute to the particle momentum measurement. In addition to its tracking capabilities, the TRT provides electron identification ability through the detection of transition radiation photons. The total number of the TRT readout channels is approximately 351,000.

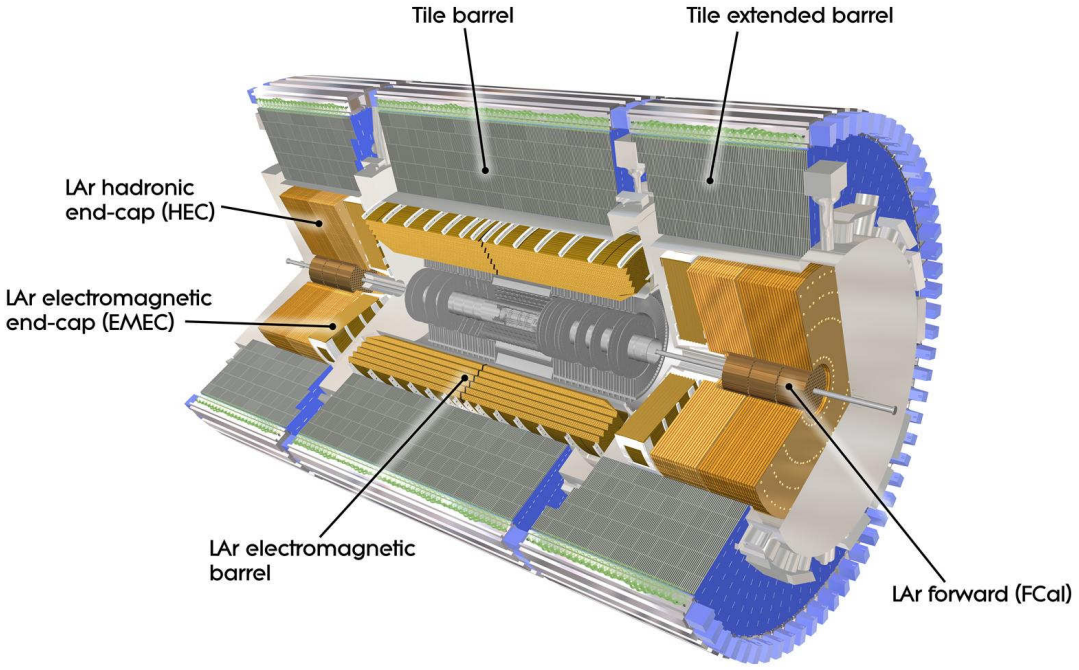


FIGURE 3.6: Cut-away view of the ATLAS Calorimeter system. Picture taken from [35].

3.2.4 The calorimeter system

The ATLAS calorimeter system [35, 39], shown in Figure 3.6, is designed to trigger on and to provide precision measurements of electrons, photons, jets, and missing E_T. It consists of electromagnetic and hadronic calorimeter subsystems covering pseudorapidity range $|\eta| < 3.2$ and $|\eta| < 4.9$, respectively.

Another important function of the calorimeter system is to provide containment for both electromagnetic and hadronic showers, stopping particles before they reach the muon spectrometer. The radiation length X_0 is used to quantify energy of electromagnetic showers absorbed by material. It is defined as the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung and as $7/9$ of the mean path for a pair production by a high-energy photon [40]. For hadronic showers, which are typically much longer, the nuclear interaction length λ is used. It is defined as a mean path length after which a fraction of $1/e$ hadrons have interacted with the traversed material.

The energy resolution of a calorimeter can be parametrised as [41]:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.6}$$

where a is stochastic term, b is the electronic noise term and c is the constant term which includes effects of detector instabilities and miscalibrations. The stochastic term takes into account the statistical fluctuations in the shower detection.

The calorimeters are housed in three cryostats at the detector radius less than 2.2 m, one barrel and two end-caps. The central cryostat contains the barrel electromagnetic calorimeter together with the central solenoid. Each end-cap cryostat contains an electromagnetic end-cap calorimeter, a hadronic end-cap calorimeter and a forward calorimeter. All these calorimeters use liquid argon as the active detector medium. At larger detector radius the tile hadronic calorimeter is used, with scintillator tiles as the sampling medium and steel as the absorber medium.

3.2.4.1 The electromagnetic calorimeters

The electromagnetic calorimeters are lead/liquid-argon detectors with accordion-shaped absorbers and electrodes. The accordion geometry provides a full coverage in ϕ without cracks and allows to have several active layers in depth, three in the pseudorapidity region devoted to precision measurements $|\eta| < 2.5$ and two in $2.5 < |\eta| < 3.2$. Mechanically, the electromagnetic calorimeters are divided into the barrel part and two end-caps.

• **The barrel EM calorimeter** consists of two half-barrels centered around z-axis with a small gap of 4 mm in between at $z = 0$, covering together the pseudorapidity region $|\eta| < 1.475$. It has three layers in depth with different granularity, as shown in Figure 3.7. Each half-barrel is divided into 16 modules, each covering $\Delta\phi = 22.5°$. The total thickness of the barrel part is minimum 22 radiation lengths X_0.

• **The end-cap EM calorimeters** consist of two wheels, one on each side of the electromagnetic barrel. Each wheel is mechanically divided into two co-axial wheels with a 3 mm wide boundary in between at $|\eta| = 2.5$. The outer wheel covers pseudorapidity region $1.375 < |\eta| < 2.5$, and the inner wheel covers the region $2.5 < |\eta| < 3.2$. Like the barrel part, the outer wheel is divided in depth into three layers, while the inner wheel has two layers. Each end-cap wheel is

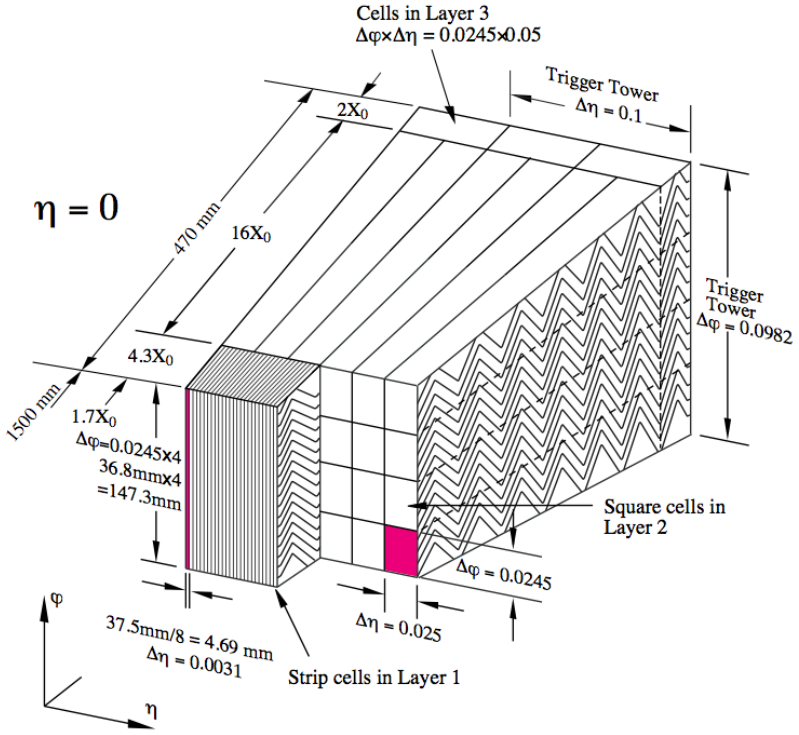


FIGURE 3.7: Sketch of an electromagnetic barrel module where different layers are visible. The granularity in η and ϕ of the cells and trigger towers is also shown. Taken from [35].

also divided into eight wedge-shaped modules. The total thickness of the end-cap calorimeter is at least 24 radiation lengths X_0.

• **The presampler** is a separate thin liquid-argon layer which is used to measure the energy lost by electrons and photons in front of the electromagnetic calorimeter. It covers the pseudorapidity region $|\eta| < 1.8$.

The total number of signal channels in the EM calorimeter is approximately 110000 and 64000 in barrel and end-cap parts, respectively. The design energy resolution of the electromagnetic calorimeter is parametrised by

$$\frac{\sigma(E[GeV])}{E[GeV]} = \frac{10\%}{\sqrt{E[GeV]}} \oplus \frac{170MeV}{E[GeV]} \oplus 0.7\%. \tag{3.7}$$

3.2.4.2 The hadronic calorimeters

The ATLAS hadronic calorimeter subsystem includes the tile calorimeter which covers the central pseudorapidity range, the liquid-argon hadronic end-cap calorimeter at the intermediate pseudorapidities and the liquid-argon forward calorimeter.

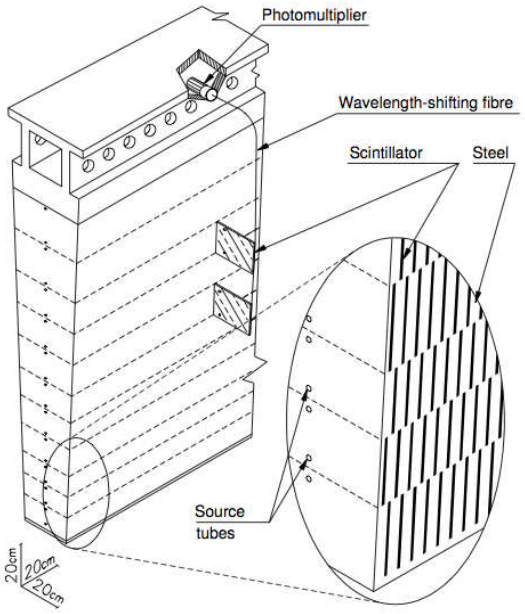


FIGURE 3.8: Sketch of a tile calorimeter module. The components of the readout are also shown. Picture taken from [35].

• **The tile calorimeter** is placed outside the EM calorimeter envelope. It is a sampling calorimeter which uses steel as the absorber and plastic scintillator as the active medium. Mechanically it is divided into a barrel part, which covers the pseudorapidity range $|\eta| < 1.0$, and two extended barrels covering the range $0.8 < |\eta| < 1.7$. Each barrel part is segmented into 64 modules of the $\Delta\phi = 5.625°$ coverage. The geometry of a tile calorimeter module is sketched in Figure 3.8. Radially the tile calorimeter is segmented into three layers, providing in total the depth of approximately 7.4 nuclear interaction lengths λ. The total number of readout channels is approximately 10000.

• **The LAr hadronic end-cap calorimeter (HEC)** is a cooper/liquid-argon sampling calorimeter. Each HEC consists of two cylindrical wheels, which are located directly behind the end-cap electromagnetic calorimeters. Each wheel is constructed of 32 wedge-shaped modules, as illustrated in Figure 3.9, and divided into two radial segments. The HEC covers pseudorapidity region $1.5 < |\eta| < 3.2$, overlapping slightly with the tile calorimeter and the forward calorimeter in order to reduce the drop in material density in the transition regions. There are in total 5632 readout channels in the HEC calorimeter.

• **The forward calorimeters (FCal)** provide both electromagnetic and hadronic coverage over the pseudorapidity region $3.1 < |\eta| < 4.9$. Each FCal

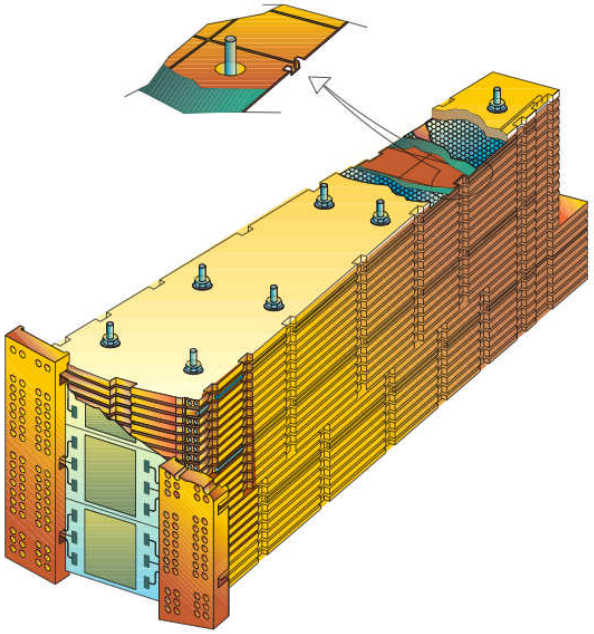


FIGURE 3.9: Schematic view of a HEC module. Picture taken from [35].

detector consists of three modules: one electromagnetic and two hadronic. For the electromagnetic layer copper is used as the absorber medium, while for hadronic modules mainly tungsten is chosen in order to provide containment and the lateral spread of hadronic showers. As mentioned above, liquid argon is chosen as the active medium. The FCal is approximately 10 nuclear interaction lengths λ deep. The total number of readout channels in the FCal is 3524.

The design energy resolution of the hadronic calorimeter is

$$\frac{\sigma(E[GeV])}{E[GeV]} = \frac{50\%}{\sqrt{E[GeV]}} \oplus 3\% \tag{3.8}$$

in the central region and about twice larger in the forward region.

3.2.5 The muon spectrometer

The muon spectrometer is designed to detect charged particles exiting the barrel and end-cap calorimeters, to measure their momentum in the pseudorapidity range $|\eta| < 2.7$ and to trigger on these particles in $|\eta| < 2.4$. The measurement is based on the deflection of the charged particles in the magnetic field provided by three large air-cone toroid magnets. Precision-tracking chambers in the barrel

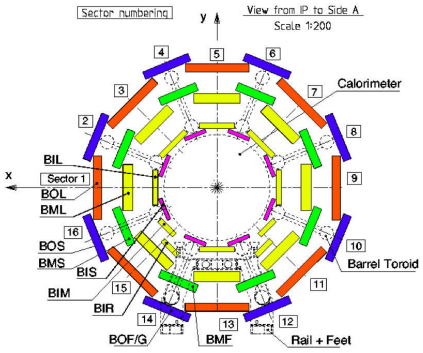


FIGURE 3.10: Cross-section of the barrel muon system in the non-bending plane. Taken from [35]

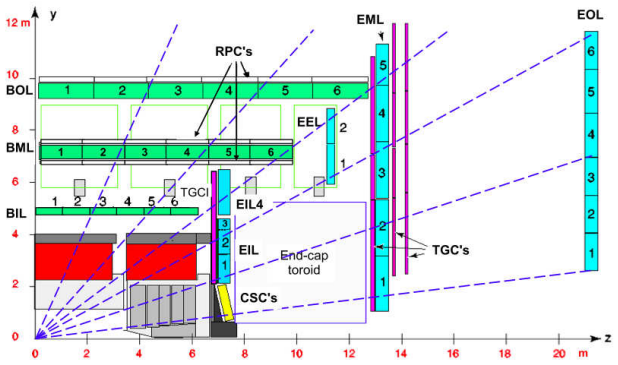


FIGURE 3.11: Cross-section of the muon system in the bending plane. Taken from [35]

region are located between and on the coils of the barrel toroid, while the end-cap chambers are in front and behind the end-cap toroids, as shown in Figure 3.11. The chambers in the barrel are arranged in three concentric cylindrical layers around the beam direction, as shown in Figure 3.10, and in the end-caps in four large wheels perpendicular to the z-axis. The momentum measurement is performed by the Monitored Drift Tube (MDT) chambers, covering the pseudorapidity range $|\eta| < 2.7$, except the innermost wheel where their coverage is limited to $|\eta| < 2.0$, and the Cathode-Strip Chambers (CSC), which are used in the region $2.0 < |\eta| < 2.7$. The precision-tracking chambers determine the coordinate of the muon track in the bending plane. To achieve higher resolution, the positions and internal deformation of MDT chambers are monitored by a high-precision optical system.

• **The MDT chambers** are made of multi-layers of pressured drift tubes with a diameter about 30 mm, filled with Ar/CO_2 gas mixture (93/7) at 3 bar. The electrons resulting from ionization are collected at the central tungsten-rhenium wire. The spacial resolution is of the order of 80 μm.

• **The CSC** are multiwire proportional chambers with the anode wires, made of tungsten with 3% rhenium, oriented in the radial direction. The cathods are segmented: one has strips parallel to wires measuring transverse coordinate, and another perpendicular, providing precision coordinate. Chambers are filled with Ar/CO_2 gas mixture (80/20). The spacial resolution is about 60 μm.

• **The alignment system** is the system of optical alignment sensors, which purpose is to continuously monitor and measure muon chambers positions and deformations during the data-taking.

In order to provide a capability of the muon system to trigger on muon tracks, the precision-tracking chambers are complemented by a system of fast trigger chambers. The Resistive Plate Chambers (RPC) are used for triggering in the barrel region $|\eta| < 1.05$, and the Thin Gap Chambers (TGC) are used in the end-cap regions $1.05 < |\eta| < 2.4$. The trigger chambers measure both η and ϕ coordinates of the track. After matching of the MDT and trigger chamber hits in the bending plane, the trigger chamber's coordinate in the non-bending plane is taken as a second coordinate of the MDT measurement.

• **The RPC** are gaseous parallel electrode-plate detectors made of two resistive phenolic-melaminic plastic plates, which are kept parallel to each other at a distance of 2 mm. The electric field between the plates is ∼4.9 kV and the volume is filled by the gas mixture $C_2H_2F_4/Iso - C_4H_{10}/CF_6$ $(94, 7/5/0.3)$. The spatial resolution is of the order of few cm and the time resolution is about 1 – 2 ns.

• **The TGC** are multiwire propotional chambers, where the wire-to-cathode distance of 1.4 mm is smaller than the wire-to-wire distance of 1.8 mm. The volume is filled with a gas mixture of CO_2/nC_5H_{12} (55/45). The time resolution is 2 – 3 ns.

3.2.6 The trigger system

The ATLAS trigger system consists of three levels of event selection: Level-1 (L1), Level-2 (L2), and event filter (EF). The L1 is completely hardware-based, while L2 and EF are software-based and form together the High-Level Trigger (HLT). The diagram of the trigger system is shown in Figure 3.12.

• **The L1 trigger** provides the initial event selection, looking for signatures of high-p_T muons, electrons, photons, jets and τ-leptons decaying into hadrons, as well as large missing E_T and large total E_T. The information used for triggering is provided by the Resistive Plate Chambers and Thin-Gap Chambers in the muon spectrometer and calorimeter subsystems. The maximum L1 accept rate is 75 kHz and the latency is about 2.5 μs.

• **The L2 trigger** is seeded by the Regions-of-Interest (ROI), which are defined by the L1 trigger as the regions with possible trigger objects. The L2 uses RoI information on coordinates, energy and type of signatures. It reduces the event rate to 3.5 kHz with an average event processing time about 40 ms.

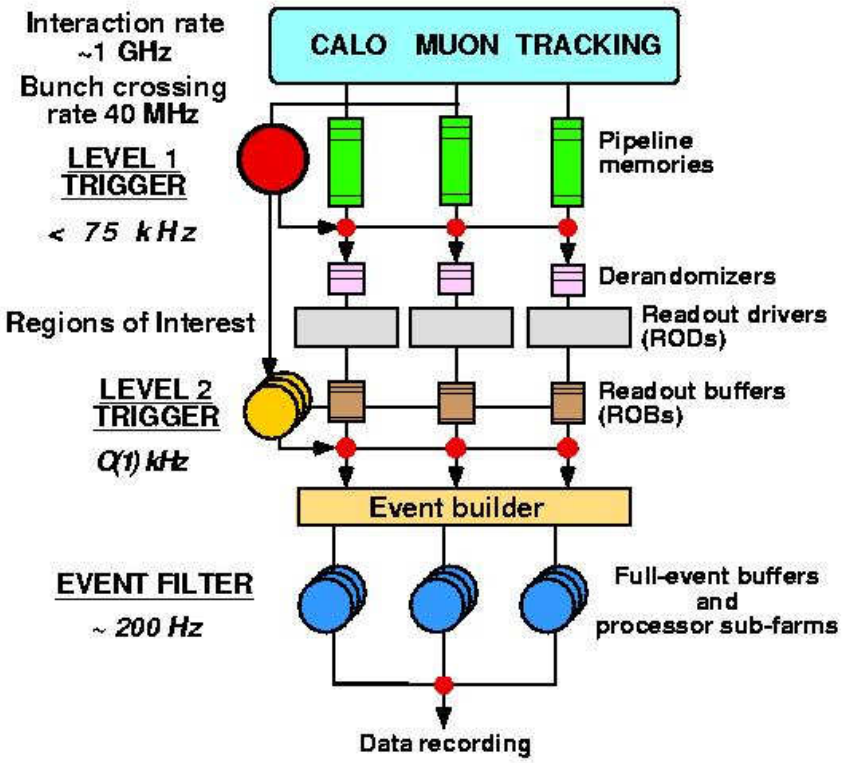


FIGURE 3.12: Block diagram of the trigger system. Taken from [35]

• **The EF** uses offline analysis algorithms on fully-built events which pass the L2 trigger. It reduces the event rate to about 200 Hz with an average time per event about 4 s.

Chapter 4

Data and Monte Carlo Samples

4.1 Data taking

During 2012 the ATLAS detector collected about 21.3 fb^{-1} of proton collision data at the center of mass energy $\sqrt{s} = 8$ TeV. Since detector conditions and beam parameters were changing during the year, the data are split into *data periods* which correspond to stable detector conditions. The list of periods with corresponding integrated luminosities is shown in Table 4.1

For physics measurements it is important to have data which satisfy the quality constrains. A set of data quality (DQ) flags includes an information about sub-detector systems and combined performance to define luminosity blocks of data which should be used for physics analysis. Such "good" luminosity blocks

Data period	Start date (day/month)	Luminosity $[fb^{-1}]$
A	4/4	0.84
B	1/5	5.3
C	1/7	1.54
D	24/7	3.37
E	23/8	2.7
G	26/9	1.3
H	13/10	1.56
I	26/10	1.06
J	2/11	2.72
L	20/11	0.89

TABLE 4.1: The luminosity per data period taken by the ATLAS detector in 2012. Numbers taken from [42].

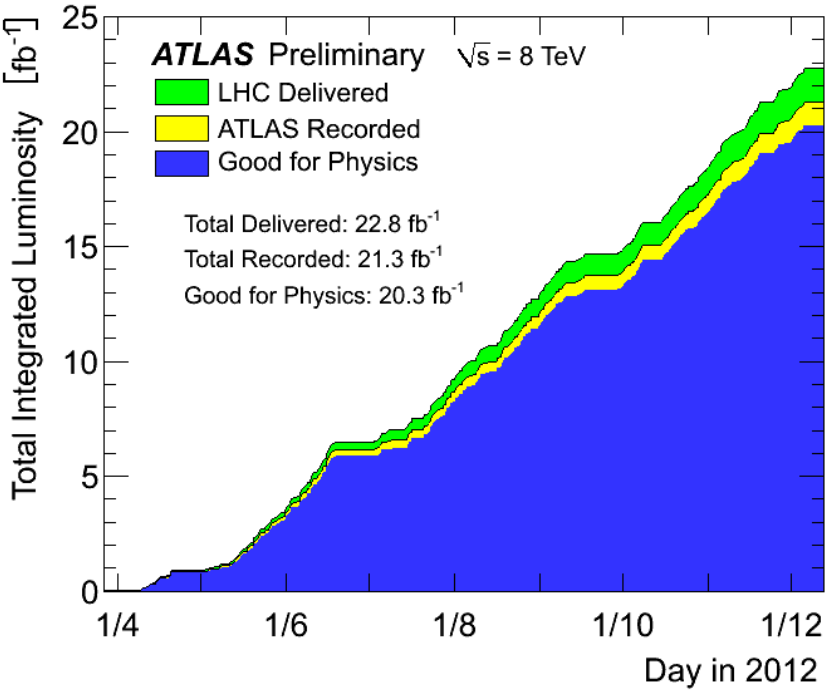


FIGURE 4.1: Total integrated luminosity and Data Quality in 2012, taken from [34]

are combined to Good Run Lists (GRL), which could be different for various analysis, depending on the specific physics requirements. For this analysis the standard GRL is used, which includes an information from all components of the detector[1]. The integrated luminosity of the 2012 data after application of this GRL is 20.3 fb^{-1} (see Figure 4.1).

4.2 Monte Carlo samples

Various Monte Carlo samples are used for this analysis in order to estimate a fraction of background from different processes, to calculate detector-to-hadron level unfolding efficiencies and matrices, and to evaluate the systematic uncertainties on the final results.

Several simulated $Z/\gamma^*(\rightarrow e^+e^-)$ + jet signal samples are used. The main one, which shows the best agreement with the data, and therefore is chosen for the unfolding procedure, (described in Section 10), is simulated with the Sherpa event

[1] data12_8TeV.periodAllYear_DetStatus-v61-pro14-02_DQDefects-00-01-00_PHYS_StandardGRL_All_Good.xml

Process	Generator	MC ID	σ [nb] (LO)	ϵ_{filter} [%]	Number of events
$Z \to ee$	Powheg+Pythia 8	147806	1.1099×10^{0}	100.00	9994580
$Z \to ee$	Sherpa	147770	1.2079×10^{0}	100.00	9999568
$Z \to ee$ $(N_p = 0)$	Alpgen+Pythia 6	147105	7.1997×10^{-1}	100.00	6298988
$Z \to ee$ $(N_p = 1)$	Alpgen+Pythia 6	147106	1.7570×10^{-1}	100.00	8184476
$Z \to ee$ $(N_p = 2)$	Alpgen+Pythia 6	147107	5.8875×10^{-2}	100.00	3175991
$Z \to ee$ $(N_p = 3)$	Alpgen+Pythia 6	147108	1.5636×10^{-2}	100.00	894995
$Z \to ee$ $(N_p = 4)$	Alpgen+Pythia 6	147109	4.0116×10^{-3}	100.00	398597
$Z \to ee$ $(N_p = 5)$	Alpgen+Pythia 6	147110	1.2592×10^{-3}	100.00	229700
$Z \to ee$ $(N_p = 0)$	Alpgen+Jimmy	107650	7.1182×10^{-1}	100.00	6604283
$Z \to ee$ $(N_p = 1)$	Alpgen+Jimmy	107651	1.5517×10^{-1}	100.00	1329994
$Z \to ee$ $(N_p = 2)$	Alpgen+Jimmy	107652	4.8745×10^{-2}	100.00	404798
$Z \to ee$ $(N_p = 3)$	Alpgen+Jimmy	107653	1.4225×10^{-2}	100.00	109998
$Z \to ee$ $(N_p = 4)$	Alpgen+Jimmy	107654	3.7595×10^{-3}	100.00	30000
$Z \to ee$ $(N_p = 5)$	Alpgen+Jimmy	107655	1.0945×10^{-3}	100.00	10000

TABLE 4.2: Monte Carlo samples used to simulate signal $Z/\gamma^*(\to e^+e^-)$ + jet process for this analysis.

generator, using the CT10 [43] PDF set. Further $Z/\gamma^*(\to e^+e^-)$ + jet samples are also used for comparison with data and for study of systematic effects. These samples are simulated with the Powheg event generator using CT10 PDFs in combination with the parton shower from Pythia8 with AU2 [44] tune, and with the Alpgen generator using CTEQ6L1 [45] PDFs in combination with the parton shower from either Pythia6 or Jimmy. The summary of $Z/\gamma^* \to e^+e^-$ samples used in this analysis is presented in Table 4.2

Several SM processes can produce events which look like $Z/\gamma^*(\to e^+e^-)$ + jet events and can pass analysis selection [46, 47]:

- $Z/\gamma^* \to \tau^+\tau^-$: contributes as background when both taus decay into electrons,
- $W \to e\nu$: contributes when a jet is misidentified as an electron,
- $W \to \tau\nu$: similarly to $W \to e\nu$, a small contribution is expected when τ decays into electron,
- WW, WZ, ZZ: contribute as background when electrons originate directly from W or Z decays and from semi leptonic charm or bottom decays,
- $t\bar{t}$, single top: contributes through final states containing two electrons, where the electrons result from W decays or semi leptonic charm or bottom decays,

Process	Generator	MC ID	σ [nb] (LO)	ϵ_{filter} [%]	Number of events
$Z \to \tau\tau$	Powheg+Pythia8	147808	1.1099×10^{0}	100.00	4999692
$W \to e\nu$	Sherpa	147774	1.1866×10^{1}	100.00	39997817
$W \to \mu\nu$	Sherpa	147775	4.7910×10^{0}	100.00	39996634
$W \to \tau\nu$	Sherpa	147776	1.1858×10^{1}	100.00	6998385
WW	Herwig	105985	3.2501×10^{-2}	38.21	2499890
WZ	Herwig	105987	1.2009×10^{-2}	30.53	1999995
ZZ	Herwig	105986	4.6915×10^{-3}	21.15	494999
$t\bar{t}$	Powheg+Pythia6	117050	2.1084×10^{-1}	54.31	49973332
t_{s-chan}	Powheg+Pythia6	110119	1.6424×10^{-3}	100.00	5999781
Wt	Powheg+Pythia6	110141	2.1478×10^{-3}	100.00	999692
t_{t-chan}	Powheg+Pythia6	110090	1.7520×10^{-2}	100.00	4994481
$\bar{t}_{t-chan}$	Powheg+Pythia6	110091	9.3932×10^{-3}	100.00	4999879

TABLE 4.3: Monte Carlo samples used to simulate various background processes for this analysis.

- multijet background includes semi-leptonic decays of heavy quarks, hadrons misidentified as leptons and electrons from conversions. The total contribution from these processes is estimated using data-driven method.

Corresponding Monte Carlo simulated samples are used in this analysis to study background contributions from the described above processes, except the multijet background. The $Z/\gamma^* \to \tau^+\tau^-$ sample, as the signal one, is simulated with Powheg using CT10 PDFs and showered by Pythia8 using AU2 tune. For W decays simulation Sherpa event generator with CT10 PDF is used. Events with diboson decays are generated by Herwig using CTEQ6L1 PDF set, interfaced to Tauola/Photos programs, and using AUET2 [48] tune. All single top and $t\bar{t}$ processes are simulated with Powheg using CTEQ6L1 PDF in combination with Pythia6 and Photos/Tauola, using Perugia2011C [49] tune. Table 4.3 summarises all simulated background processes with corresponding Monte Carlo generators and cross-section values.

Chapter 5

Event Reconstruction and Selection

5.1 Tracks and vertices

Tracks are reconstructed in the Inner Detector using the sequence of algorithms, as described in [50]. The inside-out algorithm starts from 3-point seeds in the silicon detector and adds hits moving away from the interaction point using a combinatorial Kalman filter. Ambiguities in the track candidates are resolved and tracks are extended into the TRT. The inside-out algorithm is a baseline algorithm for reconstruction of primary charged particles. As a second step, tracks are searched starting from reconstructed segments in the TRT and moving back to the interaction point by adding hits in the silicon detector. This back-tracking is mainly designed to reconstruct secondary particles. Finally, tracks with reconstructed TRT segments, but with no extension to the silicon detector are referred as TRT-standalone tracks.

Primary vertices are reconstructed using an iterative vertex finding algorithm [51]. Vertex seeds are found by looking for the global maximum in the distribution of z-coordinates of the reconstructed tracks at the beamline, computed at the point of the closest approach to the beam spot position. An iterative χ^2-fit is made, using the seed position and tracks around it. If track positions are displaced by more than 7 σ, they are used to seed new vertices. The procedure is repeated until no unassociated tracks are left in the event or no additional vertices can be found. Vertices are required to have at least two reconstructed tracks.

5.2 Electron reconstruction and identification

5.2.1 Electron reconstruction

The standard electron reconstruction procedure in the *central region* of the ATLAS detector ($|\eta| < 2.47$) is based on clusters reconstructed in the electromagnetic (EM) calorimeter, which are then associated to tracks of charged particles reconstructed in the Inner Detector. To reconstruct EM clusters sliding-window algorithm [52] is used. It proceeds in three steps: towers building, precluster finding and cluster filling.

• **Towers building.** The η - ϕ space of the EM calorimeter is divided into a grid of $N_\eta \times N_\phi = 200 \times 256$ towers of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, corresponding to the granularity of the calorimeter middle layer. The energy of the cells in all longitudinal layers are summed to get the tower energy. The energy of the cells spanning several towers is distributed according to the fractional area of the cells intersected by each tower.

• **Sliding-window precluster finding.** To reconstruct EM clusters, seed clusters of longitudinal towers with an energy above the threshold $E_T = 2.5$ GeV are searched by the sliding-window algorithm. A window of a fixed size $N_\eta \times N_\phi = 3 \times 5$ towers is moved across each element of the tower grid. If E_T of the window is a local maximum and is above an energy threshold, a precluster is formed. The size of the window and the threshold value are optimised to obtain the best efficiency of the precluster search and to limit the rate of fake preclusters due to the calorimeter noise.

The position of the precluster is computed as the energy-weighted η and ϕ barycenters of all cells within a fixed $N_\eta \times N_\phi$ size window around the tower. The size of this window is usually smaller than what is used for the tower search, to make the position computation less sensitive to the noise in the calorimeter. Duplicated clusters are removed, if the distance between two clusters is smaller then $\Delta\eta \times \Delta\phi$, only the precluster with the larger E_T is kept.

• **EM clusters filling.** Cells positioned within a rectangle of size $N_\eta \times N_\phi$, which is centered on the layer-dependent seed position, are assigned to the EM cluster. In the middle layer the positions of preclusters $\eta_{precluster}$, $\phi_{precluster}$ are used as seed positions. The barycenter coordinates are then recomputed including

cells from the middle layer. In the strip layer the position η_{middle}, ϕ_{middle} from the middle layer is used as a seed position. And, finally, in the presampler and the back layers the strip and the middle layer barycenters are used as seed positions.

Within a tracking acceptance $|\eta| < 2.5$ a track which matches the identified cluster is searched. Reconstructed tracks are matched to seed clusters by extrapolating them from their last measurement point to the middle layer of the calorimeter. The distance between the track impact point and the cluster position is required to be within $|\Delta\eta| < 0.05$. To account for bremsstrahlung losses, the size of the sign corrected $\Delta\phi$ window is 0.1 on the side where the extrapolated track bends as it traverses the solenoidal magnetic field and is 0.05 on the other side. An electron is reconstructed if at least one track is matched to the seed cluster. If several tracks are matched to the same cluster, the one with the smallest distance $\sqrt{\Delta\eta^2 + \Delta\phi^2}$ to the seed cluster is chosen considering that tracks with silicon hits have a priority over all other tracks.

The electron cluster is recomputed then using 3×7 (5×5) longitudinal towers of cells in the barrel (endcaps). These lateral cluster sizes are optimised to take into account the different overall energy distributions in the barrel and endcap calorimeters.

The four-momentum of central electrons is computed using information from both the cluster and the best matched track. The energy is taken from the cluster, while η and ϕ correspond to the track position at the vertex.

Algorithms used to reconstruct the electron candidate are assigned an *author* value, which distinguishes different selections [53].

• **author = 1** means that the object has been found by only the standard (cluster based) algorithm.

• **author = 2** means that the object has been found by only the track based algorithm. Typically it corresponds to electrons which are under the E_T threshold of the sliding window cluster algorithm.

• **author = 3** means that the object has been found by both the standard (cluster based) and the track-based algorithms.

5.2.2 Electron identification

Not all objects built by the electron reconstruction algorithms are signal electrons. Hadronic jets and electrons from photon conversions give main contributions to background. The electron identification in the central ($|\eta| < 2.47$) region is based on calorimeter, tracking and combined discriminating variables and allows to reject most of the background keeping high efficiency for signal electrons. The list and a short description of variables are given in Table 5.1.

The cut-based selections, *Loose*, *Medium* and *Tight*, are optimised in bins of electrons $|\eta|$ and E_T.

• **Loose** set of cuts provides high signal electron selection efficiency (about 95 %) but low background rejection. It includes cuts on the shower shape variables in both the strip and the middle layers and on the hadronic leakage. This selection also requires track quality hits in the pixel and silicon layers as well as loose track-cluster matching.

• **Medium** selection increases the background rejection with respect to the loose selection with signal selection efficiency decreasing down to about 85 % by adding cuts on the shower shape variable in the back layer, track transverse impact parameter, number of hits in the B-layer, number of hits in the TRT and ratio of the number of high-threshold hits to the total number of hits in the TRT.

• **Tight** set of cuts in addition to *Medium* selection tighter track-cluster matching is applied together with veto on the electron candidates matched to reconstructed photon conversions. Also cut on ratio of the cluster energy to the track momentum is added. The resulting signal electron selection efficiency is about 78 %.

5.3 Jet reconstruction

For current analysis jets are reconstructed using the anti-k_T algorithm with distance parameter R = 0.4 using FastJet [55] software package. The total four-momentum of a jet is defined as the four-momenta sum of all its constituents.

Type	Description	Name
Hadronic leakage	Ratio of E_T in the first layer of the hadronical calorimeter to E_T of the EM cluster (used in range $\|eta\| < 0.8$ or $\|\eta\| > 1.37$)	R_{Had1}
	Ratio of E_T in the hadronical calorimeter to E_T of the EM cluster (used in range $0.8 < \|\eta\| < 1.37$)	R_{Had}
Back layer of EM calorimeter	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter	f_3
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i\eta_i^2)/(\sum E_i) - ((\sum E_i\eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of the cell i and the sum is calculated in a window of 3×5 cells	$w_{\eta 2}$
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position	R_η
Strip layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i(i - i_{max})^2)/(\sum E_i)}$, where i rus over all strip in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η, and i_{max} is the index of the highest-energy strip	w_{stot}
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	E_{ratio}
Track quality	Number of hits in the B-layer	n_{Blayer}
	Number of hits in the pixel detector	n_{Pixel}
	Number of total hits in the pixel and SCT detectors	n_{Si}
	Transverse impact parameter	d_0
TRT	Total number of hits in the TRT	n_{TRT}
	Ratio of the number of high-threshold hits to the total number of hits in the TRT	F_{HT}
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the middle layer and the extrapolated track	$\Delta\phi_2$
	Ratio of the cluster energy to the track momentum	E/p
Conversion	Veto electron candidates matched to reconstructed photon conversion	isConv

TABLE 5.1: Definition of electron discriminating variables, taken from [54].

The inputs to the jet algorithm are stable simulated particles (*truth jets*), reconstructed tracks in the Inner Detector (*track jets*) or energy deposits in the calorimeter (*calorimeter jets*).

5.3.1 Calorimeter jet inputs

Fine segmentation of the ATLAS calorimeter allows reconstruct jets with high precision. To provide an input to the jet reconstruction a "topological" clustering algorithm [52] is used, the basic idea of which is to group neighbouring cells with energies above an expected noise level into a cluster. The noise level is defined as the RMS of the total noise for the current gain and conditions. The "topological" clustering algorithm consists of the following steps:

Clusters	EM	Combined
Calorimeters	EM only	All
Seed signal definition	E	$\|E\|$
Cluster cut before splitting	$E_T > 5$ GeV	$E_T > 0$ GeV
t_{seed}	6	4
$t_{neighbour}$	3	2
t_{cell}	3	0

TABLE 5.2: Parameters of the topological cluster algorithm. Taken from [52].

- **Finding seeds**. All cells with the signal-to-noise ratio higher than seed threshold t_{seed} are put into a seed list. Each seed cell forms a "proto-cluster". The expected contribution from the pile-up is added to the noise in quadrature.

- **Finding neighbors**. All cells in the seed list are ordered in decreasing signal-to-noise ratio. For each seed cell its neighbour cells are considered. If the neighbour cell is not included in the seed list, but has signal to noise ratio higher than medium threshold $t_{neighbour}$, it is added to the list of neighbour seeds and is included in the adjacent proto-cluster. If the cell is adjacent to more than one proto-cluster the clusters are merged. If the signal to noise ratio is above the low cell threshold t_{cell} but below $t_{neighbour}$, the cell is included only in the first adjacent proto-cluster, which is the one providing the more significant neighbour to this cell. When all seeds are processed the neighbour seed list becomes the new seed list. The procedure is repeated until the seed list is empty.

- **Finalise**. All resulting proto-clusters are ordered by E_T and those with E_T higher then a certain threshold are converted to clusters.

In the standard ATLAS reconstruction, two types of topological clusters are built: the electromagnetic clusters and the combined clusters. The parameters defining these two cluster types are listed in Table 5.2.

The energy of the topological cluster (topo-cluster) is at the electromagnetic (EM) scale. It can be calibrated for the presence of hadronic activity using the local cell weighting (LCW) [56]. In both cases, the mass of the topological clusters is set to zero. The topo-clusters at the EM scale calibrated topo-clusters have an average response equal to 1 for electrons/photons. In the LCW calibration the topo-clusters are classified as electromagnetic or hadronic and then a weighting

scheme corrects for the different electron-to-pion response in the calorimeters. Then, out-of-cluster correction and dead material correction are applied for the LCW topo-clusters.

5.3.2 Jet cleaning

Jets produced in proton-proton collisions must be distinguished from background jet candidates originating from other sources. The main contributions to such background jets come from the beam-gas events, in which a proton of the beam interacts with the residual gas in the beam pipe, beam-halo events, caused by interactions in the beam-line away from the detector, cosmic ray muons and calorimeter noise [57].

Various properties of event topology and jet kinematics are used to reject fake jets from background processes. These properties include the quality of the energy reconstruction at the cell level, jet energy deposits in the direction of the shower development, and tracks matched to the jets. Several sets of discrimination criteria with different background rejections and signal jet selection efficiencies are developed: *Looser, Loose, Medium* and *Tight*. The *Looser* set has the highest jet selection efficiency and the *Tight* set is the one with the best background rejection.

Some of the Tile calorimeter modules were either temporarily or permanently masked throughout the data taking periods. This can lead to the poor reconstruction of jets falling into such modules, in particular of the high-p_T jets, since they are more collimated and, therefore, can be fully contained in modules. In order to correct for events with poorly reconstructed jets, two jet properties are exploited. First, BCH_CORR_CELL, which is an approximate value of energy in a masked cell, calculated using neighboring cells. Second, the fraction of the energy of the jet deposited in the EM calorimeter, f_{EM}. Two flags can be used for the analysis purposes to reject poorly reconstructed jets: *BadMediumBCH*, which flags all jets fallling into the core of the masked region, and jets falling close to the edge of the masked regions with a given combination of f_{EM} and BCH_CORR_CELL, and *BadTightBCH*, which flags all jets which fall into the core or the edge of the masked regions.

5.3.3 Jet vertex fraction

In order to suppress jets from overlapping proton-proton collisions, jet-vertex association algorithm [58] is applied. The information from the Inner Detector about charged particle tracks associated to a jet is used in order to determine the jet origin. The algorithm uses the collections of reconstructed tracks, jets and primary vertices in the event. It proceeds in three steps:

• Tracks passing a set of quality criteria and calorimeter jets within the tracking acceptance $|\eta| < 2.5$ are selected.

• Selected tracks are associated to jets. There are two standard methods of association [59]: the simple ΔR matching, and *ghost association* [60], which is used for 2012 data analysis. In the latter method, tracks are used as inputs to the jet algorithm, but with their p_T scaled by very low number. As a result, they do not affect the reconstruction of calorimeter jets, but it is possible to identify which tracks were clustered into which jets. The ghost association procedure provides a more correct matching, since it properly accounts for jets with irregular cross-sectional shapes.

• Each associated track is required to originate from the reconstructed primary vertex in the event.

The jet vertex fraction (JVF):

$$JVF(jet_i, vertex_j) = \frac{\sum_k p_T(trk_k^{jet_i}, vertex_j)}{\sum_n \sum_l p_T(trk_l^{jet_i}, vertex_n)}, \tag{5.1}$$

where k runs over all tracks associated to a jet i and originating from a vertex j, n runs over all reconstructed primary vertices and l runs over all tracks associated to a jet i and originating from any reconstructed vertex, is defined for each jet with respect to each primary vertex. The jet i has a fraction JVF of its total matched-track momentum originated from the vertex j.

5.4 Monte Carlo reconstruction

Event simulation proceeds through generation, detector simulation and digitisation steps, and finally reconstruction. After Monte Carlo generation of proton-proton

collision event, the generated particles are propagated through the model of the ATLAS detector. The interactions of the particles with the detector material are simulated using Geant4 (GEometry ANd Tracking) software [61]. The results of interactions are digitised into the detector signals, and stored in the same format as data. Reconstruction of the Monte Carlo event is done using the same algorithms. The resulting output is referred to as *reconstructed-level* Monte Carlo events and can be directly compared to reconstructed data.

Generated Monte Carlo events are stored separately and referred to as *truth-level* events. These events are used for calibration and correction purposes, as well as for data unfolding. Main truth-level objects used in this analysis are electrons and jets. Based on the fact that charged leptons can emit photons via QED Final State Radiation (FSR), three different definitions can be used for electrons at the truth-level [62]:

• the "born" electrons are the electrons at the lowest-order diagram in the QED for the process under study, which are prior to QED FSR,

• the "bare" electrons are electrons after QED FSR,

• the "dressed" electrons are "bare" electrons which include all FSR photons in cone of $\Delta R = 0.1$ around the "bare" electron.

For this analysis "dressed" electrons are used, since this definition is the closest to the reconstructed-level electrons, where possible radiation close to the electron will be included into the electron cluster.

The jets at the particle-level are reconstructed running jet clustering algorithm on simulated stable particles. The stable particles are defined as those with a lifetime $c\tau_0 > 10$ mm. A particle is considered to be interacting, if most of its energy is expected to be deposited in the calorimeter; muons and neutrinos are considered to be non-interacting and are excluded from the list of input particles.

5.5 $Z/\gamma^*(\rightarrow e^+e^-)$ + jet event selection

$Z/\gamma^*(\rightarrow e^+e^-)$ + jet events are selected with a set of requirements ("cuts") in both data and Monte Carlo simulation. After selection of two "good" electrons, they are combined to form a Z boson. The four-vector of each electron is constructed

by taking the energy information from electromagnetic calorimeter cluster and the position information from the corresponding matched track in the Inner Detector.

Events are selected using standard Good Run List (see Section 4.1) and dielectron trigger, where both electrons required to have $p_T > 12$ GeV. Events are further required to have a primary vertex with at least three matched tracks. LAr cleaning and Object Quality cuts are applied to reject bad quality clusters or fake clusters originating from calorimeter problems. The LAr cleaning procedure identifies and removes the events with noise bursts and data integrity errors in the LAr calorimeter. The Object Quality test checks the pulse shape in the cells associated to the electron candidate.

Electrons are required to have $p_T > 20$ GeV and to be in the central rapidity region $|\eta| < 2.47$, excluding the transition region $1.37 < |\eta| < 1.52$. Each electron candidate should satisfy the *Medium* selection (see Section 5.2.2) and pass *author* 1 or 3 requirement (see Section 5.2.1). Exactly two electrons should pass the selection defined above. These electrons are required to have opposite charge and their invariant mass should be in the range $66 < M_{ee} < 116$ GeV.

Events are further required to have at least one jet with $p_T > 25$ GeV in the $|\eta| < 4.4$ range. Since leptons are also reconstructed as jets with jet algorithms, all jets are required to be isolated from the selected electrons within a distance $\Delta R = 0.35$ in $\eta - \phi$ space in order to avoid overlap. Each jet with $p_T > 50$ GeV and $|\eta| < 2.4$ (tracking acceptance region) is required to have JVF > 0.25 (see Section 5.3.3). In addition, the jet quality selection criteria, discussed in Section 5.3.2 are applied: each jet is required to pass $Medium$ criteria and not to be flagged as $BadMediumBCH$.

For quark-gluon tagging purposes the slightly different jet selection is used. In addition to the selection described above, the p_T of subleading jets in the event is required to be $p_T > \max(0.3p_T(Z), 40\text{GeV})$ in order to suppress multijet events. This requirement helps to select LO Z + 1 jet events, where the leading jet is expected to be quark-like in 80 % of cases. An isolation requirement for a leading jet is applied in order to suppress contributions from both pile-up and QCD color-connection to nearby jets: ΔR(leading jet, nearest jet) > 0.8, for nearest jets with $p_T > 15$ GeV. A balance between the Z boson and the leading jet is ensured by requiring $\Delta\phi$(leading jet, Z) > 2.0. The rapidity region is restricted to tracking acceptance region $|\eta| < 2.5$. The selection requirements are different in order to

Cut	$Z/\gamma^*(\to e^+e^-)$ + jet analysis	quark-gluon tagger
Dielectron trigger	✓	✓
GRL	✓	✓
LAr and OQ cuts	✓	✓
N_{tracks} from PV ≥ 3	✓	✓
Exactly two OS electrons	✓	✓
$66 < M_{ee} < 116$ GeV	✓	✓
$\lvert\eta_{el}\rvert < 2.47$, excl. $1.37 < \lvert\eta_{el}\rvert < 1.52$	✓	✓
$E_{T,el} > 20$ GeV	✓	✓
Author 1 or 3	✓	✓
Medium ID	✓	✓
$\lvert\eta_{jet}\rvert$	< 4.4	< 2.5
$p_T^{jet} > 25$ GeV	✓	✓
JVF > 0.25	✓	✓
$\Delta R(jet, el) \leq 0.35$	✓	✓
Jet cleaning	Medium	Loose
BCH cleaning	Medium	-
$p_T^{jet_2} < \max(0.3 \times p_T(Z), 40$ GeV)	-	✓
$\Delta\phi$(leading jet, Z)> 2.0	-	✓
ΔR(leading jet, nearest jet)> 0.8	-	✓

TABLE 5.3: Summary of the event selection for the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross-section measurement and the quark-gluon tagging study.

be consistent with the selection of other samples used for quark-gluon tagger: each jet is required to pass *Loose* criteria and no BCH cleaning is applied.

The analysis selection is summarised in Table 5.3.

Chapter 6

Data and Monte Carlo Corrections

6.1 Pile-up reweighting

Additional proton-proton collisions in the current bunch crossing (in-time pile-up) and those from past and future collisions in preceding and following bunch crossings (out-of-time pile-up) result in energy deposits in the detector systems. Each simulated event is produced under a particular detector condition and with a given number of pile-up events overlayed.

Out-of time pile-up is characterised by an average number of interactions per bunch crossing, μ. The re-weighting of the μ distribution is applied to Monte Carlo events in order to correct for differences between the μ distribution used to produce the sample and that measured in data. The pile-up related event weight correction factor for an event produced in Monte Carlo sample with a given detector condition A in the i bin of μ distribution is defined as [63]:

$$w_{pileup} = \frac{L_{A,i}/L}{N_{A,i}/N}, \tag{6.1}$$

where L is the total integrated luminosity in data, $L_{A,i}$ is the integrated luminosity of the data measured with the detector condition A in μ distribution bin i. N is the sum of generator weights in the produced Monte Carlo sample and $N_{A,i}$ is the sum of generator weights of the part of the sample produced under the detector condition A in the same bin i of μ distribution.

Additional correction factor of 1.09 is applied to μ in simulated samples to account for differences between data and Monte Carlo in the number of primary vertices in a given event, N_{PV}, which characterises in-time pile-up.

6.2 Electron corrections

6.2.1 Reconstruction, identification and trigger scale factors

The electron detection efficiency in the ATLAS detector is split into different components, namely trigger, reconstruction and identification efficiencies. The reconstruction efficiency, ϵ_{reco}, is defined as the ratio of the number of electrons, which are reconstructed as a cluster and matched to a track passing the track quality criteria, to the number of electrons, which are reconstructed as a cluster with or without a matching track. The identification efficiency, $\epsilon_{id}|_{reco}$, is calculated as the ratio of the number of electrons passing a certain identification selection, *Loose, Medium* or *Tight*, to the number of electrons with a matching track passing the track quality requirements. And the trigger efficiency, $\epsilon_{trigger}|_{reco,id}$, is calculated as the ratio of the number of electrons passing trigger selection to the number of electrons passing a given identification criterion. The full efficiency ϵ_{total} for a single electron can be written as [54]:

$$\epsilon_{total} = \epsilon_{reco} \times \epsilon_{id}|_{reco} \times \epsilon_{trigger}|_{reco,id} \tag{6.2}$$

The electron efficiencies are measured with the tag-and-probe method using $Z \to ee$ and $J/\psi \to ee$ decays. In both cases, strict selection criteria are applied on one of the two decay electrons called *tag*, and the second, *probe* electron candidate is used for the efficiency measurements.

The accuracy of the Monte Carlo based detector simulation in modelling the electron efficiency plays an important role in the cross-section measurements. The simulated samples need to be corrected in order to reproduce as close as possible the efficiencies measured in data . This is done by using the correction factors defined as the ratio of the efficiency measured in data to that in the simulation. Since the electron efficiencies depend on the transverse energy and pseudorapidity, the efficiency measurements are performed in two dimensions.

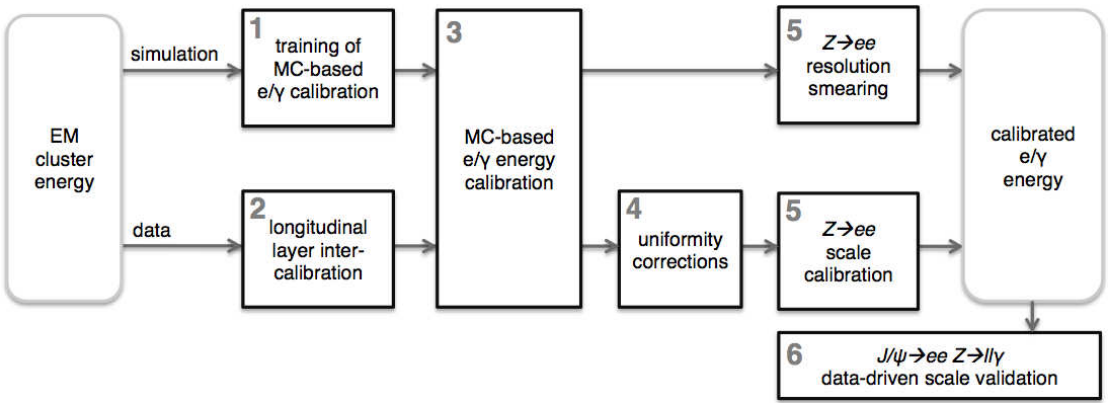


FIGURE 6.1: The scheme of the procedure used to calibrate the energy response of electrons and photons in ATLAS, taken from [64].

6.2.2 Electron energy scale and resolution

The electron and photon energy response calibration procedure is schematically shown in Figure 6.1 and described in details in [64]. As the first step of this procedure, in simulated Monte Carlo samples the EM cluster properties are calibrated to the original (truth-level) electron and photon energy using multivariate techniques with energy and cluster position related variables. In measured data the scales of the different longitudinal layers of the EM calorimeter are equalised in order to ensure the correct extrapolation of the energy response in the full p_T range. Then the Monte Carlo based e/γ response calibration is applied to the cluster energies in both data and Monte Carlo simulation. After this, additional corrections are applied to data only, in order to account for response variations in specific detector regions which are not included in the simulation, e.g. non-optimal HV regions, geometric effects such as intermodule widening or biases associated with the LAr calorimeter electronic calibration.

As the last step, the overall energy response in data is calibrated as a function of pseudorapidity, using $Z \to ee$ events. The energy miscalibration is parametrised as:

$$E^{data} = E^{MC}(1 + \alpha_i), \tag{6.3}$$

where E^{data} and E^{MC} are the electron energy in data and simulation, respectively, and α_i is a deviation from the optimal calibration in a given bin i of η. For the dielectron invariant mass the effect of this miscalibration is:

$$m_{ij}^{data} = m_{ij}^{MC}(1 + \alpha_{ij}), \quad \alpha_{ij} \sim \frac{\alpha_i + \alpha_j}{2}, \tag{6.4}$$

neglecting second-orders and considering the angle between two electrons to be perfectly known. m_{ij}^{data} and m_{ij}^{MC} are the invariant mass of the electron pair reconstructed in i and j bins of η in data and Monte Carlo simulation, and α_{ij} is a shift of the mass peak.

Electron energy resolution $\frac{\sigma_E}{E}$ is assumed to be well modelled by the simulation up to a Gaussian constant term c:

$$\left(\frac{\sigma_E}{E}\right)^{data} = \left(\frac{\sigma_E}{E}\right)^{MC} \oplus c. \tag{6.5}$$

For each electron pair with rapidities of electrons η_i, η_j, the relative electron energy and dielectron invariant mass resolutions must satisfy:

$$\left(\frac{\sigma_m}{m}\right)_{ij}^{data} = \left(\frac{\sigma_m}{m}\right)_{ij}^{MC} \oplus c_{ij} = \frac{1}{2}\left[\left(\frac{\sigma_E}{E}\right)_i^{MC} \oplus c_i \oplus \left(\frac{\sigma_E}{E}\right)_j^{MC} \oplus c_j\right], \quad c_{ij} = \frac{c_i \oplus c_j}{2}, \tag{6.6}$$

where c_{ij} is a relative invariant mass resolution correction for (η_i, η_j) electron configuration.

To determine the α and c parameters, histograms of the $Z \to ee$ invariant mass are created separately in data and Monte Carlo for all pairs of electrons with pseudorapidity configuration (η_i, η_j). The optimal values, uncertainties and correlations of α_{ij} and c_{ij} are obtained using the χ^2 minimisation method. The individual electron energy scale and resolution corrections are obtained by solving the system of equations 6.4 and 6.6. Per-electron energy scale calibration factors are applied to electron and photon candidates in data, while the resolution correction is applied to simulation to match the data. The calibrated energy scale is validated with electron candidates from $J/\psi \to ee$ events and with photon candidates from $Z \to ee\gamma$ events in data.

6.3 Jet calibration

The reconstructed jets, built from topological clusters at the EM or LCW scale, are calibrated further to account for several different effects:

- Calorimeter non-compensation: the calorimeter response to hadrons is smaller than that to electrons with the same energy.

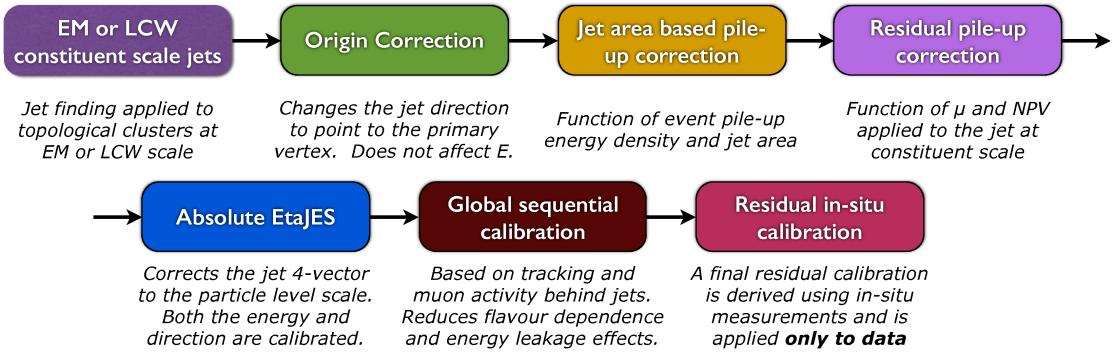


FIGURE 6.2: The scheme of the procedure used to calibrate EM and LCW jets, taken from [65].

• Dead material: some fraction of the energy is not measured due to an inactive detector material.

• Leakage of the energy outside the calorimeter.

• Out-of-cone losses: not all particles originating from the initial parton are included into the reconstructed jet due to such effects as magnetic field, which bends away charged particles.

• Energy deposits from jet constituents below calorimeter noise thresholds are not included in input topological clusters.

• Pile-up effects.

The jet calibration procedure used for 2012 data [65] is shown schematically in Figure 6.2 and described below.

Origin correction

The origin correction is the first step of the jet calibration. This correction modifies the directions of the topological clusters in such a way that they point to the event primary vertex instead of the center of ATLAS detector, as by default. The primary vertex is defined as the vertex with the highest $\sum p_T^2$ of associated tracks. The direction of the jet is recalculated then with the modified topological clusters. At this step the energy of the jet is unchanged.

Pile-up correction

As a second step, a correction to account for the energy offset caused by pile-up interactions is applied. An area based subtraction method [59, 60] is employed,

which uses information about the median of the p_T density, ρ, and the area of the jet, A. After this, a residual pile-up correction is applied. It is performed in bins of a number of primary vertices in a given event, N_{PV}, and average number of interactions in bunch crossing, μ. The pile-up subtracted p_T is given by [65]:

$$p_T^{corrected} = p_T^{jet} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \mu, \tag{6.7}$$

where α and β are derived from Monte Carlo and p_T^{jet} is the p_T of the jet before the correction.

Monte Carlo jet energy scale calibration

After the pile-up and origin corrections, the Monte Carlo based jet energy scale (JES) calibration is applied, which relates the reconstructed jet energy to the truth jet energy. The calibration factors are derived from an inclusive jet Monte Carlo sample with an additional requirement for jets to be isolated, i.e. no other jets with $p_T > 7$ GeV within a cone of $\Delta R = 1.5R$ for reconstructed and $\Delta R = 2.5R$ for truth jets.

Global sequential calibration

The global sequential calibration (GSC) [66] is derived for jets calibrated at the EM and LCW energy scale. It is used to reduce the dependence of the calorimeter response on the flavour of the parton initiating jet, and to improve the jet energy resolution. Several variables which carry information about the jet structure properties are used. For each such a variable x a correction is derived by inverting the average jet response, defined as $R = \langle p_T^{reco\ jet} / p_T^{truth\ jet} \rangle$, as a function of this variable:

$$C(x) = R^{-1}(x). \tag{6.8}$$

The average jet response is then scaled back to the initial value, such that after this correction the remaining dependence of the average jet response on x is removed without changing the average jet energy. This results in a reduction of the spread of the reconstructed calorimeter jet energy and therefore, an improvement in resolution. Corrections are applied sequentially in five stages, based on the following variables:

- the fraction of the energy deposited in the first layer of the Tile calorimeter,

• the fraction of the energy deposited in the third layer of the electromagnetic calorimeter,

• the number of tracks with $p_T > 1$ GeV associated to the jet,

• the p_T-weighted transverse width of the jet measured using tracks with $p_T > 1$ GeV associated to the jet,

• the number of segments in the muon spectrometer behind the jet.

Only the track-based and muon spectrometer corrections are applied to the LCW calibrated jets, since calorimeter calibrations are already included in the LCW calibration (see Section 5.3.1).

In-situ JES Calibration

The in-situ techniques are used as a final step of the JES calibration, using the balance between the transverse momentum of jets and well-measured physics objects, which are photons, Z bosons or sets of previously calibrated low-p_T jets. The correction which is applied to jets reconstructed in data is defined by the ratio:

$$\frac{R_{MC}}{R_{data}} = \frac{< p_T^{jet}/p_T^{ref} >_{MC}}{< p_T^{jet}/p_T^{ref} >_{data}}, \tag{6.9}$$

where p_T^{ref} is the transverse momentum of reference objects.

As a first step, dijet events are used to derive η-intercalibration, in which the energy of a jet in the forward region, $0.8 \leq |\eta_{jet}| < 4.5$, is calibrated using p_T of a balancing jet in the central region, $|\eta_{jet}| < 0.8$. This is done in order to remove any residual pseudorapidity dependence of the jet response after the Monte Carlo JES calibration. Next, JES correction factors are derived for jets in the central region $|\eta| < 0.8$, using the balance between Z boson or photon recoiling against central jets in the Z/γ + jet events. And finally, the calibration of high-p_T jets is derived using multijet events, where the highest-p_T jet is recoiling against the system of low-p_T jets, which are already calibrated by other, previously described in-situ methods.

Chapter 7

Background Extraction

As described in Section 4, several processes can contribute as backgrounds to $Z/\gamma^*(\rightarrow e^+e^-)$ + jet events selection. The contributions from electroweak processes, including top decays, are estimated using Monte Carlo simulation. The simulated samples are normalised to corresponding cross-sections in the highest available order, given in Table 7.1, and after the analysis selection are subtracted from data.

The multijet background is highly suppressed by the electrons identification requirements. However, since the jet production cross-section is about six order of magnitude larger than the $pp \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production cross-section, this process can still have significant contribution to the selected data sample. In

Process	σ [nb], higher order	Uncertainty up [nb]	Uncertainty down [nb]
$Z \rightarrow ee$	1.122 (NNLO) [67]	0.0561	0.0561
$Z \rightarrow \tau\tau$	1.122 (NNLO) [67]	0.0561	0.0561
$W \rightarrow e\nu$	12.087 (NNLO) [67]	0.60435	0.60435
$W \rightarrow \mu\nu$	12.087 (NNLO) [67]	0.60435	0.60435
$W \rightarrow \tau\nu$	12.087 (NNLO) [67]	0.60435	0.60435
WW	58.7×10^{-3} (pNNLO) [68]	3.0×10^{-3}	2.7×10^{-3}
WZ	20.3×10^{-3} (NLO) [69]	0.8×10^{-3}	0.8×10^{-3}
ZZ	7.2×10^{-3} (NLO) [70]	0.3×10^{-3}	0.2×10^{-3}
$t\bar{t}$	252.89×10^{-3} (NNLO+NNLL) [71]	13.3×10^{-3}	14.5×10^{-3}
t_{s-chan}	5.61×10^{-3} (NLO+NLL) [72]	0.22×10^{-3}	0.22×10^{-3}
Wt	22.37×10^{-3} (NLO+NLL) [73]	1.5×10^{-3}	1.5×10^{-3}
t_{t-chan}	87.8×10^{-3} (NLO+NLL) [74]	3.4×10^{-3}	1.9×10^{-3}

TABLE 7.1: Cross-sections for signal and background processes calculated to higher order in α_S expansion.

order to use the Monte Carlo prediction for the multijet production, huge amount of simulated event is needed, which makes the simulation impractical. In addition, such a complex process is difficult to model precisely, which makes the simulation unreliable. Therefore, a data-driven method is used in this analysis to estimate the multijet background contribution, as described below.

7.1 Multijet background estimation

The data-driven method allows to get a model independent description of the background processes with small statistical uncertainties. In this method a background enriched sample, referred as a *template sample*, is obtained from data using the selection close to the nominal analysis, but with suppressed signal. This selection has the following differences in comparison to the analysis selection described in Section 5.5:

• Additional two single electron triggers added to the main selection triggers through the OR condition, requiring at least one electron in the event passing *Medium* cuts and having E_T more than 24 or 60 GeV.

• Both electrons should have the same charge.

• Only one electron candidate should satisfy *Medium* but not *Tight* requirement; no ID cut on the other electron candidate applied.

Due to the single electron requirement, this selection includes also background contribution from W($\rightarrow e\nu$) + jet process. Therefore, the background from W($\rightarrow e\nu$) + jet process is estimated with data-driven method and the corresponding simulated sample is used for cross-checks only.

The number of background-like events in each bin i in the template sample is defined as:

$$N_{template,i} = N_{data,i}^{bkg\ enriched} - \sum_{j}^{MC} N_{MC,i}^{bkg\ enriched}, \tag{7.1}$$

where $N_{data,i}^{bkg\ enriched}$ and $N_{MC,i}^{bkg\ enriched}$ are the numbers of events selected from the data and Monte Carlo simulated samples using the background enriched selection, respectively. The sum runs over all simulated (signal and backgrounds) samples.

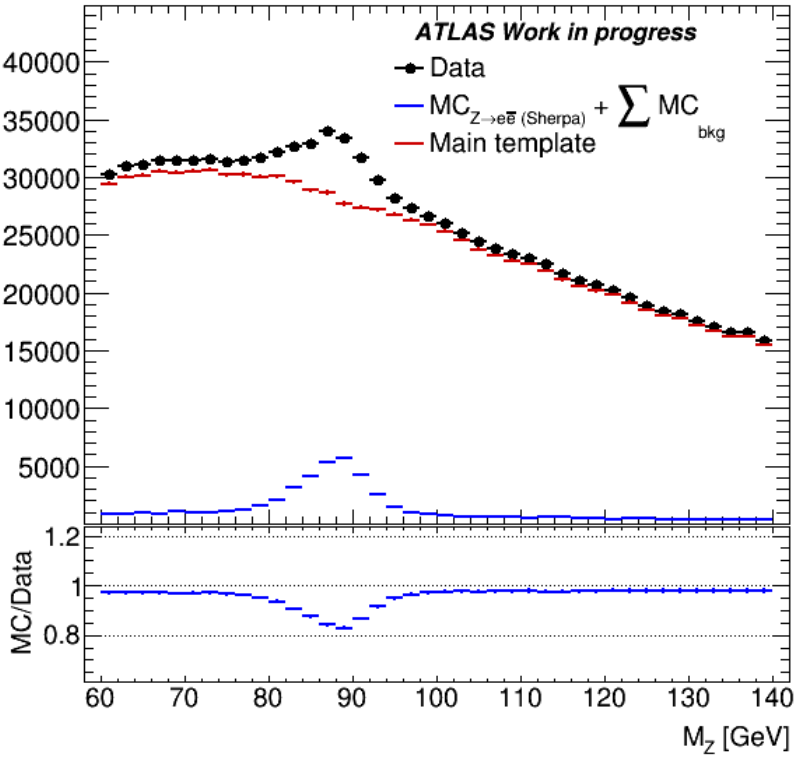


FIGURE 7.1: The selected template (upper plot) and the purity of the selected template sample (lower plot) as a function of M_{ee}.

The quality of the resulting template is estimated using purity $P_{template,i}$, defined as:

$$P_{template,i} = \frac{N_{data,i}^{bkg\ enriched} - \sum_j^{MC} N_{MC,i}^{bkg\ enriched}}{N_{data,i}^{bkg\ enriched}}. \tag{7.2}$$

The dielectron mass is the discriminant variable in this analysis, since the shape of M_{ee} distribution is different for signal and background processes and is not dependent on jet kinematics. The template purity as a function of M_{ee} variable is shown in Figure 7.1. The resulting purity is about 98 % at the tails of the M_{ee} distribution and decreases to about 82 % in the Z mass peak region.

To estimate the fraction of multijet and W($\to e\nu$) + jet background events in data, the dielectron mass distributions of the simulated signal and the background enriched template are fitted to the distribution in data in extended M_{ee} window $60 < M_{ee} < 140$ GeV. All Monte Carlo simulated backgrounds are subtracted from data before the fit is performed. The resulting M_{ee} distribution is shown in Figure 7.2. The scale factor obtained from the fit for the simulated signal sample,

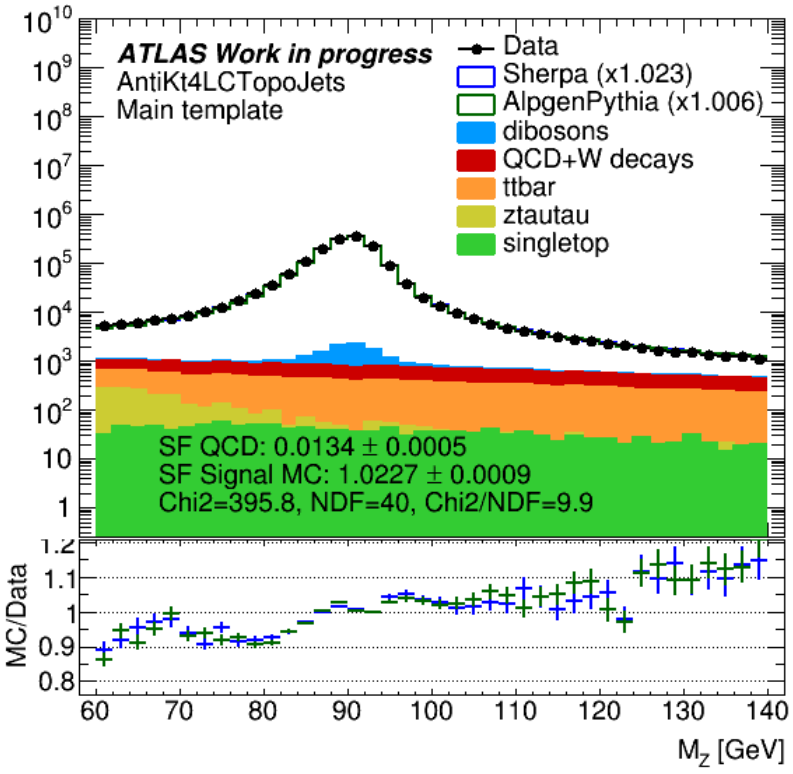


FIGURE 7.2: The M_{ee} distribution in data compared to simulated $Z(\rightarrow e^+e^-)$ + jets sample using Sherpa together with all backgrounds. Multijet background template and signal distribution are scaled by factors K_s and K_t obtained from the fit.

K_s, allows to check the compatibility of Monte Carlo predictions and data, and the scale factor for the background enriched template, K_t, is used for the normalization of other template sample distributions before subtracting them from data.

The uncertainties in the data-driven method of multijet and W($\rightarrow e\nu$) + jet backgrounds estimation are obtained in the following way:

• the background-enriched template selection is changed to select only one electron candidate with *Loose* but not *Medium* requirement,

• the fit range is varied from the nominal range of the M_{ee} window $60 < M_{ee} < 140$ GeV to $60 < M_{ee} < 100$ GeV and $80 < M_{ee} < 140$ GeV ranges,

• cross-sections of the background processes simulated by Monte Carlo are varied by their corresponding uncertainties, listed in Table 7.1, before subtracting them from the template.

The resulting uncertainty from the data-driven background estimation on the measured cross section is described in Section 11.2.3.

Chapter 8

Control Distributions

When all corrections are applied and the fraction of background processes is obtained, the description of data by simulated samples is studied. Data distributions are compared to $Z/\gamma^*(\rightarrow e^+e^-)$ + jet samples simulated by Powheg+Pythia8, Sherpa and Alpgen+Pythia6 event generators together with all backgrounds contributions.

The η and p_T distributions of the electrons coming from Z boson decay are shown in Figure 8.1. The discrepancies between data and Monte Carlo simulation are up to 10 % for Alpgen+Pythia6, tending to increase in the forward η region, while Powheg+Pythia8 and Sherpa describe data well. All three generators show

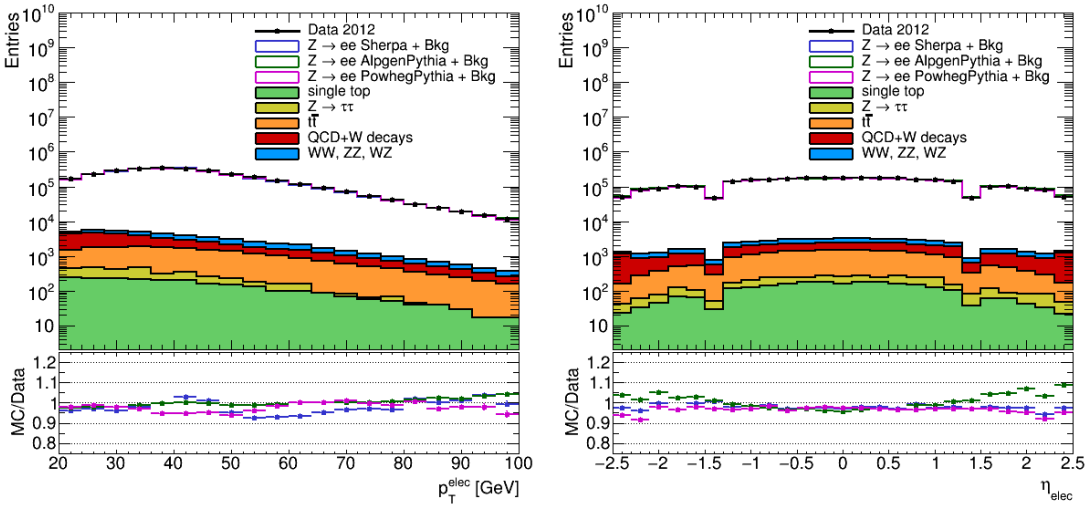


FIGURE 8.1: The distributions of p_T (left) and η (right) of electrons from Z boson decay after the $Z/\gamma^*(\rightarrow e^+e^-)$ + jet selection.

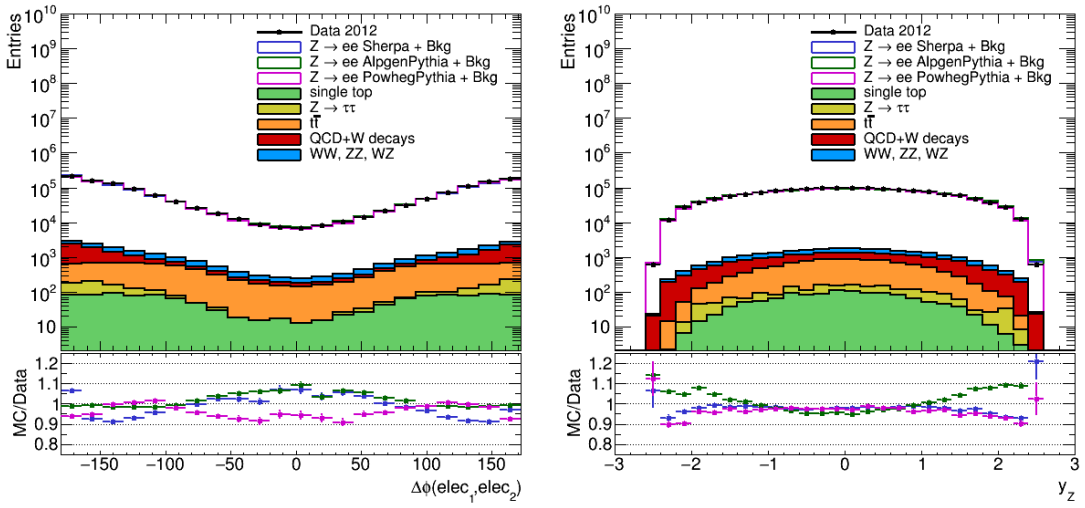


FIGURE 8.2: The distributions of $\Delta\phi$ between the electrons from Z boson decay (left) and y_Z (right) after the $Z/\gamma^*(\rightarrow e^+e^-)$ + jet selection.

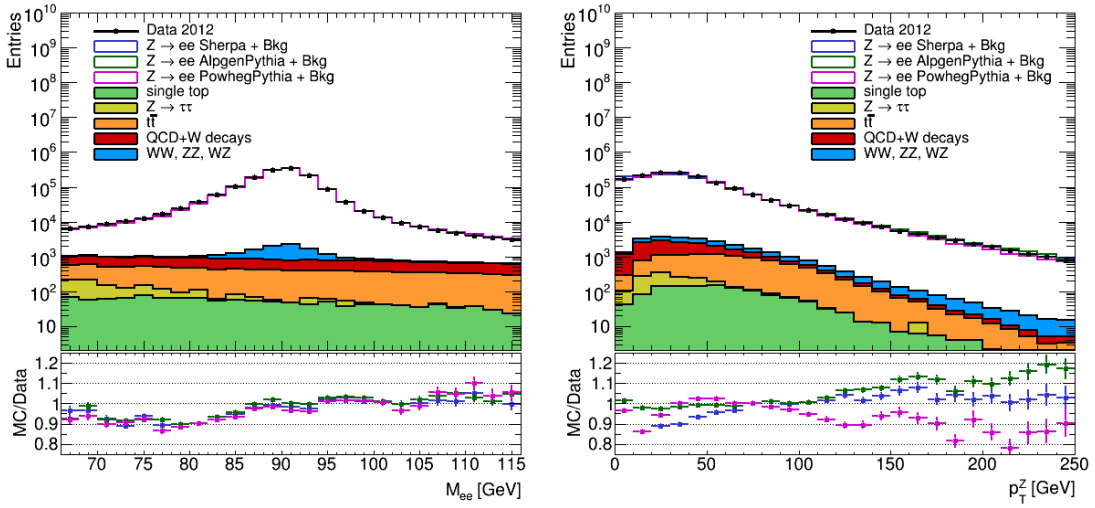


FIGURE 8.3: The distributions of M_Z (left) and p_T^Z (right) after the $Z/\gamma^*(\rightarrow e^+e^-)$ + jet selection.

up to 10 % differences in description of $\Delta\phi$ between the electrons coming from Z decay (see Figure 8.2, left plot). The central region of Z boson rapidity is described well by all three generators, while in the forward region Alpgen+Pythia6 prediction overestimates data and Sherpa and Powheg+Pythia8 predictions underestimate data by up to 10 %, as shown in Figure 8.2 (right plot).

The distributions of Z boson properties, M_{ee} and p_T^Z are shown in Figure 8.3. The difference between data and Monte Carlo is at the level of a few percent in the M_{ee} peak region and increases to about 10 % in the tails of the distribution.

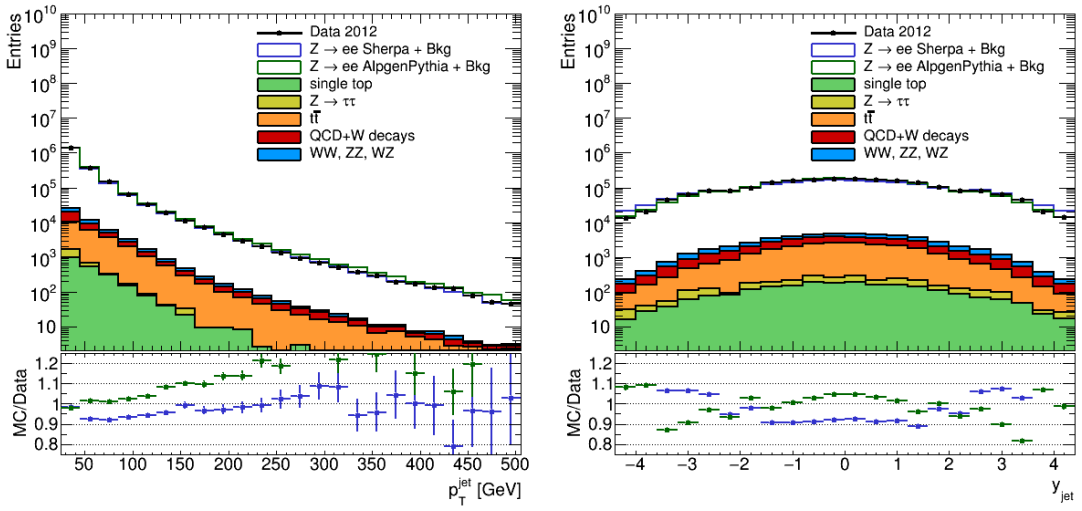


FIGURE 8.4: The distributions of p_T^{jet} (left) and y_{jet} (right) after the $Z/\gamma^*(\rightarrow e^+e^-)$ + jet selection.

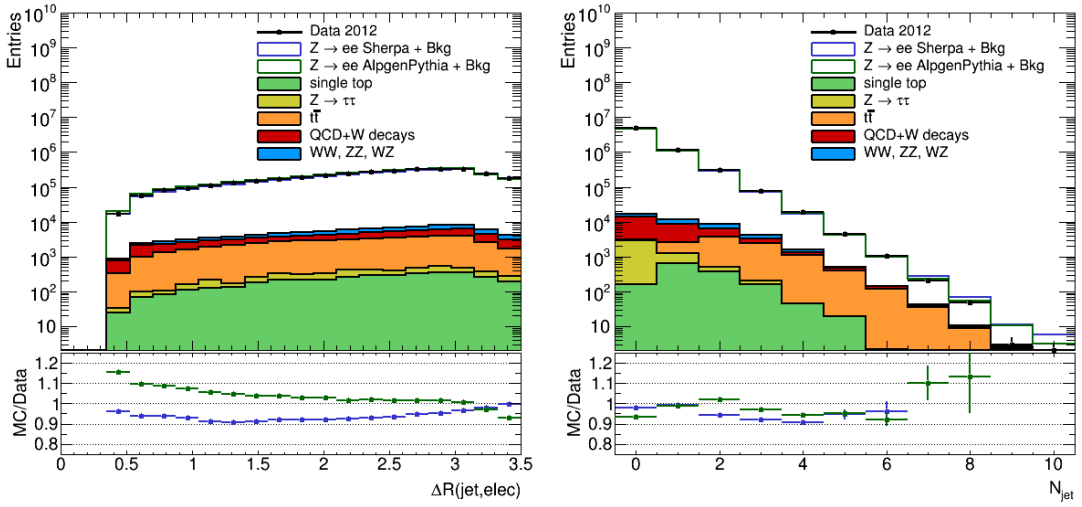


FIGURE 8.5: The distributions of ΔR between jets and the electrons from Z boson decay (left) and jet multiplicity (right) after $Z/\gamma^*(\rightarrow e^+e^-)$ + jet selection.

All three generators have differences in the p_T^Z description, which decrease in the intermediate p_T region.

The distributions of jet properties used for the differential cross sections measurements, y_{jet} and p_T^{jet}, are shown in Figure 8.4. Sherpa event generator describes the transverse momentum of inclusive jets reasonably well, while Alpgen+Pythia6 tends to overestimate data with increasing p_T^{jet}. The rapidity of inclusive jets,

y_{jet}, is described by both Sherpa and Alpgen+Pythia6 within 10 % in the central region; in the forward region the differences between data and Monte Carlo simulation increase.

The distribution of ΔR between jets and the electrons coming from Z boson decay is described by both Sherpa and Alpgen+Pythia6 within 10 – 15 %, as shown in Figure 8.5 (left plot). Both Sherpa and Alpgen+Pythia6 event generators provide good description of data in events with jet multiplicities up to $N = 6$ jets (see Figure 8.5, right plot).

Chapter 9

Z + jet Events for the Quark-Gluon Tagging Purposes

9.1 Introduction

The properties of a hadron jet depend on the initiating parton. The possibility to distinguish between the light quark-initiated and the gluon-initiated jets is especially important for beyond Standard Model searches, where many signal processes contain light quarks, while Standard Model backgrounds contain mostly jets coming from gluons.

Jets that are initiated by heavy quarks (charm and bottom) have different properties compared to those of light quark jets. Such jets are often identified by long-lived or leptonically decaying hadrons. The studies presented below do not include discrimination studies for such jets. The quark jets here and below refer to light quark initiated jets only.

The development of a quark-gluon tagger which uses information from Monte Carlo predictions only becomes complicated due to the differences in the response to quark and gluon jets in different Monte Carlo generators. However, high luminosity of the 2012 dataset allows to use high-purity quark and gluon samples from data and to reduce significantly the dependence on the Monte Carlo modeling.

9.2 Data samples

Several event samples are used for the quark-gluon tagger construction: a dijet sample, a γ + jet sample, and a Z + jet sample. The dominant LO processes for γ + jet and the Z + jet production have quark and gluon in the initial state and outgoing quark jet, while in the dijet case the process with gluon jets in the final state is the dominant. The selection of Z + jet events for quark-gluon tagger is described in Section 5, and the selection of dijet and γ + jet samples is detailed in [75]. The performance of the quark-gluon tagger is checked in bins of the leading jet p_T and y.

Other data samples are used for the validation of the extracted templates. Those are a three-jet sample and a γ + 2 jets sample, where certain phase-space regions are known to have high gluon and quark purities. The selection of these samples are detailed in [75]. The gluon purity of ∼80 % was achieved in the three-jet sample using likelihood method, and the quark purity of ∼80 % was achieved in the γ + 2 jets sample.

9.3 Discriminating variables

The choice of the quark-gluon discriminating variables is based on the fact that gluon-initiated and quark-initiated jets have different electric and color charges. Both calorimeter-based and track-based variables are used. In the former case calorimeter topological clusters are used for the variables calculations, and in the latter case tracks within $\Delta R < 0.4$ to the jet axis and satisfying the following criteria are used:

- $p_T > 1$ GeV,
- Number of pixel hits ≥ 1,
- Number of SCT hits ≥ 6,
- Reduced $\chi^2 < 3$,
- $|z_0 sin(\theta)|$ with respect to primary vertex < 1 mm,
- $|d_0|$ with respect to primary vertex < 1 mm.

The following discrimination variables are tested:

- Number of tracks in the jet: $n_{trk} = \sum_{trk \epsilon jet}$,

- Track width, defined as the p_T weighted sum of the distances between the jet constituents and the jet axis: $w_{trk} = \frac{\sum_{trk \epsilon jet} p_{T,trk} \Delta R_{trk,jet}}{\sum_{trl \epsilon jet} p_{T,i}}$,

- E_T weighted calorimeter width, defined similarly to the track width above: $w_{calo} = \frac{\sum_{cluster \epsilon jet} p_{T,cluster} \Delta R_{cluster,jet}}{\sum_{cluster \epsilon jet} p_{T,i}}$,

- Number of calorimeter clusters in the jet: $n_{calo} = \sum_{cluster \epsilon jet}$,

- Number of calorimeter clusters carrying 90% of the jet energy in the jet [76] $n^{90}_{calo} = \sum^{90\% E_{jet}}_{cluser \epsilon jet}$. In comparison to the previous variable, n_{90} is less sensitive to energy splitting over between cells, either due to QCD shower, or due to particle decays and widening of showers within the calorimeter,

- Fraction of energy carried by the largest energy calorimeter cluster: $f^{largest} = \frac{E_{largest\ cluster}}{E_{jet}}$,

- Two points energy correlation function calculated using jet calorimeter towers [77]: $C_\beta = \frac{\sum_{i,j \epsilon jet} E_{T,i} E_{T,j} (\Delta R_{i,j})^\beta}{(\sum_{i,j \epsilon jet} E_{T,i})^2}$. By choosing values of $\beta \simeq 0.2$ the correlators are able to probe small-scale collinear splittings more effectively, which is useful for quark-gluon jets discrimination,

- p_T weighted jet charge [78], defined as: $Q^k = \frac{1}{(p_T)^k} \sum_{i \epsilon tracks} p^k_{T,i} \times q_i$, where p_T in the denominator refers either to the jet p_T or to the sum of p_T of tracks in the jet, and k is a free parameter. In the first case this variable is further referred as jet charge, and in the second as track charge. For the quark-gluon tagging the value $k = 1$ provides the best discrimination power.

9.4 Quark-Gluon templates extraction

Since Monte Carlo modeling of the quark and gluon jet properties is generator dependent, data driven techniques are used. The quark-gluon tagging procedure is derived assuming that the shapes of the tagging distributions are independent from the process where quark and gluon jets are produced. Therefore, the effects from generator- and process-dependence are included in the final systematic uncertainty.

9.4.1 Two-samples extraction

With two independent pure-gluon and pure-quark samples, such as dijet and γ + jet or Z + jet samples, the inclusive shapes can be described by the following system of equations:

$$\begin{pmatrix} Data_{Dijet,i} \\ Data_{\gamma/Z+jet,i} \end{pmatrix} = \begin{pmatrix} f^q_{Dijet} & f^g_{Dijet} \\ f^q_{\gamma/Z+jet} & f^g_{\gamma/Z+jet} \end{pmatrix} \begin{pmatrix} q_i \\ g_i \end{pmatrix} + \begin{pmatrix} f^{hf}_{Dijet} \cdot hf_{Dijet,i} \\ f^{hf}_{\gamma/Z+jet} \cdot hf_{\gamma/Z+jet,i} \end{pmatrix}, \quad (9.1)$$

where i corresponds to the bin of a distribution histogram. The fractions of quark- or gluon-initiated jets are given as $f^{q/g}_{sample}$, and are taken from the Monte Carlo predictions. The inclusive shapes are given as $Data_{sample,i}$, pure quark and gluon distributions are given as q_i and g_i, respectively. Similarly, the fraction of heavy flavour (charm and bottom) jets is represented by f^{hf}_{sample}, and pure heavy flavor distributions are given as $hf_{sample,i}$. In case of the γ + jet sample, the contribution from jets with "fake" photons needs to be subtracted before solving the system of equations. For Z + jet the background fraction is very small and the final impact of its subtraction on the gluon rejection and quark acceptance is found to be up to 0.2 %, which is taken as a systematic uncertainty.

The example of quark and gluon templates extracted from data using the two-samples extraction procedure is shown in Figure 9.1. The dijet an Z + jet samples are used for the extraction, with dijet simulated with Pythia8 and Z + jet simulated with Powheg in combination with the parton shower from Pythia8. The statistics at low p_T is large enough in Z + jet sample, which allows to extend quark/gluon discrimination in the $25 < p_T < 40$ GeV region. The templates extracted from data using Pythia8/Powheg+Pythia8 samples are compared to the templates extracted from Pythia8/Powheg+Pythia8 and Herwig++/Alpgen+Jimmy using corresponding Monte Carlo samples for truth-tagged templates.

9.4.2 Three-samples extraction

The statistics of γ + jet sample is quite limited at low p_T^{jet} and in the forward region, while Z + jet sample has poor statistics at high p_T^{jet}. Using all three samples helps to cover full p_T and η range. Since adding the third sample over-constrains the system of equations, the χ^2 minimisation method is used. The χ^2 is defined as:

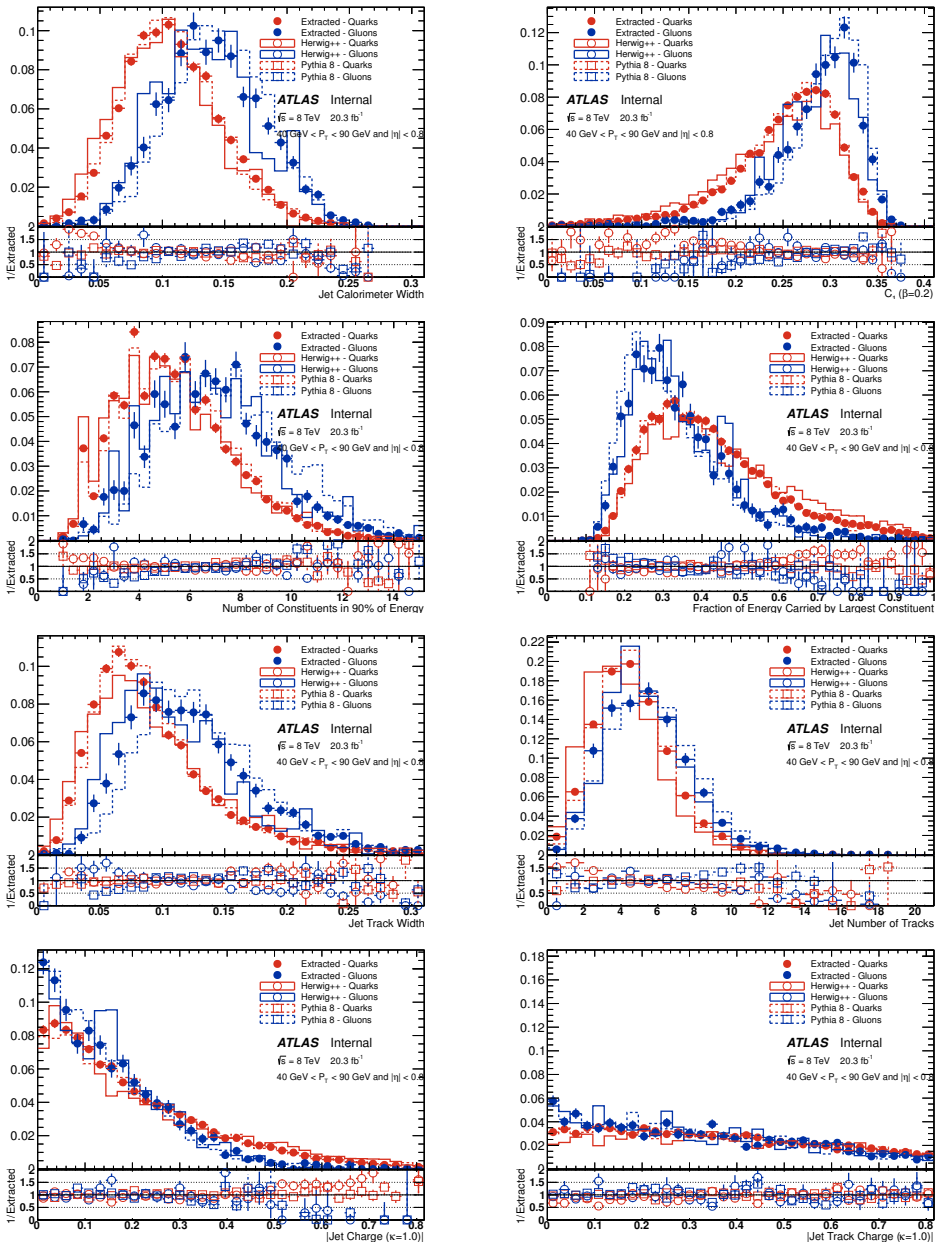


FIGURE 9.1: The extracted quark/gluon templates for different discrimination variables. Data is compared to Pythia8/Powheg+Pythia8 and Herwig++/Alpgen+Jimmy. The dijet and Z + jet samples are used for extraction in the $40 < p_T < 90$ GeV, $|\eta| < 0.8$.

$$\chi^2(q_i, g_i) = \sum_{j=0}^{N} (Data'_{j,i} - f_j^q \cdot q_i - f_j^g \cdot g_i)^2, \tag{9.2}$$

where j corresponds to the sample and i corresponds to the bin of the distribution histogram. $Data'_{j,i}$ represents inclusive distributions, already corrected by heavy flavour and fake contributions. Solving $\frac{\delta\chi^2}{\delta q} = 0$, $\frac{\delta\chi^2}{\delta g} = 0$ brings us to the system of two equations:

$$\begin{pmatrix} \sum_{j=0}^{N} Data'_{j,i} \cdot f_j^g \\ \sum_{j=0}^{N} Data'_{j,i} \cdot f_j^q \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{N} f_j^q \cdot f_j^g & \sum_{j=0}^{N} (f_j^g)^2 \\ \sum_{i=0}^{N} (f_j^q)^2 & \sum_{j=0}^{N} f_j^g \cdot f_j^q \end{pmatrix} \begin{pmatrix} q_i \\ g_i \end{pmatrix}. \tag{9.3}$$

The example of quark and gluon templates extracted from data using three-samples extraction procedure is shown in Figure 9.1. The same data and Monte Carlo samples as for the two-samples extraction procedure are used (see Section 9.4.1).

9.4.3 Variable ranking

In order to determine the most powerful for the quark-gluon discrimination variables, a likelihood value is calculated and the variables are ranked based on the gluon rejection for the fixed quark acceptance. The likelihood for a variable x is defined as:

$$\mathcal{L}(x) = \frac{f_q(x)}{f_q(x) + f_g(x)}. \tag{9.4}$$

The ROC curves for the two sample extraction with the dijet and Z + jet samples are shown in Figures 9.3 and 9.4, and the ROC curves for the templates extraction with all three samples are shown in Figures 9.5 and 9.6. It is easy to see that color-based variables in general provide better quark-gluon discrimination, than the charge-based ones. Since the statistics of the Z + jet sample is getting low at higher p_T, there are large fluctuations in the discrimination curves shapes in the last p_T bins for the two samples extraction.

The performance of the quark-gluon tagging for different discrimination variables is shown in Figures 9.7 and 9.8. The color-based variables and the charge-based variables tend to group, and color-based variables provide better tagging

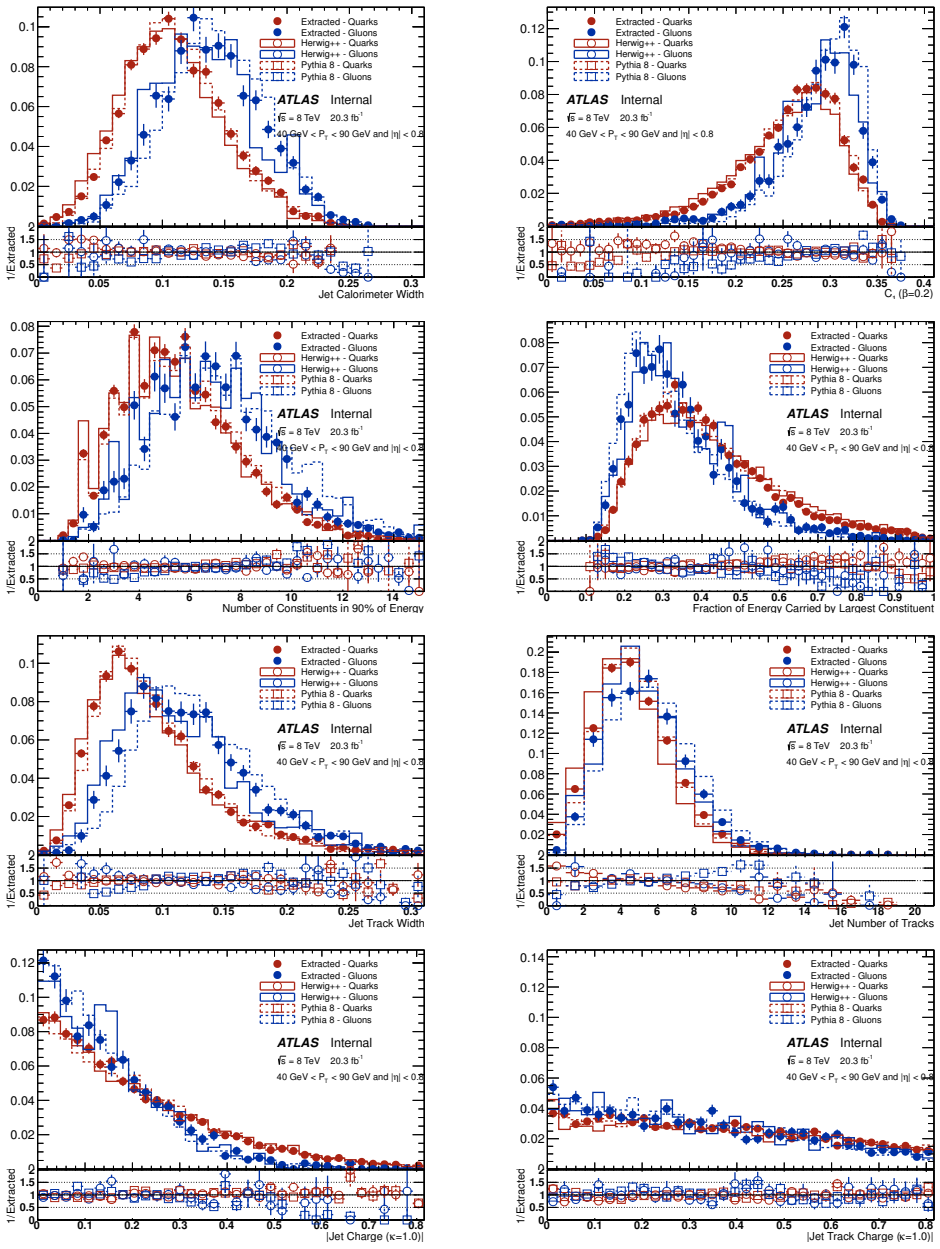


FIGURE 9.2: The extracted quark/gluon templates for different discrimination variables. Data is compared to Pythia8/Powheg+Pythia8 and Herwig++/Alpgen+Jimmy. The dijet, Z + jet and γ + jet samples are used for extraction in the $40 < p_T < 90$ GeV, $|\eta| < 0.8$.

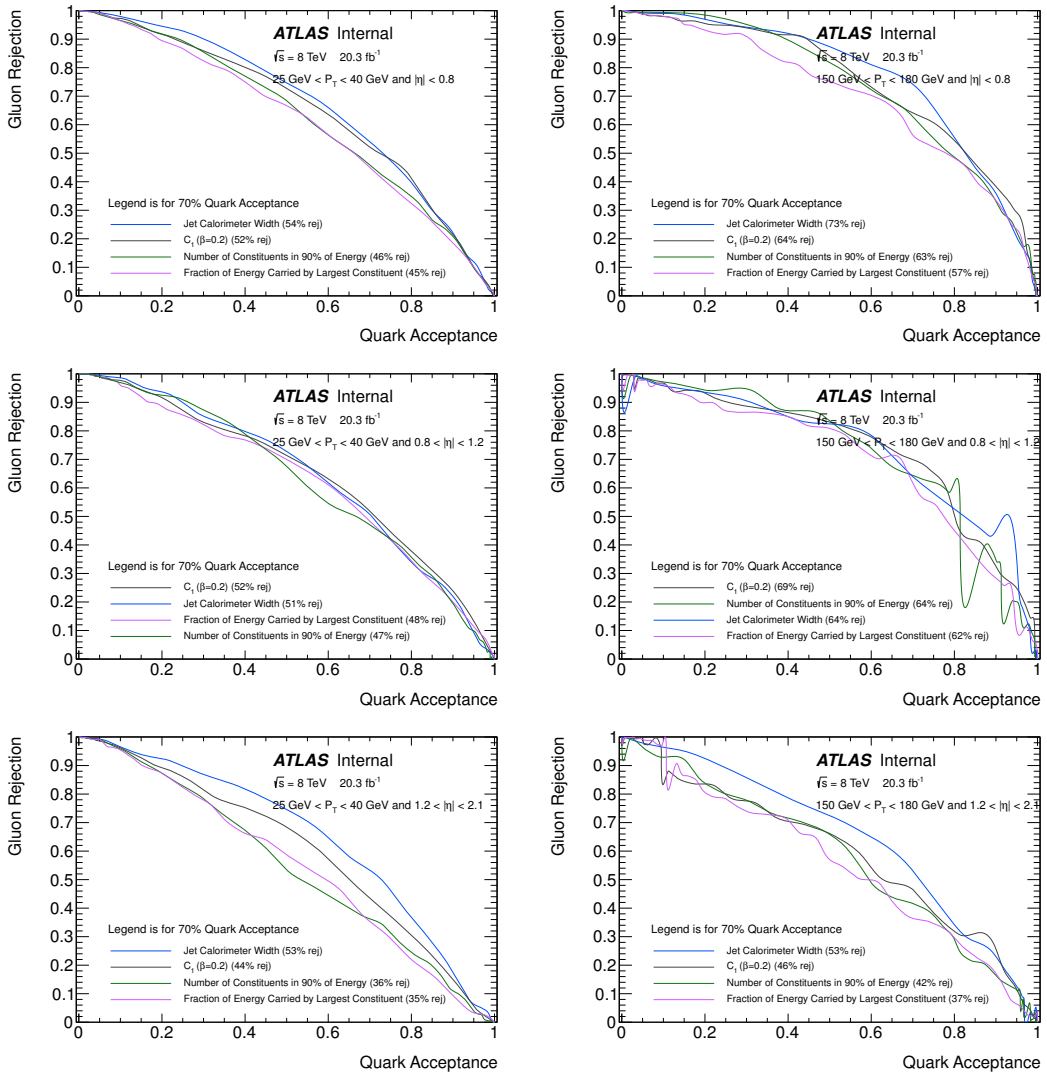


FIGURE 9.3: Gluon rejection versus quark efficiency in data for 1-D likelihoods built from extracted templates for w_{calo}, n^{90}_{calo}, $f^{largest}$, and C_β ($\beta = 0.2$) variables for jets in the $25 < p_T < 40$ GeV (left) and $150 < p_T < 180$ GeV (right) and $|\eta| < 0.8$ (top), $0.8 < |\eta| < 1.2$ (middle) and $1.2 < |\eta| < 2.1$ (bottom). Dijet and Z + jet samples were used for the extraction.

performance in the whole p_T range. The discrimination power increases with increasing p_T except of the forward region $2.1 < |\eta| < 2.5$, where statistics is very low for all samples.

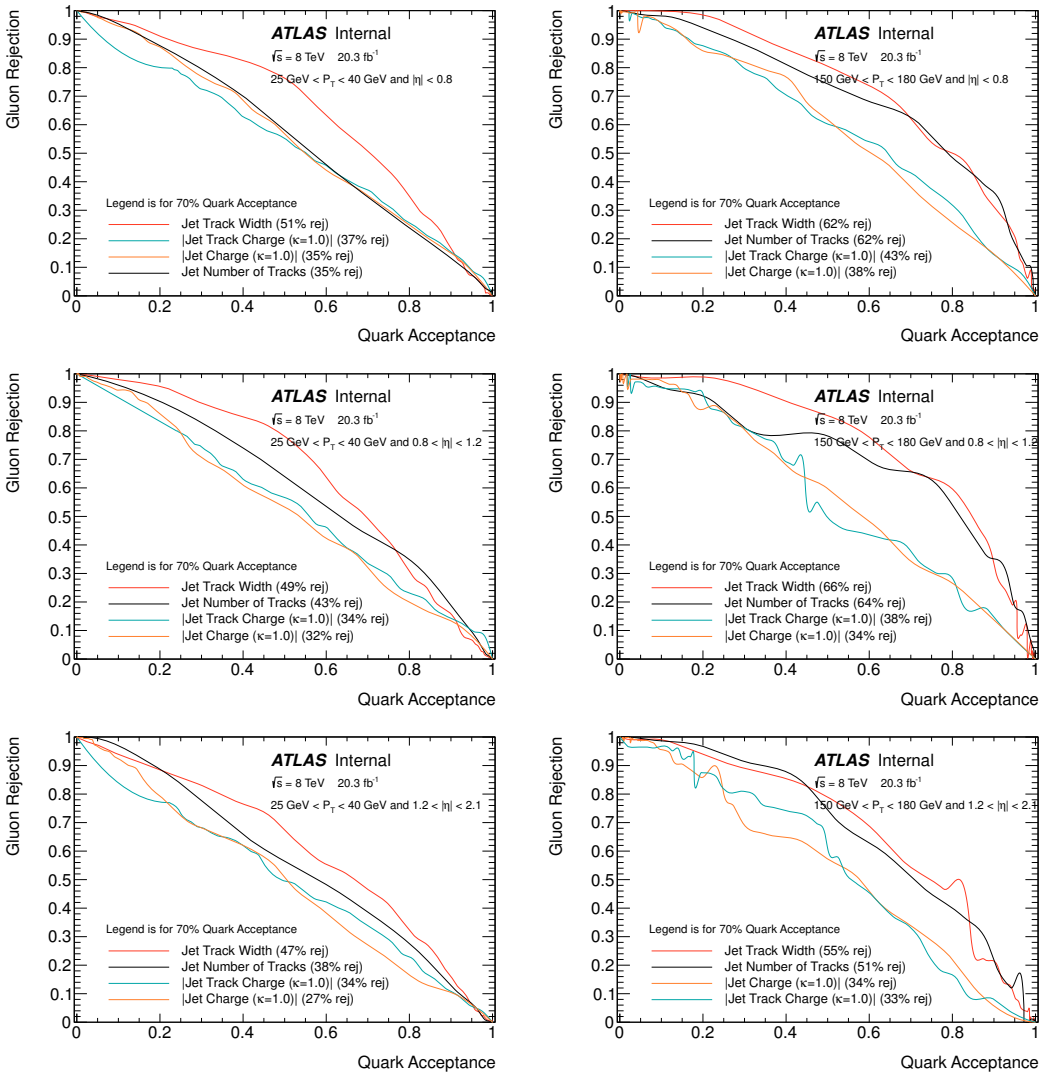


FIGURE 9.4: Gluon rejection versus quark efficiency in data for 1-D likelihoods built from extracted templates for n_{trk}, w_{trk}, jet charge, and track charge for jets in the $25 < p_T < 40$ GeV (left) and $150 < p_T < 180$ GeV (right) and $|\eta| < 0.8$ (top), $0.8 < |\eta| < 1.2$ (middle) and $1.2 < |\eta| < 2.1$ (bottom). Dijet and Z + jet samples were used for the extraction.

9.5 Uncertainties in the Quark-Gluon Jets Discrimination

All uncertainties are calculated using the same method referred to as a method of template variations. First, the nominal templates from the data are extracted and the nominal point with gluon rejection resulting in 70 % quark acceptance is defined. Then new templates are extracted by varying some of the inputs to the extraction procedure. These new templates are called template variations. The

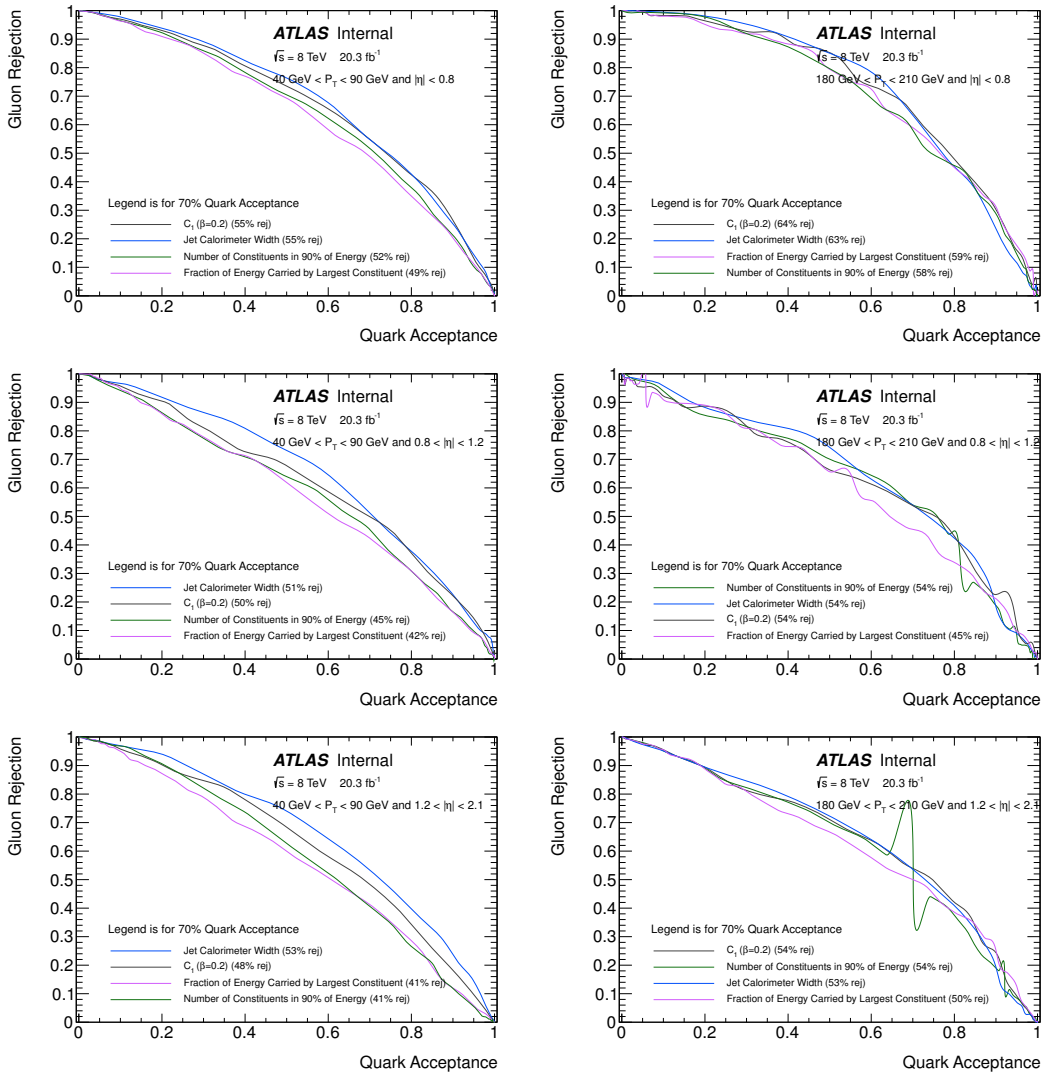


FIGURE 9.5: Gluon rejection versus quark efficiency in data for 1-D likelihoods built from extracted templates for w_{calo}, n^{90}_{calo}, $f^{largest}$, and C_β ($\beta = 0.2$) for jets in the $40 < p_T < 90$ GeV (left) and $180 < p_T < 210$ GeV (right) and $|\eta| < 0.8$ (top), $0.8 < |\eta| < 1.2$ (middle) and $1.2 < |\eta| < 2.1$ (bottom). All three samples were used for the extraction.

maximum difference between the nominal gluon rejection/quark acceptance and the variations gluon rejection/quark acceptance is taken as an uncertainty. The uncertainties are quoted separately as uncertainty on the gluon rejection and on the quark acceptance and are calculated asymmetrically.

• **Statistical uncertainty** is calculated by varying the input data distributions bin-by-bin using a Poisson distribution with the initial number of data events in each bin as a central value. The procedure is repeated thousand times. The resulting uncertainties are categorised in two groups depending on positive or

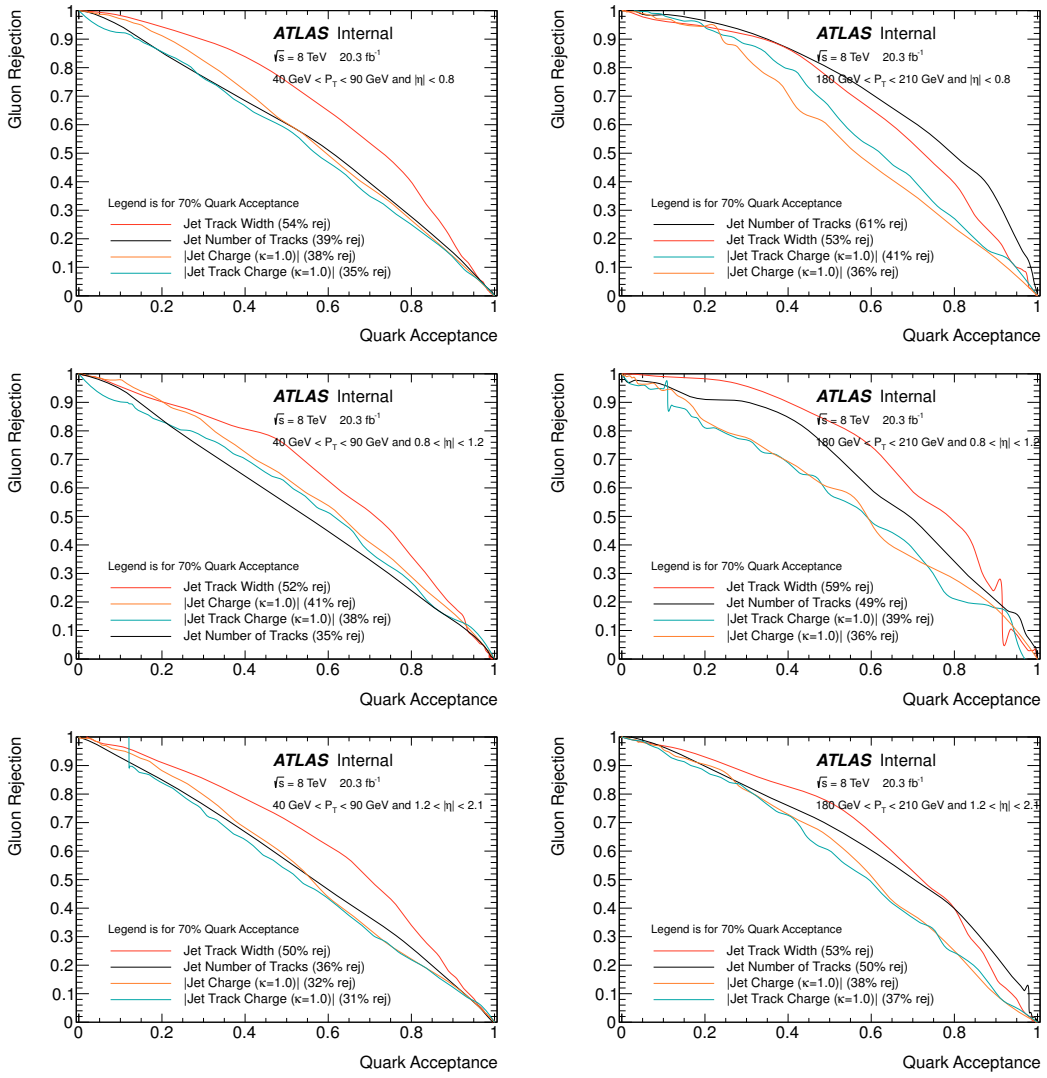


FIGURE 9.6: Gluon rejection versus quark efficiency in data for 1-D likelihoods built from extracted templates for n_{trk}, w_{trk}, jet charge, and track charge for jets in the $40 < p_T < 90$ GeV (left) and $180 < p_T < 210$ GeV (right) and $|\eta| < 0.8$ (top), $0.8 < |\eta| < 1.2$ (middle) and $1.2 < |\eta| < 2.1$ (bottom). All three samples were used for the extraction.

negative deviations from the nominal value, in order to get asymmetric uncertainties.

• The tagging variables can look differently between the dijet, γ + jet and Z + jet samples because of the color connections in the processes. However, more importantly, certain selection can bias events to the specific kinematic region. In particular, the shape of the p_T spectrum of the tagged jet is important, since shapes of the tagging variables are p_T dependent. Due to this, the p_T spectrum of the inclusive leading jet is reweighted with respect to the p_T spectrum of the leading

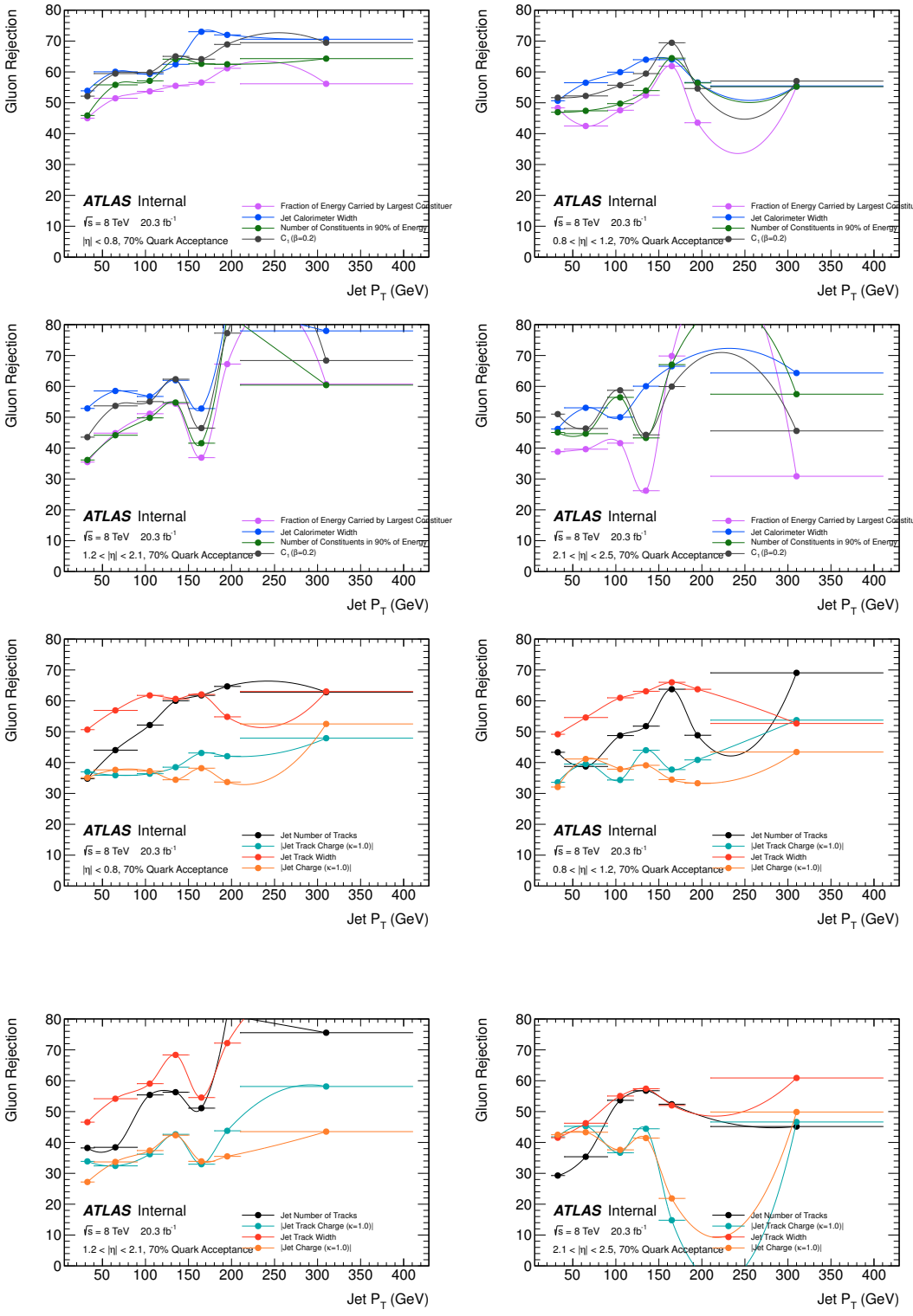


FIGURE 9.7: Gluon rejection at 70% quark efficiency dependence on jet p_T for extracted templates. The plots are for $|\eta| < 0.8$ (top-left), $0.8 < |\eta| < 1.2$ (top-right), $1.2 < |\eta| < 2.1$ (bottom-left) and $2.1 < |\eta| < 2.5$ (bottom-right). The top four plots correspond to calorimeter-based variables and the bottom four to track-based variables. Dijet and Z + jet samples were used for the extraction.

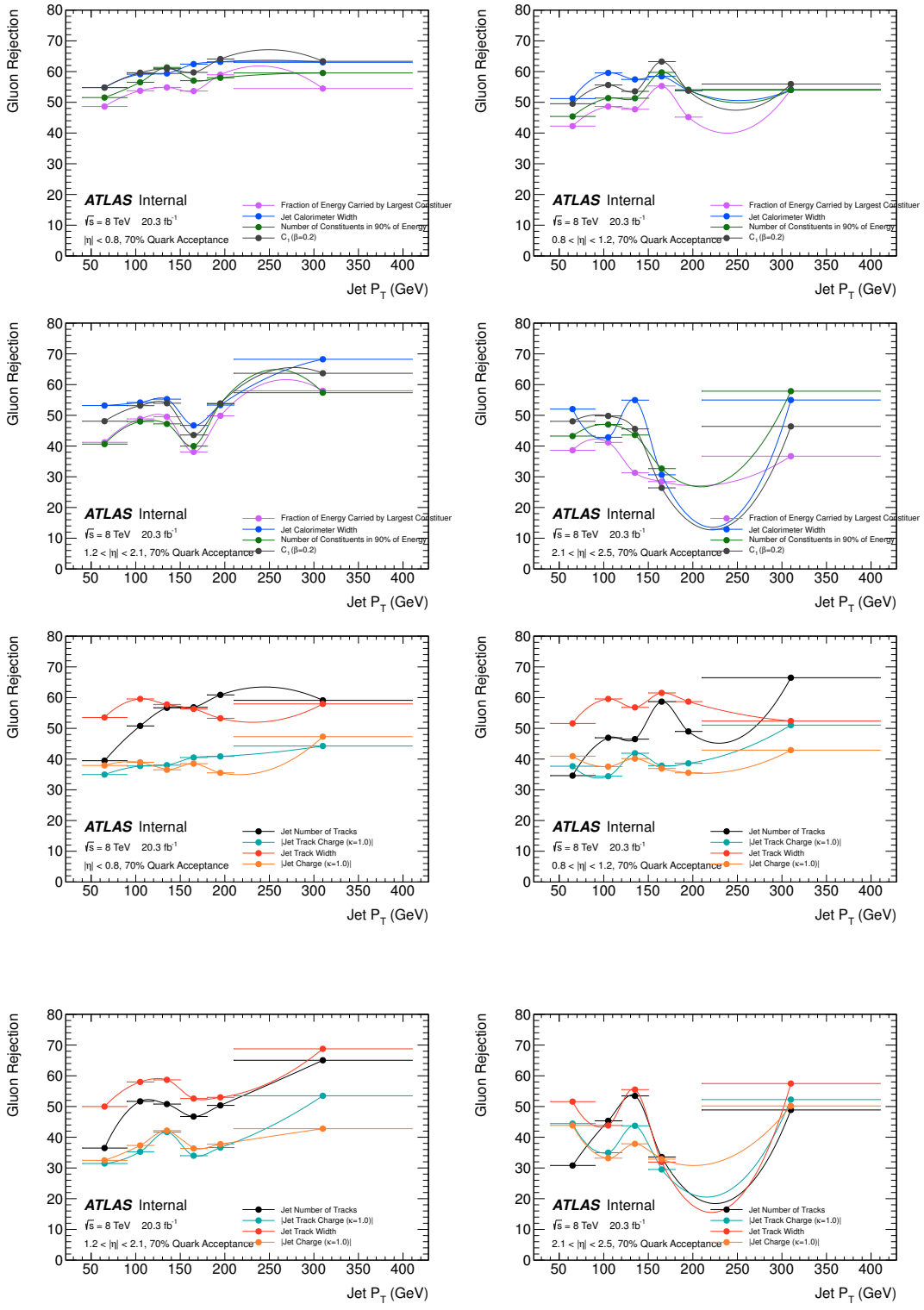


FIGURE 9.8: Gluon rejection at 70% quark efficiency dependence on jet p_T for extracted templates. The plots are for $|\eta| < 0.8$ (top-left), $0.8 < |\eta| < 1.2$ (top-right), $1.2 < |\eta| < 2.1$ (bottom-left) and $2.1 < |\eta| < 2.5$ (bottom-right). The top four plots correspond to calorimeter-based variables and the bottom four to track-based variables. All three samples were used for the extraction.

jet in the dijet sample simulated by Pythia8. **Sample dependence systematic uncertainty** covers remaining differences between the shapes of the tagging variables due to selections and physical processes in the samples.

The nominal templates are extracted from inclusive simulated distributions using truth-tagged templates from Monte Carlo, which are taken also as the template variations. The comparison is done using both Pythia8/Powheg+Pythia8 and Herwig++/Alpgen+Jimmy generators and the maximum discrepancy is taken as an uncertainty.

• **Quark/gluon jet fraction systematic uncertainty** is calculated using the fraction of quarks and gluons taken from Herwig++/Alpgen+Jimmy simulation, while bottom and charm contributions are kept to be from Pythia8/Powheg+Pythia8. The nominal templates are extracted with Pythia8/Powheg+Pythia8 fractions. The resulting uncertainties on both the gluon rejection and on the quark acceptance are less then 2% for all variables in most of p_T and η bins.

• **Systematic uncertainty due to the PDFs** covers the difference coming from the use of different PDF sets by various Monte Carlo simulations. The Monte Carlo samples are reweighted to different PDF sets using LHAPDF [79] package. The PDF variations considered are: CT10, CTEQ6L1, MSTW 2008 NLO and NNPDF2.3. The method of template variations is performed one-by-one on the new templates, providing the uncertainty due to the PDF choice. The maximum difference is taken as the systematic uncertainty. The final uncertainty is less than 0.5 % in all p_T, η bins.

• **Heavy flavour jet fraction systematic uncertainty** takes into account the uncertainty on the c/b-jet fractions in the sample. The uncertainty in the c/b-jet fraction is taken to be 20 % for dijet sample and 50 % for γ + jet and Z + jet samples. The templates variations were extracted by varying the heavy flavour jet fraction in each sample up and down by a given amount. The resulting uncertainty is up to 2 % for all variables.

• **Heavy flavour jet shapes systematic uncertainty** is calculated separately for c-initiated and b-initiated jets. For the c-jet shapes the contributions of most common c-quark decay modes, which are D^0 , $D^*(2007)^0$, D^+ and $D^*(2010)^+$, are varied by $\pm$10 % in simulated samples. The varied shapes are then used as inputs for the extraction procedure to create template variations. The resulting uncertainty from this channel is below 1 %. For the determination of the b-jet

Uncertainty source	Gluon rejection (%)	
	$\|\eta\| < 0.8$	$2.1 < \|\eta\| < 2.5$
Statistics	1 – 3	1 – 3
PDF	0.5 – 2	1 – 10
Sample dependence	5 – 20	10 – 20
Quark/gluon fraction	1 – 2	1 – 4
Heavy flavour fraction	< 1	< 1.5
Heavy flavour shape	< 1	< 1
Z + jet purity	< 1	< 1

TABLE 9.1: Systematic uncertainties on the gluon rejection for quark acceptance of 70 % for first and last $|\eta|$ bins of the two sample extraction with dijet and Z + jet in p_T range from 25 GeV to 410 GeV, except for the last $|\eta|$ bin, where p_T up to 120 GeV is considered.

shape uncertainty, the template variation is extracted using the b-jet shape taken from the Herwig++/Alpgen+Jimmy sample, while the integral of the b-jet shape distribution and all the rest of the truth-tagged distributions are still taken from Pythia8/Powheg+Pythia8 sample. The final uncertainty is considerably less than 1 %.

• **Systematic uncertainty due to the γ + jet purity estimation** is calculated by varying the definition of the "non-tight" ID on the photon. The largest difference as a result of using the method of template variations is taken to be the uncertainty due to the purity estimation. The systematic uncertainties are of the order of 1% at low p_T , where the γ + jet sample has large contribution from dijet background.

• **Systematic uncertainty due to the Z + jet purity estimation** is calculated by extracting the template variations from data with electroweak (including top decays) and multijet backgrounds subtracted, as described in Section 7. The resulting uncertainties are up to 0.2 % in p_T bins where the fraction of background in Z + jet is the largest.

The ranges of uncertainties are summarised in Tables 9.1 and 9.2 for the templates extraction with dijet and Z + jet samples combination and in Tables 9.3 and 9.4 for the three samples extraction. The sample dependence systematics is the dominant one. The size of it points out the difficulties in producing a general purpose quark-gluon tagger.

Uncertainty source	Quark acceptance (%)	
	$\|\eta\| < 0.8$	$2.1 < \|\eta\| < 2.5$
Statistics	1 – 3	1 – 3
PDF	0.5 – 2	0.5 – 4
Sample dependence	5 – 10	5 – 20
Quark/gluon fraction	1 – 2	1 – 4
Heavy flavour fraction	< 1	< 1
Heavy flavour shape	< 0.5	< 1
Z + jet purity	< 1	< 1

TABLE 9.2: Systematic uncertainties on the quark acceptance for quark acceptance of 70 % for for first and last $|\eta|$ bins of the two sample extraction with dijet and Z + jet in p_T range from 25 GeV to 410 GeV, except for the last $|\eta|$ bin, where p_T up to 120 GeV is considered.

Uncertainty source	Gluon rejection (%)	
	$\|\eta\| < 0.8$	$2.1 < \|\eta\| < 2.5$
Statistics	1 – 2	1 – 2
PDF	< 2	1 – 10
Sample dependence	2 – 20	5 – 20
Quark/gluon fraction	1 – 2	1 – 4
Heavy flavour fraction	< 1.5	< 1.5
Heavy flavour shape	< 1	< 1
γ + jet purity	1 – 4	1 – 4
Z + jet purity	< 1	< 1

TABLE 9.3: Systematic uncertainties on the gluon rejection for quark acceptance of 70 % for first and last $|\eta|$ bins of the three sample extraction in p_T range from 40 GeV to 410 GeV, except for the last $|\eta|$ bin, where p_T up to 120 GeV is considered.

Uncertainty source	Quark acceptance (%)	
	$\|\eta\| < 0.8$	$2.1 < \|\eta\| < 2.5$
Statistics	1	1 – 2
PDF	0.5 – 2	0.5 – 3
Sample dependence	2 – 5	5- – 20
Quark/gluon fraction	< 1	< 1
Heavy flavour fraction	< 1.5	< 1.5
Heavy flavour shape	< 0.5	< 0.5
γ + jet purity	< 3	< 3
Z + jet purity	< 1	< 1

TABLE 9.4: Systematic uncertainties on the quark acceptance for quark acceptance of 70 % for for first and last $|\eta|$ bins of the three sample extraction in p_T range from 40 GeV to 410 GeV, except for the last $|\eta|$ bin, where p_T up to 120 GeV is considered.

9.6 Validation of Extracted Quark/Gluon Templates

The validation of the extracted quark/gluon templates is done by comparing them with tagging variables distributions from purified data samples. The trijet sample provides high purity gluon distributions and the γ + 2 jets sample provides pure quark distributions. To avoid any differences coming from the p_T dependence, the p_T spectrum of the "pure" jet is reweighted to the leading jet spectrum in the dijet events, similarly as it is done for all input samples for extraction. Due to the lack of statistics, only the means of the distributions are compared. The mean values of the discriminating variables differ by 10 % at most between extracted and validation templates.

9.7 Conclusions

The power of different jet properties to discriminate between quark-like and gluon-like jets is studied using the data-driven method. For a fixed quark-like jet acceptance of 70 %, the colour-charge based discrimination variables are able to reject about 50 % of the gluon-like jets, and the electric-charge based variables about 40 % of the gluon-like jets. Performance is investigated using combinations of dijet, γ + jet and Z + jet samples. The latter sample is useful in the low p_T region, where γ + jet statistics is limited by trigger prescales.

The main contribution to systematic uncertainty comes from the sample dependence. Discrimination variables tend to vary between different samples, which is mainly due to the p_T dependence of these variables. Reweighting of the p_T spectrum reduces this effect but does not completely remove it.

Chapter 10

Unfolding in the Cross-Section Measurement

The experimental measurements are distorted by the detector resolution and efficiencies. Unfolding is a procedure of correcting the measured spectrum to the "truth-level" spectrum, i.e., the spectrum that would be measured with an ideal detector. For this analysis two unfolding methods are used [80]:

- bin-by-bin correction,
- the iterative Bayesian method (D'Agostini [81]).

Bin-by-bin correction is commonly used because of the simplicity. It can be described shortly in the following way. If T_i is the expected number of events in bin i, and R_i is the measured number of events in simulated spectrum, then the correction in bin i is

$$C_i \equiv \frac{T_i}{R_i}. \tag{10.1}$$

The values of T_i, R_i and hence C_i are determined from the Monte Carlo simulation of the process. And if D_i is the observed number of events in bin i in data, then the estimator of truth, U_i, is defined as:

$$U_i \equiv C_i \cdot D_i. \tag{10.2}$$

However, this method may not take into account migrations of events and correlations between adjacent bins. Therefore, this correction is valid only if the impact of migrations is negligible.

Another method, proposed by D'Agostini [81], is based on the Bayes' theorem. It takes into account migrations between bins and provides the bin-by-bin correlation information of the results. This method is chosen as the nominal method for current analysis, while the bin-by-bin method is used as a cross-check. The short description of the theoretical grounds of the iterative method and the implementation procedure is described below.

10.1 Bayes' iterative method

In terms of several independent causes C_i, $i = 1, 2, ..., n_C$, each of which can produce several possible effects E_j, $j = 1, 2, ..., n_E$, the Bayes formula is [81]:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{i=1}^{n_C} P(E_j|C_i)P_0(C_i)}. \tag{10.3}$$

where $P(E_j|C_i)$ is the conditional probability of the i^{th} cause to produce the effect j, and $P_0(C_i)$ is the probability of the causes. This equation means that, if we observe a single event or effect j, the probability that it is happening due to the i^{th} cause is proportional to the probability of this cause times the probability of this cause to produce the effect.

If $n(E_j)$ events with effect E_j are observed, the best estimate of the true number of events:

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j)P(C_i|E_j), \ \epsilon_i \neq 0, \tag{10.4}$$

taking into account that $0 \leq \epsilon_i \equiv \sum_{j=1}^{n_E} P(E_j|C_i) \leq 1$ is the efficiency of detecting the cause C_i in any of the possible effects.And the final probabilities of the causes are:

$$\hat{P}(C_i) = \frac{\hat{n}(C_i)}{\sum_{i=1}^{n_C} \hat{n}(C_i)}. \tag{10.5}$$

The agreement between the initial distribution $P_0(C)$ and the final probability $\hat{P}(C)$ depends on the consistency of the $P_0(C)$ and a true distribution. In fact, usually the distribution $\hat{P}(C)$ lies between $P_0(C)$ and the true one. This suggests to proceed iteratively through the following steps:

• choose the initial distribution of $P_0(C)$ from the best knowledge of the process under study, and hence the initial expected number of events $n_0(C) = P_0(C)N_{obs}$.

• calculate $\hat{n}(C)$ and $\hat{P}(C)$

• calculate χ^2 for $n_0(C)$ and $\hat{n}(C)$

• replace $P_0(C)$ by $\hat{P}(C)$, and $n_0(C)$ by $\hat{n}(C)$ and start again from the second step.

The optimum number of iterations depends on the problem and needs to be determined. The optimisation of the number of iterations is discussed below.

In terms of this analysis, the truth-level spectrum corresponds to the cause C, and the reconstructed-level spectrum to the effect E. The Bayesian Iterative method as implemented in the RooUnfold [82, 83] software package, is used in this measurement.

10.2 Matching efficiency, purity & stability

The unfolding algorithm uses a response matrix with the direct correspondence between the truth- and reconstructed-level observables. This matrix is calculated using the jet matching procedure in simulated events, with Z boson found on the reconstructed and truth levels. A pair of reconstructed and truth jets is matched, if the distance ΔR between the jet axes is within $R \leq 0.4$ and these jets are the closest to each other in the event.

Matching efficiencies for reconstructed and generated jets are defined as:

$$E_{reco}^{reco\_i} = \frac{N_{reco\ \&\ truth}^{reco\_i}}{N_{all\ reco}^{reco\_i}}, \quad E_{truth}^{truth\_i} = \frac{N_{reco\ \&\ truth}^{truth\_i}}{N_{truth}^{truth\_i}}, \tag{10.6}$$

where $N_{reco\&truth}^{reco\_i}$ is the number of jets in bin i of the reconstructed-level spectrum matched to the generated jets, $N_{all\ reco}^{reco\_i}$ is the number of all jets reconstructed in bin i of the reconstructed-level spectrum, $N_{reco\&truth}^{truth\_i}$ is the number of jets in bin i of the truth-level spectrum matched to the reconstructed jets, and $N_{all\ truth}^{truth\_i}$ is the number of all jets reconstructed in bin i of truth-level spectrum. Matching efficiencies for reconstructed and generated jets, are shown in Figures 10.1 and 10.2,

p_T^{jet} bin	$\lvert y_{jet}\rvert$ bins boundaries
$17-25$ GeV	0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4
$25-50$ GeV	0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4
$50-100$ GeV	0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4
$100-200$ GeV	0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 4.2 4.4
$200-300$ GeV	0.0 0.4 0.8 1.2 1.6 2.0 2.4 3.4 3.8 4.4
$300-400$ GeV	0.0 0.4 0.8 1.2 1.6 2.0 3.0 3.6 4.4
$400-1050$ GeV	0.0 0.4 0.8 1.2 2.6 3.0 3.6 4.4

TABLE 10.1: Optimised binning of $|y_{jet}|$ in bins of p_T^{jet}.

respectively. The x-axes on these and following plots is defined as the bin number, according to:

$$n_{bin} = \sum_{i_{p_T}=1}^{n_{p_T=8}} n_{|y|}^{i_{p_T}}, \tag{10.7}$$

where $n_{|y|}^{i_{p_T}}$ is the number of the absolute rapidity $|y_{jet}|$ bin in the corresponding bin of the p_T^{jet}, i_{p_T}. The binning takes into account statistics and shape of the distribution. The final binning is detailed in Table 10.1.

Matching inefficiency at the reconstructed level comes mostly from the presence of low-p_T jets at the reconstructed level only, which could be explained by the pile-up effects. Jet vertex fraction cut suppresses the selection of such jets partially, but there is a drop of efficiency at higher rapidity in the first p_T^{jet} bins, since it can be performed only within tracking acceptance region $|y_{jet}| < 2.4$. At high transverse momentum, $p_T^{jet} \geq 100$ GeV, the pile-up effect is very small and the matching efficiency is close to 1.

Matching inefficiency at the truth level is explained by the Z boson reconstruction inefficiency of about 40 %, while the impact from jets is not significant. In the underflow p_T^{jet} bin there is a small effect from the jets rejected by the selection cuts on the reconstructed level, otherwise in all bins the matching efficiency does not depend on jet properties and is on the level of 50 – 60 %.

Figure 10.3 shows the response matrix calculated using the Alpgen+Pythia6 simulation. The migration effects between $|y_{jet}|$ and p_T^{jet} bins become visible as

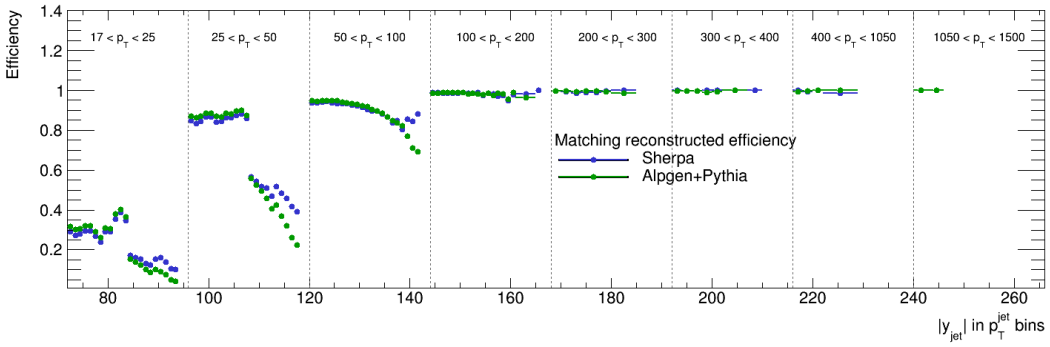


FIGURE 10.1: The matching efficiency as a function of $|y_{jet}|$ in bins of p_T^{jet} at the reconstructed level.

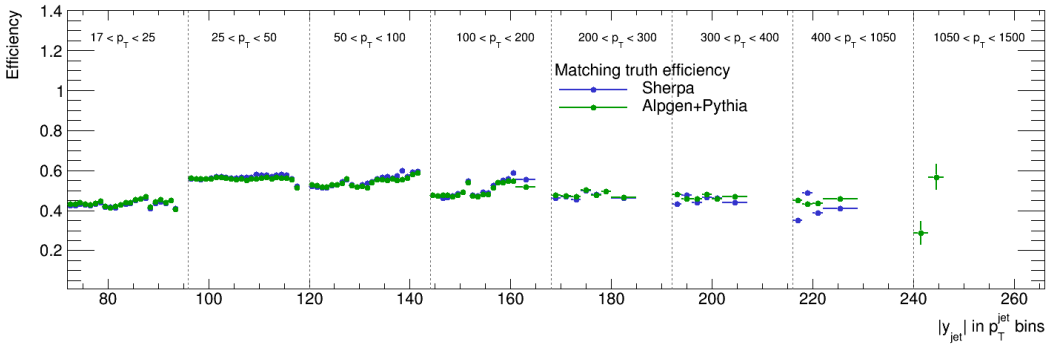


FIGURE 10.2: The matching efficiency as a function of $|y_{jet}|$ in bins of p_T^{jet} at the truth level.

the non-diagonal elements of the matrix. The migration effects are the largest between adjacent p_T^{jet} bins, while the migration between $|y_{jet}|$ bins is small. The effect of migrations between bins is described by purity and stability, defined as:

$$P^i = \frac{N^i_{reco\&truth}}{N^i_{reco}}, \quad S^i = \frac{N^i_{reco\&truth}}{N^i_{truth}}, \tag{10.8}$$

where $N^i_{reco\&truth}$ is the number of jets reconstructed and generated in bin i, N^i_{reco} is the number of jets reconstructed in bin i, and N^i_{truth} is the number of jets generated in bin i. The purity and stability are calculated using only matched jets in the denominator, in order to see effects of migration between bins, but not the matching inefficiency. The purity and stability are shown in Figures 10.4 and 10.5, respectively.

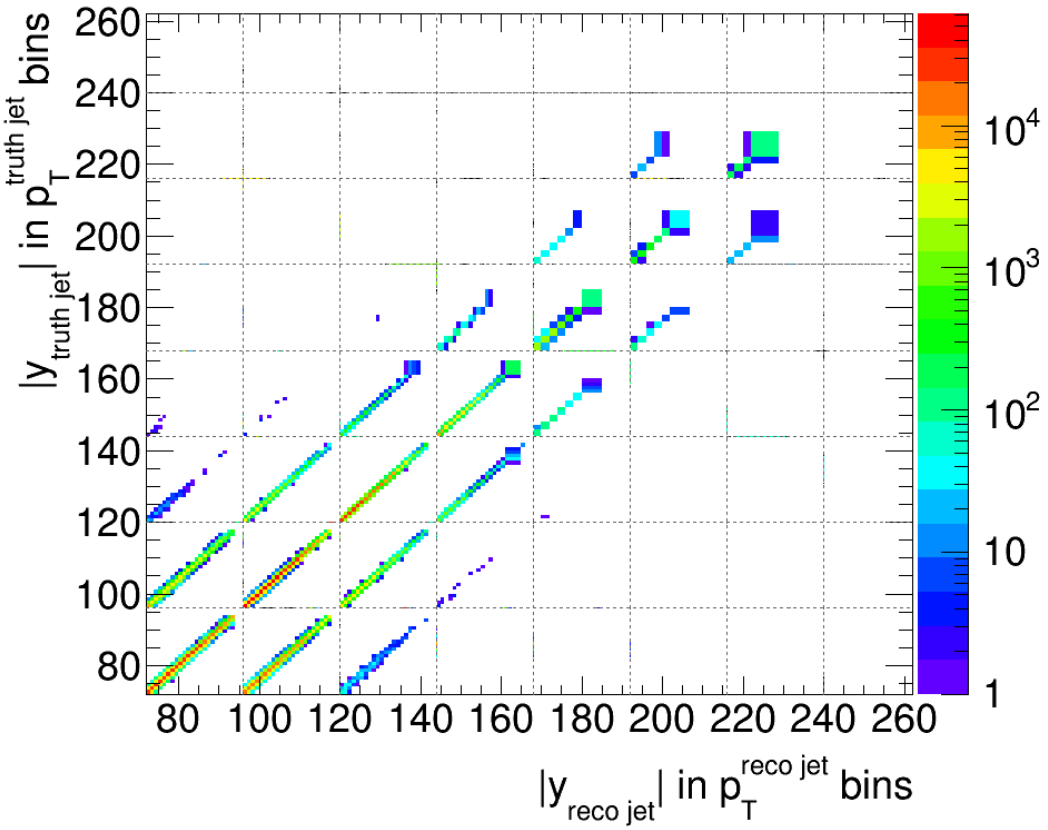


FIGURE 10.3: The response matrix calculated using Sherpa simulation. Binning on the plot is explained in Eq. (10.7).

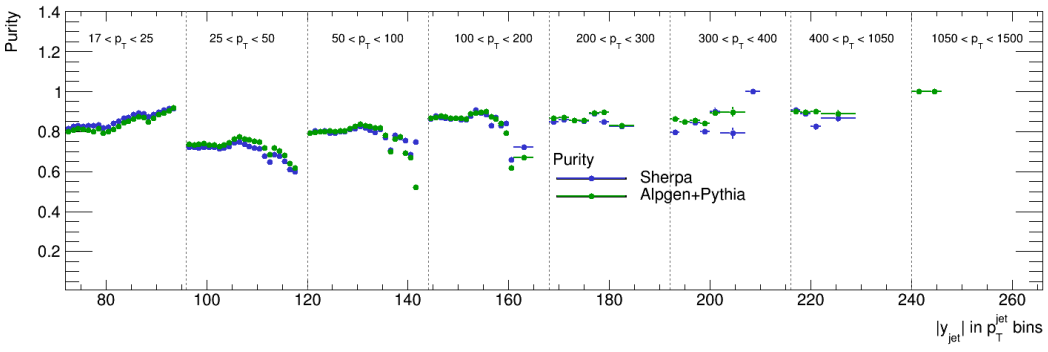


FIGURE 10.4: The purity as a function of $|y_{jet}|$ in bins of p_T^{jet}.

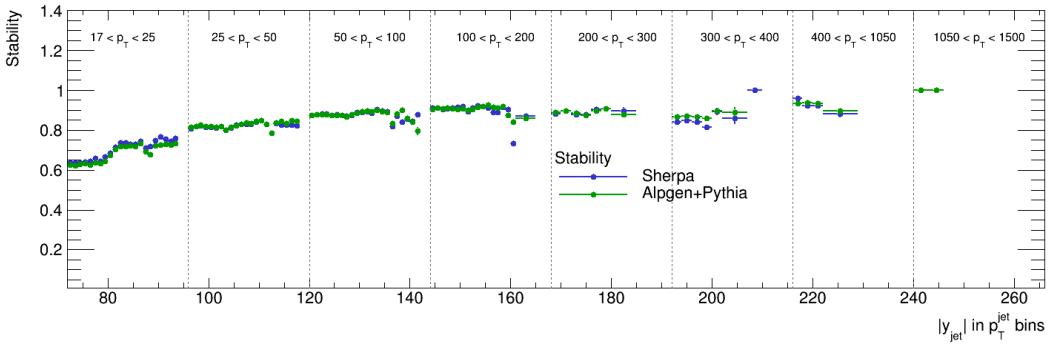


FIGURE 10.5: The stability as a function of $|y_{jet}|$ in bins of p_T^{jet}.

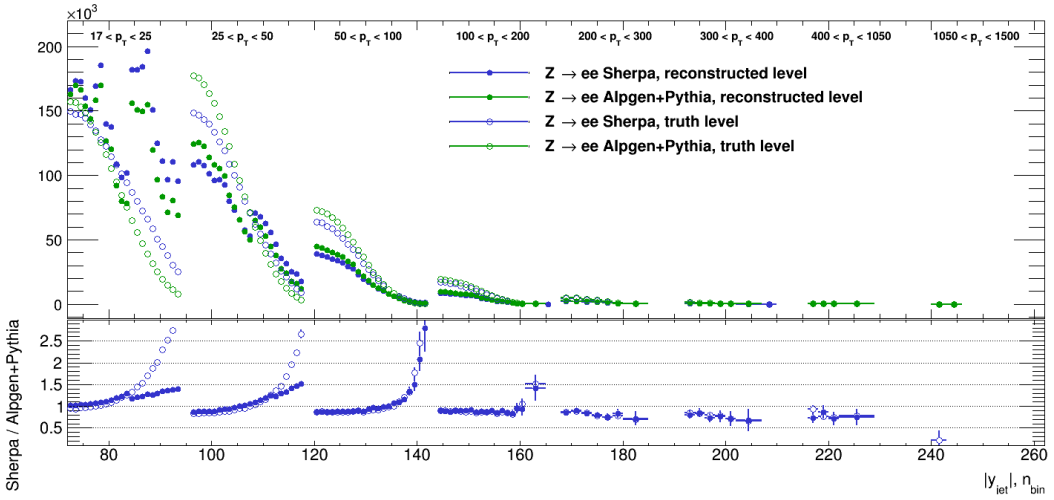


FIGURE 10.6: The reconstructed-level and the truth-level distributions predicted by Sherpa (blue) and Alpgen+Pythia6 (green).

10.3 Optimisation of number of iterations

The number of iterations in the unfolding procedure should be optimised for the best results, as mentioned in Section 10.1. Too many iterations result in the bin-by-bin fluctuations in the unfolded spectrum, while too few iterations make the result dependent on the prior distribution.

The Monte Carlo samples simulated with Alpgen+Pythia6 and Sherpa are used for optimisation of the number of iterations. The comparison between the reconstructed-level and truth-level distributions using these two Monte Carlo samples is shown in Figure 10.6. The difference between these two Monte Carlo predictions covers differences between data and Monte Carlo, therefore providing conservative estimate of the optimal number of iterations.

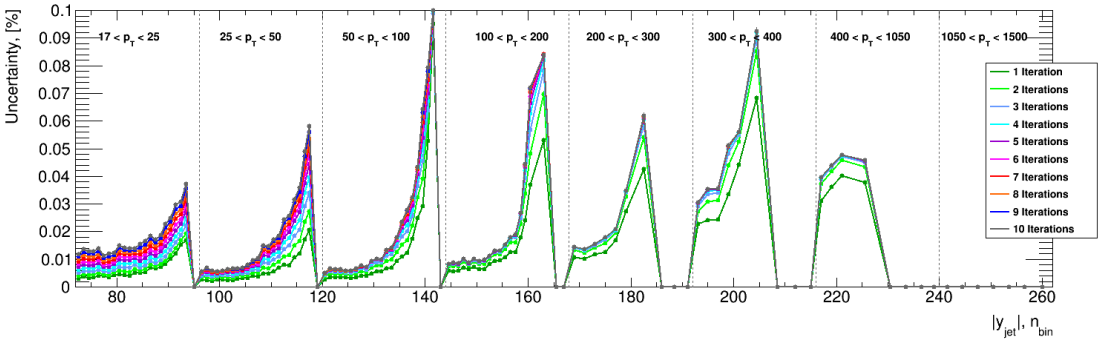


FIGURE 10.7: The statistical uncertainties, calculated using Bootstrap method, for $n_{it} = 1 - 10$.

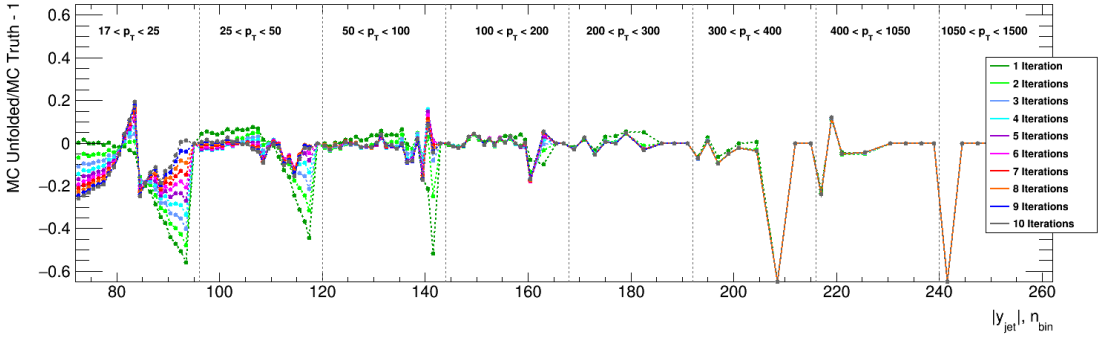


FIGURE 10.8: The bias Δ_{it}, calculated for $n_{it} = 1 - 10$.

The Sherpa spectrum is unfolded with the response matrix derived using Alpgen+Pythia6. The statistical uncertainty, σ, is calculated using Bootstrap method (see Section 11.1) with 100 replicas. For each number of iterations n_{it}, $\sigma_{n_{it}}$ is compared to the bias $\Delta_{n_{it}}$ between the unfolded spectrum and the original truth-level spectrum of Sherpa. The $\Delta_{n_{it}}$ is defined as:

$$\Delta_{n_{it}} = \frac{U_{n_{it}} - T}{T}, \tag{10.9}$$

where $U_{n_{it}}$ is the reconstructed-level Sherpa spectrum, unfolded with n_{it} iterations, and T is the truth-level Sherpa spectrum. The statistical uncertainties and the resulting $\Delta_{n_{it}}$ for $n_{it} = 1 - 10$ iterations are shown in Figures 10.7 and 10.8, respectively. Five iterations are selected to be the optimal choice, since the uncertainties and the $\Delta_{n_{it}}$ are at the same level in each p_T^{jet} bin except the underflow bin $17 < p_T^{jet} < 25$ GeV, as shown in Figure 10.9. The resulting $\Delta_{n_{it}}$ is treated as the systematical uncertainty in the unfolding procedure.

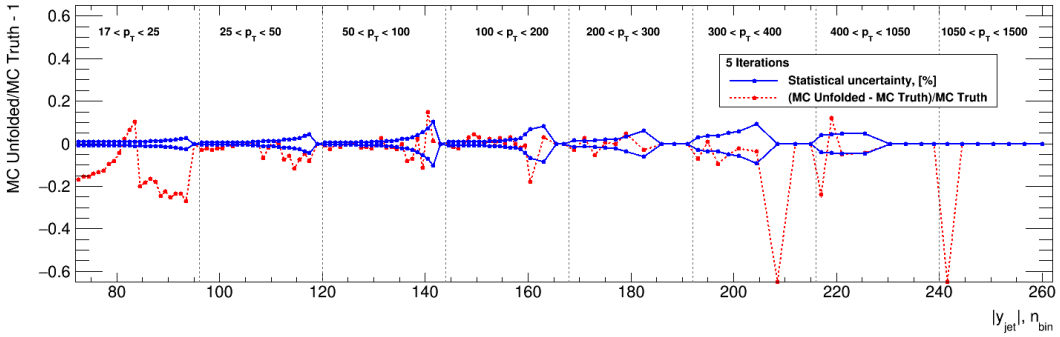


FIGURE 10.9: Bias Δ_{it} (red) versus statistical uncertainties (blue) for $n_{it} = 5$.

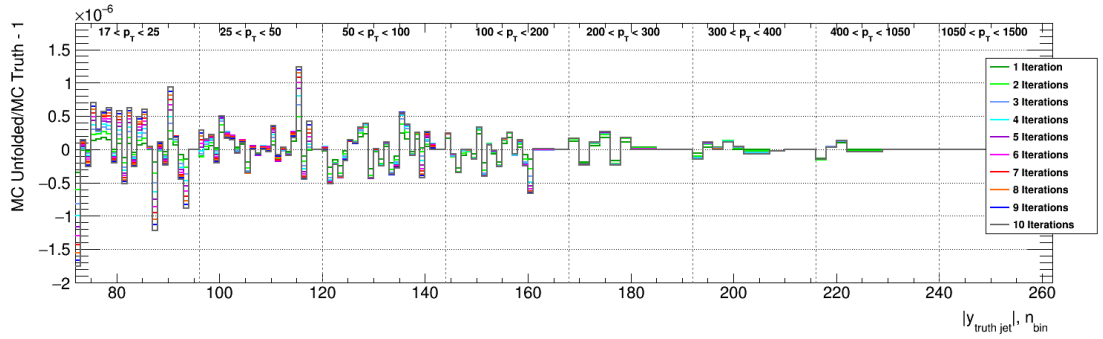


FIGURE 10.10: Comparison between the truth-level and the unfolded distributions from Sherpa. Unfolding is performed using the response matrix from Sherpa for $n_{it} = 1 - 10$.

10.4 Closure tests

Closure tests are performed in order to validate the implementation of the unfolding procedure. For these tests, same Monte Carlo predictions from Sherpa are used for the response matrix, prior distribution as well as pseudo-data. The resulting differences between the truth-level and unfolded Sherpa distributions are less than 0.0001 % for $n_{it} = 1 - 10$, as shown in Figure 10.10.

Chapter 11

Uncertainties in the Cross-Section Measurement

The cross-section measurements rely on a set of corrections and calibrations applied to Monte Carlo simulation and data. These corrections, as well as analysis methods used for the background subtraction and unfolding, have intrinsic uncertainties of statistical and systematic nature, which should be propagated to the final result.

11.1 Statistical uncertainties

The statistical uncertainties in both data and Monte Carlo simulation are derived using Bootstrap method. In each event series of pseudo-experiments are performed. In each pseudo-experiment an event weight is generated using the Poisson distribution with a mean value λ=1, as a result producing a replica of the unfolding matrix in simulation or the input distribution of $|y_{jet}|$ in p_T^{jet} bins in data and simulation. The data unfolding is performed with each of these replicas, and the final uncertainty σ in each bin i of measured observable is calculated as:

$$\sigma_i = \sqrt{\frac{1}{N}\left(\sum_{r=0}^{N} a_i^2 - \left(\sum_{r=0}^{N} a_i\right)^2\right)}, \tag{11.1}$$

where N is a number of replicas and a_i is the bin content in the bin i of the observable distribution. N should be large enough to avoid statistical fluctuations. For the provided results the value of N=100 is used.

The correlation matrix is used to investigate the correlation between different bins. It is calculated, as follows. For each two bins i and j the correlation coefficient is:

$$\rho_{ij} = \frac{1}{\sigma_i \sigma_j N} \left(\sum_{r=0}^{N} a_i b_j - \sum_{r=0}^{N} a_i \sum_{r=0}^{N} b_j \right), \tag{11.2}$$

where N is a number of replicas and a_i and b_j are bin contents in bins i and j, correspondingly. The correlation matrices are calculated using $N = 100$ Bootstrap replicas of the response matrix from Sherpa as input to the unfolding procedure. The resulting matrix for $n_{it} = 5$ unfolding iterations is shown in Figure 11.1. No strong correlation between bins of measurement is observed.

11.2 Systematic uncertainties

There are several sources of systematic uncertainties in this measurement. These uncertainty sources are categorized as electron-related, jet-related, coming from the data-driven and Monte Carlo based background estimation and the unfolding method and described below.

Electron-related and jet-related systematic uncertainties are propagated to the final result through corresponding variations of electron or jet properties in the analysis. These variations are performed only in the simulated samples, and are further split into categories: phase-space shifts, after which the full event selection has to be repeated, and event weight shifts, which re-weight samples of generated events. All jet-related uncertainties together with electron energy scale and resolution uncertainties fall into the first category, while all systematic uncertainties in the electron efficiencies are in the second category. These two categories of shifts are separated in order to save computing time for recalculation of the full event kinematics, which is not needed in the case of event weight shifts. For each systematic variation a new response matrix is calculated and data unfolding is performed. The resulting relative uncertainty in each bin i of $\frac{d^2\sigma}{dy_{jet}dp_{T,jet}}$ distribution

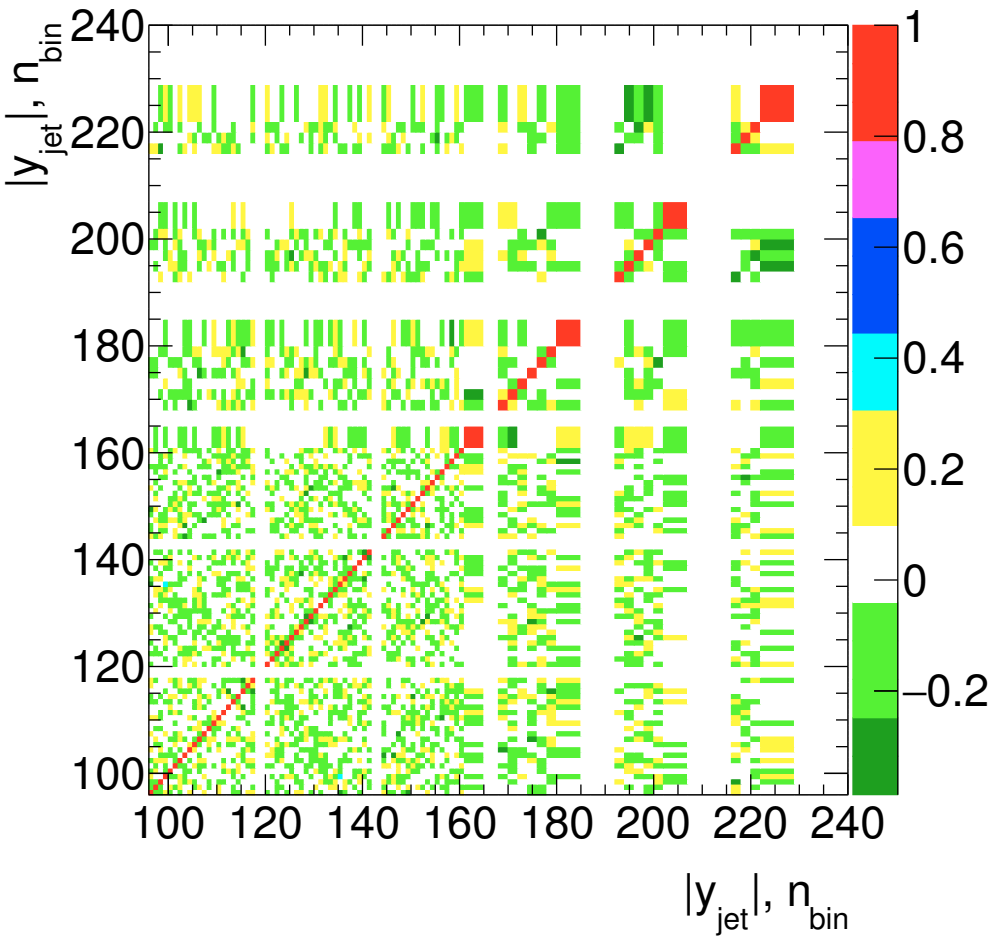


FIGURE 11.1: The correlation matrix for measured differential cross-sections in bins of $|y_{jet}|$ and p_T^{jet}.

is calculated as:

$$\Delta(\sigma_i) = \frac{(\sigma_{Nom,i} - \sigma_{Var,i})}{\sigma_{Nom,i}}, \tag{11.3}$$

where $\sigma_{Nom,i}$ is the nominal cross-section value in bin i, and $\sigma_{Var,i}$ is the cross-section value measured with the systematic shift. The uncertainties are calculated separately for up and down systematics variations, which results in asymmetric uncertainties.

The total uncertainty in each bin is calculated as:

$$\Delta^{Total}(\sigma_i) = \sqrt{\sum_j^N (\Delta^j(\sigma_i))^2}, \tag{11.4}$$

where $\sum_j^N$ runs over systematic uncertainties. It is calculated separately for all resulting up and down shifts.

11.2.1 Electron uncertainties

The following sources of systematic uncertainties related to electron measurements are studied:

Electron reconstruction, identification and trigger efficiencies uncertainties come from the measurements of electron efficiencies [54] and are propagated to the final result by varying the correction scale factors in Monte Carlo simulation by the corresponding efficiency uncertainty.

Electron energy scale uncertainty comes from the electron energy scale calibration procedure. It is evaluated by varying the electron energy scale correction (see Section 6.2.2, [64]) by its total uncertainty. The variation is performed using simulated samples only.

Electron energy resolution uncertainty comes from the uncertainties in the measurement of the resolution constant term (see Section 6.2.2, [64]), which represents discrepancies between the energy resolution in data and simulation (see Section 6.2.2). To propagate this uncertainty to the final result, the electron energy resolution correction is varied by its uncertainty in simulation .

11.2.2 Jet uncertainties

Jet energy scale uncertainties are associated with the jet energy calibration, which is detailed in Section 6.3. The total JES uncertainty is compiled from multiple sources [65]:

- 47 sources from in-situ methods used to derive the JES calibration, which are related to detector description, physics modelling and statistics/methods,
- 2 uncertainties from η-intercalibration method,
- a single-hadron response uncertainty which affects only high-p_T jets,
- Monte Carlo non-closure term, which takes into account the differences in the jet response between different Monte Carlo samples,
- 4 uncertainty sources associated to the pile-up corrections of the jet kinematics,

• 1 source of uncertainty coming from correction for high-p_T jets with energy not fully contained within the calorimeter jet ("punch-through" jets).

The sources of systematic uncertainties listed above are treated as independent from each other and result in a JES uncertainty defined in terms of 65 parameters (nuisance parameters). In this analysis a reduced set of nuisance parameters is used. Uncertainties from in-situ methods are combined into 6 nuisance parameters. For each source of JES uncertainty jet four-vector is scaled up and down by the corresponding uncertainty value. The summary of JES uncertainties is shown in Figures 11.2 – 11.3. The dominant contribution in the central region comes from in-situ JES calibration methods, giving from 3 % to 5 % uncertainty in the $|y_{jet}| < 3$ region. In the forward region the dominant is the η-intercalibration source, contributing with about 75 % uncertainty in the $25 < p_T^{jet} < 50$ GeV range, and with about 30 % in the $100 < p_T^{jet} < 200$ GeV range. Uncertainties from the pile-up jet corrections are from about 2 % in the central rapidity region up to 12 % in the forward region. Uncertainties related to the single hadron response measurements, Monte Carlo non-closure and "punch-through" correction are at the level of 0.001 % at low p_T^{jet} and up to 2 % in the last p_T^{jet} bin, $400 < p_T^{jet} < 1050$ GeV.

The quark- and gluon-initiated jets have different responses in the calorimeter, which leads to an additional source of the uncertainty in JES, if the flavour content of the jet in final state of the process under study is different compared to that in the sample used to derive the calibration. In addition, due to the differences in the jet fragmentation in different Monte Carlo, the response of the gluon-initiated jets changes when switching Monte Carlo generators. In order to take into account the residual flavour dependence of the JES, two **flavour-based sources of uncertainties** are propagated to the analysis results.

The jet energy resolution can have an impact on bin migrations, and therefore on unfolded cross-sections. The jet energy resolution has been measured as a function of jet p_T and η using the transverse momentum balance of jets in dijet events [84]. In general, it is found that the jet energy resolution in data is well reproduced by the simulation. Consequently, no JER correction is applied in this analysis, but the uncertainty due to JER is evaluated. **Jet energy resolution uncertainty** is coming from the mismodelling of the detector jet energy resolution in the Monte Carlo simulation.

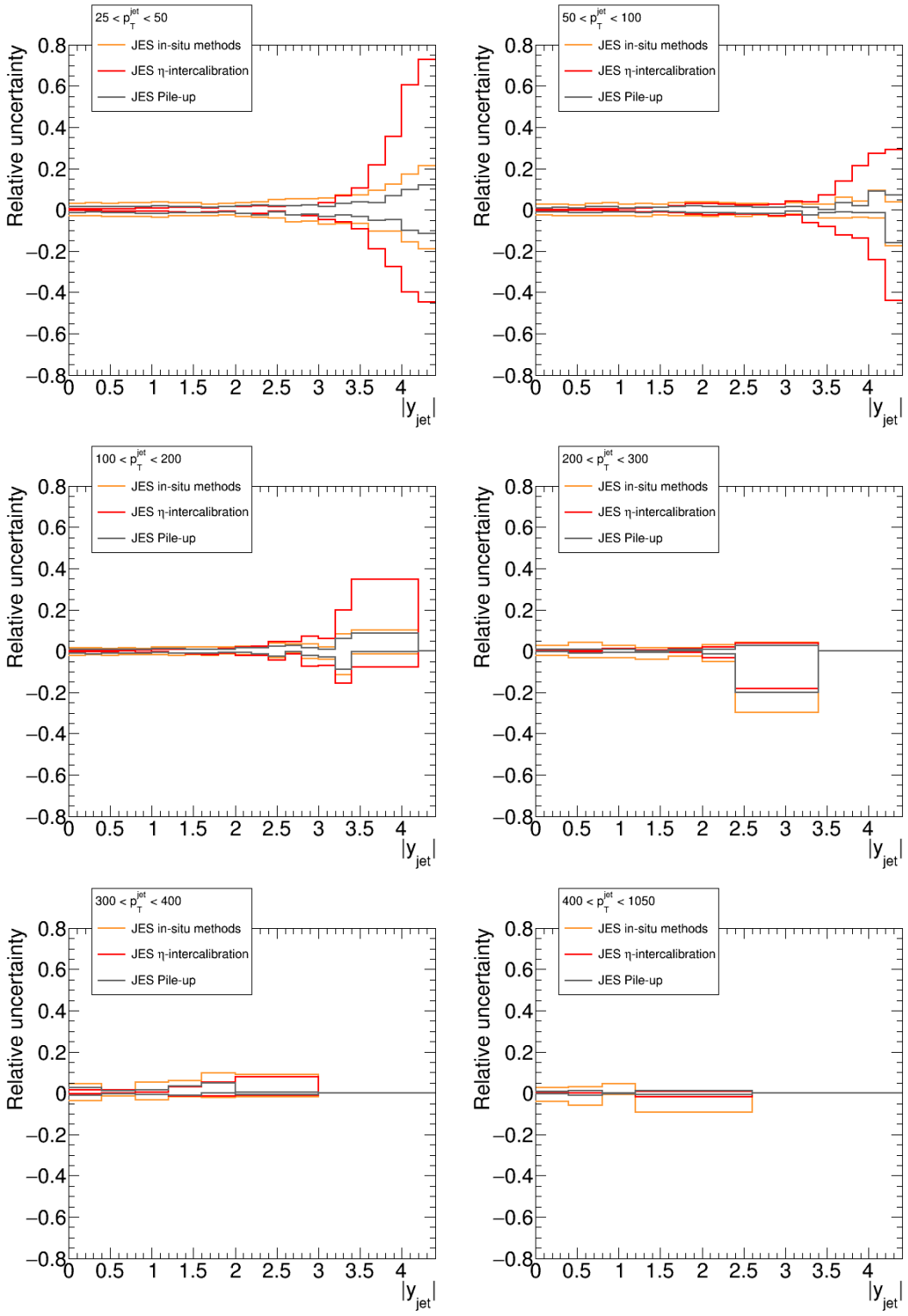


FIGURE 11.2: Systematic uncertainties from JES related sources in the measured differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production as a function of $|y_{jet}|$ in p_T^{jet} bins.

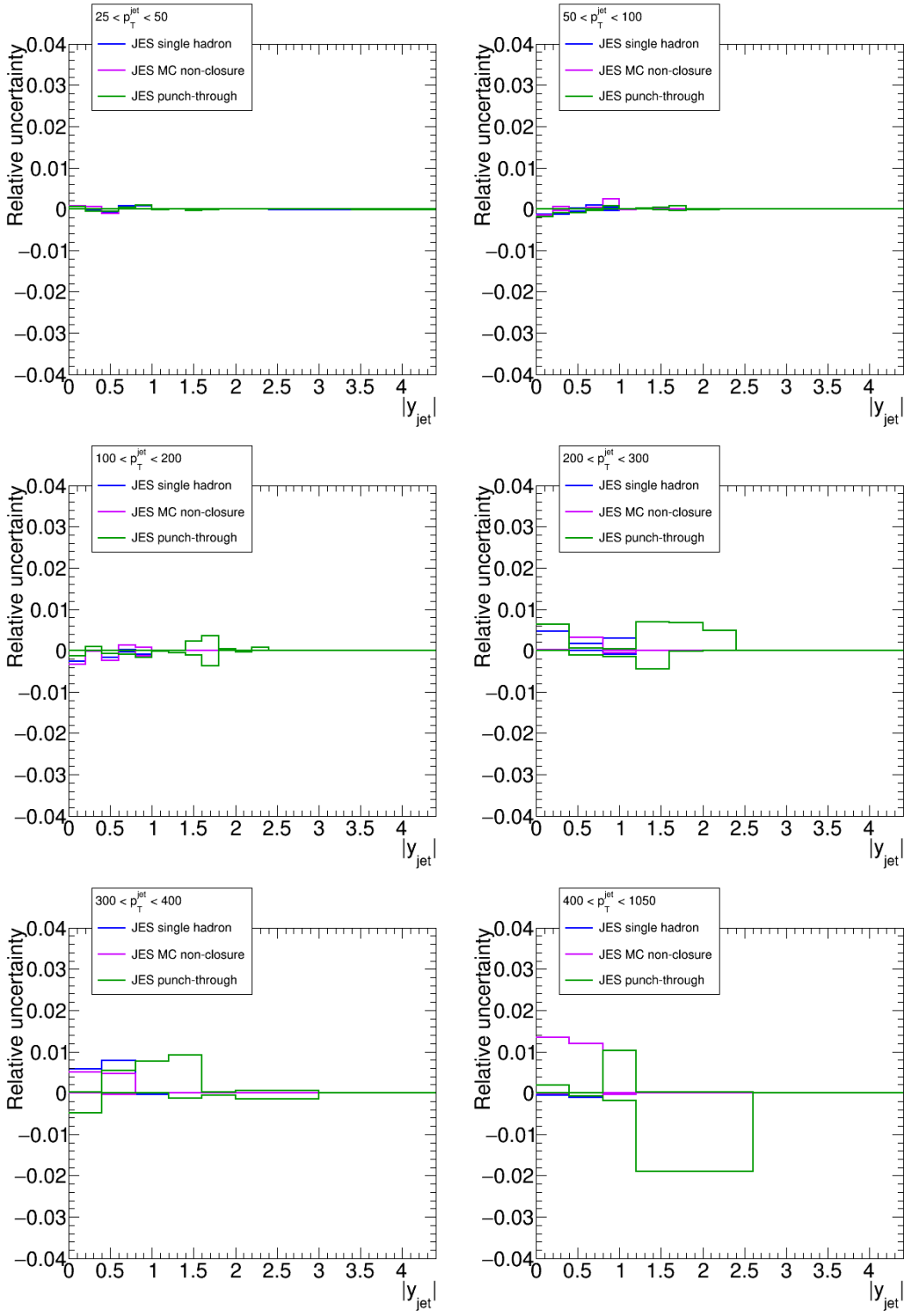


FIGURE 11.3: Systematic uncertainties from JES related sources in the measured differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production as a function of $|y_{jet}|$ in p_T^{jet} bins. The total experimental uncertainty is shown in black.

To evaluate the impact of JER systematic uncertainty effect on the measured cross-section, each jet four-vector in Monte Carlo simulation is scaled by a random number pulled from the Gaussian distribution with mean value 1 and width equals to the smearing factor, defined as:

$$SmearFactor = \sqrt{(\sigma_{MC} + \Delta\sigma)^2 - \sigma_{MC}^2}, \tag{11.5}$$

where σ_{MC} is the jet resolution taken from the Monte Carlo simulation as the width of the distribution of p_T^{reco}/p_T^{truth}, and $\Delta\sigma$ is the difference between the jet resolution in data and Monte Carlo simulation, measured with the in-situ methods. The scaling for each jet is repeated 100 times and the resulting averaged response matrix is used to unfold the data.

Jet vertex fraction uncertainty is assigned to different JVF cut efficiencies in data and Monte Carlo. This uncertainty can have an impact only on low-p_T jets in the central η region, where JVF cut is applied. The nominal value of 25 % used for this analysis is varied up and down by 3 % in Monte Carlo simulation [59].

11.2.3 Uncertainty from the data-driven background subtraction

The sources of uncertainties coming from the data-driven method of multijet and W($\rightarrow e\nu$) + jet backgrounds estimation are described in Section 7.1. For each of these sources the variation is performed and the resulting background enriched template together with its normalization is obtained. Then the subtraction of background contributions from the data is performed using normalized templates from one-by-one variations, and the unfolding procedure is repeated for each resulting data distribution. The main contribution of the data-driven background estimation method to the resulting uncertainty comes from the fit range variation.

11.2.4 Uncertainty from the Monte Carlo based background subtraction

All electroweak backgrounds (except W($\rightarrow e\nu$) + jet) are estimated using Monte Carlo simulation, as described in Section 7. The uncertainty which comes from this

procedure is estimated by varying cross-sections for each background process by corresponding cross-section uncertainties. The cross-section values and uncertainties are shown in Table 7.1. The resulting uncertainty is at the level of 0.2 – 0.8 % in all bins.

11.2.5 Uncertainty in the unfolding procedure

The resulting difference between the truth-level distribution and the reconstructed-level distribution in Sherpa, unfolded using the response matrix from Alpgen+Pythia6, is taken as the uncertainty from the unfolding procedure. The bias is symmetrised and added in quadrature to other uncertainties.

11.2.6 Summary

The summary of all uncertainties as a function of $|y_{jet}|$ in p_T^{jet} bins is shown in Figure 11.4 for jet related sources and in Figure 11.5 for electron, Monte Carlo statistics, the data-driven multijet and $W(\rightarrow e\nu)$ + jet backgrounds estimation and unfolding method. The total uncertainty is shown in both figures.

The dominant source of uncertainty comes from the JES uncertainties, giving about 5 % uncertainty in the final result in the central region and up to 80 % in the forward region in the $25 < p_T^{jet} < 50$ GeV range, which comes mainly from the η-intercalibration sources (see Figure 11.2). JES uncertainty decreases with increasing p_T^{jet} and results in about 3 – 10 % in the last p_T bins. JER uncertainty is also one of the main sources of uncertainty in the low p_T^{jet} region. However, in the $p_T^{jet} > 300$ GeV the contribution from this uncertainty becomes much smaller and is about 4 %. Jet flavour related sources give the resulting uncertainty of about 3 % in the central rapidity region in all p_T^{jet} bins, and up to 30 % in the forward region. The JVF cut variation has a valuable impact only in the first p_T^{jet} bin, and results in about 1 – 3 % uncertainty.

The uncertainty from Monte Carlo statistics is large in the last p_T^{jet} bins, increasing to about 15 – 25 %. The uncertainty in the unfolding procedure is relatively small in most bins; however, in some bins the contribution from this source increases to about 15 – 25 %. The uncertainty from the data-driven method of multijet and $W(\rightarrow l\nu)$ backgrounds subtraction is 1 – 3 % in all bins of the jet

rapidity and transverse momentum. The electron-related uncertainties contribute with about 1 % in all p_T^{jet} bins. The smallest contribution comes from the Monte Carlo based background estimation, which is less than 1 % in all p_T^{jet} bins.

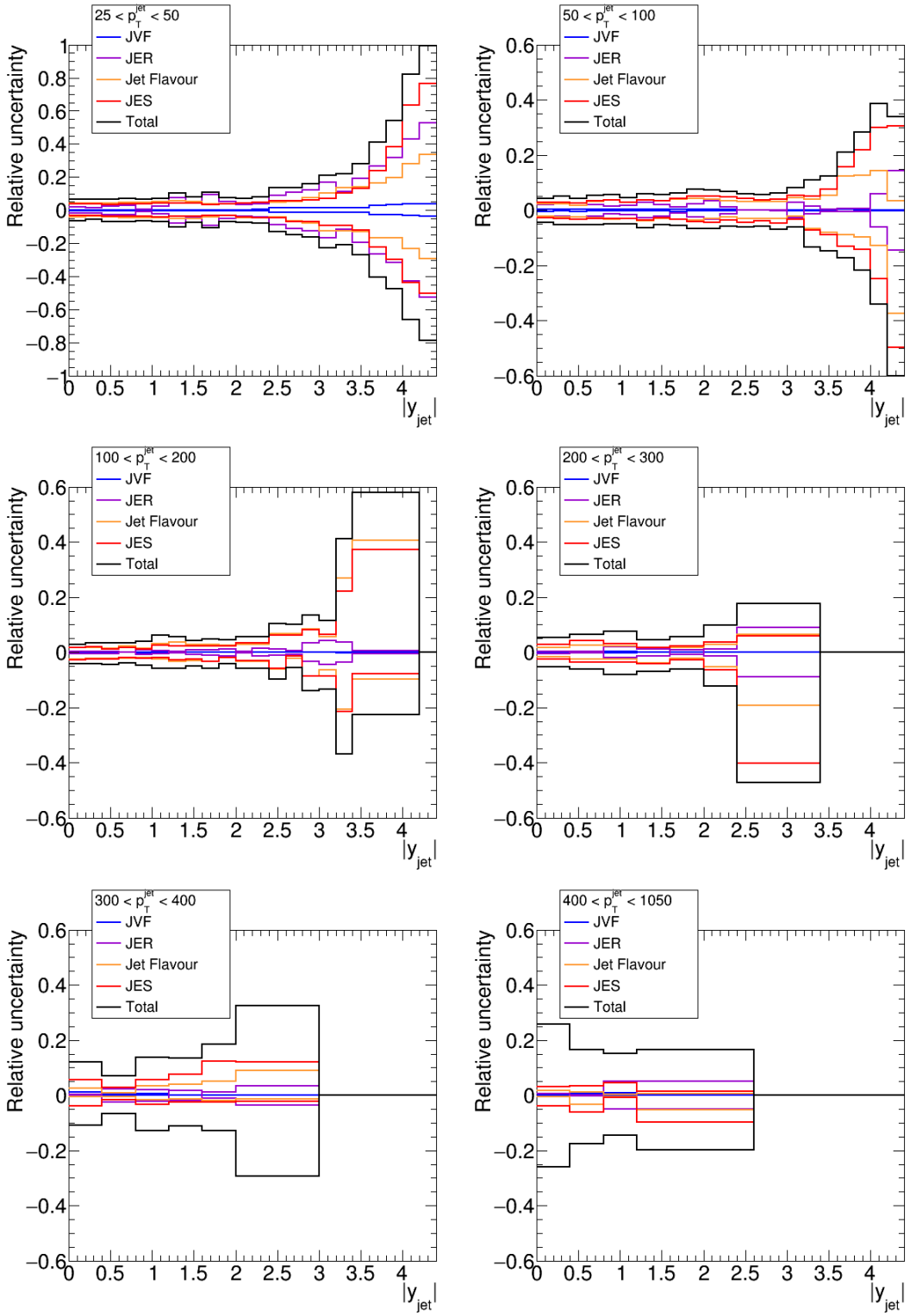


FIGURE 11.4: Systematic uncertainties from jet related sources in the measured differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production as a function of $|y_{jet}|$ in p_T^{jet} bins. The total experimental uncertainty is shown in black.

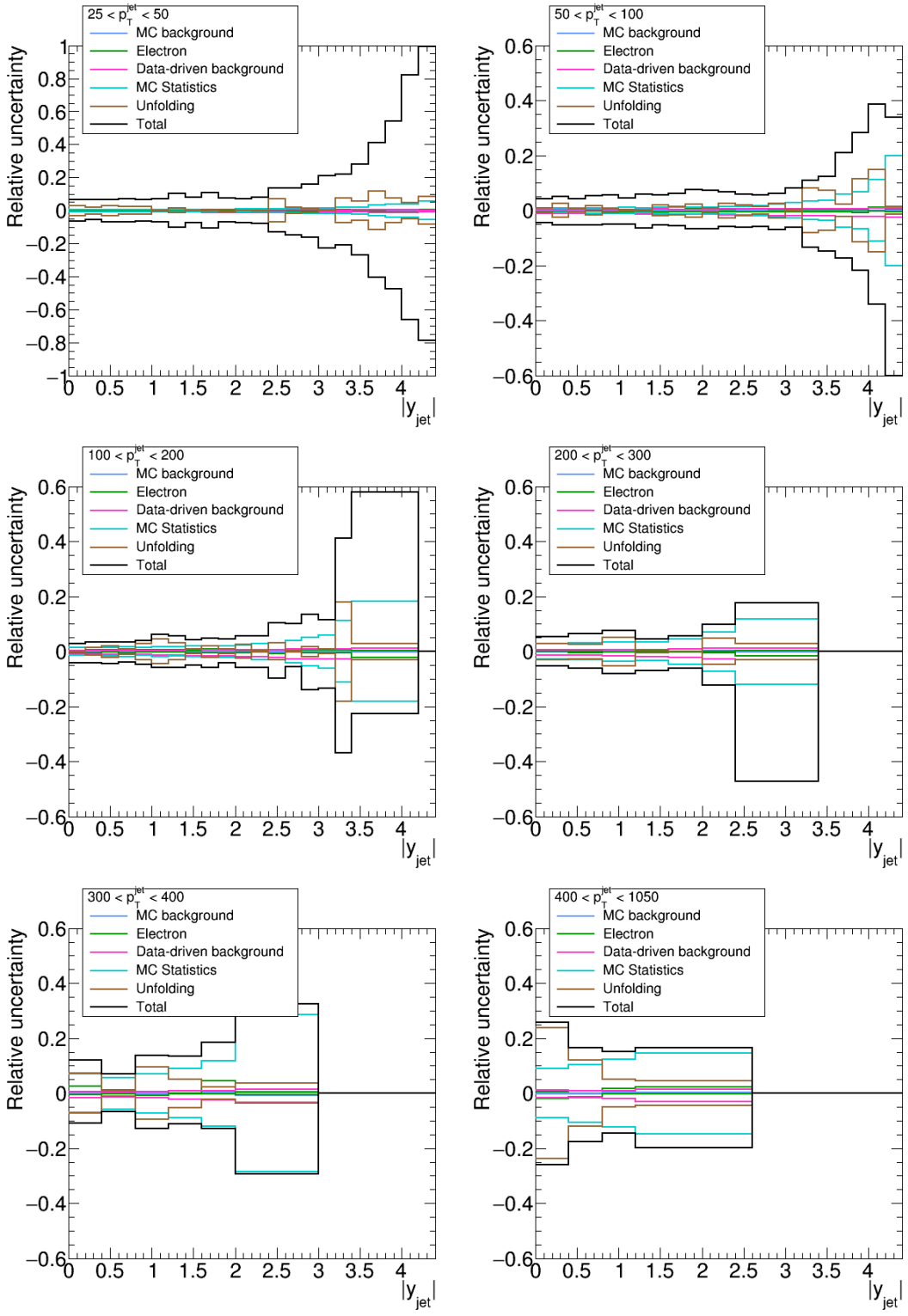


FIGURE 11.5: Systematic uncertainties from electrons, Monte Carlo statistics and unfolding method in the measured differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production as a function of $|y_{jet}|$ in p_T^{jet} bins. The total experimental uncertainty is shown in black.

Chapter 12

Results of the Cross-Section Measurement

The double differential cross section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process as a function of p_T^{jet} and $|y_{jet}|$ is calculated using formula:

$$\frac{d^2\sigma}{dp_T^{jet}|y_{jet}|} = \frac{1}{\mathcal{L}} \frac{N_{sig}}{\Delta_{bin}(p_T^{jet}, |y_{jet}|)}, \tag{12.1}$$

where $\mathcal{L}$ is the total integrated luminosity, N_{sig} is the total number of the signal events, and $\Delta_{bin}(p_T^{jet}, |y_{jet}|)$ is the width of the given bin of the two-dimensional $(p_T^{jet}, |y_{jet}|)$ distribution. N_{sig} is obtained by subtracting all background contributions from data and unfolding the resulting distribution to the particle level using Monte Carlo simulated samples (see Section 10).

The results of the cross section measurement as a function of $|y_{jet}|$ in different p_T^{jet} bins together with the final uncertainty in the measurement are shown in Figures 12.1 – 12.6. The resulting systematic and Monte Carlo statistics uncertainties are included into the uncertainty band, and the data statistics uncertainty is shown on the data points. The measured cross section is compared with the predictions from the LO matrix-element Monte Carlo generators Sherpa and Alpgen, where the latter is showered using Pythia6.

The measured cross section as a function of $|y_{jet}|$ in the lowest p_T^{jet} bin, $25 < p_T^{jet} < 50$ GeV, is shown in Figure 12.1. The total systematic uncertainty is from 7 % in the central region up to 100 % in the most forward $|y_{jet}|$ bin. Main

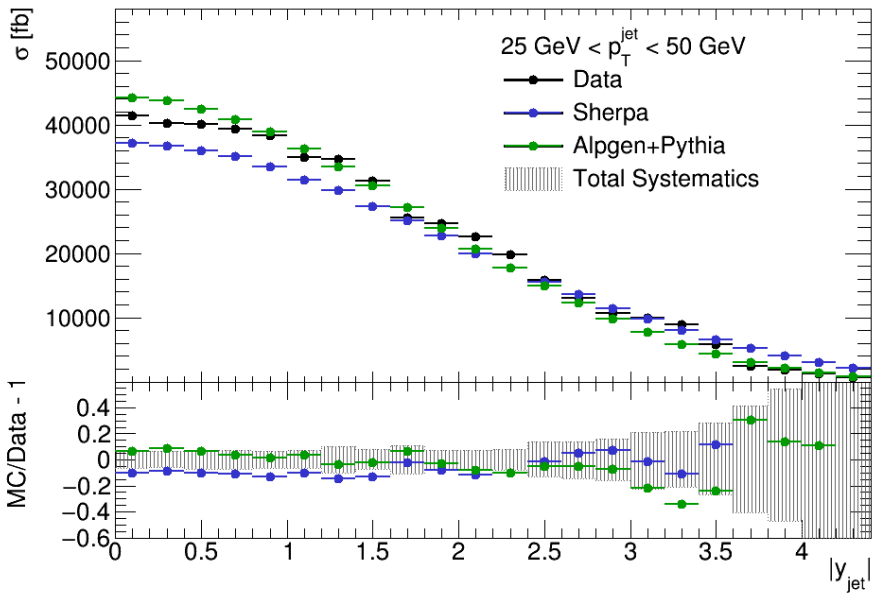


FIGURE 12.1: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $25 < p_T^{jet} < 50$ GeV.

contribution to this uncertainty comes from the JES calibration, in particular the large uncertainty in the forward region is connected to the η-intercalibration. The JER and jet flavour related uncertainties also have large impact in this p_T^{jet} bin. Data and Alpgen+Pythia6 prediction are in a good agreement in the central rapidity region, while in the forward region there are some discrepancies, which are covered by systematic uncertainties in most of the $|y_{jet}|$ bins. Prediction from Sherpa agrees with data slightly worse compared to Alpgen+Pythia6 in the central region and in the very forward region it tends to overestimate data.

In the next p_T bin, $50 < p_T^{jet} < 100$ GeV (see Figure 12.2), the total systematic uncertainty becomes smaller, varying from 5 – 6 % in the central rapidity region to 60 % in the last $|y_{jet}|$ bin. The largest contribution to the total uncertainty comes from the JES calibration together with significant contributions from jet flavour and, in the forward region, from the unfolding method and Monte Carlo statistics. The JER uncertainty does not belong to the dominant contributions to the total uncertainty in this p_T^{jet} region, except for a few particular $|y_{jet}|$ bins, where it fluctuates up. Again, Alpgen+Pythia6 provides better description of data than Sherpa, being in agreement within uncertainties in most of the bins, while Sherpa prediction overestimates data in the very forward region and slightly underestimates the data in the central region.

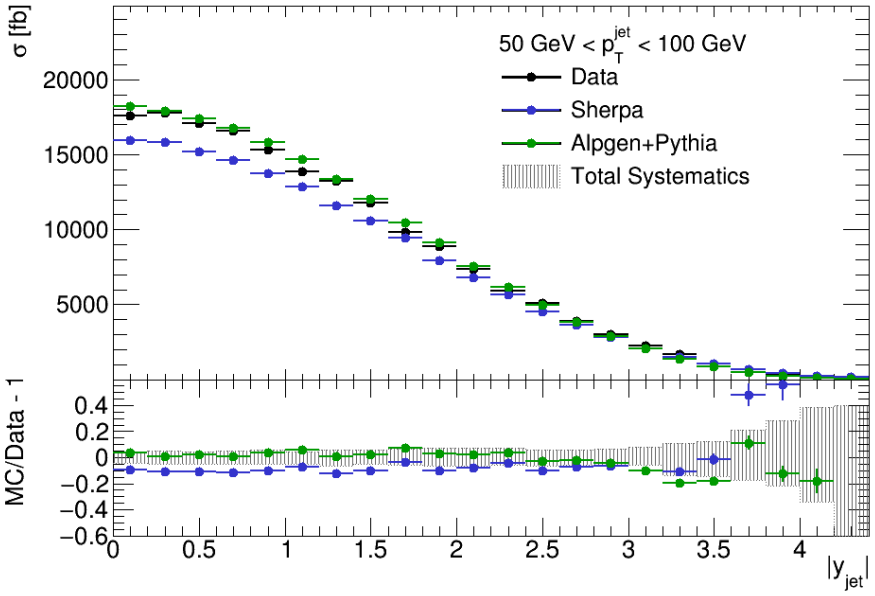


FIGURE 12.2: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $50 < p_T^{jet} < 100$ GeV.

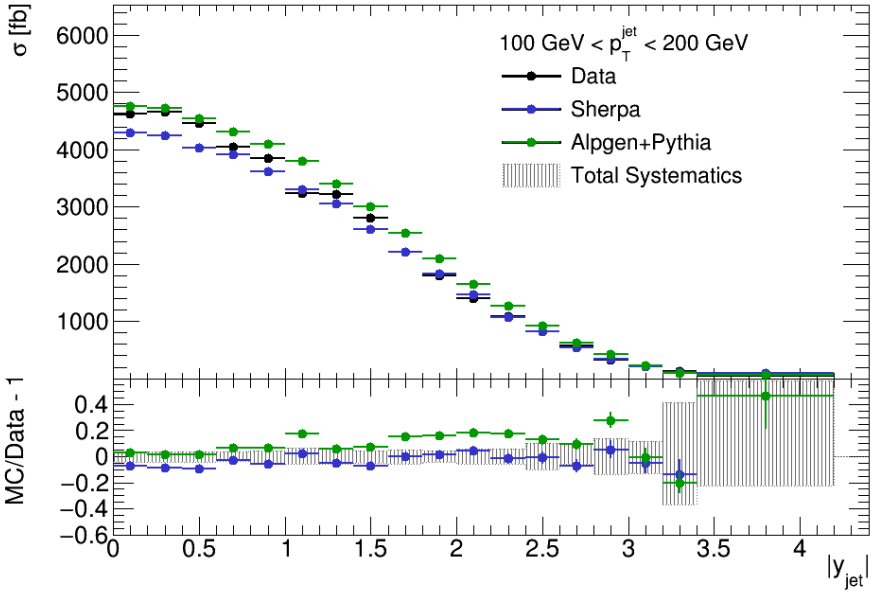


FIGURE 12.3: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $100 < p_T^{jet} < 200$ GeV.

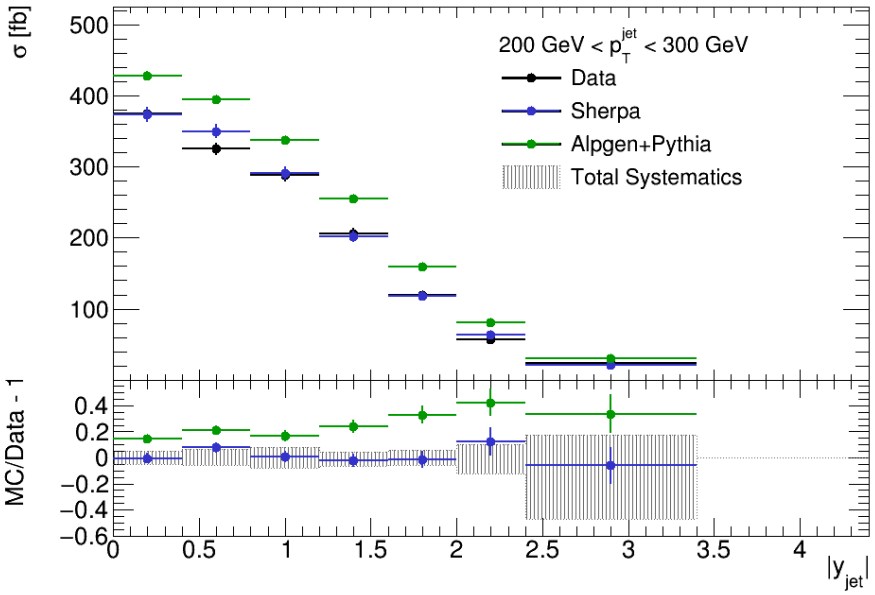


FIGURE 12.4: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $200 < p_T^{jet} < 300$ GeV.

For the cross section measured in the next p_T^{jet} bin, $100 < p_T^{jet} < 200$ GeV, shown in Figure 12.3, the total uncertainty decreases further and results in about 4 % in the central region and up to 40 – 55 % in the last two $|y_{jet}|$ bins. The dominant sources of uncertainties are the JES calibration and jet flavour, while the uncertainties from the unfolding method and Monte Carlo statistics are contributing significantly in the forward $|y_{jet}|$ bins. Sherpa describe the data well in this p_T^{jet} region, being in most of the $|y_{jet}|$ bins within systematic uncertainty band, while Alpgen+Pythia6 tends to overestimate the data.

The $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross section in the $200 < p_T^{jet} < 300$ GeV bin (see Figure 12.4) is measured with the total uncertainty varying from 5 – 6 % in the central region up to 45 % in the last $|y_{jet}|$ bin, $2.4 < |y_{jet}| < 3.4$. Jet related uncertainties, coming from the JES calibration, jet flavour and JER sources give the main contribution to the total uncertainty, together with Monte Carlo statistics uncertainty in the last $|y_{jet}|$ bins. Predictions from Sherpa describe data well, while predictions from Alpgen+Pythia6 overestimate data.

In the last two p_T^{jet} bins, shown in Figures 12.5 – 12.6, the dominant uncertainties are coming from the Monte Carlo statistics and the unfolding method, while the jet related systematic uncertainties become relatively small. The resulting uncertainty changes from 8 % up to 30 % depending on $|y_{jet}|$ bin. In

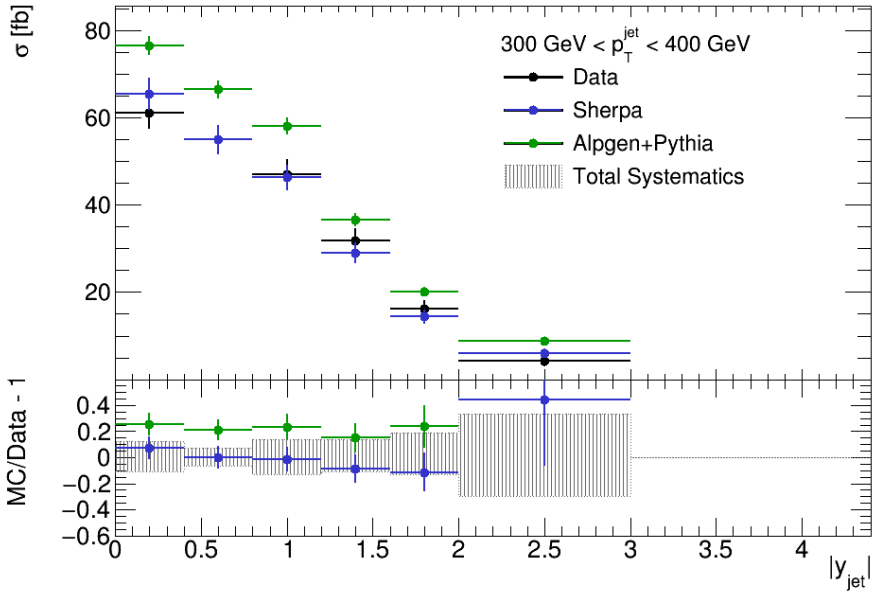


FIGURE 12.5: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $300 < p_T^{jet} < 400$ GeV.

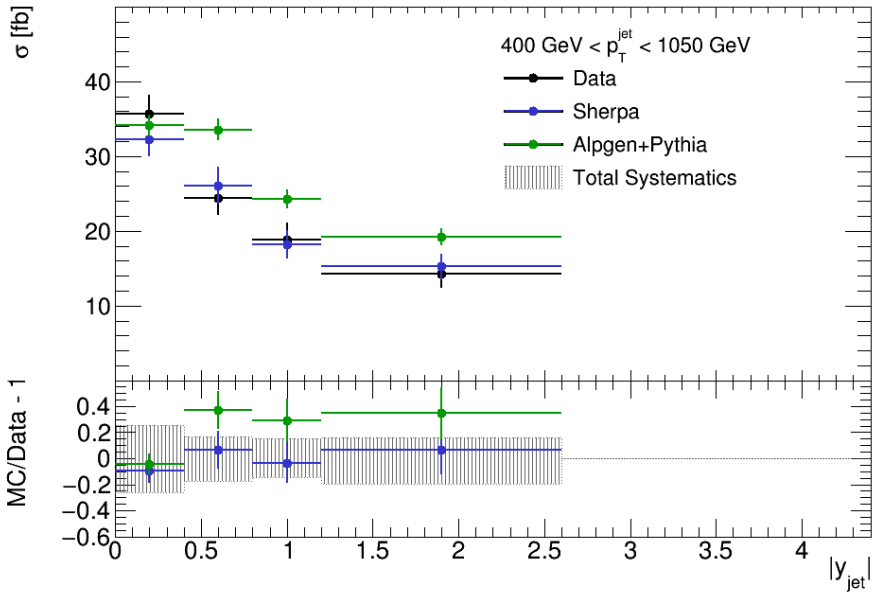


FIGURE 12.6: The comparison of the measured double-differential cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet process and prediction from Sherpa and Alpgen+Pythia6 event generators as a function of $|y_{jet}|$ in $400 < p_T^{jet} < 1050$ GeV.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.2	41416.90	0.48	6.72	6.53
0.2 - 0.4	40286.14	0.43	6.47	5.65
0.4 - 0.6	40024.77	0.45	6.73	7.00
0.6 - 0.8	39397.48	0.47	6.99	7.11
0.8 - 1.0	38297.98	0.45	6.59	6.62
1.0 - 1.2	34997.66	0.45	7.02	6.68
1.2 - 1.4	34712.93	0.58	10.48	10.09
1.4 - 1.6	31246.97	0.5	8.07	7.5
1.6 - 1.8	25513.13	0.57	10.65	10.74
1.8 - 2.0	24579.48	0.52	7.55	6.96
2.0 - 2.2	22574.02	0.57	7.04	7.2
2.2 - 2.4	19742.15	0.61	7.85	7.8
2.4 - 2.6	15778.84	0.8	13.66	13.04
2.6 - 2.8	13005.97	0.92	13.71	14.65
2.8 - 3.0	10659.27	0.92	15.9	16.29
3.0 - 3.2	9896.68	0.85	20.98	22.79
3.2 - 3.4	8914.09	0.97	22.0	21.04
3.4 - 3.6	5797.2	1.25	28.1	27.02
3.6 - 3.8	2369.79	1.82	41.29	40.5
3.8 - 4.0	1828.41	2.11	54.18	47.39
4.0 - 4.2	1218.96	2.27	81.19	66.14
4.2 - 4.4	714.67	3.17	99.65	79.15

TABLE 12.1: The measured $pp \to Z/\gamma^*(\to e^+e^-) +$ jet production cross-section as a function of $|y_{jet}|$ in the $25 < p_T^{jet} < 50$ GeV range.

most of the $|y_{jet}|$ bins the Alpgen+Pythia6 predictions tend to overestimate the data, while Sherpa agrees with the data in all bins within systematic and statistics uncertainties.

The results of the measurement are summarised in Tables 12.1 – 12.6.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.2	17615.01	0.61	4.38	4.5
0.2 - 0.4	17780.99	0.79	5.05	5.25
0.4 - 0.6	17077.59	0.83	4.26	5.14
0.6 - 0.8	16583.29	0.78	5.43	5.29
0.8 - 1.0	15293.47	0.77	5.59	4.9
1.0 - 1.2	13846.05	0.78	4.54	5.02
1.2 - 1.4	13248.96	0.89	5.92	6.24
1.4 - 1.6	11757.98	0.83	5.77	5.31
1.6 - 1.8	9796.62	1.01	6.31	5.52
1.8 - 2.0	8854.49	0.97	7.57	6.63
2.0 - 2.2	7372.75	1.02	7.34	6.62
2.2 - 2.4	5936.18	1.18	6.9	5.74
2.4 - 2.6	5062.51	1.44	5.79	6.18
2.6 - 2.8	3895.69	1.6	5.52	5.71
2.8 - 3.0	2988.91	1.71	6.22	6.8
3.0 - 3.2	2272.48	1.74	8.04	5.94
3.2 - 3.4	1686.89	2.4	11.02	13.44
3.4 - 3.6	1061.75	2.58	12.43	14.75
3.6 - 3.8	450.085	4.18	21.07	17.28
3.8 - 4.0	282.96	5.99	28.46	21.88
4.0 - 4.2	127.05	9.26	38.71	34.14
4.2 - 4.4	28.68	18.54	39.89	67.39

TABLE 12.2: The measured $pp \to Z/\gamma^*(\to e^+e^-) +$ jet production cross-section as a function of $|y_{jet}|$ in the $50 < p_T^{jet} < 100$ GeV range.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.2	4620.76	1.32	2.96	4.21
0.2 - 0.4	4658.44	1.28	3.39	4.09
0.4 - 0.6	4456.11	1.32	3.47	4.39
0.6 - 0.8	4042.74	1.47	3.5	3.86
0.8 - 1.0	3849.75	1.46	4.0	4.75
1.0 - 1.2	3235.92	1.72	6.33	5.9
1.2 - 1.4	3225.1	1.38	5.6	5.77
1.4 - 1.6	2807	1.4	4.27	4.85
1.6 - 1.8	2213.2	1.92	4.85	5.72
1.8 - 2.0	1803.97	1.91	4.44	4.18
2.0 - 2.2	1397.93	2.36	5.62	5.87
2.2 - 2.4	1081.69	2.63	5.51	5.81
2.4 - 2.6	821.36	3.33	10.32	9.87
2.6 - 2.8	573.45	3.39	10.16	5.55
2.8 - 3.0	328.32	4.14	13.49	13.98
3.0 - 3.2	219.06	5.72	11.44	13.25
3.2 - 3.4	123.19	8.56	41.11	36.81
3.4 - 4.2	38.09	15.82	58.07	22.37

TABLE 12.3: The measured $pp \to Z/\gamma^*(\to e^+e^-) +$ jet production cross-section as a function of $|y_{jet}|$ in the $100 < p_T^{jet} < 200$ GeV range.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.4	375.13	2.6	5.35	5.09
0.4 - 0.8	325.19	2.48	6.55	6.13
0.8 - 1.2	289	3.31	7.69	8.02
1.2 - 1.6	206.01	3.74	4.51	6.88
1.6 - 2.0	119.76	4.53	5.48	6.08
2.0 - 2.4	56.43	6.65	9.89	12.27
2.4 - 3.4	22.8	9.75	17.71	47.41

TABLE 12.4: The measured $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production cross-section as a function of $|y_{jet}|$ in the $200 < p_T^{jet} < 300$ GeV range.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.4	61.09	6.21	12.09	10.86
0.4 - 0.8	54.91	5.83	7.08	6.68
0.8 - 1.2	47.03	7.09	13.71	12.78
1.2 - 1.6	31.76	8.72	13.58	10.98
1.6 - 2.0	16.22	11.82	18.62	12.78
2.0 - 3.0	4.15	30.12	33.2	29.86

TABLE 12.5: The measured $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production cross-section as a function of $|y_{jet}|$ in the $300 < p_T^{jet} < 400$ GeV range.

$\|y_{jet}\|$ bin	σ [fb]	Stat. unc. [%]	Syst. unc. up [%]	Syst. unc. down [%]
0.0 - 0.4	35.68	6.95	25.74	25.9
0.4 - 0.8	24.47	9.77	16.34	17.36
0.8 - 1.2	18.82	12.2	15.13	14.38
1.2 - 2.6	14.28	13.59	16.0	19.28

TABLE 12.6: The measured $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production cross-section as a function of $|y_{jet}|$ in the $400 < p_T^{jet} < 1050$ GeV range.

Chapter 13

Summary

This thesis presents the measurement of the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet differential cross-section and the study of the jet quark-gluon decomposition. The cross-section of the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production is measured double-differentially as a function of the absolute rapidity and the transverse momentum of inclusive jets. The study of the jet quark-gluon decomposition is performed in bins of the transverse momentum and the absolute rapidity of the highest-p_T jet in the event. The data collected with the ATLAS detector at the LHC in 2012 at a center-of-mass energy $\sqrt{s} = 8$ TeV are used.

Jets for both the cross-section measurement and the jet quark-gluon decomposition study are defined using the anti-k_T algorithm with radius R = 0.4 using the topological clusters as input at the hadron level and stable simulated particles at the particle level.

The study of the jet quark-gluon decomposition is focused on the discrimination between quark-initiated and gluon-initiated jets using different jet properties, which is studied using data-driven method. The quark-like and the gluon-like jet enriched samples, which are dijet, γ + jet and $Z(\to e^+e^-)$ + jet, are used. The $Z(\to e^+e^-)$ + jet event selection is optimised to get the high quark-like jet purity. The resulting quark-gluon discrimination using different variables provides about 40 – 50 % rejection of the gluon-like jet for a fixed quark-like jet acceptance of 70 %. The dominant systematic uncertainty on the quark acceptance and gluon rejection comes from the sample dependence.

The cross-section of $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production is measured in the fiducial volume defined by the detector geometry and resolution. Both final state electrons coming from Z decay are required to be detected in the central detector within $|\eta_{elec}| < 2.5$ and to have $p_T^{elec} > 20$ GeV. Jets kinematic range of $|y_{jet}| < 4.4$ and $p_T^{jet} > 25$ GeV is selected. The contributions from background processes for the $pp \to Z/\gamma^*(\to e^+e^-)$ + jet production cross-section measurement are estimated using both simulation and data-driven methods. The total background fraction in the selected $Z/\gamma^*(\to e^+e^-)$ + jet sample is about 3.2 %, where the dominant background contributions are coming from the mulitjet and $t\bar{t}$ processes. The measured cross section is corrected for detector effects using the Bayes' unfolding procedure.

Main sources of the systematic uncertainties in the measured cross-section are checked. The dominant contributions for $p_T^{jet} < 300$ GeV come from the JES calibration, in particular from the η-intercalibration in the forward region, jet flavour related sources, and the uncertainty from JER in the lowest p_T^{jet} bin, $25 < p_T^{jet} < 50$ GeV. For $p_T^{jet} > 300$ GeV the dominant contributions come from the limited Monte Carlo samples statistics, together with the uncertainty in the unfolding procedure. For $p_T^{jet} < 300$ GeV the total uncertainty on the measured cross-section is about 4 – 7 % in the central region and about 45 – 100 % in the forward region, while for $p_T^{jet} > 300$ GeV the uncertainty is from 10 to 30 % in most of the $|y_{jet}|$ bins.

The measured cross-section is compared to the predictions from Sherpa and Alpgen+Pythia6 Monte Carlo generators, which are based on the LO matrix elements supplemented by the parton showers. The predictions are normalised to the inclusive $pp \to Z/\gamma^* \to e^+e^-$ NNLO cross-section. In the central region the Sherpa prediction agrees within the uncertainty with the measured cross-section, while in the forward region it overestimates data. The prediction from Alpgen+Pythia6 describes data well for $p_T^{jet} < 200$ GeV, while at higher p_T^{jet} it tends to overestimate data.

The measured data thus have a discrimination power versus different Monte Carlo predictions. The measured $pp \to Z/\gamma^*(\to e^+e^-)$ + jet cross-section can be used for the PDF fits, as well as for validation of NNLO QCD prediction for Z + jet process.

Bibliography

[1] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.* **22**, *(1961) 579–588.* doi: 10.1016/0029-5582(61)90469-2.

[2] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.* **19**, *(1967) 1264–1266.* doi: 10.1103/PhysRevLett.19.1264.

[3] A. Salam. Renormalizability of gauge theories. *Phys. Rev.* **127** *(1962) 331.* doi: 10.1103/PhysRev.127.331.

[4] Wikipedia. Elementary particle — wikipedia, the free encyclopedia, 2015. URL https://en.wikipedia.org/w/index.php?title=Elementary_particle&oldid=684282599. [Online; accessed 19-November-2015].

[5] G. Sterman J. C. Collins, D. E. Soper. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.* **5** *(1989) 1.* doi: 10.1142/9789814503266_0001.

[6] J. M. Campbell, J. W. Huston, W. J. Stirling. Hard Interactions of Quarks and Gluons: a Primer for LHC Physics. *Rept. Prog. Phys.* **70** *(2007) 89.* doi: 10.1088/0034-4885/70/1/R02.

[7] J. C. Collins. Renormalization: general theory. *arXiv:hep-th/0602121*, 2006.

[8] D. A. Britzger. *Regularized Unfolding of Jet Cross Sections in Deep-Inelastic ep Scattering at HERA and Determination of the Strong Coupling Constant.* PhD thesis, Universität Hamburg, 2013. URL http://www-h1.desy.de/psfiles/theses/h1th-831.pdf.

[9] G. Dissertori. The Determination of the Strong Coupling Constant. *arXiv:hep-ex/1506.05407*, 2015.

[10] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C* **75**, *no. 5, 204 (2015).* doi: 10.1140/epjc/s10052-015-3397-6.

[11] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D* **48** *(1993) 3160.* doi: 10.1103/PhysRevD.48.3160.

[12] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber. Better jet clustering algorithms. *JHEP* **9708** *(1997) 001.* doi: 10.1088/1126-6708/1997/08/001.

[13] M. Cacciari and G. P. Salam and G Soyez. The $anti - k_t$ jet clustering algorithm. *JHEP* **0804** *(2008) 063.* doi: 10.1088/1126-6708/2008/04/063.

[14] G. P. Salam and G. Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP* **0705** *(2007) 086.* doi: 10.1088/1126-6708/2007/05/086.

[15] DELPHI Collaboration. Investigation of the Splitting Quark and Gluon Jets. DELPHI 97-72 CONF 58, 1997.

[16] T. Schörner-Sadenius (Ed.). *The Large Hadron Collider: Harvest of Run 1.* Springer International Publishing Cham, 2015. doi: 10.1007/978-3-319-15001-7.

[17] K. A. Olive et al. (Particle Data Group). The Review of Particle Physics. *Chin. Phys. C,* **38**, *090001*, (2014) and 2015 update.

[18] A. Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.* **504** *(2011) 145.* doi: 10.1016/j.physrep.2011.03.005.

[19] B. R. Webber. Fragmentation and hadronization. *Int. J. Mod. Phys. A* **15S1** *(2000) 577.* doi: 10.1142/S0217751X00005334.

[20] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP* **0902** *(2009) 007.* doi: 10.1088/1126-6708/2009/02/007.

[21] S. Catani, F. Krauss, R. Kuhn and B. R. Webber. QCD Matrix Elements + Parton Showers. *JHEP* **0111** *(2001) 063.* doi: 10.1088/1126-6708/2001/11/063.

[22] S. Frixione, P. Nason and C. Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP* **0711** *(2007) 070.* doi: 10.1088/1126-6708/2007/11/070.

[23] S. Frixione, P. Nason and G. Ridolfi. The POWHEG-hvq manual version 1.0. *arXiv:0707.3081 [hep-ph]*.

[24] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP* **0307** *(2003) 001*.

[25] M. L. Mangano, M. Moretti, F. Piccinini and M. Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP* **0701** *(2007) 013*. doi: 10.1088/1126-6708/2007/01/013.

[26] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour and B. R. Webber. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons. *JHEP* **0101** *(2001) 010*. doi: 10.1088/1126-6708/2001/01/010.

[27] J. M. Butterworth, J. R. Forshaw and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys. C* **72** *(1996) 637*. doi: 10.1007/s002880050286.

[28] T. Sjostrand, S. Mrenna and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP* **0605** *(2006) 026*, . doi: 10.1088/1126-6708/2006/05/026.

[29] T. Sjostrand, S. Mrenna and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.* **178** *(2008) 852*, . doi: 10.1016/j.cpc.2008.01.036.

[30] S. Jadach, J. H. Kuhn and Z. Was. TAUOLA: a library of Monte Carlo programs to simulate decays of polarized τ leptons. *Comput. Phys. Commun.* **64** *(1990) 275*. doi: 10.1016/0010-4655(91)90038-M.

[31] E. Barberio, B. van Eijk and Z. Was. PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays. *Comput. Phys. Commun.* **66** *(1991) 115*. doi: 10.1016/0010-4655(91)90012-A.

[32] F. Marcastel. CERN's Accelerator Complex. La chaîne des accélérateurs du CERN. Oct 2013. URL http://cds.cern.ch/record/1621583. General Photo.

[33] ATLAS Collaboration. Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C* **73** *(2013) 8, 2518*, . doi: 10.1140/epjc/s10052-013-2518-3.

[34] ATLAS Public Results. Luminosity Public Results. URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[35] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3** *(2008) S08003.*, . doi: 10.1088/1748-0221/3/08/S08003.

[36] J. J. Goodson. *Search for Supersymmetry in States with Large Missing Transverse Momentum and Three Leptons including a Z-Boson*. PhD thesis, Stony Brook University, May 2012. Presented 17 Apr 2012.

[37] ATLAS Collaboration. ATLAS magnet system : Technical Design Report. ATLAS-TDR-6; CERN-LHCC-97-018, 1997. URL https://cds.cern.ch/record/338080.

[38] ATLAS Collaboration. ATLAS Inner Detector : Technical Design Report, Volume 1. ATLAS TDR 4; CERN-LHCC-97-16, 1997. URL https://cds.cern.ch/record/331063.

[39] ATLAS Collaboration. ATLAS Liquid Argon Calorimeter : Technical Design Report. ATLAS-TDR-2 ; CERN-LHCC-96-041, 1996. URL http://cds.cern.ch/record/331061.

[40] K. A. Olive et al. (Particle Data Group). The Review of Particle Physics. *Phys. Rev. D* **86**, *010001*, 2012.

[41] F. Cavallari. Performance of calorimeters at the LHC. *Journal of Physics: Conference Series* **293 (2011) 012001**.

[42] ATLAS Collaboration. Monitoring and data quality assessment of the ATLAS liquid argon calorimeter. *JINST* **9** *(2014) P07024*, . doi: 10.1088/1748-0221/9/07/P07024.

[43] H. L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C.-P. Yuan. New parton distributions for collider physics. *Phys. Rev. D* **82** *(2010) 074024*. doi: 10.1103/PhysRevD.82.074024.

[44] ATLAS Collaboration. Summary of ATLAS Pythia 8 tunes. ATL-PHYS-PUB-2012-003, 2012. URL https://cds.cern.ch/record/1474107.

[45] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung. New generation of parton distributions with uncertainties from

global QCD analysis. *JHEP* **0207** *(2002) 012*. doi: 10.1088/1126-6708/2002/07/012.

[46] ATLAS Collaboration. Measurement of the W $\to$ lν and Z/γ $\to$ ll production cross sections in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *JHEP* **1012** *(2010) 060*, . doi: 10.1007/JHEP12(2010)060.

[47] ATLAS Collaboration. Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s}$ = 7TeV with the ATLAS detector. *JHEP* **1307** *(2013) 032*, . doi: 10.1007/JHEP07(2013)032.

[48] ATLAS Collaboration. First tuning of HERWIG/JIMMY to ATLAS data. ATL-PHYS-PUB-2010-014, ATL-COM-PHYS-2010-858, 2010. URL https://cds.cern.ch/record/1303025.

[49] P. Z. Skands. The Perugia Tunes. *arXiv:0905.3418 [hep-ph]*, 2009.

[50] ATLAS Collaboration. Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment. ATLAS-CONF-2012-042, 2012. URL http://cdsweb.cern.ch/record/1435196.

[51] ATLAS Collaboration. Performance of primary vertex reconstruction in proton–proton collisions at $\sqrt{s}$ = 7 TeV in the ATLAS experiment. ATLAS-CONF-2010-069, 2010. URL http://cdsweb.cern.ch/record/1281344.

[52] W. Lampl et.al. Calorimeter Clustering Algorithms: Description and Performance. ATL-LARG-PUB-2008-002, 2008. URL https://cds.cern.ch/record/1099735.

[53] ATLAS Internal. Electron reconstruction. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronReconstruction.

[54] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data. ATLAS-CONF-2014-032, 2014. URL http://cdsweb.cern.ch/record/1706245.

[55] M. Cacciari, G. P. Salam and G. Soyez. FastJet User Manual. *Eur. Phys. J. C* **72** *(2012) 1896*. doi: 10.1140/epjc/s10052-012-1896-2.

[56] P. Giovannini, on behalf of the ATLAS Liquid Argon Calorimeter Group. Local hadron calibration with ATLAS. Journal of Physics: Conference Series 293 (2011) 012057, 2011.

[57] ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C* **75** *(2015) 17*, . doi: 10.1140/epjc/s10052-014-3190-y.

[58] ATLAS Collaboration. Jet-Vertex Association Algorithm. ATL-COM-PHYS-2008-008, 2008. URL https://cds.cern.ch/record/1082880.

[59] ATLAS Collaboration. Pile-up subtraction and suppression for jets in ATLAS. ATLAS-CONF-2013-083, 2013. URL http://cds.cern.ch/record/1570994.

[60] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B* **659** *(2008) 119*. doi: 10.1016/j.physletb.2007.09.077.

[61] S. Agostinelli *et al.* [GEANT4 Collaboration]. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A* **506** *(2003) 250*. doi: 10.1016/S0168-9002(03)01368-8.

[62] ATLAS Collaboration. Proposal for particle-level object and observable definitions for use in physics measurements at the LHC. ATL-PHYS-PUB-2009-001, 2015. URL https://cds.cern.ch/record/2022743.

[63] ATLAS Collaboration. Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data. ATL-COM-SOFT-2015-119, in preparation.

[64] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C* **74** *(2014) 10, 3071*, . doi: 10.1140/epjc/s10052-014-3071-4.

[65] ATLAS Collaboration. Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS. ATLAS-CONF-2015-037, 2015. URL https://cds.cern.ch/record/2044941.

[66] ATLAS Collaboration. Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV. ATLAS-CONF-2015-002, 2015. URL https://cds.cern.ch/record/2001682.

[67] ATLAS Collaboration. Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV. ATL-COM-PHYS-2010-695, 2010. URL https://cds.cern.ch/record/1287902.

[68] ATLAS Collaboration. Measurement of W^+W^- production cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. ATLAS-CONF-2014-033, 2014. URL http://cdsweb.cern.ch/record/1728248.

[69] ATLAS Collaboration. A Measurement of $W^{\pm}Z$ Production in Proton–Proton Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. ATLAS-CONF-2013-021, 2013. URL http://cdsweb.cern.ch/record/1525557.

[70] ATLAS Collaboration. Measurement of the total ZZ production cross section in proton–proton collisions at $\sqrt{s} = 8$ TeV in 20 fb^{-1} with the ATLAS detector. ATLAS-CONF-2013-020, 2013. URL http://cdsweb.cern.ch/record/1525555.

[71] ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *Eur. Phys. J. C* **74** *(2014) 10, 3109*, . doi: 10.1140/epjc/s10052-014-3109-7.

[72] ATLAS Collaboration. Search for s-channel single top-quark production in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Lett. B* **740** *(2015) 118*, . doi: 10.1016/j.physletb.2014.11.042.

[73] ATLAS Collaboration. Measurement of the cross-section for associated production of a top quark and a W boson at $\sqrt{s} = 8$ TeV with the ATLAS detector. ATLAS-CONF-2013-100, 2013. URL http://cdsweb.cern.ch/record/1600799.

[74] ATLAS Collaboration. Measurement of the Inclusive and Fiducial Cross-Section of Single Top-Quark t-Channel Events in pp Collisions at $\sqrt{s} = 8$ TeV. ATLAS-CONF-2014-007, 2014. URL http://cdsweb.cern.ch/record/1668960.

[75] ATLAS Collaboration. Discrimination of Light Quark and Gluon Jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. ATL-COM-PHYS-2015-350, in preparation. URL https://cds.cern.ch/record/2013509/.

[76] J. Pumplin. How to tell quark jets from gluon jets. *Phys. Rev. D* **44** *(1991) 2025.* doi: 10.1103/PhysRevD.44.2025.

[77] A. J. Larkoski, G. P. Salam and J. Thaler. Energy Correlation Functions for Jet Substructure. *JHEP* **1306** *(2013) 108.* doi: 10.1007/JHEP06(2013)108.

[78] D. Krohn, M. D. Schwartz, T. Lin and W. J. Waalewijn. Jet Charge at the LHC. *Phys. Rev. Lett.* **110** *(2013) 21, 212001.* doi: 10.1103/PhysRevLett.110.212001.

[79] M. R. Whalley, D. Bourilkov and R. C. Group. The Les Houches Accord PDFs (LHAPDF) and Lhaglue. *arXiv:0508110 [hep-ph].*

[80] G. Choudalakis on behalf of the ATLAS Collaboration. Unfolding in ATLAS. *arXiv:1104.2962 [hep-ex]*, 2011.

[81] G. D' Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth. A* **362** *(1995) 487.* doi: 10.1016/0168-9002(95)00274-X.

[82] The RooUnfold package and documentation are available from:. URL http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html.

[83] T. Adye. Unfolding algorithms and tests using RooUnfold. *arXiv:1105.1160 [physics.data-an]*, 2011.

[84] ATLAS Collaboration. Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at sqrt(s) = 8 TeV. ATLAS-CONF-2015-017, 2014. URL https://cds.cern.ch/record/2008678.

Acknowledgements

I am really grateful to all those people who supported and helped me during my PhD research. Special thanks to my supervisor Dr. Alexander Glazov for his guidance during the years of my PhD, answering questions and giving invaluable advices. I would like to thank my committee members, Prof. Dr. Johannes Haller, Prof. Dr. Daniela Pfannkuche, Prof. Dr. Elisabetta Gallo and Dr. Krisztian Peters, for making my defense possible.

I thank my colleagues Dr. Pavel Starovoitov, Dr. Ulla Bluhmenschein, Aliaksei Hrynevich, Dr. Lauren Tompkins, Karol Krizka and others, for productive collaboration, interesting and inspiring discussions, their help and explanations regarding the analysis, which made this thesis possible.

I would like to thank also Dr. Pavel Starovoitov and Dr. Ewelina Lobodzinska, for proofreading and providing useful feedback on this thesis.

And finally I would like to thank my husband, for his incredible support and patience.