Università degli studi di Napoli
"Federico II"

Scuola Politecnica e delle Scienze di Base
Area Didattica di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Fisica

Tesi di Laurea Specialistica
Tesi Sperimentale

Studio delle proprietà degli eventi W+jets con l'esperimento ATLAS ad LHC

Study of properties of W+jets events with the ATLAS experiment at the LHC

Relatori:
Prof. M. Alviggi
Dott. P. Iengo
Dott. A. Tricoli

Candidata:
Noemi Calace
Matr. N94/148

Correlatore:
Prof. F. Ambrosino

Anno Accademico 2012/2013

A mia nonna. Il mio angelo custode.

Abstract

The aim of this work is the study of W boson production in association with jets using $\sqrt{s} = 7$ TeV proton-proton collisions of the 2011 ATLAS dataset with an integrated luminosity of 4.64 fb^{-1}.
This thesis describes the event's selection of W decaying to electrons and muons and the main sources of background in both channels focusing on QCD multijet and $t\bar{t}$. They represent a large fraction of background to both electron and muon channels: QCD is relevant at low jet multiplicity while the background from semileptonic top decays becomes very important at high jet multiplicity. Moreover, because of few statistics in MC samples and large uncertainties on multijet and $t\bar{t}$ cross-sections, these backgrounds are estimated using data-driven methods. The control sample selection and the normalisation of QCD multijet and $t\bar{t}$ templates are discussed including an investigation of the QCD template bias in muon channel and correlation studies between the top templates and data. Then, the evaluation of all the systematic uncertainties derived at detector level is described. The comparison between experimental data and theoretical predictions will be provided for a few distributions, such as transverse momentum and rapidity of the leading jet, angular distance between the two leading jets and inclusive event quantities (e.g. S_T and H_T)
These studies are a critical ingredient towards the measurement of the W+jets cross-section and will contribute to the ATLAS forthcoming publication of the W+jets cross-section measurement.

Contents

List of Figures

List of Tables

Introduction

The LHC began high energy proton-proton collisions in 2010, and since then has accumulated high statistics datasets at centre of mass energies of $\sqrt{s} = 7$ TeV and from 2012 $\sqrt{s} = 8$ TeV. These high statistics samples allow for precision measurements of the Standard Model of particle physics, as well as searches for physics beyond the Standard Model.
Of particular interest is the vector boson production in association with jets. Many "new physics" processes, for example the production of heavy quarks, supersymmetric particles and Higgs bosons, can be mimicked by the production of vector bosons in association with jets. It is therefore important to estimate these backgrounds accurately. In addition, these processes provide a unique test-bench for QCD: the rate of V+multijets production is dependent on the strong coupling constant and the kinematic distributions of the jets probe the theory of the underlying scattering process. This is particularly important in the kinematic region accessible at the LHC in order to understand the physics at or above the electroweak symmetry-breaking scale. Since the individual measurements of kinematic observables in V+jets events are affected by experimental and theoretical systematic uncertainties the measurement of ratios of observables was designed to exploit the cancellation of theoretical and experimental uncertainties, building the foundations of a high precision test of the Standard Model. In the ratio of the production cross-sections of V+jets with different jet multiplicities, the full information on the dynamics of the hadronic final state is retained, while the systematic uncertainties are significantly reduced. The similarity between W and Z production can also be exploited by performing the ratio of W and Z cross-sections for a given jet multiplicity, to achieve a large cancellation of theoretical and experimental uncertainties and provide an accurate test of

QCD. In addition this measurement provides model-independent sensitivity to new physics coupling to leptons and jets.
Within this contest, my thesis describes and analyses W+jets processes in order to understand QCD and electroweak higher order effects.
This thesis is organized as follows. Chapter 1 describes the formalism of the Standard Model of particle physics. Chapter 2 outlines the main features of the LHC and the ATLAS detector which form the experimental setup of this thesis. The reconstruction of physics objects and quantities are reviewed in Chapter 3 while the generation of simulated event samples is the subject of Chapter 4. Then, in Chapter 5 the approach to calculate the cross-section for the production of a vector boson in association with hadronic jets is described. Chapters 6 and 7 describe the main topics of this thesis: the selection of W+jets events with the vector bosons decaying to electron or muon channels and the estimation of QCD multijet and $t\bar{t}$ backgrounds in both channels, the evaluation of all the systematic uncertainties derived at detector level. In Chapter 8 detector level distributions for several observables, *i.e.* transverse momentum and rapidity of the leading jet, angular distance between the two leading jets and inclusive events quantities (e.g. S_T and H_T), have been measured and compared to ALPGEN and SHERPA's theoretical predictions. The last chapter shows the conclusions of my work.
The work and the results presented in this thesis contributed to the measurement of the W+jets cross-section and will be part of the ATLAS forthcoming publication of the W+jets cross-section measurement.

1 Introduction and theoretical background

This chapter introduces the theoretical framework of the Standard Model, the theory that attempts to describe Nature at the most fundamental level.

1.1 The Standard Model

The Standard Model (SM) of elementary particles and their interactions [1, 2, 3] aims at describing Nature at the most fundamental level. Since its formulation in the 1970s, more than three decades of experiments have tested it in meticulous details, accurately confirming its predictions up to the TeV energy scale. Although the SM provides a satisfactory explanation of nearly all known microscopic physical phenomena, some open questions still remain: SM does not provide an explanation for why there are three fermion generations, or why their masses have values spreading over more than twelve orders of magnitude. Moreover, it does not describe the dark-matter content of the Universe and does not include gravity. These and other unanswered questions suggest that a more fundamental theory is likely to exist. The SM is considered by many as an effective theory corresponding to the low-energy approximation of a more fundamental theory.
Several theories that extend the SM have been proposed. They differ in both their particle content and interactions and are designed to reduce to the SM at energies below the TeV scale while offering solutions to its shortcomings and predicting new physical phenomena at higher energies, where the effects of the added degrees of freedom become detectable. These predictions are compared with experimental results to check their compatibility.

1.1.1 The particle content and interactions of the Standard Model

The SM is a quantum field theory [4] that describes the fundamental constituents of matter and the interactions among them. The model is defined by the symmetries of the Lagrangian and by the representations of the particles under these symmetries. The Lagrangian of the SM is based on the gauge symmetry

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1.1}$$

where the $SU(3)_C$ and the $SU(2)_L \times U(1)_Y$ groups describe strong and electroweak interactions, respectively. Strong interactions are the subject of the quantum chromodynamics and electroweak interactions are the subject of the electroweak theory.

The particle content of the SM is as follows (see Tables 1.1 and 1.2 and Figure 1.1): the fermions of spin $1/2$ form the matter particles, whereas the spin 1 bosons are the exchange quanta of the force fields. In the fermion sector, there are six quark (q) flavours, namely up, down, charm, strange, top, and bottom (*u, d, c, s, t, b*) quarks, which make three generations (or families) of doublets. In addition to their electroweak charge each of the quarks appears in three versions with different color charge. No quark q can exist freely, but they form baryons made up of three quarks (qqq) or mesons of one quark and one anti-quark ($q\bar{q}$). The three generations of leptons, namely, electron (e), muon (μ) and tauon (τ), together with their respective neutrinos, have electroweak charges but no color charge. For each particle, an antiparticle with identical properties but opposite charges exists. Matter and antimatter annihilate immediately in the presence of each other and the universe is dominated by matter. Nevertheless, antiparticles can be created in particle collisions.

The force carrier particles are the massless photons (γ), the massive weak interaction gauge bosons $W^{\pm}$ and Z, and the eight massless gluons (g) that mediate the strong force. In addition, the favoured mechanism for breaking the electroweak symmetry necessitates the existence of at least one spin 0 boson, the Higgs boson (H), the possible discovery of which was announced on 4 July 2012 [5, 6] and then confirmed on 6 March 2013. The fourth elementary force known, gravity, has negligible influence at the microscopic scale and is not included in the SM.

		Generation			Charge	Weak Isospin
		First	Second	Third	Q[e]	I_3
Leptons	mass	ν_e < 2 eV	ν_μ < 0.19 MeV	ν_τ < 18.2 MeV	0	$+1/2$
	mass	e^- 511 keV	μ^- 105.66 MeV	τ^- 1.777 GeV	-1	-1/2
Quarks	mass	u 2.3 MeV	c 1.275 GeV	t 173.5 GeV	$+2/3$	$+1/2$
	mass	d 4.8 MeV	s 95 MeV	c 4.18 GeV	-1/3	-1/2

Table 1.1: The fermionic particle content of the SM of Particle Physics with the particles' masses, charges and the weak isospins.

	Electroweak			QCD
	γ	Z^0	$W^\pm$	g
electrical charge[e]	0	0	± 1	0
mass[GeV]	-	91.19	80.43	-

Table 1.2: The force-mediator bosons of the SM of Particle Physics with their masses and charges.

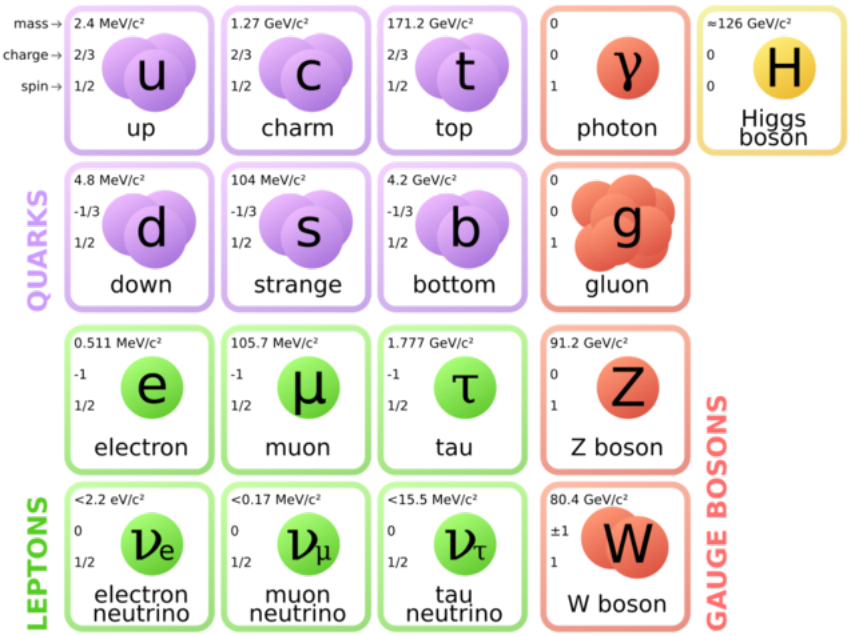


Figure 1.1: Schematic view of the SM particles.

1.1.2 Electroweak theory

The electromagnetic interactions are described in a theory called Quantum Electrodynamics (QED), which can be united with the theory of weak interactions in the electroweak theory (sometimes called "quantum flavordynamics", QFD), based on a $SU(2)_L \times U(1)_Y$ gauge symmetry [9]. It is a chiral theory in the sense that it affects right-handed and left-handed fields differently. All right-handed fermionic fields are electroweak singlets, whereas the left-handed fields are doublets. This forbids mass terms for the fermions, and they are reintroduced into the SM together with $W^{\pm}$ and Z masses by the mechanism of electroweak symmetry breaking. The standard way to introduce the symmetry breaking is via the *Higgs mechanism*.
Historically, the basic structure of electroweak theory was formulated by Sheldon Glashow (1961) [1], but without the Higgs mechanism, and the complete form was found by Steven Weinberg (1967) [2] and Abdus Salam (1968) [3]. The fundamental vertices of electroweak interactions are shown in Figure 1.2. In the weak interactions only the left-handed components couple to the $SU(2)_L$ gauge field. It acts on the doublets

$$\begin{aligned} \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L &= \frac{1}{2}\left(1-\gamma^5\right)\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} = \frac{1}{2}\left(1-\gamma^5\right)\Psi_e, \\ \begin{pmatrix} u \\ d' \end{pmatrix}_L &= \frac{1}{2}\left(1-\gamma^5\right)\begin{pmatrix} u \\ d' \end{pmatrix} = \frac{1}{2}\left(1-\gamma^5\right)\Psi_u, \end{aligned} \tag{1.2}$$

and on the doublets of the remaining two families of leptons and quarks

$$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \qquad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \qquad \begin{pmatrix} c \\ s' \end{pmatrix}_L, \qquad \begin{pmatrix} t \\ b' \end{pmatrix}_L, \tag{1.3}$$

where $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the product of all four Dirac matrices. The d, s and b components have been denoted with a prime, because mixing among weak-eigenstates will occur via the CKM matrix [7, 8].

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.4}$$

This $SU(2)_L$ symmetry is called weak isospin, its generators are denoted τ_1, τ_2 , τ_3, and the corresponding quantum numbers of the lepton doublets are

$$t = \frac{1}{2}, \qquad t_3 = \pm\frac{1}{2}. \tag{1.5}$$

	t	t_3	Y	Q
$\nu_{eL},\nu_{\mu L},\nu_{\tau L}$	$1/2$	$+1/2$	-1	0
$\bar{\nu}_{eL},\bar{\nu}_{\mu L},\bar{\nu}_{\tau L}$	-	-	-	-
$\nu_{eR},\nu_{\mu R},\nu_{\tau R}$	-	-	-	-
$\bar{\nu}_{eR},\bar{\nu}_{\mu R},\bar{\nu}_{\tau R}$	$1/2$	$-1/2$	$+1$	0
e_L^-,μ_L^-,τ_L^-	$1/2$	$-1/2$	-1	-1
e_L^+,μ_L^+,τ_L^+	0	0	$+2$	$+1$
e_R^-,μ_R^-,τ_R^-	0	0	-2	-1
e_R^+,μ_R^+,τ_R^+	$1/2$	$+1/2$	$+1$	$+1$
u_L,c_L,t_L	$1/2$	$+1/2$	$+1/3$	$+2/3$
$\bar{u}_L,\bar{c}_L,\bar{t}_L$	0	0	$-4/3$	$-2/3$
u_R,c_R,t_R	0	0	$+4/3$	$+2/3$
$\bar{u}_R,\bar{c}_R,\bar{t}_R$	$1/2$	$-1/2$	$-1/3$	$-2/3$
d'_L,s'_L,b'_L	$1/2$	$-1/2$	$+1/3$	$-1/3$
$d'_L,\bar{s}'_L,b'_L$	0	0	$+2/3$	$+1/3$
d'_R,s'_R,b'_R	0	0	$-2/3$	$-1/3$
$d'_R,\bar{s}'_R,b'_R$	$1/2$	$+1/2$	$-1/3$	$+1/3$

Table 1.3: Quantum numbers of fermions and anti-fermions. The prime symbol is for the the CKM-mixed state (see §1.1.2) and L and R are respectively for left-handed and right-handed field.

The right-handed fields have $t = 0$ and are singlets, *i.e.* invariant under weak isospin transformation

$$e_R, u_R, d'_R \qquad \mu_R, c_R, s'_R \qquad \tau_R, t_R, b'_R \tag{1.6}$$

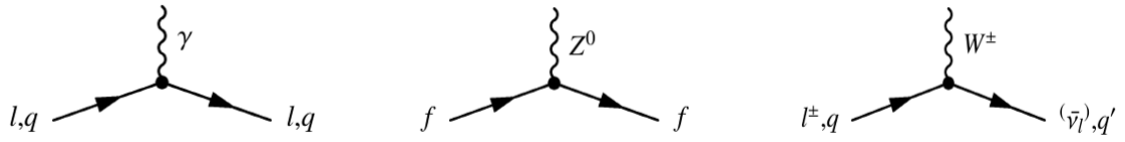
Note that there are no right-handed neutrinos in the SM. This is due to the fact that they are treated as massless particles[1].
The other symmetry group $U(1)_Y$ is associated with the weak hypercharge Y, which is related to the electric charge Q by the *Gell-Mann-Nishijima relation*:

$$Q = t_3 + \frac{Y}{2}. \tag{1.7}$$

Gauge fields and Higgs mechanism

The gauge group $SU(2) \times U(1)$ has gauge bosons W^i_μ, $i = 1, 2, 3$, and B_μ for the $SU(2)$ and $U(1)$ factors, respectively, and the corresponding gauge coupling constants g and g'. These are not the physical gauge bosons of

[1]It is now established that neutrinos have a mass, however small.



(a) QED

(b) Weak neutral current interactions

(c) Weak charged current interactions

Figure 1.2: The fundamental vertices of electroweak interactions. in QED (a) the photon couples to any charged fermion. The weak neutral current interaction (b) is mediated by the Z^0 boson. In this process f stands for any lepton or quark (including neutrinos). (c) shows the weak charged current interaction mediated by the $W^{\pm}$ bosons. Here a (anti-)lepton converts into its corresponding (anti-)neutrino or a (anti-)quark q into its partner (anti-) quark q' of the same generation. The missing charge is carried away by the charged bosons.

electroweak interactions. Indeed, charged current weak interactions are mediated by bosons $W^{\pm}$, which are linear combinations of W_1 and W_2, while electromagnetic interactions and neutral current weak interactions are mediated respectively by the photon and Z^0, which are linear combinations of W_3 and B. A complex scalar Higgs doublet,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.8}$$

is added to the model for mass generation through spontaneous symmetry breaking with potential given by,

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda^2}{2} (\phi^\dagger \phi)^2. \tag{1.9}$$

For μ^2 negative, ϕ develops a vacuum expectation value, $v/\sqrt{2}$, where $v \sim 246.22$ GeV, breaking part of the electroweak gauge symmetry, after which only one neutral Higgs scalar, H, remains in the physical particle spectrum. In non-minimal models there are additional charged and neutral scalar Higgs particles [10].

After the symmetry breaking the Lagrangian for the fermion fields, ψ_i, is

$$\begin{aligned}\mathcal{L}_F = &\sum_i \overline{\psi}_i \left(i\partial - m_i - \frac{g m_i H}{2M_W} \right) \psi_i \\ &- \frac{g}{2\sqrt{2}} \sum_i \overline{\Psi}_i \gamma^\mu (1-\gamma^5) \left(T^+ W_\mu^+ + T^- W_\mu^- \right) \Psi_i \\ &- e \sum_i q_i \overline{\psi}_i \gamma^\mu \psi_i A_\mu \\ &- \frac{g}{2\cos\theta_W} \sum_i \overline{\psi}_i \gamma^\mu (g_V^i - g_A^i \gamma^5) \psi_i Z_\mu .\end{aligned} \tag{1.10}$$

where:

- $\theta_W \equiv \tan^{-1}(g'/g)$ is the weak angle;
- $e = g \sin\theta_W$ is the positron electric charge;
- $A \equiv B\cos\theta_W + W^3 \sin\theta_W$ is the photon field (γ);
- $W^\pm \equiv (W^1 \mp i W^2)/\sqrt{2}$ and $Z \equiv -B\sin\theta_W + W^3 \cos\theta_W$ are the charged and neutral weak boson field, respectively.

The Yukawa coupling of H to ψ_i in the first term in $\mathcal{L}_F$, which is flavor diagonal in the minimal model, is $g m_i / 2M_W$. The boson masses in the electroweak sector are given (at tree level, *i.e.* to lowest order in perturbation theory) by $M_H = \lambda v$, $M_W = \frac{1}{2} g v =$, $M_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\, v$; as expected in the Higgs mechanism the photon remains massless.
The first term in Eq. 1.10 also gives rise to fermion masses. The interaction of the Higgs boson with the matter fields of mass m_i, trough Yukawa coupling, allows to inroduce mass terms for the matter particles. The Higgs mechanism is an elegant way of introducing mass terms in the SM Lagrangian for both force carriers and matter particles without spoiling the gauge invariance and the renormalisability of the theory.
The second term in $\mathcal{L}_F$ represent the charged-current weak interaction, where T^+ and T^- are the weak isospin raising and lowering operators.
The third term in $\mathcal{L}_F$ describes electromagnetic interactions (QED), and the last is the weak neutral-current interaction. The vector and axial-vector couplings are

$$g_V^i \equiv t_{3L}(i) - 2 q_i \sin^2\theta_W, \tag{1.11}$$

$$g_A^i \equiv t_{3L}(i), \tag{1.12}$$

where $t_{3L}(i)$ is the weak isospin of fermion i (+1/2 for u_i and ν_i ; -1/2 for d_i and e_i) and q_i is the charge of ψ_i in units of e.

1.1.3 Quantum chromodynamics

The strong interactions are described by Quantum Chromodynamics (QCD) [12], a $SU(3)_C$ gauge theory. The index C stands for color and mirrors the fact that the quarks and gluons interact with each other via a color charge associated to this gauge symmetry. Each of the quarks is described by a triplet of color states

$$q = (q_1, q_2, q_3)^T \qquad with\ q = u, d, s, c, b, t. \tag{1.13}$$

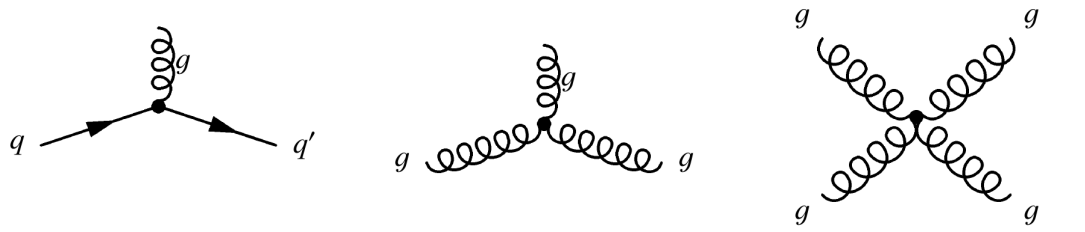
The Lagrangian of QCD is given by

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab})\psi_{q,b} - 1/4 F^A_{\mu\nu} F^{A\mu\nu}, \tag{1.14}$$

where repeated indices are summed over. The $\psi_{q,a}$ are quark-field spinors for a quark of flavor q and mass m_q , with a color-index a that runs from $a = 1$ to $N_c = 3$, *i.e.* quarks come in three "colors". Quarks are believed to be in the fundamental representation of the $SU(3)$ color group. The $\mathcal{A}^C_\mu$ correspond to the gluon fields, with C running from 1 to $N_c^2 - 1 = 8$, *i.e.* there are eight kinds of gluon. Gluons are believed to be in the adjoint representation of the $SU(3)$ color group. The t^C_{ab} correspond to eight 3×3 matrices and are the generators of the $SU(3)$ group. They encode the fact that a gluon's interaction with a quark rotates the quark's color in $SU(3)$ space. The quantity g_s is the QCD gauge coupling constant.
Finally, the field tensor $F^A_{\mu\nu}$ is given by

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu \qquad \left[t^A, t^B\right] = i f_{ABC} t^C, \tag{1.15}$$

where the f_{ABC} are the structure contant of the $SU(3)$ group.
The fundamental vertex of QCD, a gluon coupling to a quark, is shown in Figure 1.3(a). Only particles carrying color (quarks and gluons) participate in the strong interactions. Color is conserved in each interaction. Hadrons are color-singlet (*i.e.* color-neutral) combinations of quarks, anti-quarks, and gluons. A unique feature of non-abelian theories in comparison to the electromagnetic theory is the self-interaction of the exchange bosons. Thus gluons participate in the strong interaction in addition to mediate it because they carry color charge themselves. Other striking features of QCD are *confinement* and *asymptotic freedom*, *i.e.* the impossibility of observing free quarks outside of bound hadron states and the asymptotically vanishing of the coupling for interactions with high momentum transfer (deep inelastic processes).



(a) Quark gluon coupling (b) Three gluon self coupling (c) Four gluon self coupling

Figure 1.3: The fundamental vertices of QCD: $q \to q + g$ (a), three-gluon vertex (b) and four-gluon vertex (c).

Renormalization and running couplings

For a given physical process, the lowest-order Feynman diagrams (the ones with the fewest vertices) are those in which one fermion line couples to a vector boson propagator through the gauge coupling constant corresponding to the occurring interaction ($\alpha \propto g^2/(2\pi)$ for the QED case, $\alpha' \propto g'^2/(2\pi)$ for the weak case and $\alpha_s \propto g_s^2/(2\pi)$ for the QCD case). Using Feynman rules people can calculate physical observables at Leading Order (LO). However, the smaller the probe's wavelength, the deeper the interaction's substructure is probed, *i.e.* more and more much complicated Feynman diagrams must be taken into account. In this way, it is needed to correct the LO calculation introducing higher-order processes. When higher-order contributions are taken into consideration, the gauge coupling constants get a Q^2 scale dependence, *e.g.* $g_c(Q^2)$ for the QCD case. More specifically, when we include higher order contribution in the propagator, ultraviolet divergences start affecting the theory through terms $\propto \ln(Q^2)$ for $Q^2 \to \infty$. The infinities in the coupling constant introduce irremovable divergences in the matrix elements of measurable cross sections. For these reason the renormalization theory has been introduced first to remove the ultraviolet divergences in QED calculations and then was successfully applied to QCD. The renormalization is a technique that re-parameterises divergent physical observables in such a way to absorb divergences in the new definition of charges, masses and wave-functions. Thus the coupling constants are not actually constants but "run", *i.e.* they depend on the energy scale Q at which the interaction takes place. The energy dependence is regulated by the renormalization group equation. Focusing on the QCD framework, we choose and arbitrary renormalization scale μ_R such that $\alpha_{s,R}(\mu_R^2)$ is finite and absorbs the cutoff dependent $\ln(\mu_R^2/Q^2)$ term and the matrix element terms become finite and

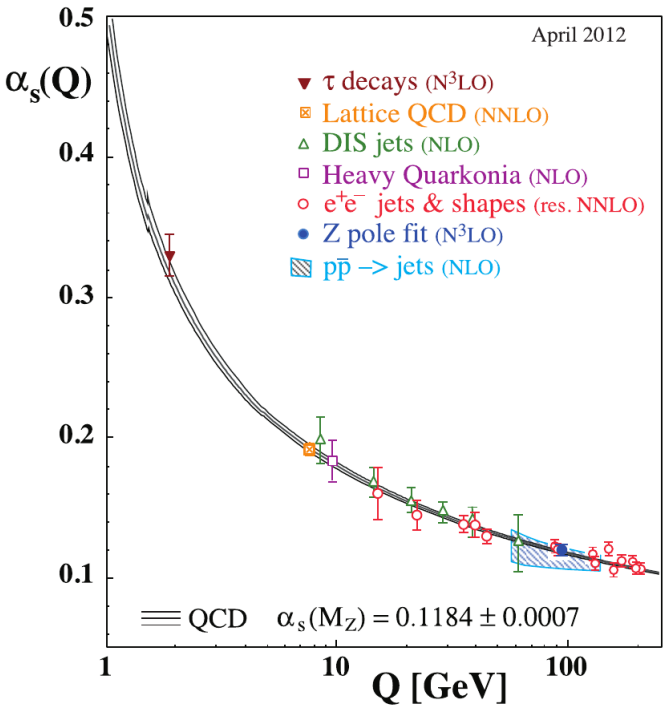


Figure 1.4: Summary of measurements of α_s as a function of the respective energy scale Q. The respective degree of QCD perturbation theory used in the extraction of α_s is indicated in brackets (NLO: next-to-leading order; NNLO: next-to-next-to leading order; res. NNLO: NNLO matched with resummed next-to-leading logs; N3LO: next-to-NNLO) [12].

μ_R-dependent. The renormalised strong coupling constant now becomes

$$\alpha_{s,R}(Q^2/\mu_R^2) = \frac{\alpha_{s,R}(\mu_R^2)}{1 + \alpha_{s,R}(\mu_R^2)\beta_0 \ln \frac{Q^2}{\mu_R^2}}. \tag{1.16}$$

This equation gives the evolution of $\alpha_{s,R}$ from a known scale μ_R^2 to a different scale Q^2, with $\beta_0 > 0$ being the first term in the expansion of the β-function of the renormalisation group equation. The renormalization scale μ_R, which slipped into the calculation as a free parameter of the renormalization procedure, is an arbitrary mass and, as such, our physical observable must be independent of its arbitrarily chosen value: the dependence of the observables, such as the matrix element, on μ_R must be exactly cancelled by the μ_R-dependence of $\alpha_{s,R}$.

The renormalization procedure introduces a Q^2 dependence in the coupling constants of QED, QCD and QFD. The typical feature of the non-abelian

theories, of which QCD is the best example, is the decrease of the coupling constant for increasing values of Q^2 (see Figure 1.4), *i.e.* decreasing length scales. In the asymptotic regime $Q^2 \to \infty$ the strong coupling constant is so small that the quarks can be considered free (asymptotic freedom). On the other hand, for small Q^2 (*i.e.* at long distances), α_s becomes larger and larger and the region of confinement is reached, where perturbative expansions are no longer valid. Since the calculable part of the strong interaction theory, *perturbative QCD* (pQCD), predicts only the running of α_s as a function of Q^2 but not its absolute value, we can use M_Z as the fundamental scale. The experimental determination of $\alpha_s(M_Z)$ fixes the absolute normalisation of $\alpha_s(Q^2)$. The current knowledge of α_s at the Z mass is $\alpha_s(M_Z) = 0.1184 \pm 0.0007$ [12].
Confinement also results in the observation of jets in hadron collisions, which are narrow streams of hadronic particles created in hard parton collisions. Indeed, quarks and gluons coming from hard processes *hadronize*, *i.e.* cluster themselves into hadrons. Hadronization occurs when the coupling become stronger and stronger making possible the creation of new colourless quark-antiquark pairs from the vacuum when trying to separate bound quarks in a hard interaction. The collection of these hadrons will then form a jet of particles moving in the direction highly correlated to that of the initial boosted parton. These jets of hadrons are what are observed in the detectors.

1.2 Physics beyond the Standard Model

Despite the extraordinary agreement between the SM predictions and the experimental data, there are many other questions that remain unanswered in the current theoretical framework, making physicist believe that the SM is only an effective theory valid in a low energy regime, probably up to the TeV scale. First of all, by construction the SM describes only three of the four fundamental forces, making no predictions about gravity. So far no theory of quantum gravity has found yet experimental evidences, but its effect will become important at very high energy scales. All these open questions call for extensions or alternative models to the SM [11].
A second peculiar problem of the SM is referred to as the hierarchy problem. The mass of the Higgs Boson is measured to be around 126 GeV [5, 6] as was also expected from the scale of the electroweak symmetry breaking ($\mathcal{O}$(100 GeV)). The corrections to this scale, however, are several orders of magnitude larger (Planck scale, $\sim 10^{19}$ GeV). In order to reach a physical mass for the Higgs Boson as small as $\sim$ 126 GeV, the bare Higgs mass - the parameter of the Lagrangian - should be set to a large value such that

the resulting physical mass stays small, in agreement with observations. The mass parameter is therefore fine tuned to the 17^{th} decimal place in order to get the observed mass from a 10^{19} GeV correction. This fine tuning is not natural and is considered as a severe lack of predictability of the SM theory [11].

Furthermore, the SM assumes many free parameters, more than 19, whose values are taken from data. It gives no explanation to why the ordinary (baryonic) matter accounts for only 4.9% of the mass-energy content of the observable Universe and why the almost totality of the mass-energy content of the observable Universe consists of dark matter ($\sim 27\%$) and dark energy ($\sim 68\%$).

One approach of addressing the fine tuning problem is to apply an additional symmetry to the SM Lagrangian. In supersymmetric (SUSY) theories a symmetric transformation between a fermion and a boson is postulated. The symmetry is necessarily broken in order to reconcile the absence of boson-fermion partners occurring at the same mass. By allowing the symmetry to be broken, however, the super-partners of the current particles could be higher in mass and can escape detection at present experiments. SUSY models tame the radiative corrections of the Higgs (and other particles) as contributions from super-partners exactly cancel those of each other. For this cancellation to be effective the super-partners cannot be considerably more massive than the original particle and hence SUSY would suggest a rich discovery ground for new particles at the TeV scale. In SUSY there can be up to 105 free parameters in addition to the SM ones. A few self-consistent SUSY frameworks explain the supersymmetry breaking mechanism and introduce further constraints on the number of free parameters, which can be reduced to 5 or 6. In the so called *R-parity* conservation assumption, which forbids super-particles to decay into particles, all SUSY particles decay through several steps into the stable Lightest Supersymmetrical Particle (LSP). The LSP has been suggested as a candidate for explaining dark matter in the universe.

Another possible extension of the SM is motivated from the formalism itself which relies heavily on group theory. The gauge theories of QED, the weak interaction and QCD all are based on Lie groups ($U(1)$, $SU(2)$ and $SU(3)$) which can be part of a larger group (*e.g.* $SU(5)$). Grand unified theories (GUTs) aim to combine all the groups of the SM into a single (larger) group which at a certain scale (the GUT scale) is responsible of all interactions. This is motivated by the fact that, in GUTs as in all new models that postulate new physics at intermediate scales, coupling constants for the interactions run with the scale of interaction and would appear, if extrapolated, to

approximately converge at high $Q^2 \sim 10^{16}$ GeV^2. In SM the couplings do not cross exactly. The structure of GUTs also naturally leads to an explanation for consistency of electrical charge quantization between leptons and quarks. GUTs, although theoretically appealing, suffer from a lack of testable predictions and early attempts at theories led to a prediction of a proton lifetime which was too short when compared to experimental observations.
The Extra-Dimensions (ED) models are considered plausible scenarios, which can be complementary or alternative to SUSY in solving the hierarchy problem and unifying the SM forces and gravitational interaction, explaining the reason of the weakness of the gravitational force with respect to the other three forces.
Additional theories state that the generations of particles found in the SM could be in part explained if they were considered to be composite as opposed to fundamental. The constituents of these particles "preons" would be held together by very strong forces interacting with hypercolor (or metacolor) charges. The higher mass generations of the SM could then be interpreted as excitations of the lowest mass state. There is, however, currently no experimental evidence for compositeness and theoretical problems are encountered in its formalism.
Other theories have been proposed, such as lepto-quark, techni-colour, etc., but experimental data either disagree with their predictions or have put so tight limits on their parameters spaces that they are commonly considered disfavoured.
To make progress with the questions surronding the SM experimental data are required. For this reason a lot of researchers are waiting for new runs at LHC after the first long shutdown (see §2).

[2]Also in supersymmetric scenarios, thanks to the introduction of new physics at intermediate scale, the SM couplings cross. This is an important theoretical property in support SUSY.

2 The ATLAS experiment at the LHC

This chapter provides some generalities about the Large Hadron Collider and a description of the ATLAS detector.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [14] is a proton-proton collider at the Swiss-French border on the outskirts of Geneva, Switzerland. It occupies the underground tunnel located at a medium depth of 100 meters[1] that was built for LEP, CERN's previous big accelerator. The LHC is part of the CERN's complex of accelerators, shown in Figure 2.1. The protons are taken from hydrogen atoms stripped of their electrons, accelerated by a linear accelerator (Linac2 in Figure 2.1) to an energy of 50 MeV and injected in the Proton Synchrotron Booster ring (PSB). The booster accelerates the protons to 1.4 GeV before sending them to the Proton Synchrotron (PS), where the energy is increased to 25 GeV. The Super Proton Synchrotron (SPS) receives the beam of protons from PS and accelerates it to 450 GeV before finally transferring it to the LHC. The protons arrive at the LHC in bunches, direct consequence of the acceleration procedure by means of radio-frequency (RF) cavities. A charged particle traversing a RF cavity can only be accelerated when the oscillating electric field in the cavity has the correct orientation, which happens at a well-defined moment of the RF cycle. The LHC uses eight superconducting RF cavities per beam each delivering 2 MV at 400 MHz and operating at a temperature of 4.5 K (-268.7°C). In order to bend the trajectory of the high energy beams around the 27 km ring of the LHC,

[1] Due to geological and cost considerations the tunnel ring is tilted, its depth varying from 175 to 50 meters.

1232 dipole magnets provide a total magnetic field of 8.3 T. Each dipole is 14.3 m long and weighs around 35 tonnes. About every fourth magnet is a quadrupole magnet, whose task is to focus the beam. Additionally, higher order magnets such as sextupoles, octupoles, decapoles, etc. are used to correct the trajectory and are mostly embedded in the cold mass of the dipole or quadrupole magnets. A total of 9593 magnets is used to steer the protons' trajectories. In order to achieve a magnetic field of this magnitude is necessary to use superconducting technology in the magnet system. The niobium-titanium (NbTi) cables used in the magnets become superconducting below a temperature of 10 K. The dipoles operate at 1.9 K, a temperature lower than the temperature of outer space, obtained by pumping super fluid helium in the world's largest cryogenic system.

Inside the LHC at the nominal operation, bunches of up to 10^{11} protons (p) will collide 40 million times per second to provide 14 TeV proton-proton collisions. The LHC can also collide heavy ions, in particular lead nuclei, at 5.5 TeV per nucleon pair. In 2011, the LHC reached a center-of-mass energy ($\sqrt{s}$) equal to 7 TeV (with 3.5 TeV each beam) and upgraded to $\sqrt{s} = 8$ TeV in 2012. During the First Long Shutdown (LS1) between 2013 and 2015, LHC will be upgraded to work in 2015 at $\sqrt{s} = 13$ TeV. The design center-of-mass energy $\sqrt{s} = 14$ TeV will be eventually reached after the Second Long Shutdown (LS2) in 2018.

The instantaneous luminosity is a measure of how many interactions occur per units of area and time. For a collider as the LHC this quantity depends on the number of particles per colliding bunch, n_1 and n_2, the number of colliding bunches per beam, N_c, the area of the cross-section of the beam, given by its dispersions σ_x and σ_y in the transverse plane and lastly on the frequency f of the collisions:

$$\mathcal{L} = f\frac{N_c n_1 n_2}{4\pi\sigma_x\sigma_y}. \tag{2.1}$$

Integrated over time $\int \mathcal{L}dt$, it is commonly used to express the size of a dataset in inverse picobarn pb^{-1} (10^{36} cm^{-2}) or inverse femtobarn fb^{-1} (10^{39} cm^{-2}). In Figure 2.2(a) the cumulative luminosity delivered by the LHC during 2010, 2011 and 2012 is shown. The cross section σ of a physics process is related to the number of expected occurrences N of this process via

$$N = \sigma \int \mathcal{L}dt. \tag{2.2}$$

The design goal for the LHC for pp collisions is to reach a bunch spacing of 25 ns, which implies a frequency of 40 MHz, and an instantaneous luminosity of 10^{34} cm^{-2} s^{-1}. For heavy ions collisions the design luminosity is 10^{27} cm^{-2}s^{-1}.

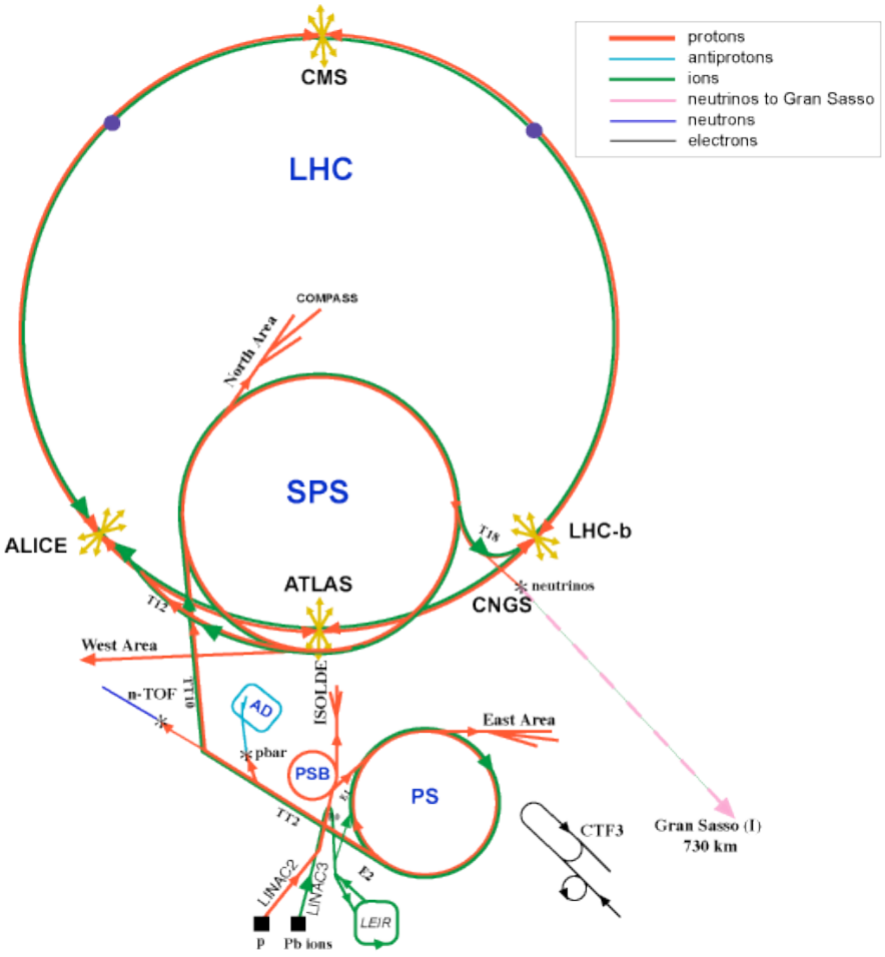
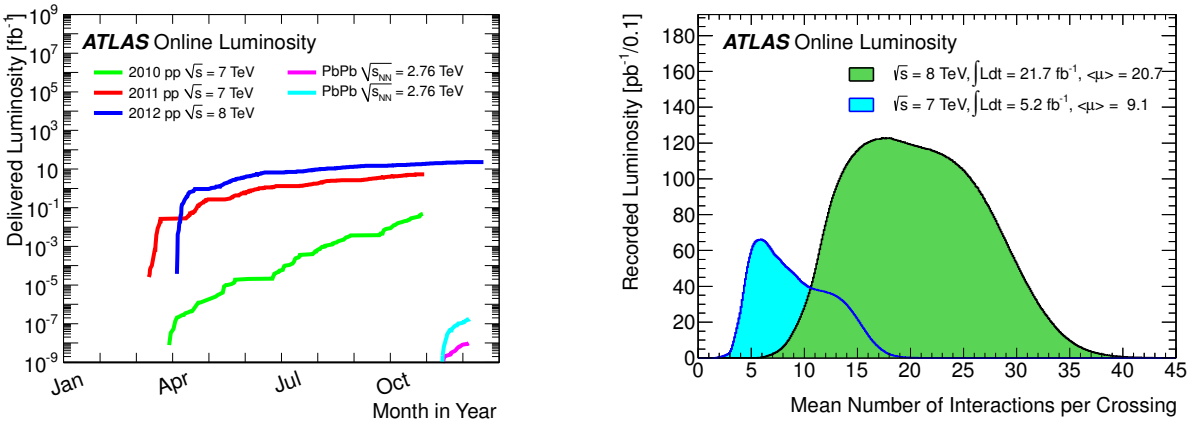


Figure 2.1: The accelerator complex of CERN.



(a) Integrated luminosity as a function of time

(b) Number of interactions per crossing

Figure 2.2: Performance of the LHC. (a) The integrated luminosity delivered by the LHC in 2010, 2011 and 2012. The last months of each running period were devoted to collisions of lead-ions. (b) The luminosity-weighted distribution of the mean number of interactions per bunch crossing for the 2011 (light blue shaded area) and 2012 (dark green shaded area) pp data.

The head-on collision between two bunches circulating in opposite directions is referred to as *bunch crossing*. In each bunch crossing one or more pairs of protons may collide. The probability to observe a certain number of proton interactions follows a Poissonian distribution. Multiple proton collisions from the same bunch crossing are referred to as (in-time[2]) *pile-up* and the mean of the Poissonian is generally denoted by μ. The value of μ at a fixed number of bunches is mainly determined by the number of proton per bunch and their collimation and is proportional to the instantaneous luminosity. In 2011, the mean number of interactions $< \mu >$ per bunch crossing was 9.1. $< \mu >$ increased significantly in 2012, when on average 20.7 interactions occurred within the same recorded event (see Figure 2.2(b)).

At four points in the ring, the two beam lines are crossed, allowing for particle collisions. This is where the four main LHC experiments are located. The two multi-purpose experiments ATLAS [13] and CMS [15] are located across from each other (see Figure 2.1). ALICE [16] and LHCb [17], are both operated with a specialized physics program and are located on both sides of the ATLAS detector close to CERN's main site in Meyrin. ALICE is a dedicated detector to study heavy ion collisions. LHCb is specialized

[2]Out-of-time pile-up refers to the residual effects in the detector due to events occurred in the previous bunch crossings.

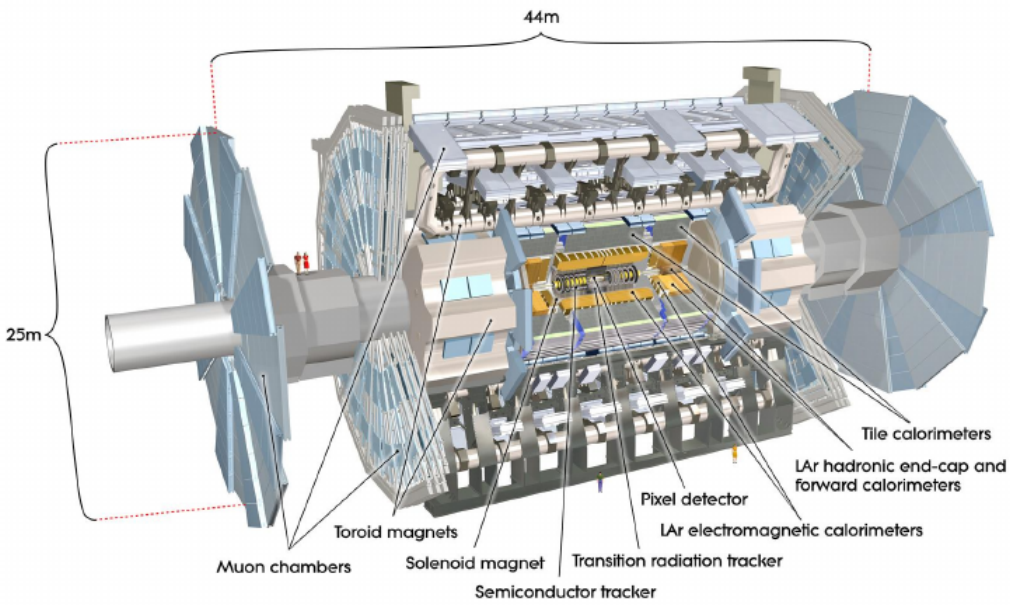


Figure 2.3: A three-dimensional view of the ATLAS detector. Indicated are the individual subdetectors and the dimensions of the detector.

in precision measurements of CP violation and rare decays of b-hadrons as well as the search for indirect evidence of new physics in these processes. In addition, four smaller experiments, TOTEM [18], LHCf [19], ALFA [20] and MoEDAL [21] with very specialized physics programs are installed.

2.2 The ATLAS detector

ATLAS [13] is a multi-purpose detector. The complete detector layout is shown in Figure 2.3 including its dimensions (25 m in height and 44 m in length) and a labeling of the most important detector components. The layout follows an onion-like structure with the inner tracking detectors, providing space points for the reconstruction of particle trajectories, at the center. They are surrounded by the electromagnetic calorimeter, which measures primarily the energy of electrons and photons. Further out, the hadron calorimeter allows for measurements of the energies of hadronic particles. The outermost part of ATLAS consists of the muon spectrometer, whose task is to detect muons. Before going into detail of the ATLAS detector components, the ATLAS coordinate system is discussed [13].

2.2.1 The ATLAS coordinate system

The origin of the coordinate system is chosen to be the nominal interaction point. The positive x-axis is defined as pointing from the interaction point to the centre of the LHC ring and the positive y-axis is defined as pointing upwards. The side-A of the detector is defined as that with positive z (towards the location of the LHCb experiment) and side-C is that with negative z (towards the ALICE experiment). The azimuthal angle ϕ is measured as usual around the beam axis, and the polar angle θ is the angle from the beam axis. Instead of θ, the pseudo-rapidity, defined as $\eta = -\ln\tan(\theta/2)$, is used, since in the massless particle approximation the difference in pseudo-rapidity between a pair of particles is invariant under the boost of the center of mass along the beam direction[3]. In case of massive particles such as jets, rapidity, defined as $y = 1/2 \ln\left[(E + p_z) / (E - p_z)\right]$, is used. In the massless particle approximation rapidity tends to pseudo-rapidity.
Cylindrical coordinates are also frequently used: again with z pointing along the beam line, $R = \sqrt{x^2 + y^2}$, and $\phi = \tan^{-1}\frac{y}{x}$.
The transverse momentum p_T , the transverse energy E_T for a particle, and the missing transverse energy E_T^{miss} in an event are defined in the $x-y$ plane unless stated otherwise. The distance ΔR in the pseudo-rapidity-azimuthal angle space is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.3}$$

The complete ATLAS detector is split into Barrel Region (BR) sections, where detector layers are positioned on cylindrical surfaces around the beam axis, and End-cap Regions (ER) sections, where detector layers are positioned in planes of constant z perpendicular to the beam pipe.

2.2.2 The ATLAS magnet system

The ATLAS magnet system [22], schematically shown in Figure 2.4, is used to bend the trajectory of charged particles within the detector, and thus allows for a momentum measurement to be obtained from the inner detector or the muon spectrometer.
The ATLAS magnet system is comprised of three distinct superconducting magnet systems: the barrel solenoid, the barrel toroid, and the two end-cap toroids.

[3]This property of pseudo-rapidity is very useful at hadron colliders. Since the proton's interacting constituents (gluon or quarks) carry only a variable fraction of the proton momentum, the centre of mass of the colliding constituents is boosted along the beam direction by a variable amount.

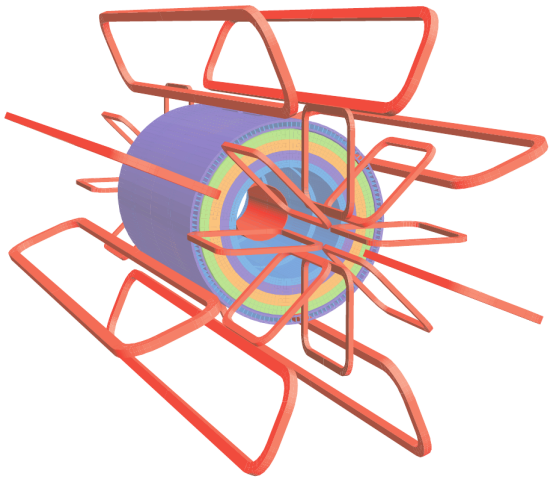


Figure 2.4: Schematic of the ATLAS magnet system. Geometry of magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid winding lies inside the calorimeter volume.

The barrel solenoid [23, 24], between inner detector and electromagnetic calorimeter, provides a 2 T axial magnetic field at its center, and is used to bend charged particles within the acceptance of the inner detector. To achieve the desired calorimeter performance, the layout was carefully optimised to keep the material thickness in front of the calorimeter as low as possible. The magnetic field produced by the barrel solenoid is parallel to the beam pipe, thus causing charged particles to curve radially away from the interaction point. The inner and outer diameters of the solenoid are 2.46 m and 2.56 m and its axial length is 5.8 m.
The cylindrical volume surrounding the calorimeters and both end-cap toroids is filled by the magnetic field of the air-core barrel toroid [25], which consists of 8 rectangular toroid magnets located just outside of the barrel calorimeter system. The overall size of the barrel toroid system as installed is 25.3 m in length, with inner and outer diameters of 9.4 m and 20.1 m, respectively. The magnetic field provided by the barrel toroid system is between 0.15 and 2.5 T in the barrel depending on the position within the system. The resulting field extends in the ϕ direction, encircling the entire detector, and causes a muon's trajectory to be curved in the polar, or η direction.
Finally, each end-cap air-core toroids [26] system follows the same design as the barrel toroid, being comprised of 8 individual rectangular toroid magnets. The strength of the magnetic field varies by position from 0.2 to 3.5 T.

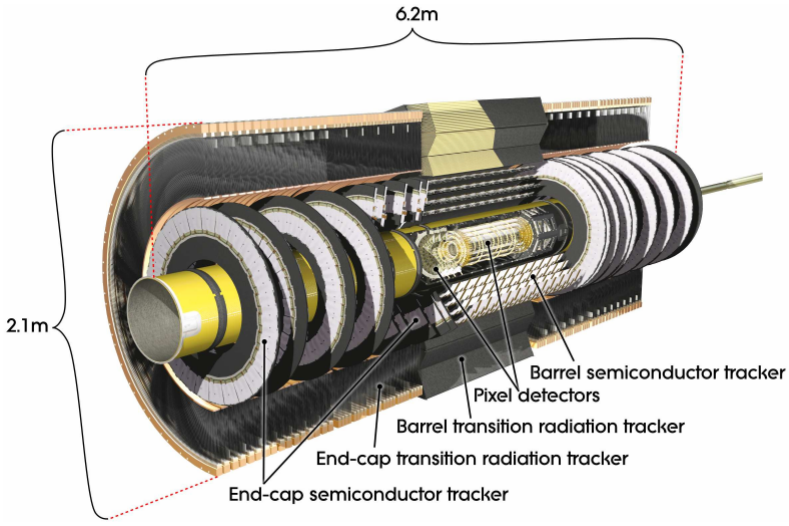


Figure 2.5: Cut-away view of the ATLAS inner detector.

2.2.3 The inner detector

The ATLAS Inner Detector (ID) [27] is designed to provide hermetic and robust pattern recognition, excellent momentum resolution and both primary and secondary vertex measurements for charged tracks above a given p_T threshold (nominally 0.5 GeV) and within the pseudo-rapidity range $|\eta| < 2.5$. It also provides electron identification over $|\eta| < 2.0$ and a wide range of energies (between 0.5 GeV and 150 GeV). The layout of the ID is illustrated in Figure 2.5. It is immersed in a 2 T magnetic field generated by the central solenoid and is composed of three nested sub-detectors: the *Pixel Detector* is the innermost followed by the *Semiconductor Trackers* (SCT) and the *Transition Radiation Tracker* (TRT). The ID extends from a radius of about 50 mm to a radius of 1.082 m, with a length of 6.2 m (see Figure 2.6).

The pixel detector

The pixel detector [28] has the highest spatial granularity and is closest to the interaction point. This provides the best resolution closest to the beam pipe, in order to get a precise location of the vertices. The pixel detector is constructed from 1744 modules about 63 mm long and 19 mm wide. Each module consists of a reverse biased silicon $p-n$ junctions. Each of them contains 328×144 pixel cells of size 50 μm $\times$ 400 μm that are connected to the front-end electronics. When a charged particle passes through the depletion layer in the junction, it produces carriers. The collection of this charge from

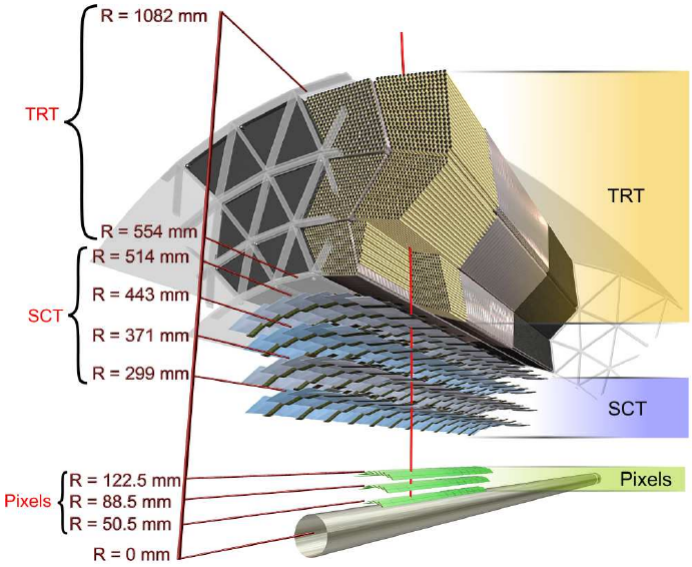


Figure 2.6: Drawing showing the sensors and structural elements traversed by a charged track of 10 GeV p_T in the barrel inner detector ($|\eta| = 0.3$). The track traverses successively the beryllium beam-pipe, the three cylindrical silicon-pixel layers with individual sensor elements of 50×400 μm^2, the four cylindrical double layers (one axial and one with a stereo angle of 40 mrad) of barrel silicon-microstrip sensors (SCT) of pitch 80 μm, and approximately 36 axial straws of 4 mm diameter contained in the barrel transition-radiation tracker modules within their support structure.

a single sensing element forms the basis of a "hit" that is a space point.
In the barrel modules are arranged on three concentric cylinders around the beam axis. In the end-caps they are arranged into six disks perpendicular to the beam axis, three on each side of the interaction point. Each track typically crosses three pixel layers. The single module intrinsic accuracy of the hit position is 10 μm in the R-ϕ plane and 115 μm in the R-z plane. The high pixel granularity also guarantees low occupancy, with a low probability of "double-hits", a crucial requirement especially for high-luminosity runs of the LHC. Due to the large amount of ionizing radiation and neutrons that the pixel detector must withstand at such short distance from the interaction point, the modules have a limited lifetime. Additional layers have been inserted at a smaller radius during the LS1.

The Semiconductor Tracker (SCT)

The semiconductor tracker (SCT) [29] complements the three pixel layers by four concentric barrels of silicon-strip detectors and nine end-cap disks on each side of the detector. The sensitive elements span radial distances from $R = 299$ mm to $R = 560$ mm, covering the same η and ϕ-regions as the pixel detector. Most of the 4088 modules contain four silicon strip sensors. On each side of the module two sensors are daisy-chained together. On the backside another pair of strip-sensors is glued at a stereo angle of 40 mrad, thus providing space points. The resolution reached with the $\sim$6.3 million readout channels is 17 μm in R-ϕ and 580 μm in the R-z plane.

The Transition Radiation Tracker (TRT)

The outermost part of the ATLAS ID is the transition radiation tracker (TRT) [30] consisting of 298304 gas-filled ($Xe/CO_2/O_2$) straw tubes (proportional drift tubes), 4 mm in diameter each. The length of the straw tubes vary from barrel to end-cap, from 144 cm arranged parallel to the beam axis to $\sim$40 cm positioned radially around the beam axis. As the name suggests, the TRT fulfils two purposes. On the one hand, it is a tracking detector that delivers on average 30 space points per charged track originating from the primary vertex. On the other hand, it is equipped with radiator material (foils in the barrel and foam in the end-caps) that causes electrons to emit transition radiation. The emerging low energy photons are absorbed by the gas mixture and induce large signals. The gas mixture is chosen for the good absorption of transition radiation photons in Xe and the high operation stability of mixtures with more than 6% CO_2. The purpose of transition radiation is to identify electrons and in particular to discriminate them from

pions. For this purpose the TRT readout signals are discriminated against two thresholds, one lower threshold (300 eV) intended for tracking and one higher threshold (6-7 keV) for electron identification, thanks to characteristic large energy deposits due to the absorption of transition radiation photons [31]. The TRT does not measure the z coordinate in the barrel but has a resolution of 118 μm in R-ϕ in the barrel and of 132 μm in z-ϕ in the end-caps.

2.2.4 The calorimeters

The primary objective of the calorimeter system is to stop all particles (except muons and neutrinos), and measure their energy. Particles are stopped using a dense absorber material which causes the particles to shower. A shower results when a particle interacts with a material and produces secondary particles which in turn produce more secondary particles, leading to a cascade. Electromagnetic showers involve the production from photon of electron-positron pairs (photon conversion) and electron bremsstrahlung (radiation of photons) as a result of photons' and electrons' interaction with atomic electrons and nuclei. Hadronic showers result from hadrons interacting with nuclei to produce mainly charged and neutral pions, some of which interact with further nuclei, and some of which decay to muons, neutrinos, electrons, and photons, where the electrons and photons shower electromagnetically.
For these reasons the ATLAS calorimetry is made of two types, the Electromagnetic (EM) Calorimeter and the Hadronic Calorimeter (HC), the former inside the latter, mainly based on two technologies: a liquid argon (LAr) and a tile calorimeter (see Figure 2.7), covering the pseudo-rapidity region $|\eta| < 4.9$. Within the η-range of the inner detector, the EM calorimeter is finely granulated to be able to perform precision measurements of electrons and photons. For the rest of the calorimeters a coarser granularity is sufficient to reconstruct jets and determine E_T^{miss}, *i.e.* the energy escaping detection in the transverse plane[4].
Calorimeters must provide good containment for electromagnetic and hadronic showers, and must also limit punch-through[5] into the muon system. Hence, calorimeter depth is an important design consideration. Together with the large η-coverage, this thickness will also ensure a good E_T^{miss} measurement,

[4]For the definition of E_T^{miss} see §2.2.7 and §3.7.

[5]Hadronic showers from very energetic hadrons may be not stopped in the calorimeter. Some of their daughter particles can reach the muon spectrometer. This process is called punch-through.

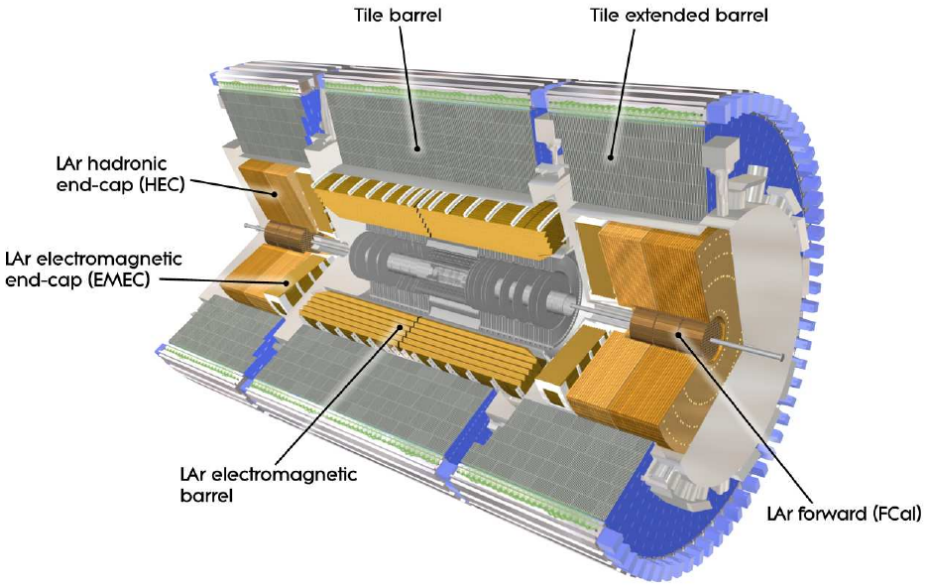


Figure 2.7: Cut-away view of the ATLAS calorimeter system.

which is important for many physics signatures, *e.g.* LSP in R-parity conservation assumption in SUSY searches (see §1.2).

Electromagnetic calorimeter

The ATLAS EM calorimeter [32] have alternating layers of liquid argon, which represents the active material, and lead absorbers arranged in an accordion geometry. Such a geometry provides naturally a full coverage in ϕ without cracks and a fast extraction of the signal. Liquid Argon is chosen for several reasons: it has a high electron mobility, allowing for quick measurements and it is radiation hard. Moreover, Argon is a noble gas which does not capture free electrons, thus minimizing signal losses. The EM calorimeter is divided into a barrel part ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$), each housed in their own cryostat. The position of the central solenoid in front of the EM calorimeter demands optimisation of the material in order to achieve the desired calorimeter performance. As a consequence, the central solenoid and the LAr calorimeter share a common vacuum vessel, thereby eliminating two vacuum walls. A small gap of 4 mm separates the barrel LAr-calorimeter at $z = 0$. The end-caps are divided into two coaxial wheels from $1.375 < |\eta| < 2.5$ and from $2.5 < |\eta| < 3.2$.
Figure 2.8 shows the three sampling layers of the LAr calorimeter at $\eta = 0$.

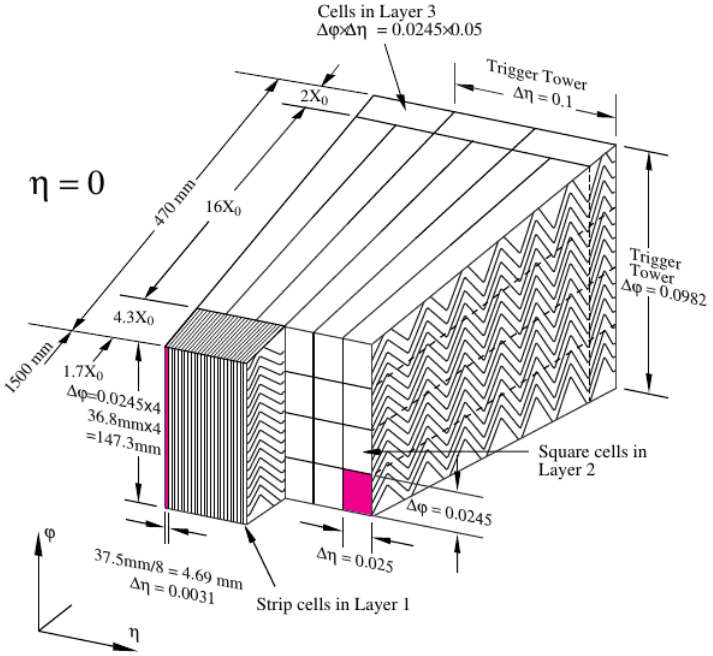


Figure 2.8: The three sampling layers of the ATLAS LAr calorimeter at $\eta = 0$.

The electromagnetic showers start in Layer 1 which has a radiation length[6] of 4.3 X_0. It is shallow in depth but very finely grained in η in order to produce precise electron and photon direction measurements, and to aid in the detection of photon conversions and in the discrimination between pions and photons by distinguishing the double peak structure of $\pi^0 \to \gamma\gamma$ from the single peak energy deposit from single isolated pothons. The second Layer, corresponding to 16 X_0, has the largest material depth and so contains the majority of the shower's energy, and is finely grained in order to aid in the analysis of electromagnetic shower shape. Only a small fraction of the shower is expected to reach Layer 3 which is only installed at $|\eta| < 2.5$ and measures the tails of the elecromagnetic shower. The region $|\eta| < 1.8$ before the first layer additionally includes a thin "presampler" LAr layer which identifies showers that have begun in the 2-6 X_0 because of infrastructure and detector material before reaching the calorimeters, and which therefore aids in the correction for energy loss in pre-calorimeter material. The energy resolution

[6]The radiation length X_0 defines the average distance over which the energy of an electron is reduced by a factor of $1/e$ due to radiation losses. It is found to be $X_0 \approx \frac{716.4A}{Z(Z+1)\ln(287/\sqrt{Z})}$ g/cm^2 where A is the atomic mass number of the absorber material and Z is its atomic number. The absorption length of a photon can be approximated as $X_p = \frac{9}{7}X_0$.

in the barrel was measured using an electron test-beam calibrated to various energies and angles:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \tag{2.4}$$

Hadronic calorimeter

The hadronic calorimeter [33] uses two technologies. The barrel and the extended barrel (see Figure 2.7) are covered by the tile calorimeter based on scintillating active material while in the forward region LAr technology is used.

- **Tile Calorimeter.** The tile calorimeter is placed directly outside the EM calorimeter envelope. Its barrel covers the region $|\eta| < 1.0$, and its two extended barrels the range $0.8 < |\eta| < 1.7$. It is a sampling calorimeter using steel as the absorber and scintillating tiles as the active material. Radially, the tile calorimeter extends from 2.28 m (inner radius) to 4.25 m (outer radius). It is segmented in depth in three layers, approximately 1.5, 4.1 and 1.8 interaction lengths[7], λ_I, thick for the barrel and 1.5, 2.6, and 3.3 λ_I for the extended barrel. The total detector thickness at the outer edge of the tile-instrumented region is 9.7 λ_I at $\eta = 0$.

- **LAr hadronic end-cap calorimeter.** The Hadronic End-cap Calorimeter (HEC) consists of two independent wheels per end-cap, located directly behind the end-cap EM calorimeter and sharing the same LAr cryostats. The wheels are made of copper plates interleaved with 8.5 mm LAr gaps, providing the active medium for this sampling calorimeter. The HEC extends out to $|\eta| = 3.2$ overlapping with the forward calorimeter, in order to reduce drop in material density at the transition between the end-cap and the forward calorimeter (around $|\eta| = 3.1$). Its η range, $1.5 < |\eta| < 3.2$, overlaps also with the tile calorimeter ($|\eta| < 1.7$). Each wheel is divided into two segments in depth, for a total of four layers per end-cap. The outer and the inner diameters of the copper plates in the wheels are respectively 2.03 m and 0.475 m (except in the overlap region with the forward calorimeter where this radius becomes 0.372 m).

- **LAr forward calorimeter.** The Forward Calorimeter (FCal) is integrated into the end-cap cryostats, as this provides clear benefits in

[7]The nuclear interaction length defines the mean length an hadronic particle would travel before undergoing an inelastic collision and it is given very roughly by: $\lambda_I = 35A^{\frac{1}{3}}$ g cm^{-2}, where A are the atomic mass number of the material the particle is passing through.

Calorimeter	Tile	LAr hadronic end-cap	LAr forward
Electrons	-	$\frac{(21.4\pm0.2)\%}{\sqrt{E}}$	$\frac{(28.5\pm1.0)\%}{\sqrt{E}}\oplus(3.5\pm0.1)\%$
Pions	$\frac{(56.4\pm0.4)\%}{\sqrt{E}}\oplus(5.5\pm0.1)\%$	$\frac{(70.6\pm1.5)\%}{\sqrt{E}}\oplus(5.8\pm0.2)\%$	$\frac{(94.2\pm1.6)\%}{\sqrt{E}}\oplus(7.5\pm0.4)\%$

Table 2.1: Fractional energy resolutions σ_E/E of the hadronic calorimeter components as determined from testbeams.

terms of uniformity of the calorimetric coverage as well as reduced radiation background levels in the muon spectrometer. In order to reduce the amount of neutron albedo[8] in the inner detector cavity, the front face of the FCal is recessed by about 1.2 m with respect to the EM calorimeter front face. This severely limits the depth of the calorimeter and therefore calls for a high-density design. The FCal is approximately 10 λ_I deep, and consists of three type of modules in each end-cap: the first, made of copper, is optimised for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions. Each module consists of a metal matrix, with regularly spaced longitudinal channels filled with the electrode structure consisting of concentric rods and tubes parallel to the beam axis. The LAr in the gap between the rod and the tube is the sensitive medium. This geometry allows for excellent control of the gaps, which are as small as 0.25 mm in the first section, in order to avoid problems due to ion build-up.

The fractional energy resolution has been determined from test-beam data separately for each of the components. The results are summarized in Table 2.1.

2.2.5 The muon system

The conceptual layout of the muon spectrometer [34] is shown in Figure 2.9. It is composed of several sub-detectors fulfilling two distinct purposes: the reconstruction of muon tracks and the triggering on events containing muons. The Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC) provide precision space points of muon tracks in the field of the large barrel toroid for $|\eta| < 1.4$ or of the two smaller end-cap toroid magnets at $1.6 < |\eta| <$

[8]Neutrons backscattering by the material in the calorimeter.

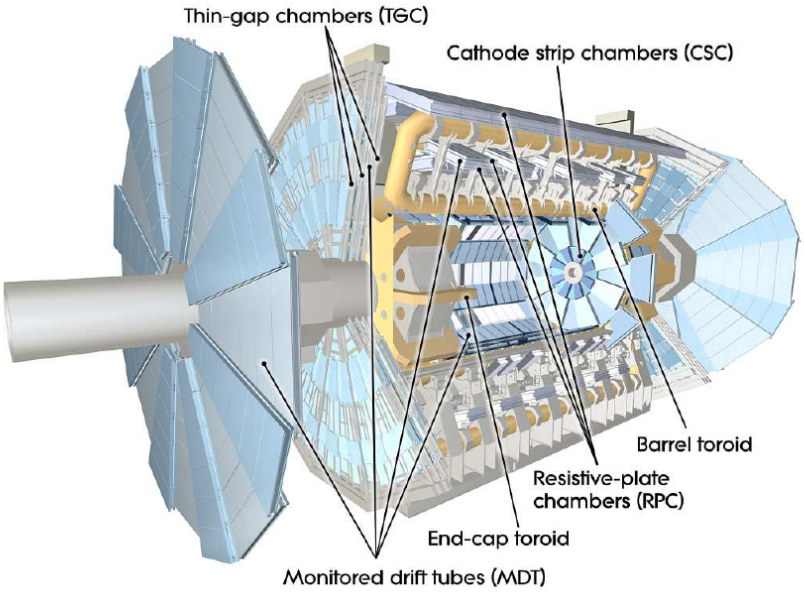


Figure 2.9: Cut-away view of the ATLAS muon system.

2.7. In the transition region $1.4 < |\eta| < 1.6$ a superposition of both magnetic fields deflect the charged particle tracks. In this manner, the magnetic fields are mostly orthogonal to the muon trajectories, thus minimizing the effects of multiple scattering. Moreover, dedicated trigger chambers, the Resistive Plate Chambers (RPC) at $|\eta| < 1.05$ and the Thin Gap Chambers (TGC) at $1.05 < |\eta| < 2.4$ provide bunch-crossing identification, well-defined p_T thresholds and measure both coordinates of the track of the muon candidate orthogonal to the tracking chambers, one in the bending (η) plane and one in the non-bending (ϕ) plane. These trigger chambers deliver tracks within ~ 10 ns after the interaction with the particle, hence allowing for fast muon trigger decisions.

The Monitored Drift Tubes (MDTs)

The MDTs are aluminium tubes with a diameter of 3 cm. A 50 μm diameter anode wire made of tungsten and rhenium run through the center of the tube. They are filled with an Ar/CO_2 (93/7) gas mixture kept at a pressure of 3 bar. A single tube resolution of 80 μm is reached, whereas the combination of tubes in a given chamber sums up to a resolution of 35 μm. The tubes are arranged in multilayer chambers, and the chambers are organized into stations. In the barrel region, there are three concentric cylinders, at radii of 5.0, 7.5, and 10.0 m, respectively, while in the end-cap they are arranged

in three wheels, separated by roughly 7 m. The chamber shapes vary from barrel to end-cap, from rectangular to trapezoidal.

The Cathode Strip Chambers (CSCs)

The CSCs are multi-wire proportional chambers, which are read out via cathode planes segmented orthogonally into strips. Via the induced charge distribution both coordinates can be measured with a resolution of 40 μm in the bending plane and $\sim$5 mm in the transverse direction. They are filled with a gas mixture composed of 80% argon and 20% carbon dioxide. In the chambers the wires are oriented parallel to a central wire pointing in the radial direction. The finer granularity of the CSCs is needed in the forward region, where the track density is higher.

The Resistive Plate Chambers (RPCs)

The RPCs are gaseous detectors that, instead of a wire, are operated with parallel electrode-plates at a distance of 2 mm. The volume is filled with $C_2H_2F_4$/Iso-C_4H_{10}/SF_6 (94.7/5/0.3) gas and is permeated by an electric field of 4.9 kV/mm. The RPCs form the barrel muon trigger system with three concentric cylindrical layers (called trigger stations) consisting of two independent detector layers each. A resolution of 10 mm in both ϕ and z direction is achieved. The timing resolution is 1.5 ns.

The Thin Gap Chambers (TGCs)

The TGCs are multi-wire proportional chambers similar to the CSCs operated in saturation mode to guarantee high timing resolution needed for the tagging of the beam-crossing with $\geq$ 99% efficiency. The TGC system is composed of seven detector layers arranged in two doublets and one triplet unit. The circular disks of detectors are arranged in two concentric rings. The outer (end-cap) ring covers the pseudo-rapidity range of $1.05 \leq |\eta| \leq 1.92$, while the inner (forward) ring covers $1.92 \leq |\eta| \leq 2.4$. Each of the three TGC's planes consists of a wheel divided into eight ϕ-octants, which are in turn divided radially into the forward and end-cap region. Radial information is determined from the anode wires arranged in azimuthal direction, while orthogonal readout strips complement the measurement with ϕ information. The achieved resolution in ϕ ranges from 3-7 mm and in R from 2-6 mm. The timing accuracy is 4 ns.

Muon system		MDT	CSC	RPC	TGC
Point Resolution	ϕ	-	5 mm	10 mm	3-7 mm
	z	35 μm	-	10 mm	-
	R	-	40 μm	-	2-6 mm
	time	-	7 ns	1.5 ns	4 ns

Table 2.2: Summary of the achieved resolutions of the ATLAS Muon System.

2.2.6 Forward detectors

Three smaller detector systems cover the ATLAS forward region. One of the main function of the first two systems is to determine the luminosity delivered to ATLAS. LUCID [20](LUminosity measurement using Cerenkov Integrating Detector), at $\pm$17 m from the interaction point and $|\eta| \simeq 5.8$, detects inelastic pp scattering in the forward direction in order to measure the integrated luminosity and to provide on-line monitoring of the instantaneous luminosity and beam conditions. The second detector is ALFA [20](Absolute Luminosity For ATLAS) located at $\pm$240 m and consisting of scintillating fibre trackers located inside Roman pots which are designed to approach as close as 1 mm to the beam. It measures the elastic pp Coulomb scattering at very small angles. The optical theorem connects the elastic scattering amplitude in the forward direction to the total cross section and can thus be used to extract the absolute luminosity. The third system is the Zero-Degree Calorimeter (ZDC) [35]. It plays a key role in determining the centrality of heavy-ion collisions. It is located at $\pm$140 m from the interaction point, just beyond the point where the common straight-section vacuum-pipe divides back into two independent beam-pipes. The ZDC modules consist of layers of alternating quartz rods and tungsten plates which will measure neutral particles, neutrons and photons, at pseudo-rapidities $|\eta| \geq 8.2$.

2.2.7 The trigger system

The nominal bunch-crossing frequency in the LHC is roughly 40 MHz with bunches of protons colliding in ATLAS every 25 ns (75-50 ns in 2011 and 50 ns in 2012 data-taking periods), and with the luminosity reached in 2012 each bunch-crossing results in an average of 20 pp collisions. It is not possible to record each event that results from each bunch-crossing. It was ultimately decided to limit the events recorded per second to $\sim 10^2$ Hz (increased approximately to 300 Hz in 2011 and up to 600 Hz in 2012).
In order to achieve large event rejection with minimal losses of interesting physics events, the ATLAS trigger system is designed to quickly and roughly

discriminate between events with potential physic interest in order to decide which events to keep and which events to reject. The trigger system is designed to select objects such as low- or high-p_T jet, low- or high-p_T electron, etc. and can select multi-objects like multiple low- or high-p_T jets, 2 medium-p_T leptons, and so on, where low-, medium- and high-p_T correspond to several programmable thresholds provided by the logic. The trigger menu is the set of all the requirements, known as trigger items, that an event has to satisfy in order to be selected by the trigger. Since most trigger items within the trigger menu consistently consume a too large fraction, they must be prescaled: their request to select an event is periodically ignored. The trigger prescales are chosen to not only ensure that the total trigger budget is not exceeded, but also to ensure that sufficient statistics is collected to perform the large range of analyses that are part of the complex ATLAS physics program.
The trigger consists of three levels of event selection: Level-1 (L1), Level-2 (L2), and Event Filter (EF). The L2 and EF together form the High-Level Trigger (HLT). The L1 trigger is hardware-based using custom-made electronics, while the HLT is software-based and uses commercially available computers and networking hardware. A schematic representation of the ATLAS trigger system is shown in Figure 2.10.

Level-1 trigger

The L1 trigger [36] is a hardware-based trigger designed to rapidly reject events or accept them for further analysis. The maximum L1 accept rate was 65 kHz[9] during the machine's first three-year running period (Run 1) and the L1 decision must reach the front-end electronics within 2.5[10] μs after the bunch-crossing with which it is associated. For this reason it uses reduced-granularity information from a subset of detectors: the Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC) for high-p_T muons, and all the calorimeter sub-systems for electromagnetic clusters, jets, τ-leptons, E_T^{miss}, and large total transverse energy. Here, the transverse energy is defined by the sum of all vectored energy depositions $\vec{E}_T$ in the transverse plane. Since the initial transverse momentum of the incoming protons is approximately zero, the total vector sum over the transverse energy of the final states has to balance. Hence, the missing transverse energy is defined as:

$$\vec{E}_T^{miss} = -\sum \vec{E}_T. \tag{2.5}$$

[9]This will be upgraded to 100 kHz after the LS1.
[10]About 1 μs of this time is accounted for by cable-propagation delays alone.

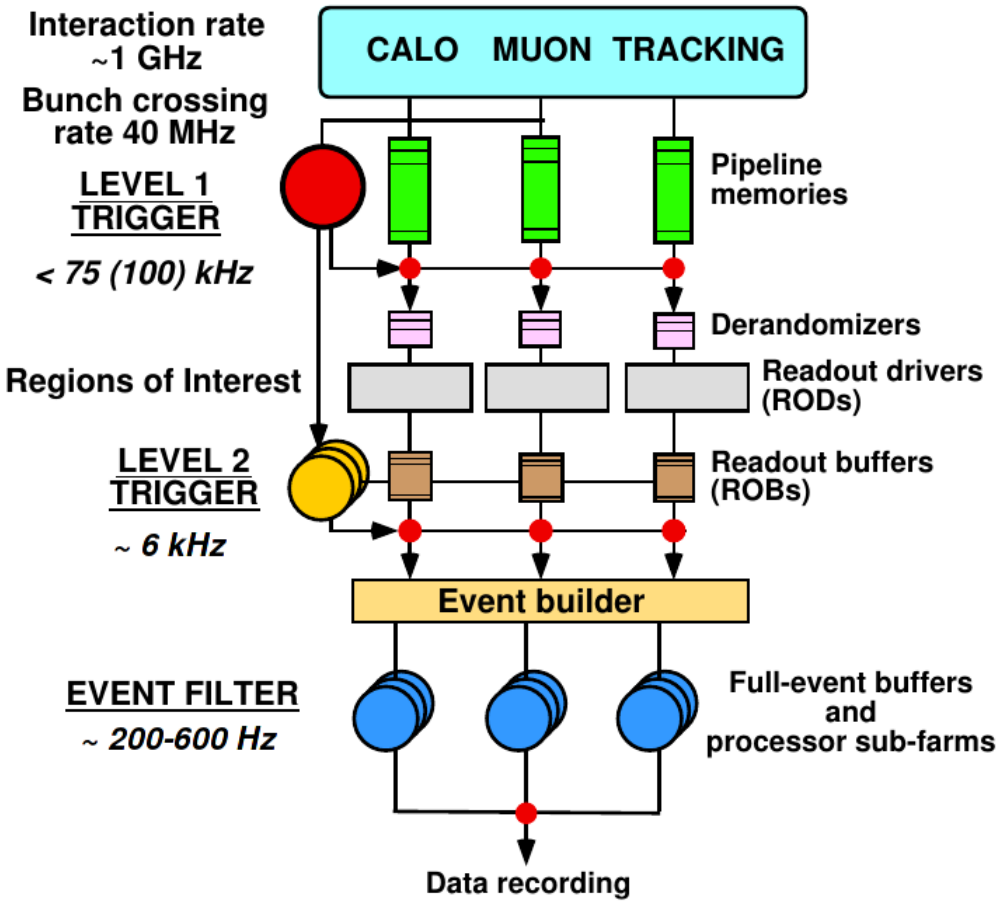
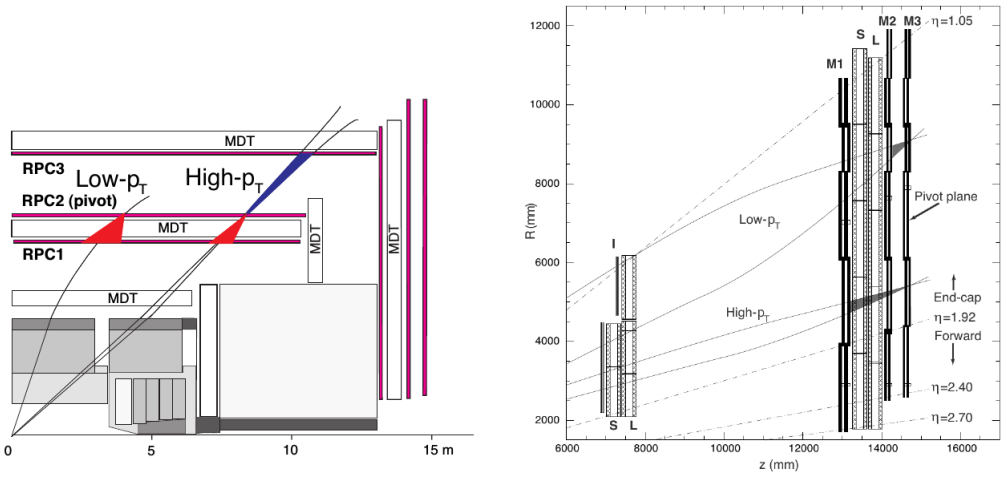


Figure 2.10: Block diagram of the ATLAS three-level trigger and data acquisition system.



(a) Layout of the RPC system

(b) Layout of the TGC system

Figure 2.11: Sketch of the RPC and TGC detector trigger system. Indicated are the different layers RPC1, RPC2, RPC3 or M1, M2, M3 respectively and the according pivot planes.

Low and high momentum muons are identified by reconstructing the track segments in the RPCs and the TGCs using coincidence windows (several windows are opened for several p_T thresholds) for discrimination. A schematic view of the muon trigger system is shown in Figure 2.11: starting from the hit on the central RPC station (RPC2) in the barrel and the last TGC station (M3) in the end-caps, also known as *pivot* planes, a correlation window is opened on the RPC1 and RPC3 layers and on M1 and M2 layers. If a good hit(*i.e.* hits in both η and ϕ and in time with the hit on the pivot plane) is found on the RPC1 layer in the barrel or on the M2 layer in the end-caps then a low-p_T muon candidate is found. The same algorithm is applied using RPC3 and M1 planes to look for high-p_T muon candidates. Finally, the classification of the muon transverse momentum is achieved by using *look-up tables* of track hits containing information of transverse momentum thresholds. Once a muon candidate is found the object is propagated through the other trigger levels.

The L1 Calorimeter Trigger (L1Calo, Figure 2.12) is based on low resolution information from all the ATLAS calorimeters (electromagnetic and hadronic; barrel, end-cap and forward). It is designed to search for high transverse momentum calorimeter object such as electrons, photons, jets and τ-leptons decaying into hadrons, as well as large missing and total transverse energy. The identification of jets/electrons/photons/τ-leptons is accomplished by considering only the sum of calorimeter signals in 0.1×0.1 $\eta \times \phi$ regions, called

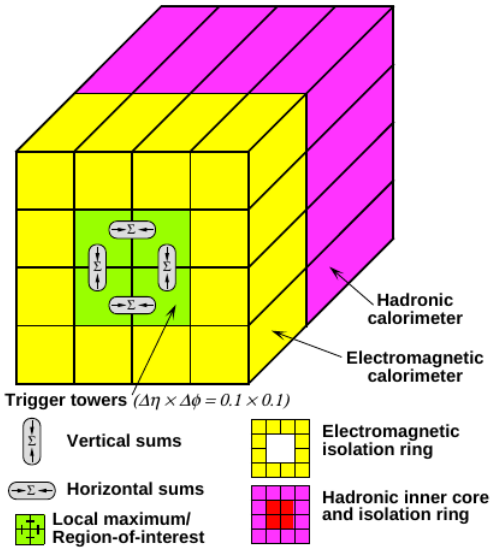


Figure 2.12: Electron/photon and τ trigger algorithms.

trigger towers, that exceed certain thresholds. Once a relevant amount of energy is detected the total transverse energy, $\sum E_T$, in a little 2×2 cluster is measured[11] (green area in Figure 2.12), and the isolation[12] with respect to electromagnetic (yellow area in Figure 2.12) and hadronic activity (pink area in Figure 2.12, *e.g.* due to electrons coming from heavy quark decay) are computed in order to discriminate between electrons/photons and τ-leptons. If these three parameters ($\sum E_T$, electromagnetic and hadronic isolation) fulfil the requirements, then the object is accepted as a good calorimetric object. The L1Calo decision is based on multiplicities and energy thresholds of object identified in the calorimeter.
The overall L1 accept decision is made by the Central Trigger Processor (CTP), which combines the information for different object types. The trigger decision, together with the 40.08 MHz clock and other signals, is distributed to the detector front-end and readout systems via the Timing, Trigger and Control (TTC) system.

[11] A "sliding-window" algorithm (see §3.3) is used to locate the maximum of the energy deposit. It is an efficient tool for precisely reconstructing electromagnetic showers and jets from tau-lepton decays and allows for a very precise cluster energy calibration.

[12] Isolation implies that the energetic particle must have a minimum separation from any significant energy deposit in the same trigger. Isolation-veto thresholds are set for the 12-tower surrounding ring in the EM calorimeter, as well as for the 2×2 hadronic-tower core sum behind the cluster and the 12-tower hadronic ring around it. All these thresholds are programmable.

While the L1 trigger decision is based only on the multiplicity of trigger objects, information about the geometric location of trigger objects is retained in the muon and calorimeter trigger processors. After a positive Level-1 Accept decision, events are read out from the pipeline memories of the detectors and stored in read-out buffers until they are processed by the second level trigger, L2. The information about the geometric location of trigger objects is sent as an $\eta-\phi$ region of the detector in which interesting activity has been detected, called Regions-of-Interest (RoI), to the L2 trigger where it is used to seed the selection performed by the HLT. Thus, L1 trigger system sends to L2 the ROIs' coordinate, together with a rough estimate of the transverse momentum of the candidate object and the event energy sums.

The High-Level Trigger (HLT)

The L2 and Event Filter together form the High-Level Trigger (HLT) [37]. With an increased processing latency of 10 ms due to an input rate reduced by L1, L2 trigger system accesses data from the read-out buffers (ROBs), if necessary with full precision and granularity. This is a software-based trigger that runs on dedicated computing farm. Despite the fact that L2 has more time than L1 to make its decision, the selection algorithms at this stage still have to be kept fast and efficient. It is expected that L2 reduces the rates to $\sim$ kHz (up to 6 kHz in Run 1).
The L2 Calorimeter Trigger basically refines the calorimetric measurement, accessing the full granularity of the calorimeters and studying the shape of the energy deposit (*e.g.* π^0/γ separation), and includes the data of the inner tracking system. At this level a calorimetric object may become an electron if an ID track consistent with it is found. Since the measurements are more precise at this level, tighter conditions on the quality and the kinematic features of the electrons candidates can be required.
At the L2 the muon track is reconstructed for the first time: algorithms run to reconstruct the muon track in the ID and in the MS separately and then combine them in order to determine p_T, η and ϕ. Then the calorimetric activity around the track is measured, in order to apply the isolation requirement.
The event filter (EF) is the final level of on-line selection. It is a software trigger that uses the full event information collected by the read-out buffers. The EF also performs tasks such as vertex reconstruction and final track fitting. The algorithms applied at this stage are similar to those used in the subsequent off-line reconstruction of selected events. The EF reduces the event rate to $\mathcal{O}$(100 Hz) (up to 600 Hz in Run 1), with an average single-event processing time of about four seconds.

Readout drivers and data acquisition system

After an event is accepted by the L1 trigger, the data from the pipe-lines are transferred off the detector to the readout drivers (RODs). Digitised signals are formatted as raw data prior to being transferred to the data-acquisition (DAQ) system. The first stage of the DAQ, the readout system, receives and temporarily stores the data in local buffers. It is subsequently solicited by the L2 trigger for the event data associated to RoIs. Those events selected by the L2 trigger are then transferred to the event-building system and subsequently to the EF for final selection. Events selected by the EF are moved to permanent storage at the CERN Data Centre, the Tier-0.
The Tier-0 facility is responsible for the first-pass processing of the raw data received from the detector, for the archival of raw and derived data on the Tier-0 mass storage system, and for the distribution of the data from the Tier-0 to the 11 Tier-1 centres around the world for further processing and analysis. From Tier-1s data are copied to other permanent storages (Tier-2s, Tier-3s) where they are stored and analysed by physicists of more than one hundred institutions around the world (there are around 140 Tier-2 sites around the world). Tier-0, 1s, 2s, and 3s form the four level of the Worldwide LHC Computing GRID (WLCG).

3 Object reconstruction and identification

This chapter describes the generic procedure used by ATLAS to reconstruct and identify physics object from detector signals. In particular, attention will be given to the algorithms used in the rest of this thesis.

3.1 Introduction

The output of the digitization process of the detector signals generated by the particles produced in real/simulated pp collisions are processed by a series of algorithms that reconstruct and identify physics objects. This step is not done in real time and thus is known as off-line event reconstruction. The result is a set of physics objects with four-momenta that can be used directly in analyses. The reconstruction of tracks, electrons, muons, jets and missing transverse energy will be described in the following sections. Photons and tau leptons' reconstruction will be not discussed since these objects are not used in this thesis.

3.2 Track reconstruction in the inner detector

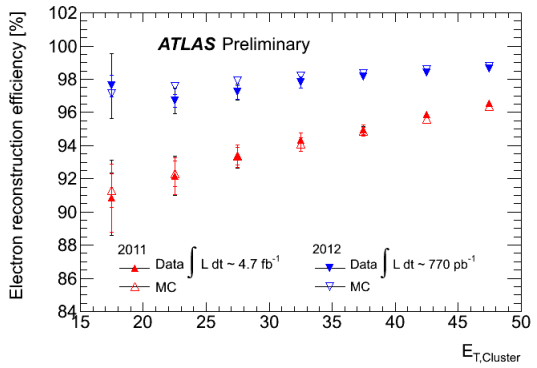
The reconstruction of tracks in the inner detector is divided into three main steps. In the first step the hits in the pixel and SCT are converted into clusters and transformed into space points and the timing information in the TRT is converted into calibrated drift circles. A track-finding stage follows: track seeds are first formed from the space points in the three pixel layers and the first SCT layer. Track candidates are obtained by extending these seeds through the outer SCT layers and fitted excluding outlier clusters. Good tracks are required to have a limited number of holes (*i.e.* no hits

in a given layer) and clusters shared with other tracks. The selected tracks are extended into the TRT and refitted with the full information of all three sub-detectors. The quality of the refitted tracks is compared to the previous track candidates of the silicon detectors: the hits on the extended tracks that result in a bad fit are labelled as outliers. At an advanced stage of event reconstruction dedicated vertex algorithms using exclusively inner detector information are used to reconstruct primary and secondary vertices. The primary vertex is selected requiring the largest transverse momentum sum calculated from the transverse momenta of all the tracks associated to it.

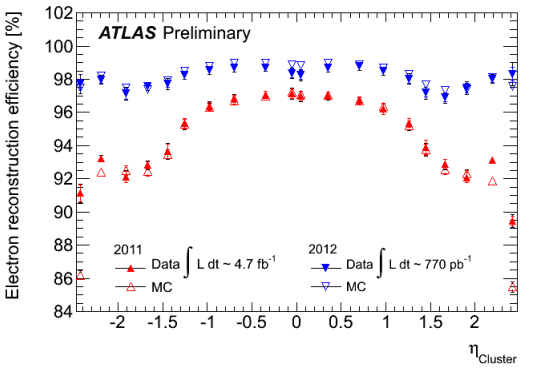
3.3 Electron reconstruction and identification

The electrons leave a clear signature in the ATLAS detector: a track measured by the inner detector and a shower in the electromagnetic calorimeter, providing two independent measurements. These two measurements are the main ingredients to electron reconstruction.
Electron reconstruction begins by defining seed clusters in the second layer of the EM calorimeter using a "sliding-window" algorith [38]. The algorithm begins by dividing the $\eta - \phi$ space of the calorimeters into a grid of $\eta \times \phi$ elements (200×256), and summing all the energy in each element into a single calorimeter "tower". A window of fixed size $N\eta \times N\phi$ (for electrons and converted photons[1] in the barrel the energy is collected over an area of 3×7 cells whereas in the end-caps this area corresponds to 5×5 cells), is moved across the tower grid in discrete steps of $\Delta\eta = 0.025$ and $\Delta\phi = 0.025$. If the sum of the transverse energy (E_T) of the towers within a window is a local maximum and above an energy threshold ($E_T^{thresh} = 2.5$ GeV), a seed cluster is formed. The seed cluster's position in the $\eta - \phi$ plane is determined using the energy weighted barycenter. Once the seed clusters have been formed in the calorimeter, an attempt is made to match these to tracks coming from the inner-detector [39]. All tracks with a transverse momentum above 0.5 GeV are extrapolated to the EM calorimeter and matched to any seed cluster within a region $\Delta R = \sqrt{\Delta\eta \times \Delta\phi} = 0.1$. In the case where several tracks are matched to the same cluster, tracks with the most silicon hits are preferred, and the one with the smallest ΔR is chosen. All seed clusters with a matched track above all the thresholds are considered candidate electrons. Electrons located outside the acceptance of the inner detector, $|\eta| > 2.47$, where no tracking information is available are reconstructed only from energy deposits in the calorimeters by grouping neighbouring cells in three dimensions, based on the significance of their energy content with respect to the expected noise,

[1] see §2.2.4.



(a) Efficiency reconstruction as function of the cluster E_T. Electrons reconstructed in the calorimeter transition region $1.37 < |\eta| < 1.52$ are not taken into account.



(b) Efficiency reconstruction as function of the cluster η for electrons with transverse energy between 30 and 50 GeV.

Figure 3.1: Electron reconstruction efficiency in the data collected in 2011 (red) and 2012 (blue) and in the simulation, as a function of the electron transverse energy (a) and pseudorapidity (b) [40].

and defining topological clusters (see §3.5.2).
For a transverse energy above 20 GeV, the efficiency of the reconstruction algorithm is $\sim 95\%$: only 0.9% of the electrons from $Z \to ee$ decay are not reconstructed and only 2.1% of the electrons are mistakenly reconstructed as photons [41]. The reconstruction of the electrons has been improved in 2012 with respect to previous years by using Gaussian Sum Filter methods, which better reconstruct the tracks of the charged particle, in particular in case of conversion and bremsstrahlung [42]. Figure 3.1 shows the electron reconstruction efficiency in the data collected in 2011 (4.7 fb^{-1}) and 2012 (0.8 fb^{-1}) and in the simulation, as a function of the electron transverse energy and pseudorapidity. Once the object is flagged as an electron, a classification of its reconstruction quality takes place. The reconstruction of electrons is not immune against background processes in which fake-electron objects are reconstructed. Often a high fraction of fake electrons is found among the reconstructed candidates. In order to separate the true electrons from background, a set of identification cuts are applied. The characteristic properties of electrons, such as the amounts of energy deposited in the electromagnetic and the hadronic calorimeters, can be exploited for example:

- Since the electrons deposit most of their energy in the EM calorimeter, this property is a powerful discriminant against jets.
- The lateral and longitudinal shower extensions in the EM calorimeter,

the shower width, is expected to be narrow for electrons due to the small lateral leakage.

- The fine granularity of the first layer of the EM calorimeter can be exploited, because the electron shower is expected to be narrow in pseudo-rapidity with no second maximum, while jets contain several particles that can possibly cause a second maximum.

- The presence of a good quality inner detector track consistent with the EM cluster is also important.

The electron candidates are required to pass a certain number of those identification criteria, which have been optimized in η (defined by calorimeter geometry, detector acceptances and regions of increasing material in the inner detector) and in transverse energy from 5 GeV to 80 GeV.
Three sets of cuts are used in ATLAS: they are called *loose*, *medium* and *tight* [39, 43] selection according to their signal efficiency and jet rejection power they correspond to.

- The loose selection uses shower shape variables of the EM calorimeter middle layer and hadronic leakage variables.

- In the medium selection, variables from the EM calorimeter strip layer, track quality requirements and track-cluster matching are added to the loose.

- The tight selection adds particle identification using the TRT, discrimination against photon conversions via a b-layer hit requirement, information about reconstructed conversion vertices and evaluation of the ratio E/p for the track-cluster matching (E is the cluster energy and p the track momentum).

In order to cope with the trigger rates and the pile up, three further re-optimized selection were redefined from 2011 as the *loose++*, *medium++*, and *tight++* operation points.

- At loose++ level, as for the loose selection, shower shape variables of the second layer of the EM calorimeter and leakage fraction in the hadronic calorimeter are used adding a loose track-cluster matching.

- The medium++ selection combines the previous shower cuts with tighter cluster-track matching in η and with a loose cut on the TRT signal.

- The tight++ selection introduced an asymetric cut on the cluster-track match variable following ϕ and stronger cuts on TRT variables.

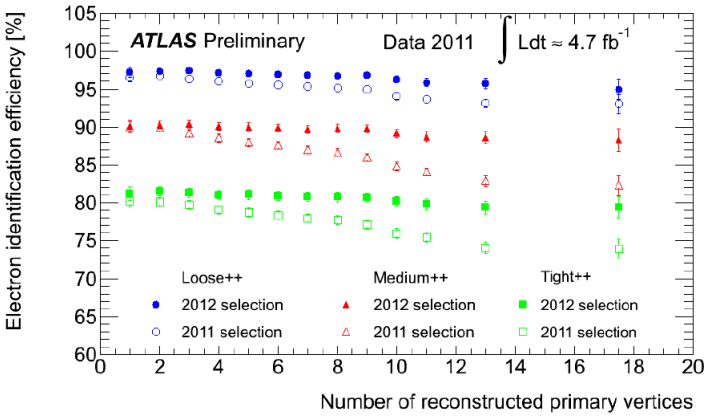


Figure 3.2: Evolution of the identification efficiency for electrons as a function of the number of reconstructed primary vertices for the 2011 and 2012 identification optimizations, using data collected in 2011 [40].

Figure 3.2 shows the evolution of the identification efficiency for electrons as a function of the number of reconstructed primary vertices for the 2011 and 2012 identifications.
In the analysis discussed in the following chapters of this thesis loose, medium and tight++ electron identification requirements as defined for 2011 data-taking (as electron identification changed from 2011 to 2012) are used.

3.3.1 Electron energy scale and resolution

The energy scale and resolution can be determined both from data using methods based on the di-electron mass of $Z \rightarrow e^+e^-$ and $J/\Psi \rightarrow e^+e^-$ decays and from E/p measurement in $W \rightarrow e\nu$ decays [39]. The well-known masses of the Z boson and the J/Ψ resonance can be used to test the linearity of the response of the EM calorimeter. Alternatively, the ratio of the energy measured by the EM calorimeter is compared to the momentum measured by the inner detector. This allows to determine the energy scale taking advantage of the larger statistics of $W \rightarrow e\nu$ decays. Any residual mis-calibration in the i-th bin in η is parametrised by

$$E_i^{meas} = E_i^{true}(1 + \alpha_i) \tag{3.1}$$

where E_i^{true} is the true electron energy, E_i^{meas} is the energy measured by the calorimeter including Monte Carlo (MC) based energy scale corrections and α_i measures the residual mis-calibration in i-th bin. The α_i energy scale

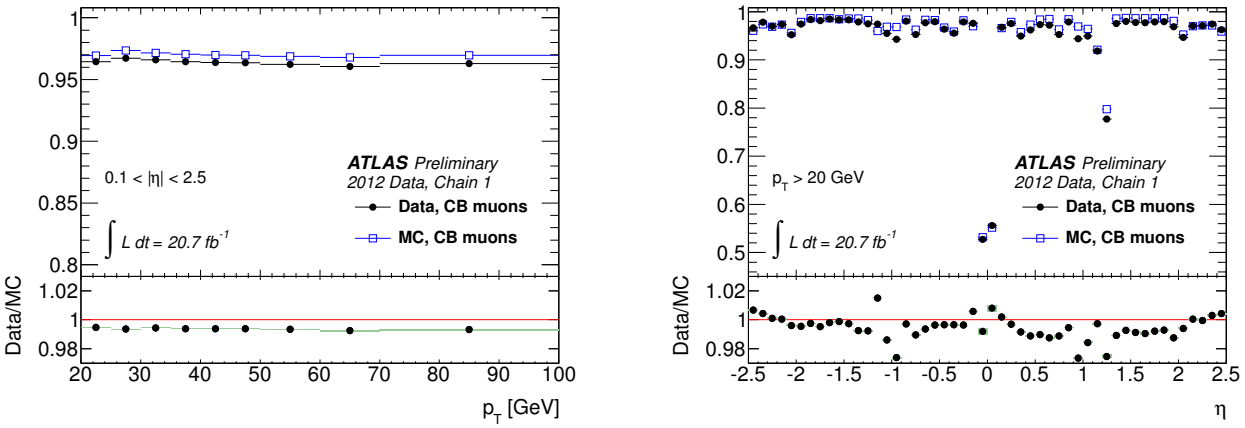
correction factors are determined by a fit to the invariant mass distribution or E/p distribution for $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ and $W \to e\nu$ events. The resulting α_i values are within $\pm 2\%$ and $\pm 5\%$ in the barrel and forward regions, respectively.
The fractional energy resolution in the calorimeter is parametrised as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{3.2}$$

with η-dependent parameters. The parameter a is the sampling term and b is the noise term. Both are evaluated from MC simulation. The constant term c is determined from fits to the simulated and measured invariant mass distributions of $Z \to e^+e^-$ decays using a Breit-Wigner distribution convoluted with a Crystal Ball function. The width of the Breit-Wigner distribution is fixed to the measured Z boson width and the experimental resolution is described by the Crystal Ball function. The resolution predicted by MC simulation is slightly better compared to the measured Gaussian components of the experimental resolution. Therefore, the transverse momentum of electrons is smeared in MC simulation to model the data distributions correctly.

3.4 Muon reconstruction and identification

Muons are the only charged particles that can travel through the entire ATLAS detector. Since muons are charged, they leave a track in the ID. Muons can also pass through the calorimeters almost without any energy loss, and are detected again by the MS in the outermost layer of the ATLAS detector. Since muons are the only known charged particles that the calorimeters do not stop, their identification relies primarily on the MS measurement.
Muon identification is performed according to several reconstruction criteria (leading to different muon "types"), according to the available information from the ID, the MS, and the calorimeter sub-detector systems [44].
Muons used in this analysis are STACO combined muons reconstructed with the *combined reconstruction* (CB) procedure. When a muon is inside the inner detector acceptance ($|\eta| < 2.5$), it will leave a track in both the ID and the MS. An independent track reconstruction is performed for both, and then the two are combined into a single track. The combination of a MS track with the ID information improves the momentum resolution and ensures very low fake rates. The combination of the tracks is done using a statistical combination based on the covariance matrices of the parameters measured in the ID and in the MS.



(a) Efficiency as function of muon p_T.

(b) Efficiency as function of muon η.

Figure 3.3: Reconstruction efficiencies for `STACO` combined muons. The efficiency, obtained from data (dots) and Monte Carlo simulation (open squares) including backgrounds, are shown in the upper part of each figure. The corresponding data/simulation ratios are shown in the lower part.

A χ^2 match is defined by the two vectors[2] of track parameters and weighted by their combined covariance matrix:

$$\chi^2_{match} = (T_{MS} - T_{ID})^T (C_{MS} + C_{ID})(T_{MS} - T_{ID}) \tag{3.3}$$

where T is a vector of five track parameters, C is the corresponding covariance matrix, and the subscripts ID and MS represent inner detector and muon spectrometer respectively. A cut on χ^2_{match} selects good pairs of tracks to be combined and a statistical combination of the two tracks is then performed:

$$T_{combined} = (C^{-1}_{MS} - C^{-1}_{ID})^{-1}(C^{-1}_{MS} T_{MS} + C^{-1}_{ID} T_{ID}). \tag{3.4}$$

The most recent measurement of the efficiency of the ATLAS muon reconstruction was studied on 2012 data at $\sqrt{s} = 8$ TeV using the so-called "tag-and-probe"[3] method on muons coming from $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ decays. Is is found to be well above 92% [45] (see Figure 3.3).

[2]A track is described by a vector $\vec{p}$ with 5 parameters: $\vec{p} = (d_0, z_0, \phi_0, \cot(\theta), Q/p_T)$, where d_0 is the transverse impact parameter, z_0 is the longitudinal impact parameter, ϕ_0 is the angle between the track the x-axis, $\cot(\theta)$ is the inverse slope of the track in the $R - z$ plane, and Q/p_T is the electric charge over the transverse momentum.

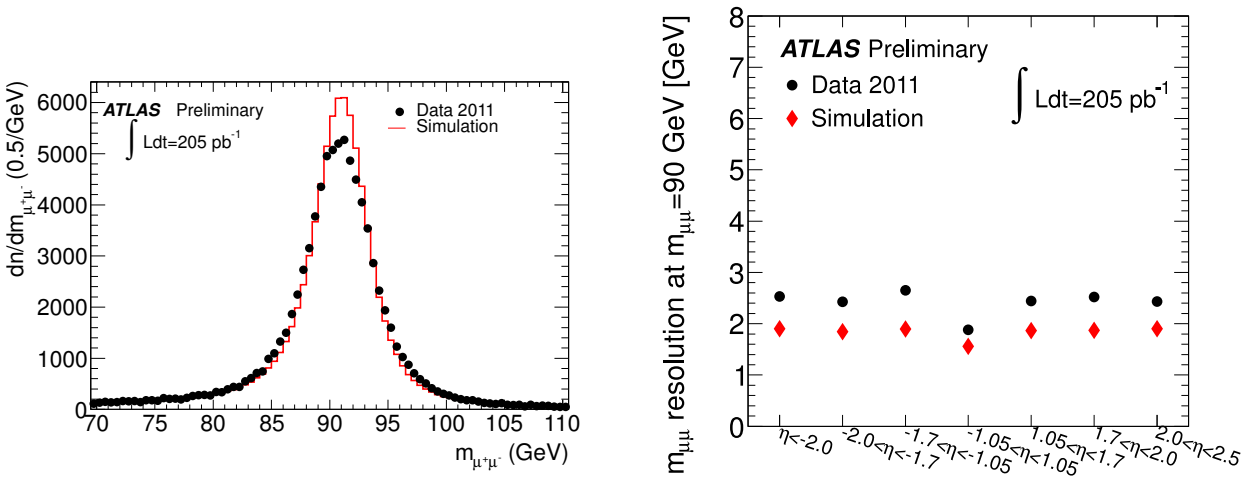
[3]The "tag-and-probe" method is a data driven way to measure the efficiencies from the data, without relying on MC only.

3.4.1 Muon energy scale and resolution

The energy scale and resolution in data is extracted from the width of the di-muon mass distribution in $Z \to \mu^+\mu^-$ decay [46]. The relative resolution of the momentum measurement depends on different effects related to the amount of material that the muon traverses, the spatial resolution of the individual track points and the knowledge of internal alignment of the two subsystems. The ATLAS MS is designed to provide an uniform momentum resolution as a function of the pseudo-rapidity. The di-muon invariant mass distribution obtained for combined muons is shown in Figure 3.4(a) The distribution is fitted using a convolution of the Z boson lineshape and two Gaussian functions modelling the detector resolution effects. The value of the fitted invariant mass at the Z boson pole is sensitive to possible shifts in the momentum scale. Therefore, the accuracy of the momentum scale is probed by measuring the average deviation of the measured invariant mass from the Z boson mass world average value. The resulting resolutions obtained by performing the fit on data and MC simulation in different pseudo-rapidity regions and averaged over all muon p_T values are shown in Figure 3.4(b). Data and MC prediction show a similar behaviour but the performance in simulation is better compared to data. In order to account for this difference and the observed shift in energy scale, the transverse momentum is smeared in the Monte Carlo simulation.

3.5 Jet reconstruction and calibration

Final state partons from the hard scattering processes in a proton-proton collision are colour states which undergo fragmentation and hadronisation processes before reaching the detector. During fragmentation, the initial parton radiates extra partons, which will form the final state hadrons during the hadronisation process. The collection of these hadrons will then form a jet of particles moving in the direction highly correlated to that of the initial parton. These jets of hadrons are what are observed in the detector. Since they carry the information about the initial partons, understanding the properties of jets is extremely important. Jet finding algorithms can be run on various physical objects in the calorimeter such as topological calorimeter clusters and calorimeter towers. In the following sub-sections, various jet finding algorithms and different types of inputs are explained. Then the methods used to calibrate jets are detailed, and finally jet cleaning procedure to recognise poorly reconstructed or non-physical jets is described.



(a) (b)

Figure 3.4: Dimuon invariant mass distribution for oppositely charged muon pairs with transverse momentum above 20 GeV, $|\eta| < 2.5$ and calorimeter isolation (sum of calorimeter cell energies < 2 GeV in a cone of $\Delta R = 0.3$) and (b) dimuon mass resolution of combined muons in different pseudorapidity regions [47]

3.5.1 Jet Algorithms

Jet algorithms are used to combine spatially related energy deposits into a jet. A jet reconstruction algorithm clusters groups of energy deposits into a single jet. This process is governed by a distance parameter R: the distance in $\eta - \phi$ space between two energy depositions. In physical terms, R governs how far a soft parton can be from the primary jet axis, and still be included in the jet. The jet reconstruction algorithms must have the following properties:

- Infrared safety: the emission of soft partons in the final state should not affect the topology and number of jets found by the algorithm.
- Collinear safety: a jet should be reconstructed independent of the fact that a particle carrying a certain fraction of the jet p_T is split into two collinear particles or not.
- Type independence: the algorithm should result in the same jet final states regardless of the type of the input (parton, particle, or detector level inputs such as towers or topological clusters).
- The algorithm should be robust in the presence of pile-up and underlying event in the collision.

The procedure used to combine the jet inputs in ATLAS is based on the four-vector sum of the inputs:

$$(\vec{p}, E)_{jet} = \sum_{i}^{N_{inputs}} (\vec{p}_i, E_i) \tag{3.5}$$

In the following the sequential recombination algorithm, the now-a-days most widespread in ATLAS, will be described.

Sequential Recombination Algorithm

The sequential recombination algorithm is based on the combination of pairs of inputs into a single constituent if they satisfy a minimum distance criterion. Pair merging is repeated till no further such combination is possible. The ATLAS recombination jet finder is known as the "k_t algorithm" [48]. In this algorithm the combination of a pair of inputs ij, depends on the value of the parameter d_{ij}, where:

$$d_{ij} = min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \tag{3.6}$$

where $k_{T,i}$ is the transverse momentum with respect to the incoming beams, and R is a free parameter, controlling the size of the jet. R is 0.4 or 0.6, corresponding to a narrow or wide cone jet. Then, it is possible to define a "beam" distance $d_{iB} = k_{T,i}^{2p}$. One identifies the smallest of all the d_{ij} and d_{iB}, and if it is a d_{ij}, then i and j are merged into a new pseudo-particle (with some prescription, a recombination scheme, for the definition of the merged four-momentum). If the smallest distance is a d_{iB}, then i is removed from the list of particles and called a jet if its transverse momentum is above a given threshold $p_{T,min}$. The parameter p in Eq. 3.6 determines the kind of algorithm: $p = 1$ corresponds to the (*inclusive-*) k_t algorithm, $p = 0$ defines the *Cambridge-Aachen* algorithm, while for $p = -1$ we have the *anti-*k_t algorithm [48]. All these variants are infrared and collinear safe to all orders of perturbation theory. Whereas the former two lead to irregularly shaped jet boundaries, the latter results in cone-like boundaries (Figure 3.5). This is due to the fact that the (*inclusive-*) k_t algorithm depends on the combination of low-p_T constituents which becomes problematic in the high pile up environment. In case of pileup, there are many random constituents not associated with the real jet which could result in additional jets being found. The *anti-*k_t algorithm behaves oppositely to the k_t algorithm in that high p_T inputs are merged first. It is much less affected by pile up, as it starts building the jet from the hard constituents and gives the jets a regular

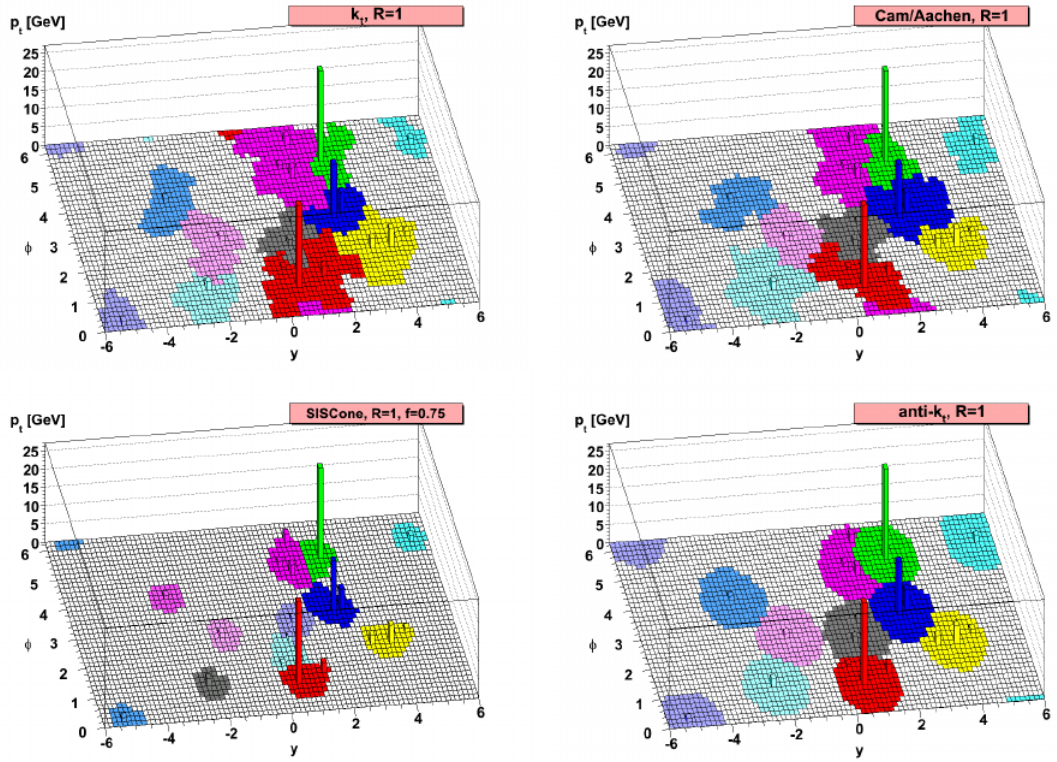


Figure 3.5: Example of clustering particles in the same event to form jets using different algorithms [48]. Note the additional parameter f in the SISCone result, which indicates the fraction of momentum that must be shared by two overlapping cones for them to be merged into a single jet candidate.

area. The *anti-k_t* algorithm is used in my analysis to reconstruct and select jets.

3.5.2 Jets input

The input constituents used to build a jet should reflect the energy deposition of the hadrons forming the jet. In the following, two types of inputs are discussed: calorimeter towers and topological calorimeter clusters.

Calorimeter towers

The calorimeter towers have a bin size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the region $|\eta| < 4.9$ and $-\pi < \phi < \pi$. This gives a total of 6400 towers in the calorimeter. The energy deposition in the towers is assumed to represent

massless particles. The total energy of each tower is the sum of the energies of the calorimeter cell constituents in the tower. These can be negative because of the pedestal subtraction of the noise in the calorimeter electronics. Simply ignoring these towers results in an enhancement of the contribution of positive noise fluctuations. To avoid this problem noise suppression is done: towers with negative energy are combined with nearby positive-energy towers, such that the combined four-momentum has $E > 0$.

Topoclusters

Another type of jet input constituents are the calorimeter topological clusters which constitute the best estimator of the shower of a particle in the calorimeter. Its purpose is to attempt to recreate a three-dimensional energy-deposit representing the shower development of a single particle entering the calorimeter. A cluster is seeded by a cell with a signal to noise[4] ratio, $\Gamma = \frac{|E_{cell}|}{\sigma_{noise,cell}}$, above a threshold of $\Gamma > 4$. All direct neighbours of the seed in all three dimensions are added to the cluster. Neighbours of the neighbours are added only if they have $\Gamma > 2$. Finally, a ring of border of cells with $\Gamma > 0$ are added to the cluster. Once all such clusters are formed, a cluster is split in the case where more than one local energy maximum is found within it. Topoclusters are the preferred jet input constituents in ATLAS as the calorimeter towers do not make best use of the fine granularity of the calorimeter. Moreover, topoclustering intrinsically applies the noise suppression, and includes fewer cells in each jet, resulting in less noise contribution to the jet. However, topoclusters are sensitive to the noise modelling in the calorimeter.

3.5.3 Jet calibration

Jets are reconstructed at the electromagnetic (EM) scale, which is the basic signal scale for the ATLAS calorimeters. It accounts correctly for the energy deposited in the calorimeter by electromagnetic showers. This energy scale is established using test-beam measurements for electrons in the barrel and end-cap calorimeters. The fact that jets also contain hadrons make the calibration of the jet energy necessary. The hadronic jet energy scale is on average restored using corrections and calibration constants derived from the comparison of the reconstructed jet kinematics to the one of the corresponding truth level jet in MC. Jets which fall below the reconstruction threshold of 7 GeV are not calibrated.

[4]The noise includes the expected electronic noise and calorimeter activity due to pile-up events.

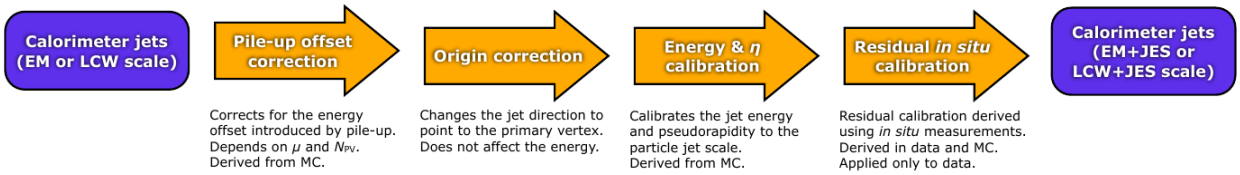


Figure 3.6: Overview of the ATLAS jet calibration scheme used for the 2011 dataset. The pile-up, absolute JES and the residual "in-situ" corrections calibrate the scale of the jet, while the origin and the η corrections affect the direction of the jet.

Figure 3.6 presents an overview of the ATLAS calorimeter jet calibration scheme used for the 2011 dataset, which restores the jet energy scale to that of jets reconstructed from stable simulated particles (truth particle level). This procedure consist of four steps as described below.

- **Pile-up correction.**
  Jets calibrated at the EM scale are affected by energy deposits arising from pile-up (see §2.1). Pile-up can be either in-time or out-of-time, *i.e.* occurring in the same or in different bunch-crossing respectively. The ATLAS calorimeter response is such that the time integral of the signal corresponding to a single particle is zero: in other words a sharp positive peak is followed by a longer but lower amplitude trough. This ensures that the calorimeter is not saturated when large numbers of particles arrive in close succession but also means that, while in-time pile-up provides a positive contribution to the calorimeter signal, out-of-time pile-up will provide negative contribution. A correction to remove the average effects of these additional proton-proton interactions, derived using minimum bias data, is applied at the EM scale: the average additional E_T per calorimeter tower, measured as a function of η and of the number of reconstructed primary vertices N_{PV}, is subtracted from each jet.

- **Jet origin correction.**
  The calorimeter clusters used for jet reconstruction are assumed to originate from the geometrical centre of ATLAS. The jet origin correction first corrects each calorimeter cluster to point back to the primary vertex with the highest p_T^{track} of the event; the beam spot is used if no primary vertex is found. The kinematics of each calorimeter cluster are recalculated using the direction from the primary vertex to the centroid of the cluster. The raw jet four momentum is then redefined as the four

vector sum of the clusters. This correction improves the angular resolution while the jet energy is unaffected. A small improvement in jet p_T resolution is introduced due to the changing jet direction, although this is rarely larger than 1%. Most of the effect of the correction comes from the z-position of the primary vertex.

- **Jet energy scale.**
  The final part of the jet calibration involves applying a jet energy scale (JES) correction to account for calorimeter non-compensation, energy losses in inactive regions, out-of-cone showering effects as well as inefficiencies in the calorimeter clustering and jet reconstruction. The jet energy calibration is derived as a simple correction relating the reconstructed jet energy to the truth jet energy.

- **"In-situ" correction.**
  A residual correction is applied to account for the accuracy of the MC simulation. A combination of Z+jets, γ+jet, and multijet balance techniques are used to compare central jet response in data and MC simulations. The central calibration is transferred to forward jets using η-intercalibration, balancing the p_T of a forward jet with a central jet. The difference in the calibrations derived using data and MC simulations is taken as a residual correction, generally around 2%.

Jets is this analysis are calibrated with the EM+JES scheme [49, 50]. The EM+JES calibration is derived from simulated events, specifically the PYTHIA di-jet sample. It is primarily dependent on energy, since the calorimeter response is energy-dependent, and the jet direction, due to the changing calorimeter technology and to the varying amounts of dead material in front of the calorimeters. To derive the correction factors in MC, isolated particle jets, reconstructed using final-state particles, are matched with isolated detector level jets, reconstructed using the full calorimeter level information. The particle jet energy is then divided by the EM scale energy of the matching calorimeter jet in order to obtain the appropriate correction factor. Following this, a small η-dependent correction is applied to remove a bias in the reconstructed η of jets that occurs when jets fall in poorly instrumented regions of the calorimeter that have a lower response than the regions around it. The reconstructed direction of the jet will be biased since the clusters that fall in these regions have a lower response when their four-vectors are added up to build the jet four-vector, and hence a smaller overall weight. As a consequence of this, the jet is pulled toward the region with the higher response. This η-correction is parameterized as a function of jet energy and

pseudorapidity, and is small, $\Delta\eta < 0.01$, in most regions of the calorimeter, although larger in the crack regions: up to $\Delta\eta = 0.07$ for low p_T jets in the HEC-FCal transition region. The EM+JES calibration restores the jet energy scale within 2% for the full kinematic range. The EM+JES systematic uncertainty is evaluated combining "in-situ", single pion test-beam measurements and variations in the MC simulation.
Other MC based calibrations use hadronic cell calibrations or the topology of the jet constituents to reduce fluctuations in the jet response, therefore improving the jet energy resolution (see §3.5.4) [51].
The Global Cell Weighting (GCW+JES) scheme attempts to compensate for the different calorimeter response to hadrons and electromagnetic particles by weighting each jet constituent cell according to its energy density (electromagnetic showers in the calorimeter leave more compact energy depositions than hadronic showers with the same energy). Based on the properties of clusters (such as their energy density, isolation and depth in the calorimeter) the Local cluster weighting (LCW+JES) scheme calibrates them individually before applying jet reconstruction. To both GCW+JES and LCW+JES schemes, a final jet energy scale correction is applied to achieve response linearity. The Global Sequential (GS) calibration scheme starts from jets calibrated with the EM+JES and uses longitudinal and transverse properties of the jet structure sequentially to improve the resolution leaving the jet energy scale unchanged.
These jet calibration schemes are validated by comparing data and MC using "in-situ" techniques such as direct p_T balance techniques in γ+jet or di-jet events.

3.5.4 Jet Energy Resolution

A precise measurement of jets also requires a precise knowledge of the jet energy resolution (JER) and a good agreement between data and simulation [52]. The performance in resolution for the four different jet calibration strategies in ATLAS, EM+JES, GCW, LCW and GS, have been measured in MC simulation and data using "in-situ" techniques. Results using the bisector and di-jet balance "in-situ" methods[5] are shown in Figure 3.7. The jet resolution agreement between 2011 data and MC is within 10% for $p_T > 30$ GeV. To evaluate the uncertainty on JER three kind of systematic uncertainties are considered:

- experimental uncertainties that affect the "in-situ" measurements;

[5]The two in situ methods are described in [52].

Jet radius	Rapidity range	Total systematic uncertainty		
		Low p_T	Medium p_T	High p_T
$R = 0.4$	$0 \geq \|y\| < 0.8$	17%	15%	11%
	$0.8 \geq \|y\| < 1.2$	20%	18%	14%
	$1.2 \geq \|y\| < 2.1$	20%	18%	14%
	$2.1 \geq \|y\| < 2.8$	20%	18%	18%

Table 3.1: Relative systematic uncertainties on JER at low ($\sim$ 50 GeV), medium ($\sim$ 150 GeV) and high ($\sim$ 400 GeV) p_T, for four rapidity regions. The uncertainty are similar for the four calibration schemes [52].

- the method uncertainties, that is the precision with which the "in-situ" methods in data describe the resolution in MC;
- the MC resolution uncertainty due to event modelling in the MC simulation.

The systematic uncertainties in Table 3.1 for jets with $R = 0.4$ are dominated by the contribution from the closure test. They decrease with p_T and are constant for the highest three rapidity bins. The systematic uncertainties for jets receive comparable contributions from closure and data/MC agreement. They tend to increase with rapidity and are slightly lower in the medium p_T range. The uncertainty increases at high p_T for the end-cap, $2.1 \leq |y| < 2.8$, because of the limited number of events in this region.

3.5.5 Jet data quality and cleaning

Jets in the calorimeters that are not associated with energy depositions of particles from collisions can be either due to calorimeter electronic noise or from non-collision events such as energetic cosmic muons showering in the calorimeter, or beam halo events. In order to remove such objects, the jet cleaning cuts are applied to all the jets with $p_T > 20$ GeV. Jets failing these criteria are flagged as either "bad", likely to be fake, or "ugly", likely to be mismeasured due to falling into less well instrumented regions. Three main issues are addressed, with a dedicated set of selection criteria for each:

- **Single-cell jets in the HEC.** Most misreconstructed jets arise from noise bursts in the HEC. This results in jets with most of their energy coming from single calorimeter cells.
- **Bad quality jets in the EM calorimeter.** Noise bursts in the EM calorimeter, although rarer than in the HEC, result in jets with most

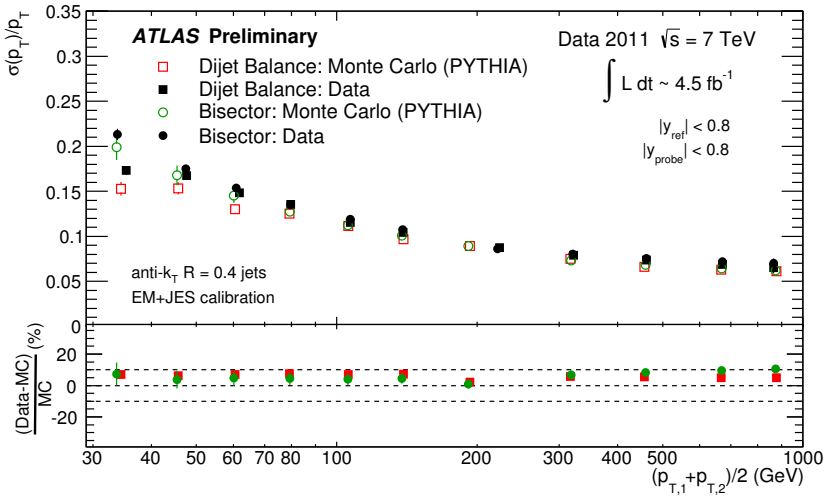
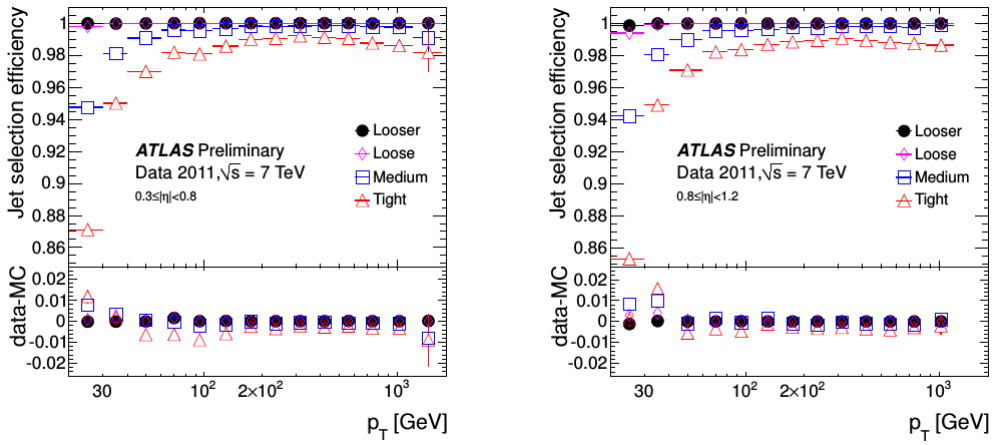


Figure 3.7: Fractional jet energy resolution as a function of the average jet transverse momenta measured with the di-jet balance (squares) and bisector (circles) "in-situ" techniques using the EM+JES calibration in anti-k_T $R = 0.4$ cluster jets. The bottom plot shows the relative difference between data results (black) and Monte Carlo simulation for each method. The dotted lines indicate a relative difference of $\pm 10\%$. Only statistical errors are shown [53].

of their energy coming from the EM calorimeter and whose cells have bad reconstruction "quality".

- **Out-of-time jets.** When large out-of-time energy deposits appear in the calorimeter, possibly from photons produced by cosmic rays, jets will be reconstructed with timing that is incompatible with the event time.

Four levels of bad jet rejection have been determined by the ATLAS Jet/Etmiss Working Group. The most lenient of these is termed "looser", "loose" cleaning, with "medium" and "tight" successively applying stricter criteria in identifying a jets as "bad". The efficiencies of the jet selection are measured using a "tag-and-probe" method on QCD di-jet events. The resulting efficiencies for anti-k_t jets with $R = 0.4$ for all selection criteria as a function of p_T^{jet} in some η ranges are shown in Figure 3.8. At higher p_T, the selection efficiency goes from 100% for looser selection to $\sim 98\%$ for tight selection.

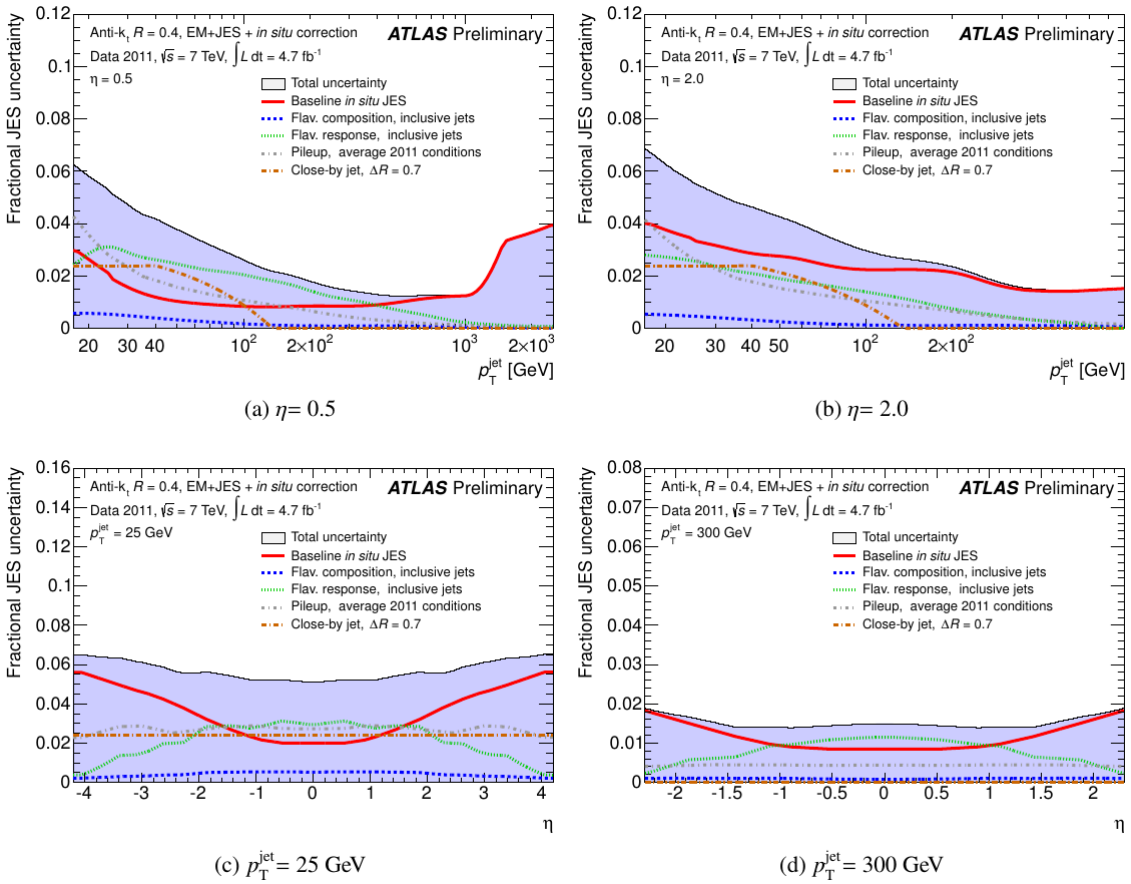


(a) $0.3 \leq |\eta| < 0.8$ (b) $0.8 \leq |\eta| < 1.2$

Figure 3.8: Jet quality selection efficiency for anti-k_t jets with $R = 0.4$measured with a "tag-and-probe" technique as a function of p_T^{jet} in some η ranges, for the four sets of selection criteria. Only statistical uncertainties are shown. Differences between data and MC simulation are also shown.

3.5.6 Jet Energy Scale uncertainty

As discussed in §3.5.3, the jet energy scale (JES) is derived in MC and is validated using a series of "in-situ" measurements. Evaluating the JES uncertainty therefore necessitates combining uncertainties arising from each of these sources: "in-situ" and single pion test-beam measurements, uncertainties on precise details of material distribution in the ATLAS detector, the MC modelling used in event simulation and electronic noise.
Important individual sources of uncertainty include: non-closure when the JES correction is applied to reconstructed MC; the single particle calorimeter response determined from "in-situ" measurements, as described in §3.5.4; the accuracy of detector simulations obtained by varying calorimeter noise in MC samples; the uncertainty associated with physics modelling, which is obtained from comparing the detector response in different MC generators and finally the relative jet calibration obtained through η-intercalibration.
The uncertainty of the jet energy measurement is evaluated for jets with calibrated transverse momenta $p_T^{jet} > 20$ GeV and pseudo-rapidities $|\eta| < 4.5$. For central jets ($|\eta| < 1.2$) the smallest uncertainty of less than 1% is found for jets with $55 \leq p_T^{jet} < 500$ GeV. For jets with $p_T^{jet} = 20$ GeV and with $p_T^{jet} \geq 1$ TeV the JES uncertainty is about 3%. The resulting uncertainty for forward jets is largest for low-p_T jets at $|\eta| = 4.5$ and amounts to 6%.



(a) η= 0.5 (b) η= 2.0

(c) p_T^{jet}= 25 GeV (d) p_T^{jet}= 300 GeV

Figure 3.9: Sample-dependent fractional jet energy scale systematic uncertainty as a function of p_T^{jet} (a,b) and jet pseudo-rapidity (c,d) for anti-k_T jets with distance parameter of $R = 0.4$ calibrated using the EM+JES calibration scheme. The uncertainty shown applies to inclusive QCD jets with average 2011 pile-up conditions, and does not include the uncertainty on the jet energy scale of b-jets.

Other sources of uncertainties due to multiple proton-proton interactions, specific event topologies (*e.g.* close-by jets), flavour composition and flavour effects in the jet response (*e.g.* differences in quark and gluon initiated jets fragmentation) are also taken into account.
In the case of a sample of inclusive QCD assuming average 2011 pile-up conditions, the total JES uncertainty accounting for all effects is below 3.5% for $60 \leq p_T^{jet} < 1000$ GeV using the EM+JES calibration scheme.
The level of total JES uncertainty is shown in Figure 3.9 as a function of jet p_T and jet pseudo-rapidity.

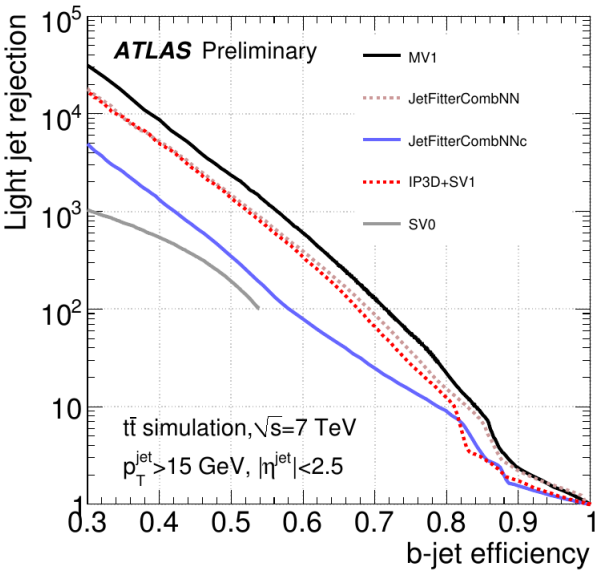


Figure 3.10: The rejection of light-flavour jets as a function of the b-jet efficiency for various b-tagging algorithms in a sample of simulated $t\bar{t}$ events [55].

3.6 b-tagging

Another aspect of jets to be discussed is known as "b-tagging" or identification of jets which stem from the hadronisation of the heavy b quark. The ability to distinguish and identify b-jets from c-jets, light quark, or gluon jets takes advantage of the relatively long lifetime of b hadrons (hadrons containing a b quark), which is of the order of 1.5 ps ($c\tau \sim 450\mu$m). This means that a b hadron within a jet will travel, on average, $\mathcal{O}$(mm) in the transverse plane before decaying. The distance of these new decay vertices from the primary vertex is well within the vertex resolution of the inner detector. These displaced vertices can be reconstructed using both the transverse impact parameter, d_0, defined as the distance of closest approach of the track to the primary vertex point in the $R-\phi$ plane, and the longitudinal impact parameter, z_0 , defined as the z coordinate of the track at the point of closest approach in the $R-\phi$ plane. While the idea behind b-tagging is quite simple and straightforward, extremely complex algorithms have been developed in order to achieve high b-tagging efficiency. The analysis presented in this thesis makes use of the multi-variate b-tagging algorithm known as ATLAS MV1 b-tagging algorithm [54] which is the most commonly used b-tagging algorithm in ATLAs and provides the best rejection of light-flavour jets for a given b-jet efficiency (see Figure 3.10) [55]. It is based on a neural network

that combines the information from three high-performance taggers: IP3D, SV1 and JetFitter. These three tagging algorithms use a likelihood ratio technique in which input variables are compared to smoothed normalized distributions for both the b- and background (light- or in some cases c-jet) hypotheses, obtained from MC simulation. The IP3D tagger takes advantage of the signed transverse and longitudinal impact parameter significances. The SV1 tagger reconstructs an inclusive vertex formed by the decay products of the b-hadron and relies on the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet and the number of two-track vertices. The JetFitter tagger exploits the topology of the primary, b- and c-vertices and combines vertex variables with the flight length significance. A Kalman filter[6] is used to find a common line on which the primary vertex and the b- and c-vertices lie.
To use b-tagging in physics analyses, the efficiency ϵ_b with which a jet originating from a b-quark is tagged by a b-tagging algorithm needs to be measured, as well as the mis-tagging rate, *i.e.* the fraction of taggable jets not labelled as b which are actually tagged as b-jets. For historical reasons the jet rejection is used instead: this is simply the inverse of the mis-tagging rate.
For each b-tagging algorithm a set of operating points is defined, based on the inclusive b-tag efficiency in a simulated sample of $t\bar{t}$ events. Each efficiency value corresponds to a different weight of the b-tagged jet. The ATLAS working points for the MV1 b-tagger are $\epsilon_b = 60\%$, 70%, 75%, and 80%. Figure 3.11 compares the b-tagging efficiencies obtained using data and MC for the MV1 algorithm at 70% b-tagging efficiency working point in the single lepton $t\bar{t}$ analyses [55]. The measured efficiency is greater than 70% for $p_T^{jet} > 100$ GeV.

3.7 Reconstruction of missing transverse energy

As stated in section 2.2.1, most kinematic variables at hadron colliders are only considered in the transverse plane, because the net momentum in the transverse plane is approximately zero, and it is therefore simple to enforce momentum conservation. For instance, if a W boson decays within the detector to an electron and a neutrino, the electron will be detected in the calorimeter and in the inner detector, and a transverse momentum and transverse energy can be associated to it. The neutrino however, will leave the

[6]The Kalman filter is an algorithm used to estimate the state of a dynamical system using a series of measurements over time depending on the state of the system and containing noise (random variations) and other inaccuracies.

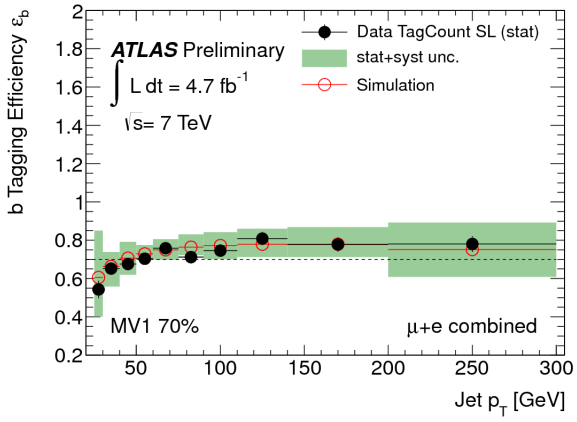


Figure 3.11: The measured b-tagging efficiency in data compared to that in simulation for the MV1 algorithm at 70% b-tagging efficiency working point in the single lepton $t\bar{t}$ analyses [55].

detector without interacting at all, but will still make its presence known through the imbalance of transverse energy within the detector, known as missing transverse energy (E_T^{miss}) [56]. E_T^{miss} is the final quantity to be considered, and in some sense the most difficult, because it relies on all the other objects' reconstructed information. The E_T^{miss} algorithm begins by making a list of all the topoclusters (see §3.5.2) reconstructed in the calorimeter. A preliminary E_T^{miss} calculation is performed by simply taking the negative sum of the topocluster's 4-vectors in the $x-y$ plane and adding the x and y components in quadrature:

$$E_{x(y)}^{miss,topo} = -\sum_{i=0}^{N_{topo}} E_{i,x(y)}^{topo} \tag{3.7}$$

$$E_T^{miss,topo} = \sqrt{\left(E_x^{miss,topo}\right)^2 + \left(E_y^{miss,topo}\right)^2} \tag{3.8}$$

It then uses all other reconstructed objects to perform a cell level overlap removal. A list of cells in all reconstructed electrons, jets, and muons is compared with a list of cells in the topoclusters. All cells already associated to a reconstructed object are removed from the topocluster sum, and replaced by the negative 4-vector of the reconstructed object. This procedure is done first with electrons, then jets, and finally with muons to ensure that any cell potentially belonging to two or more reconstructed objects is only counted once. Finally, any cells belonging to a topocluster, but not associated to a reconstructed object are summed up and collectively referred to as the

"cellout" contribution to the E_T^{miss} [13]. The E_T^{miss} calculation is shown in the following equation:

$$E_{x(y)}^{miss} = -\sum_{electrons}^{N_{electrons}} E_{i,x(y)}^{electron} - \sum_{jets}^{N_{jets}} E_{i,x(y)}^{jet} - \sum_{muons}^{N_{muons}} E_{i,x(y)}^{muon} - \sum_{cells}^{N_{cells}} E_{i,x(y)}^{cell} \quad (3.9)$$

$$E_T^{miss} = \sqrt{\left(E_x^{miss}\right)^2 + \left(E_y^{miss}\right)^2} \quad (3.10)$$

Important requirement is the minimization of the impact of limited detector coverage, finite detector resolution, the presence of dead regions and different sources of noise that can produce fake E_T^{miss}.
Figure 3.12 shows the resolution of the missing transverse energy measured in $Z \to ee$ and $Z \to \mu\mu$ events, for which no missing energy is expected. The quantity plotted is the resolution along an axis which is parallel and perpendicular to the transverse component of the Z boson momentum and it is parametrized in terms of the transverse calorimeter activity defined as:

$$\sum E_T = \sum_{i=1}^{N_{cell}} E_i \sin\theta_i \quad (3.11)$$

where E_i and θ_i are the energy and the polar angle of the cells associated to energy clusters.
The optimal reconstruction of E_T^{miss} in the ATLAS detector is complex and many studies have been made on data to achieve the best performance [56]. The overall systematic uncertainty on E_T^{miss} scale, calculated by combining the uncertainties on the various terms entering the full E_T^{miss} calculation, is estimated to be, on average, 2.6% in events with a W decaying to a lepton and a neutrino. This uncertainty is larger at large $\sum E_T$.

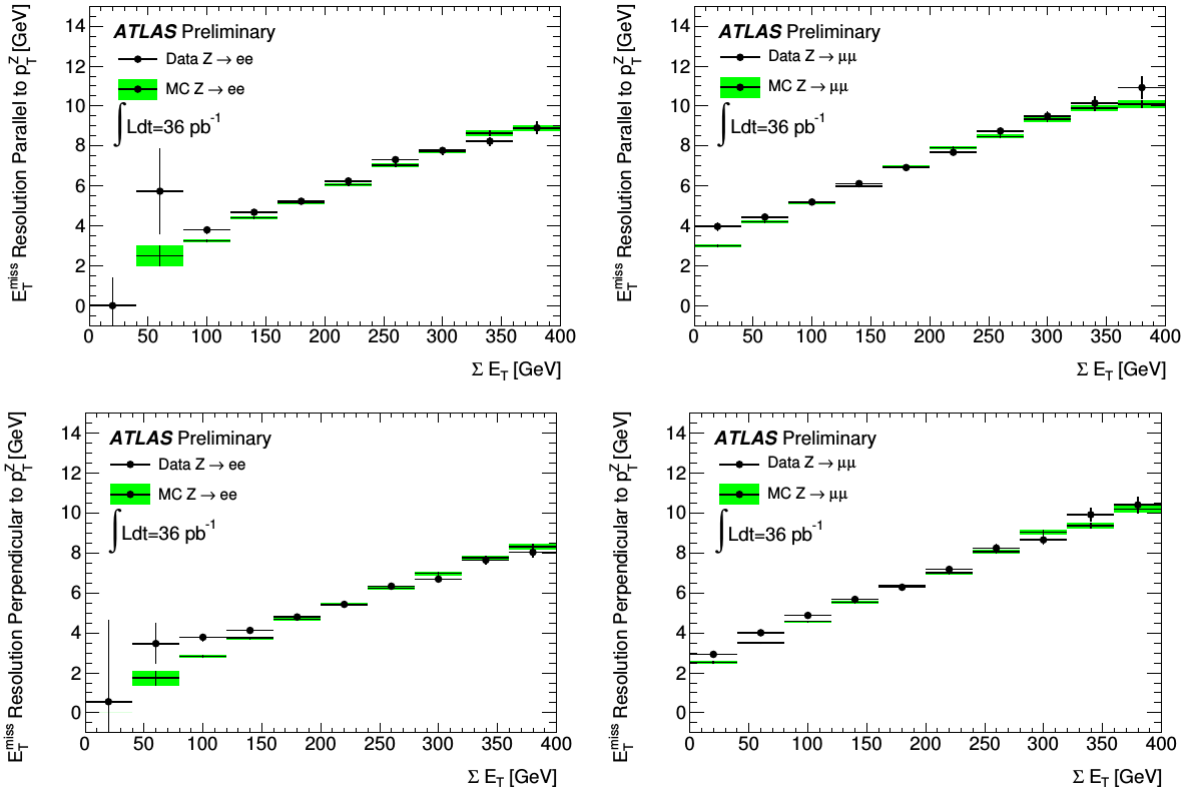


Figure 3.12: Resolution of E_T^{miss} in $Z \to ee$ (left) and $Z \to \mu\mu$ (right) events projected along an axis parallel (top) and perpendicular (bottom) to the Z boson transverse momentum as a function of the total calorimeter transverse energy of the event [57]

4 proton-proton collisions simulation

This chapter is dedicated to the description of the simulation of proton-proton collisions and of the Monte Carlo generators used in this thesis. The Monte Carlo samples used in the analysis will be described in the next chapter.

4.1 Monte Carlo simulation

Monte Carlo (MC) simulations are an essential part of data analyses in experimental high energy physics. Implementing the state-of-the-art knowledge of physics and experimental detection, MC simulations try to emulate the outcome of particle collision as realistically as possible. Common MC simulations in particle physics consist of event generation and detector simulation. Then the simulated events are reconstructed using the same procedure as applied to data (see §3). Various algorithms for each of the three parts are implemented in the ATLAS software framework, called Athena [58], and will be discussed briefly in this chapter.

4.1.1 Event generation

In the most common MC generators the method used to estimate cross sections in the pQCD regime (see §1.1.3) is to make the calculations of the matrix elements (ME) expressed by the Feynman diagrams of the particular process; combining the matrix element with the process phase-space it is possible to get the desired cross section. Fixed order ME calculations lead to divergences when soft and collinear partons (gluons and quarks) are produced. However, it is possible to obtain an approximate description of the cross-section for processes with several gluons and quarks in the final state,

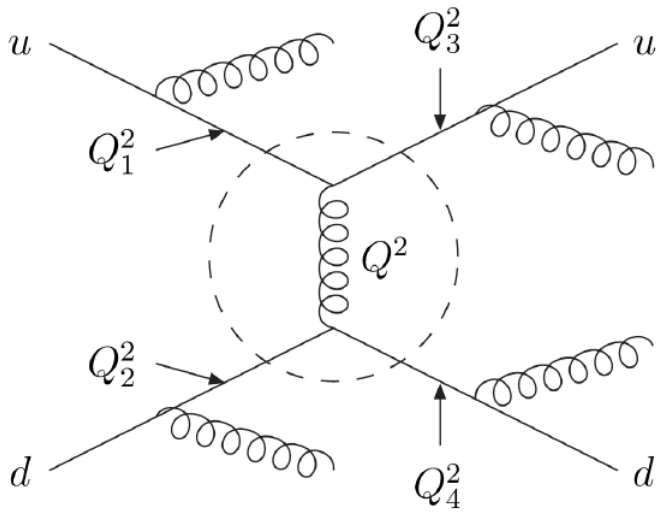


Figure 4.1: The "factorization" of a $2 \to n$ process.

using the *parton shower* approach. Consider for example the process of quark scattering in Figure 4.1, where many radiative processes occur. The starting point is to decompose a complex $2 \to n$ process, where n represents a large number of partons in the final state, into a simple core process convoluted with showers.
The hard process is the part of the event (the dashed circle in Figure 4.1 and the inner square in Figure 4.2) where large momentum transfer occurs and it determines the main characteristics of the event. The partons participating in the hard scattering are asymptotically free quanta and their interaction can be described by perturbation theory. The partonic content of the protons is described by the *parton distribution functions* (PDFs, see §5.1).
Thus the hard subprocess (whose scale is Q^2) cross section is calculated first at LO (or NLO if possible) with matrix element expression. All the higher-order effects, such as initial- and final-state radiation (ISR and FSR), both from QED and QCD are simulated later by the parton shower.
As the event evolves downwards in momentum scale, it ultimately reaches a scale at which QCD becomes strongly interacting. At this stage the perturbative evolution must be replaced by a non-perturbative hadronisation model that describes the confinement of the system of colored partons into colorless hadrons. The hadrons produced in this hadronisation phase are typically unstable resonances. Their decays into lighter particles long-lived enough to be detected are also simulated.

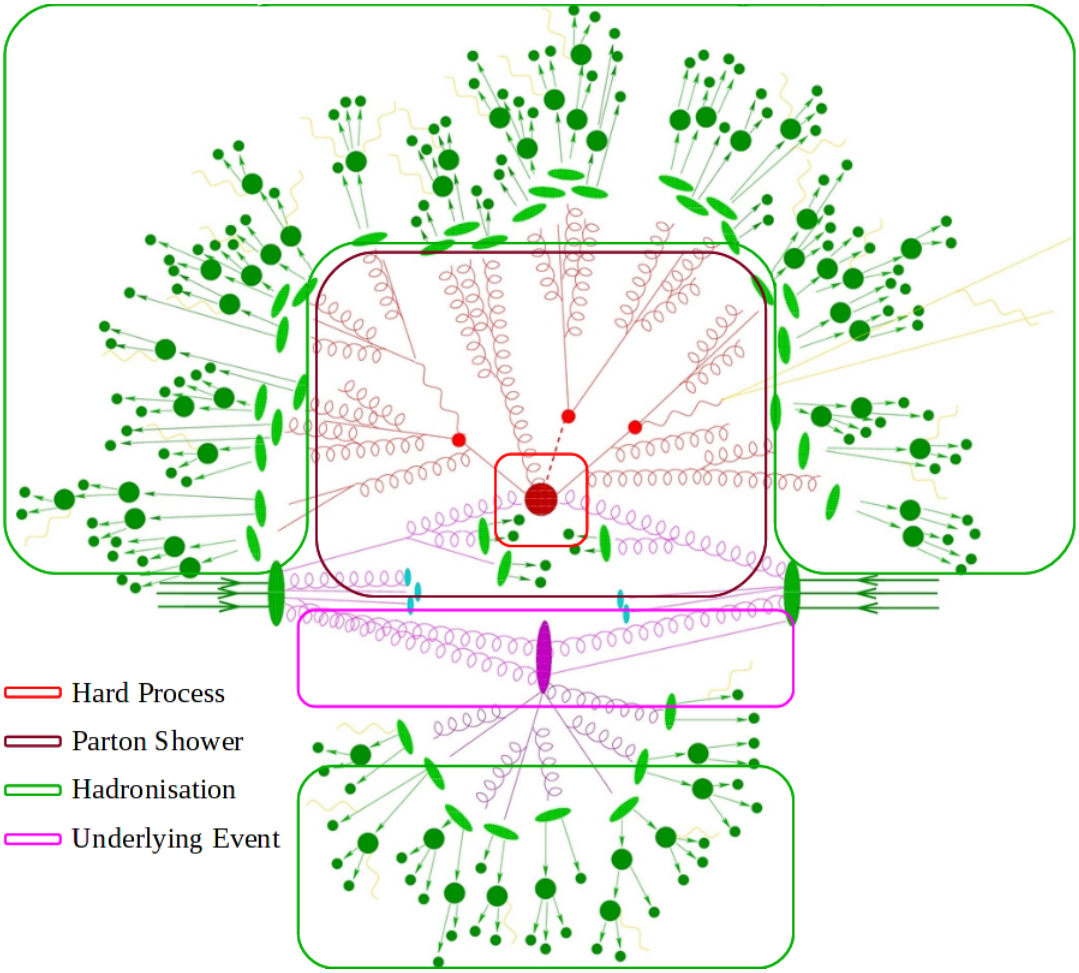


Figure 4.2: Schematic representation of a proton-proton collision [59].

Moreover it is possible that in a given proton-proton collision more than one pair of partons may interact (multi parton interactions, MPIs) with each other thus producing additional partons which may contribute to the final state. The beam remnants are all that is left in the colliding protons after the partons participating in the hard scattering. MPIs together with the beam remnants form the so-called underlying event (UE).
Finally additional soft proton-proton collisions can occur from simultaneous collisions in the same LHC proton-proton bunch crossing. These pile-up events (see §2.1) are overlaid on top of the hard event. Additional beam background comes from occasional interaction of the beams with gas in the beam pipe, from backgrounds due to the interaction of the beams with the collimators (beam halo) or from noise due to cavern background which is primarily made up of low-energy photons and neutrons generated from interactions in the cavern walls or in the detector shielding.

MC generators

To take into account all these effects, general-purpose MC generators are used providing fully exclusive modeling of high-energy collisions.
The PYTHIA [60] Monte Carlo generator implements leading-order matrix elements from perturbative QCD for $2 \to 2$ processes, followed by p_T-ordered parton showers, calculated in the leading-logarithm approximation and finally the Lund string model for hadronisation [61]. The underlying event in PYTHIA consists mainly of multiple-parton interactions interleaved with the initial state parton shower. In the ATLAS collaboration PYTHIA Minimum Bias (MB) events are overlaid to the main production process to reflect the number of pile-up events seen in data.
The HERWIG [62] generator uses similar leading order matrix elements to PYTHIA, but applies an angular-ordered parton shower and a clustering hadronisation model. For the underlying event, HERWIG is interfaced to JIMMY [63] to provide multiple partonic interactions. HERWIG++ [64], the latest version of HERWIG, directly implements a JIMMY-like approach to the underlying event.
One possible approach to address higher-order corrections in perturbation theory, pursued by so-called Tree Level Matrix Element generators such as SHERPA [65] and ALPGEN [66], is to calculate fixed-order diagrams corresponding to the emission of real particles, thereby neglecting all virtual contributions from loop diagrams. This approach has the disadvantage of introducing a cut on the parton level, necessary to avoid soft and collinear singularities in the tree-level matrix elements, that influences the physical observables. Commonly, Tree Level Matrix Element generators are used in

combination with showering and hadronisation generators, taking separate care of hard process computation and the soft and collinear emission, respectively.
ALPGEN is a specialized tree matrix-element generator for hard multi-parton processes $2 \to n$ ($n \leq 9$) in hadronic collisions. It is interfaced to HERWIG or PYTHIA to produce angular-ordered or p_T-ordered parton showers, respectively. Parton showers are matched to the matrix element with the MLM [67] matching scheme.
SHERPA is able to generate complete hadronic final states using a self-designed parton shower algorithm. It uses the CKKW scheme [67] to match the parton showers to the matrix element.
A more advanced approach that combines a NLO Matrix Element generator, taking into account both real and virtual emissions at NLO, and a parton shower programme, is pursued by the MC@NLO [68], POWHEG [69] and SHERPA MEPS@NLO [70] event generators. Special care is taken while merging the soft and collinear and the hard emission, to avoid double counting of portions of phase space, due to overlaps between higher-order processes and parton showers.

4.1.2 Detector simulation and reconstruction

For generated MC events the response of the ATLAS detector is simulated via the GEANT4 simulation toolkit [71], embedded within the Athena framework. It provides a precise description of the detector geometry (position, dimension and material of all detector parts) as well as a detailed simulation of the various physics processes, caused by the interaction of the particles with the detector material. This includes processes ranging from energies of a few eV, such as the ionisation in gases, up to TeV energies, to provide a detector-response model as realistically as possible. In a subsequent digitisation step, the previously obtained information is processed in order to emulate the detector electronics output that one expects from the experiment. The offline reconstruction processes the space points from simulated events in the same way as those recorded by the detector to find tracks (trajectories of charged particles in the detector) in the inner detector and the muon system, determine energy deposits in the calorimeters, etc., following the procedures already described in Chapter 3.

5 The production of vector bosons

In this section the approach to calculate the cross-section for the production of a vector boson in association with hadronic jets is described.

5.1 W+jets and Z+jet associated production

At LHC, proton-proton collisions are performed. To be exact, since protons are made of quarks and gluons (partons), partons collide at variable energies and the effective center-of-mass is the one between the interacting partons. It was first pointed out by Drell and Yan [72] that parton model ideas developed for deep inelastic scattering could be extended to certain processes in hadron-hadron collisions. The paradigm process is the production of a lepton pair $l\bar{l}$ with a large invariant mass-squared, $M^2 = (p_l + p_{\bar{l}})^2 \gg 1$ GeV^2, by quark-antiquark annihilation (the Drell-Yan process illustrated in Figure 5.1):

$$pp \to q\bar{q} \to l\bar{l} + \chi \tag{5.1}$$

where χ is any generic hadronic final state consistent with energy and momentum conservation.
The calculation of the cross-sections when the colliding hadrons are composite particles cannot be treated by perturbative QCD. To be able to calculate it the interaction is separated into an interaction of the whole particles composition (i.e. the protons) on a soft binding energy scale (long distance) and a partons collision on a hard energy (short distance) scale. This approach, with the soft part factored into *parton distribution functions* (PDFs), is called factorisation. Following the factorisation theorem [73, 74], the cross-section

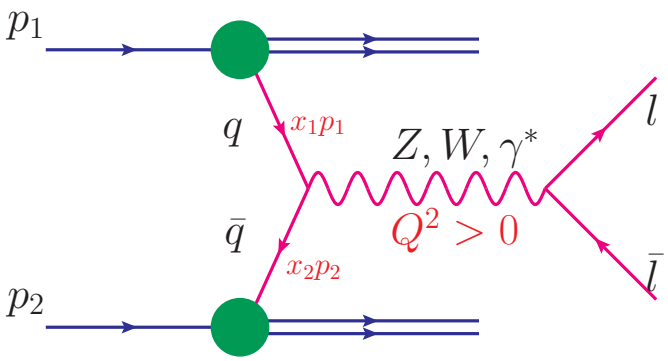


Figure 5.1: The Drell-Yan process $pp \to l\bar{l} + \chi$

for a process such as $pp \to l\bar{l} + \chi$ can be expressed as a convolution of the PDFs $f_q(x_1)$ and $f_{\bar{q}}(x_2)$ for the colliding protons with momenta p_1 and p_2 (see Figure 5.1) and the hard parton-parton cross-section $\hat{\sigma}_{q\bar{q}\to l\bar{l}}$ of all possible quark-antiquark combinations in the protons. Symbolically, one can write:

$$\sigma^{DY} = \sum_q \int dx_1 dx_2 f_q(x_1) f_{\bar{q}}(x_2) \hat{\sigma}_{q\bar{q}\to l\bar{l}}(\hat{s}, \mu_R^2) \tag{5.2}$$

where $\hat{s}$, defined as $\hat{s} \equiv (x_1p_1 + x_2p_2)^2 \approx x_1x_2s$, represents the centre-of-mass energy of the colliding partons squared, x_1 and x_2 the partons momentum fractions, μ_R^2 denotes the renormalisation scale and s is the square of the centre-of-mass energy at the hadron level, i.e. of the colliding protons. $f_{q,p_1}(x_1)$ and $f_{\bar{q},p_2}(x_2)$ are extracted mostly from deep inelasting scattering experiments and Tevatron and recently included LHC data. The formal domain of validity of Eq. 5.2 is the asymptotic "scaling" limit $M^2, s \to \infty$, $\tau = M^2/s$ fixed, analogus to the Bjorken limit of deep inelastic scattering. In Figure 5.2 proton's PDF are shown: for high x, the PDF is dominated by u-quarks while for low x it is dominated by gluons with a small contribution of virtual $q\bar{q}$ pairs in the proton known as sea quarks.

At lowest order in perturbation theory and assuming $M \ll M_z$, the cross-section of the sub-process represented in Figure 5.1 is given by the annihilation process $q\bar{q} \to \gamma^* \to l^+l^-$. Its total cross section is [72]:

$$\hat{\sigma}(q(p_q)\bar{q}(p_{\bar{q}}) \to l^+l^-) = \frac{4\pi\alpha^2}{3\hat{s}} \frac{1}{N} Q_q^2 \tag{5.3}$$

where $p_q \equiv x_1p_1$ and $p_{\bar{q}} \equiv x_2p_2$ are the parton's four-momentum, N the number of colours, Q_q the quark fractional charge and α is the electromagnetic coupling constant.

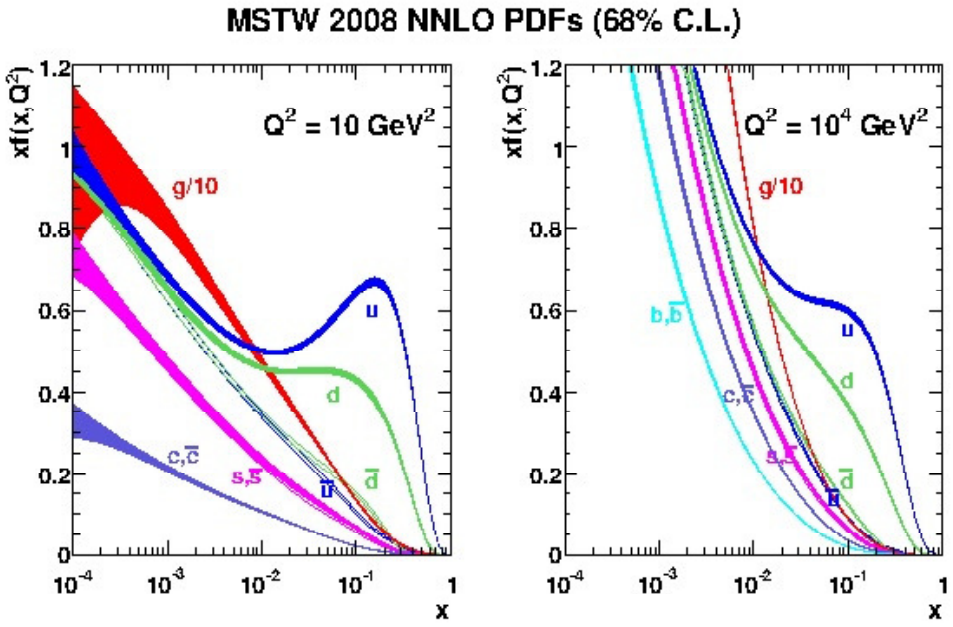


Figure 5.2: Proton's parton distribution functions (theorical model MSTW) with $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right) [75].

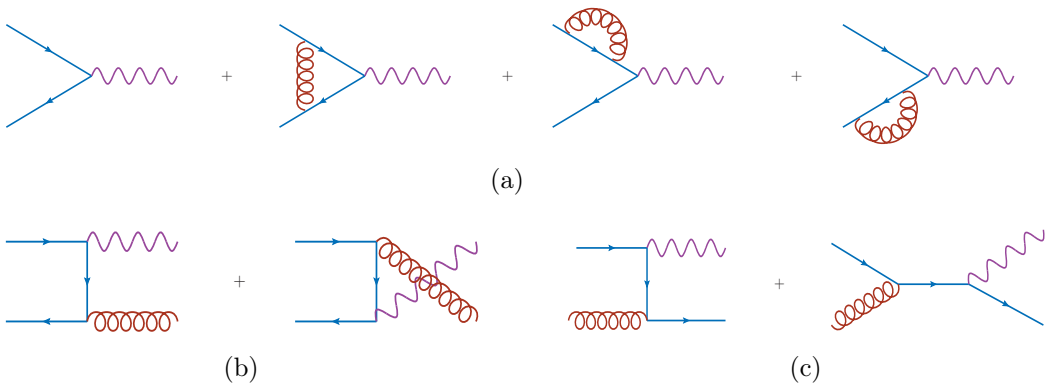


Figure 5.3: The leading-(a) and next-to-leading-order(b,c) diagrams for the Drell-Yan process.

Cross sections calculated at a fixed order, like in this case, leads to divergences that can be cured adding higher order corrections. The calculations of the perturbative QCD corrections to the leading order (LO) cross section in Eq. 5.3 are not shown here but there are at least three classes of diagrams that contribute at $\mathcal{O}(\alpha_s)$: virtual gluon corrections to the LO contribution (Figure 5.3(a)), real gluon corrections (Figure 5.3(b)), quark-gluon scattering process together with the corresponding $\bar{q}g$ contribution (Figure 5.3(c)). The effect of such corrections is that the PDF acquire a (calculable) logarithmic Q^2-dependence, and $\mathcal{O}(\alpha_s)$ "K-factor" appears, so that Eq. 5.2 becomes

$$\sigma^{DY} = K \sum_q \int dx_1 dx_2 f_q(x_1, Q^2) f_{\bar{q}}(x_2, Q^2) \hat{\sigma}_{q\bar{q}\to l^+l^-}. \tag{5.4}$$

The Q^2-dependent PDF are available[1], so the theory is completely predictive. Also, calculations at the higher next-to-next-to-leading order (NNLO) are available.

To calculate the cross-sections for Z and W boson production and their decays in their leptonic channels, it is sufficient to substitute the $q\bar{q} \to \gamma^* \to l^+l^-$ subprocess cross-section (Eq. 5.3) with the cross-sections $q\bar{q} \to Z \to l^+l^-$ and $q\bar{q} \to W \to l\nu$ (taking into account also the interference between $Z \to l^+l^-$ and $\gamma^* \to l^+l^-$):

$$\sigma_{q\bar{q}\to Z\to l^+l^-} = \sigma_{q\bar{q}\to Z} \cdot BR(Z \to l^+l^-)$$

$$\sigma_{q\bar{q}\to W\to l\nu} = \sigma_{q\bar{q}\to W} \cdot BR(W \to l\nu)$$

where $\sigma_{q\bar{q}\to Z\to l^+l^-}$ and $\sigma_{q\bar{q}\to W\to l\nu}$ are the production cross-sections for Z and W boson respectively, and $BR(Z \to l^+l^-)$ and $BR(W \to l\nu)$ their branching-ratios to leptons. Since the decay rate of Z and W bosons are small ($\Gamma_Z = 2.50$ GeV and $\Gamma_W = 2.08$ GeV, respectively) compared to their masses, it is sufficient to consider the production of effectively stable particles [72]. The production cross-sections can therefore be approximated as Z and W were on-shell:

$$\sigma_{q\bar{q}\to Z} = \frac{\pi}{3}\sqrt{2} G_F M_Z^2 \left(V_q^2 + A_q^2\right) \delta\left(\hat{s} - M_Z^2\right) \tag{5.5}$$

$$\sigma_{q\bar{q}\to W} = \frac{\pi}{3}\sqrt{2} G_F M_W^2 \left|V_{q\bar{q}}\right|^2 \delta\left(\hat{s} - M_W^2\right) \tag{5.6}$$

where V_q^2 and A_q^2 are associated with the vector and axial vector coupling constants of the neutral current interaction, and $\left|V_{q\bar{q}}\right|^2$ is the appropriate CKM matrix element.

[1]The Q^2-dependent PDFs are calculated by DGLAP evalution equations [76] at any value of Q^2 grater than an initial scale $Q_0 = \mathcal{O}(1$ GeV)

5.1.1 Multijet production

The total W and Z production cross sections can be decomposed in multijet cross sections of increasing order in α_s, and summed over. For example for the W boson [72]:

$$\sigma_W = \sigma_{W+0jets} + \sigma_{W+1jets} + \sigma_{W+2jets} + ... \tag{5.7}$$

where, in turn, each cross-section can be expanded as follows:

$$\begin{aligned} \sigma_{W+0jets} &= a_0 + \alpha_s a_1 + \alpha_s^2 a_2 + ... \\ \sigma_{W+1jets} &= \alpha_s b_1 + \alpha_s^2 b_2 + ... \\ \sigma_{W+2jets} &= \alpha_s^2 c_2 + ... \\ &... \ . \end{aligned} \tag{5.8}$$

The coefficient a_i, b_i, c_i, ... in these expansions are in general functions of the jet-definition parameters, for example the cone size used to cluster the partons into jets, and the transverse momentum, rapidity and separation cuts imposed on jets/clusters. The sum at each order in perturbation theory, i.e.

$$\begin{aligned} a_0 &= \delta_0 \\ a_1 + b_1 &= \delta_1 \\ a_2 + b_2 + c_2 &= \delta_2 \\ &... \end{aligned} \tag{5.9}$$

is independent from the jets parameters and corresponds simply to the perturbative expansion in powers of α_s of the total cross-section (δ_i in Eq. 5.9 are linked to the expansion of the "K-factor" in Eq. 5.4).
The largest contribution to the vector boson plus jets cross-section is given by the sum of the very first coefficients of every "exclusive" multijet cross-section a_0, b_1, c_2, ... It is possible to calculate these coefficients from the Feynman diagrams of partonic processes $xy \to V + j_1...j_n$, where x, y, j_i are quarks and gluons, and V is a vector boson, Z or W. The explicit calculations of these coefficients were carried out by Berends and Giele for a multiplicity of jets up to $N_{jets} \leq 4$ [77]. Berends and Giele also studied the ratio between the cross section with N jets and the cross section with $N-1$ jets in the final state

$$f_N(V) = \frac{\sigma_{V+N_{jets}}}{\sigma_{V+(N-1)_{jets}}} \tag{5.10}$$

and found that this ratio is approximatively constant. Therefore $f_N(V)$ can be parametrized as:

$$f_N(V) = \alpha + \beta N_{jets}.$$

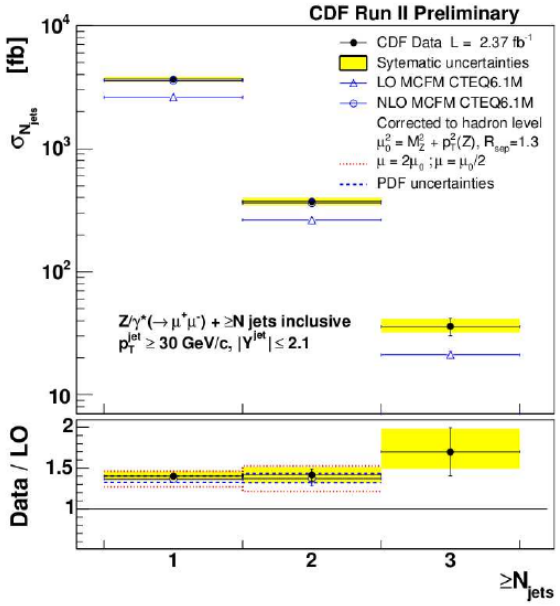


Figure 5.4: Measured total cross section for inclusive jet production in $Z/\gamma^* \to \mu\mu$ events, as a function of the number of jets. The data are compared to LO and NLO perturbative QCD predictions.

This scaling has been tested in various experiments, for example Figure 5.4 shows the cross section for the production of a Z boson in association with N jets ($\sigma_{N_{jets}}$) versus the number of jets as measured by the CDF collaboration at Tevatron, for a center-of-mass energy of 1.8 TeV [78].
The last few years have seen continued progress in the perturbative description of high-multiplicity processes. The NLO QCD corrections for $W + 4$-jet and $Z + 4$-jet production at hadron colliders were completed [79, 80, 81] and new results are reached for the $W+5$-jet process. These fixed-order QCD predictions have been obtained using of BlackHat and Sherpa with a novel technique based on unitarity rather than on diagram calculations at NLO. Other programs for QCD corrections calculation, *e.g.* HEJ [82], are based on BFKL (Balitsky-Fadin-Kuraev-Lipatov) approximation. Recent calculation using LoopSim [83] were able to compute approximate NNLO QCD corrections for W/Z + jets measurements.
Within this context is important to state that theoretical calculation of QCD corrections beyond NLO is a thriving field of theoretical research. Indeed, last years have seen great developments in MC generators (*e.g.* Sherpa MePs@Nlo) merging multi-leg approach with NLO ME calculation.

6 W+jets physics

In this chapter the main steps of the study of the associated production of a W boson and hadronic jets are presented. The analysis will be described with its peculiarities and difficulties.

6.1 Study of W+jets and Z+jets associated production at LHC

The study of the production of a Z or W boson with accompanying hadronic jets is one of the most important SM processes in high-energy hadron-hadron collisions. Many "new physics" processes, for example the production of heavy quarks, supersymmetric particles and Higgs boson, can be mimicked by the production of vector bosons in association with jets. It is therefore important to estimate these backgrounds accurately. In addition, these processes provide a unique test-bench for QCD: the rate of multijet production is dependent on the strong coupling constant and the kinematic distributions of the jets probe the theory of the underlying scattering process. This is particularly important in the kinematic region accessible at the LHC in order to understand the physics at or above the electroweak symmetry-breaking scale.
Direct measurements of the production cross sections of W+jets [84] or Z+jets [85] suffer from inherent theoretical and experimental uncertainties. Uncertainties in the theoretical SM predictions are dominated by the PDF and the scale (identified by Q in §5.1), initial- and final-state radiation (ISR and FSR), EW higher order corrections, and the non-perturbative evolution of partons into on-shell particles that could be detected as jets.
From the experimental point of view, W+jets and Z+jets events are af-

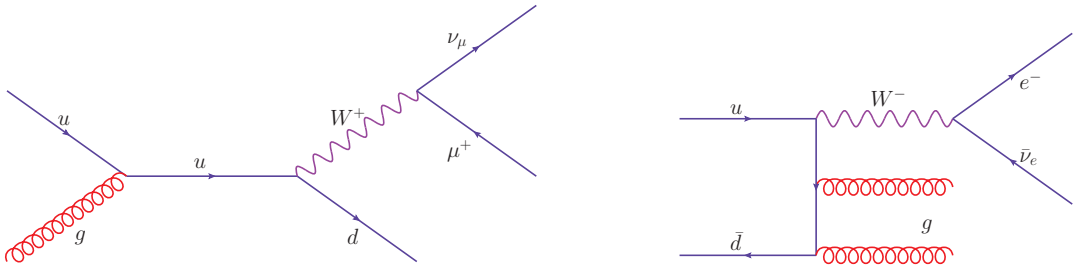
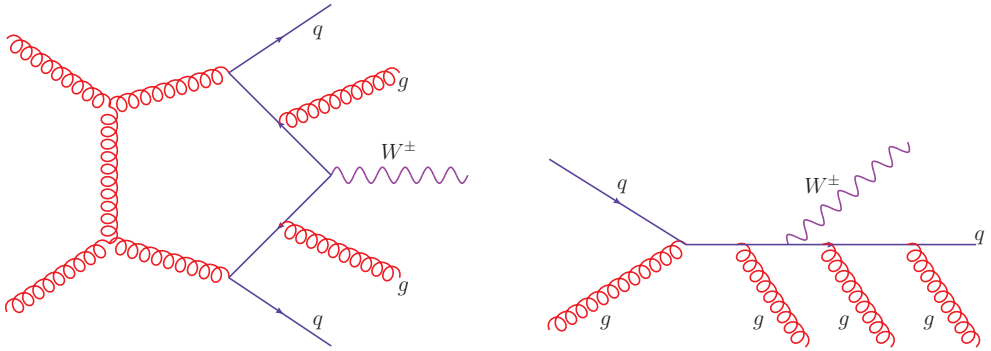


Figure 6.1: Tree level Feynman diagrams for W+jets production. a) Starting from an up quark and gluon and resulting in muon, neutrino, and down quark; b) Starting from an up quark and down antiquark and resulting in muon, neutrino, and two gluon jets.

fected by several systematic uncertainties, such as those affecting the single physics objects, like leptons and jets coming from MC mis-modelling and limited knowledge of the detector response to signal and background processes. Among all these kind of systematic uncertainties, the most relevant, as dominating the cross-section measurements, are the jet energy scale (JES) and jet energy resolution (JER) uncertainties already described in §3.5.6 and §3.5.4.

Thus, since the individual measurements of kinematic observables in W+jets and Z+jets events are affected by experimental and theoretical systematic uncertainties the measurement of ratios of observables [86] was designed to exploit the cancellation of theoretical and experimental uncertainties, building the foundations of a high precision test of the SM. In the ratio of the production cross-sections of V+jets with different jet multiplicities or the same jet multiplicity, the full information on the dynamics of the hadronic final state is retained, while the systematic uncertainties are significantly reduced. The similarity between W and Z production can also be exploited by performing the ratio of W and Z cross-sections for a given jet multiplicity, to achieve a large cancellation of theoretical and experimental uncertainties and provide an accurate test of QCD. In addition this measurement provides model-independent sensitivity to new physics coupling to leptons and jets.

Figures 6.1 and 6.2 show some examples of Feynman diagrams for W+jets production where the complexity increases with the number of jets.



(a) (b)

Figure 6.2: Feynman diagrams for W+4 jets production: a loop diagram on the left and a tree-level diagram on the right.

6.2 Analysis strategy

Let's consider the process $W \to \mu\nu$; its fiducial cross-section[1] is given by:

$$\sigma^{fid}_{meas}(W + jets) \cdot BR(W \to \mu\nu) = \frac{N_{obs} - N_{bkg}}{\mathcal{L}_{int} \cdot \epsilon} \tag{6.1}$$

where

- N_{obs} is the number of observed events in data passing the selection;
- N_{bkg} is the estimated background events passing the selection;
- $\mathcal{L}_{int}$ is the integrated luminosity ($\mathcal{L}_{int} = \int \mathcal{L} dt$, see §2.1). The selections adopted in this analysis corresponds to an integrated luminosity of 4.64 fb^{-1} with an uncertainty of $\delta\mathcal{L}/\mathcal{L} = \pm 1.8\%$ [87].
- $BR(W \to \mu\nu)$ is the branching fraction of W bosons decaying to $\mu\nu$;
- ϵ is the efficiency correction that takes into account trigger, identification and reconstruction efficiencies and the efficiency of selection cuts.

To determine the signal yield with its uncertainty at particle level ($(N_{obs} - N_{bkg})/\epsilon$ in Eq. 6.1), *i.e.* correcting for all the detector effects, an unfolding procedure [88] is used. The unfolding corrections account for the inefficiency and resolution effects both of the jets selection and of the boson selection. The fiducial phase space used in this thesis is reported in Table 6.1.

[1] Cross section measured in a restricted phase space, termed a *fiducial volume*.

Selection	Acceptance cuts
$W \to \mu\nu$	1 μ with $p_T > 25$ GeV and $\|\eta\| < 2.4$ 1 ν with $p_T > 25$ GeV $m_{T_{\mu\nu}} > 40$ GeV
$W \to e\nu$	1 e with $p_T > 25$ GeV and $\|\eta\| < 2.4$ 1 ν with $p_T > 25$ GeV $m_{T_{e\nu}} > 40$ GeV
Jets	$p_T > 30$ GeV and $\|y\| < 4.4$ $\Delta R(l, jet) > 0.5$

Table 6.1: Fiducial regions of the cross-section measurement for the different channels.

An accurate knowledge of the value of all the quantities in Equation 6.1 together with their statistical and systematic uncertainties is important to measure correctly the cross-section of the process we are interested in. Moreover, the fiducial cross-section has smaller theoretical uncertainties than the total cross-section, since in extrapolating from the experimentally measured fiducial cross-section to the total cross-section additional uncertainties arise due to uncertainties on the PDF and the factorisation and re-normalisation scales.
In this thesis, the production cross-section for W+jets events is not measured but attention is given to the determination of the number of events of two main backgrounds to W+jets events: QCD multijet events and $t\bar{t}$. They represent a large fraction of background to both electron and muon channels: QCD is relevant at low jet multiplicity while the background from semileptonic top decays becomes very important at high jet multiplicity. Moreover, because of few statistics in MC samples and large uncertainties on multi-jet and $t\bar{t}$ cross-sections, the background processes are estimated using data-driven methods.
In next sections the event's selections for $W \to \mu\nu$ and $W \to e\nu$ are discussed. Then an overview on the main sources of background is presented focusing on QCD multijet events and $t\bar{t}$. For both, the template selection and its normalisation will be discussed including an investigation of the QCD template bias in muon channel and correlation studies between the top templates and data. These studies are a critical ingredient towards the measurement of the W+jets cross-section and will contribute to the ATLAS forthcoming publication of the W+jets cross-section measurement.

6.3 Data and MC samples

6.3.1 Data samples

The analysis is performed using proton-proton collision data collected in 2011 with ATLAS experiment at a center-of-mass energy of 7 TeV. Data quality criteria are required to select useful luminosity blocks[2] and usable data for analysis based on electrons, muons, jets and missing transverse energy. The selections adopted in this analysis corresponds to an integrated luminosity of 4.64 fb^{-1}.

6.3.2 MC samples

MC simulations are used to subtract some of the background contamination to the signal and to correct measured distributions for detector effects. The used samples are summarised in Table 6.2.

- Events from W+jets production decaying to all three lepton flavours, e, μ, τ, are generated using ALPGEN interfaced with HERWIG for parton shower and fragmentation and to JIMMY to model underlying event contributions. Alternative signal samples of W+jets processes for comparison are generated with SHERPA.
- Top-quark pair samples are generated with the same generators as for the signal samples, *i.e.* ALPGEN.
- Single-top events are generated with ACERMC, interfaced with PYTHIA.
- Diboson processes (WW, WZ and ZZ) are simulated with HERWIG.
- The QCD samples are produced entirely by using PYTHIA.

The generated samples are passed through a full simulation of the ATLAS detector based on GEANT4. The simulated events are then reconstructed and analysed with the same analysis chain as for the data, and the same trigger and event selection criteria.
During 2011, detector and pile-up conditions changed considerably from few additional interactions up to $\sim$ 10 and the Monte Carlo events were reweighted to reproduce the average number of interaction per bunch-crossing observed in data.

[2]The Luminosity Block (LB) is the atomic unit of ATLAS data. One LB contains roughly 1 minutes of data taking, but this can vary due to run conditions and other operational issues.

Process	Generator	Cross Section(nb)
$W \to e\nu$	ALPGEN (+ HERWIG)	10.46
$W \to \mu\nu$	ALPGEN (+ HERWIG)	10.46
$W \to e\nu$	Sherpa, v 1.4	10.46
$W \to \mu\nu$	Sherpa, v 1.4	10.46
$W \to \tau\nu$	ALPGEN (+ HERWIG)	10.46
WZ	HERWIG	$5.42 \cdot 10^{-3}$
ZZ	HERWIG	$6.49 \cdot 10^{-3}$
WW	HERWIG	$4.50 \cdot 10^{-2}$
$t\bar{t}$	ALPGEN (+ HERWIG)	0.1668
single top, t-channel	ACER + PYTHIA	$6.97 \cdot 10^{-3}$
single top, s-channel	ACER + PYTHIA	$5.0 \cdot 10^{-4}$
single top, W top	ACER + PYTHIA	$1.57 \cdot 10^{-2}$
QCD $b\bar{b}$ with at least one muon with $p_T^{\mu} \geq 15$ GeV	PYTHIAB	73.9
QCD $c\bar{c}$ with at least one muon with $p_T^{\mu} \geq 15$ GeV	PYTHIAB	28.4

Table 6.2: Samples of simulated events used in the analysis. The cross sections quoted are the ones used to normalise estimates of expected number of events.

The signal samples are normalized to next-to-next-to-leading order (NNLO) pQCD calculation of inclusive Drell-Yan predictions.
In the next sections the analysis in the electron and muon channels will be treated separately. For each of them the event's selection and the background's estimation will be described as well as the studies of possible sources of bias related to data-driven methods.

6.4 Event selection

To select events with a leptonically decaying W in electron or muon channel the events must pass a series of different requirements summarised in Table 6.3.
The event's pre-selection is the same for both channels and requires the event to have at least one reconstructed primary vertex with at least 3 associated reconstructed tracks. The pre-selection removes all the events with noise bursts in LAr. Then the selection proceeds separately for electron and muon channels.

$W \to e\nu$ candidates are selected by requiring events to pass the following cuts:

- The events are triggered by the medium single electron trigger with a threshold in $p_T^e = 20 - 22$ GeV for different data taking periods.
- Events are required to have one tight++ reconstructed electron with $|\eta| < 2.47$ (excluding $1.37 < |\eta| < 1.52$) and $p_T^e > 25$ GeV. The W reconstructed electron has also to pass calorimeter and track isolation requirements considering the calorimeter deposits within $\Delta R < 0.2$ and the ID tracks within $\Delta R < 0.4$.

$W \to \mu\nu$ candidates are selected by requiring events to pass the following cuts:

- The events are triggered by the single muon trigger with a threshold in $p_T^\mu = 18$ GeV.
- Events are required to have one STACO combined reconstructed muon (see §3.4) with $|\eta| < 2.4$ and $p_T^\mu > 25$ GeV.
- The leading selected muon has also to pass a relative track isolation[3] ($<$0.1) requirement considering the ID tracks within $\Delta R < 0.2$ and a further requirement on the impact parameter. Considering the ratio $|d_0/\sigma_{d_0}|$, where d_0 is the impact parameter and σ_{d_0} its uncertainty, the events are accepted if $|d_0/\sigma_{d_0}| < 3$ in order to remove all the muons coming from secondary vertices. The ratio $|d_0/\sigma_{d_0}|$ is know as d_0 significance.
- Quality requirements are applied to the muon's track in the ID and in the MS: the events are accepted if the muon's track has a limited numbers of holes and dead sensors crossed in the pixels detector, SCT and TRT.

Additional cuts are then applied to both channels:

- The events are rejected if they have more than one reconstructed lepton (*Z veto*).
- E_T^{miss} cleaning cuts are applied, *i.e.* events are rejected if they have looser "bad" jets (see §3.5.5) with $p_T^{jet} > 20$ GeV and $|\eta| < 4.5$.

[3]The relative track isolation is defined as the ratio between the sum of the transverse momenta of all the tracks in a cone of radius R around the muon trajectory and the muon p_T.

- $E_T^{miss} > 25$ GeV.
- $m_T > 40$ GeV

The jets' selection is the same for both channels and is summarised in Table 6.4.
Jets are reconstructed using topoclusters (see §3.5.2) as inputs for the anti-k_T algorithm (see §3.5.1) with $\Delta R = 0.4$ and are calibrated using the EM+JES scheme (see §3.5.3). Events are rejected if they contain at least one jet meeting all of the following criteria:

- Falling into dead front end boards in the LAr calorimeter[4] (*LAr hole veto*).
- Overlapping with a selected $W \to l\nu$ lepton within $\Delta R < 0.5$.
- p_T^{jet}(EM+JES)<30 GeV.
- $|\eta^{jet}| < 4.4$

Within the tracker acceptance, jets from pile-up interactions were removed by a cut on the jet-vertex fraction (JVF) which was computed for each jet in the event. The JVF is defined as the sum of the p_Ts of all matched-tracks from a given vertex divided by the total jet-matched track p_T. It is defined for each jet with respect to each primary vertex and is close to 0 for pile-up jets, close to 1 for hard scattering signal jets and -1 for jets without associated tracks. Therefore, jets for which JVF < 0.75 are rejected. Jets which fall outside of the fiducial tracking region ($|\eta| > 2.4$) or which have no matching tracks are not considered for the JVF cut.
In the next sections the main sources of background will be described.

6.5 Backgrounds estimation

The main backgrounds of this analysis are QCD and $t\bar{t}$. They are both estimated using data-driven methods and will be described separately in the next two sections.
Electroweak (EWK) background comprises $W \to \tau\nu$ and diboson production processes (WW, WZ, ZZ). $W \to \tau\nu$ can be source of background when hadronic tau decays fake a lepton or the τ lepton decays leptonically. Diboson production processes (WW, WZ, ZZ) mimic $W \to l\nu$ when leptons escape detection. Also $Z \to ee$ and $Z \to \mu\mu$ are relevant backgrounds, respectively

[4]The problematic region is $-0.1 < \eta < 1.5$ and $-0.9 < \phi < -0.5$

Pre-selection	
Vertex	at least one primary vertex with at least 3 associated reconstructed tracks
remove events affected by LAr noise burst	
Electron selection	
Triggers	medium single electron trigger with $p_T^e = 20 - 22$ GeV for different data taking periods
Electron identification and reconstruction	Tight++ $\|\eta\| < 2.47$ (excluding $1.37 < \|\eta\| < 1.52$) $p_T^e > 25$ GeV
Electron Isolation	W reconstructed electron passes the calorimeter isolation requirement within $\Delta R < 0.2$ and the track isolation requirement within $\Delta R < 0.4$
Muon selection	
Triggers	single muon trigger with $p_T^\mu = 18$ GeV
Muon identification and reconstruction	STACO combined muon $\|\eta\| < 2.4$ $p_T^\mu > 25$ GeV
Quality	muon's track with a limited numbers of holes and dead sensors crossed in pixels detector, SCT and TRT
Isolation*	relative track isolation within $\Delta R < 0.2$ less than 0.1
Impact parameter*	$\|d_0/\sigma_{d_0}\| < 3$
**Isolation and impact parameter selections only applied to leading selected muon. The variable used to define the impact parameter selection is called d_0 significance and is indicated as $\|d_0/\sigma_{d_0}\|$*	
$W \to l\nu$ event selection	
Z veto	no second selected lepton
MET cleaning	MET cleaning cuts (event rejected if any looser "bad" jets found with $p_T > 20$ GeV and $\|\eta\| < 4.5$)
Missing energy	$E_T^{miss} > 25$ GeV
Transverse mass	$m_T > 40$ GeV

Table 6.3: Event selections for electron and muon channels.

W event selection: jets' variables	
LAr Hole	Veto events with jet in LAr hole
Jet isolation	$\Delta R(l, jet) > 0.5$ (jet veto, not event)
Jet rapidity	$\lvert\eta^{jet}\rvert < 4.4$
Pileup-jet rejection	$JVF > 0.75$ if $\lvert\eta^{jet}\rvert < 2.4$
Transverse momentum	$p_T^{jet} > 30$ GeV

Table 6.4: Jet selection.

for $W \to e\nu$ and $W \to \mu\nu$, as one single lepton is missed due to reconstruction and identification inefficiencies or because the lepton falls outside the detector acceptance. These events will lead to fake E_T^{miss} in the event and provides a single lepton signature. Another small background source that increases with the jet multiplicity is represented by single top processes.
All these background sources are estimated from Monte Carlo simulations. All the MC samples employed are listed in Table 6.2; they are subjected to the same selection cuts as the data and normalized to the data integrated luminosity.

6.6 Data-driven QCD background

QCD background on the $W \to e\nu$ is principally caused by mis-measured jets which fake an electron. Since the jet is mis-measured it is possible to get a spurious E_T^{miss} signal as well. In muon channel, the hadronic background is dominated by heavy flavour di-jet events where a b-hadron decays into a muon, as the isolation requirement rejects most light flavour multijet events. Hadronic backgrounds are at the level of $\sim 13\%$ for one or more jets in the event after all selection requirements. A data-driven technique is employed to model this background because of several reason: large uncertainties on the jet cross-section, large MC sample that would be required to provide adequate statistics given the large cross-section for jet production and large uncertainty on MC modelling of jets faking an electron or muons in a jet reconstructed as isolated muons.

6.6.1 The fit procedure

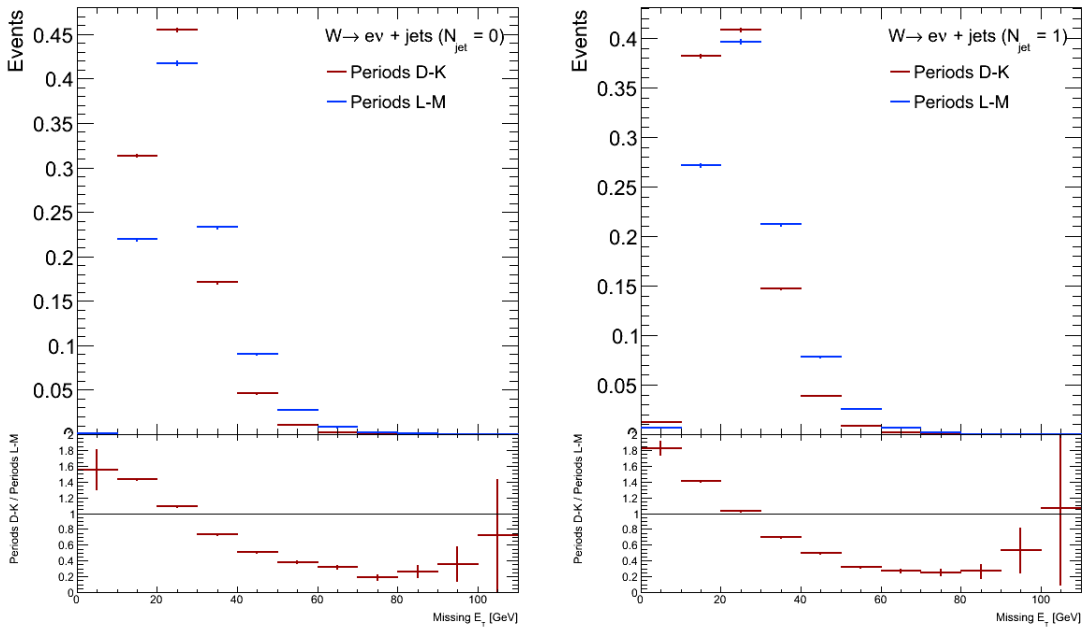
An estimate of the shape of this background is derived from data by selecting a QCD-enriched phase space. Then the normalization of the multijet background proceeds in the following way. A kinematic variable, such as E_T^{miss} which discriminates between QCD and W events is used to derive two

templates. The template for QCD is derived using the data-driven control sample, described above, while the templates for the W signal and EWK background are taken from MC. These two templates are fit to the data using a maximum likelihood fit. The fit procedure is performed using the `TFractionFitter` class from `ROOT` [91].
The fit takes as input the QCD template, after removing contamination from W signal and EWK based on MC simulation, and a signal template that includes the signal and all backgrounds as predicted by MC simulation and varies the normalisation of each to obtain the best fit with the data. The fit is carried out in the E_T^{miss} [92] variable after all selection cuts applied, excluding $E_T^{miss} > 25$ GeV cut to take advantage of the large difference in shape at low E_T^{miss} values. Indeed, QCD events contain mis-measured jets and tend to peak at low values of E_T^{miss} while the signal and other backgrounds peak at higher values. As a final step, the QCD estimation is calculated after the additional cut on E_T^{miss} in order to provide the level of contamination of this background in the signal sample after the full event selection. The fit is performed in exclusive jet bins with a separate normalisation factor derived for each jet bin. This scale factor is propagated to each distribution to derive the QCD background contamination for each of them. For $N_{jets} > 5$ there are insufficient statistics in the QCD template to perform a stable fit, thus an inclusive fit for $N_{jets} \geq 5$ is carried out.

6.6.2 Template selection and fit results in $W \to e\nu$

A QCD-enriched sample in electron channel is selected by using relaxed triggers and electron identification (ID) requirements ("loose" ID and a subset of the "medium" ID cuts while failing the "tight++" ID used in the signal selection). Finally the electrons were required to be anti-isolated against close-by energy deposits: the relative calorimeter isolation, defined as ratio of the sum of the transverse energy of all the deposits (excluding the reconstructed electron) in the EM calorimeter within $\Delta R < 0.3$ around the reconstructed electron with respect to the transverse energy of the reconstructed electron, is required to be > 0.2.
The prescales of triggers used to select multijet background to $W \to e\nu$ event changed over the course of the 2011 run and pileup conditions changed dramatically across the 2011 run. All these effects influence the QCD background shape and rate due to pileup interactions. In order to take care of the time dependence of this effect, data were divided in two different periods[5] the

[5]In ATLAS data taking is split in periods of constant LHC and detector conditions.



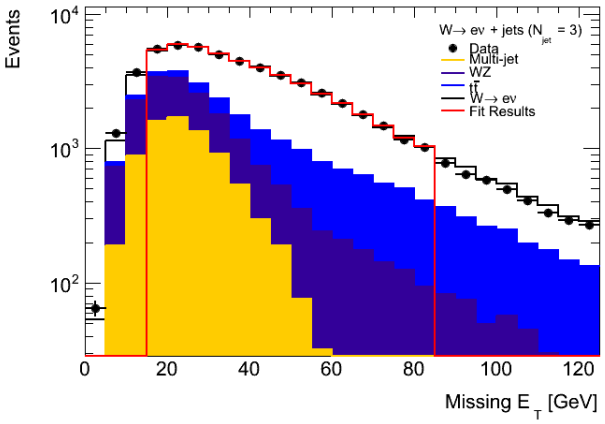
(a) $N_{jet} = 0$ (b) $N_{jet} = 1$

Figure 6.3: The E_T^{miss} distribution (normalised to the area) for events passing the QCD template selection divided in periods D-K and L-M for 0- and 1-jet events in the electron channel. Bottom plots show the ratios between E_T^{miss} distribution in D-K with respect to the one in period L-M.

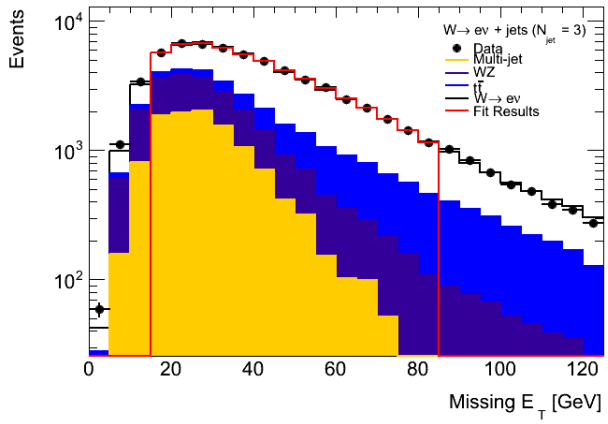
early 2011 data (periods D-K – from April to August) and the later periods (L and M – from September to October) (See figure 6.3). Thus, the fit of QCD background in the electron channel is performed in the two different data periods templates and then combined into a single QCD estimation. The combination is carried out in the following way. Once the fits are performed separately in periods D-K and L-M, both QCD templates are scaled to their luminosity fraction with respect to the full dataset. Then, they are added together to build the combined QCD template. The combined QCD fractions is therefore defined as the ratio between the number of events in the combined template and the number of events in data. The fit range in E_T^{miss} is [15; 80] GeV for both data periods.
The results of these fits can be seen in Table 6.5 and the plots in Figure 6.4. Table 6.5 shows the QCD fractions obtained in period D-K and L-M separately as results of the fits and their combination.

Period	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
D-K	$(2.9 \pm 0.2)\%$	$(7.9 \pm 0.2)\%$	$(8.1 \pm 0.5)\%$	$(8.3 \pm 1.2)\%$	$(6 \pm 2)\%$	$(7 \pm 4)\%$
L-M	$(5.7 \pm 0.7)\%$	$(14.6 \pm 0.6)\%$	$(13.4 \pm 1.5)\%$	$(13 \pm 3)\%$	$(9 \pm 5)\%$	$(5 \pm 7)\%$
Combined	$(4.1 \pm 0.6)\%$	$(10.9 \pm 0.5)\%$	$(10.6 \pm 1.3)\%$	$(10.5 \pm 3)\%$	$(7 \pm 5)\%$	$(6 \pm 7)\%$

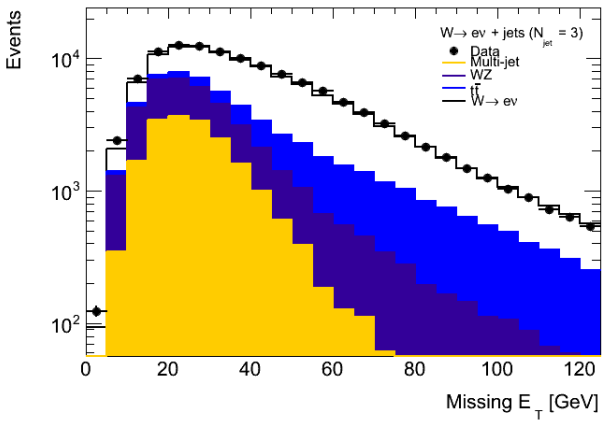
Table 6.5: QCD fraction in electron channel (ratio between the estimated number of QCD events over the total number of events observed in data) versus jet multiplicity. Only the statistical uncertainty returned by the fit are quoted.



(a) Period D-K



(b) Period L-M



(c) Combined result

Figure 6.4: Template fits to determine the QCD fraction for the electron channel for 3-jets events in periods D-K (a) and L-M (b) and the combined result (c).

6.6.3 Template selection and fit results in $W \to \mu\nu$

To determine the QCD template in muon channel, requirements on impact parameter d_0, d_0 significance and the relative track isolation are used. d_0 and d_0 significance give us a measure of how far the track is from the primary vertex and which is the uncertainty on it. They are used again to separate isolated μ from W decay from μ in jets. The combination of different requirements on these observables allows to determine an hadronic-enriched sample. Several W QCD control samples are investigated, as shown in the following list using a pseudonym for each of them to make easier the reading and avoid repetitions:

- `AntiDOWithIso`: muon failing $|d_0/\sigma_{d_0}| < 3$ but passing signal isolation requirement. The cut on d_0 significance takes into account that background muons are mainly non-prompt from b-hadron decays while the signal isolation requirement reduces the bias of the QCD template with respect to the signal sample.

- `AntiDOWithIso_DOWindow`: muon failing $|d_0/\sigma_{d_0}| < 3$ but passing signal isolation cut and additional $0.1 < |d_0| < 0.4$ cut. The additional requirement on the impact parameter further remove contamination from μ coming from W decays [92].

- `RestrAntiIso`: muon passing inverted isolation window, *i.e.* with relative track isolation in the range $]0.1; 0.5[$. This selection collect anti-isolated muons without applying requirements on d_0 or d_0 significance.

- `RestrAntiIsoWithDO`: muon passes inverted isolation window, *i.e.* with relative track isolation in the range $]0.1; 0.5[$ and passing $|d_0/\sigma_{d_0}| < 3$ cut. The requirement on d_0 significance reduces the bias of the QCD template with respect to the signal sample.

All these choices define QCD templates orthogonal to the signal sample.
Table 6.6 shows the QCD fraction returned by the fits to each of these templates, the amount of W signal contamination in the template as estimated by MC and the total contamination of the template derived from the sum of $t\bar{t}$, diboson and signal MC. The bias on observables due to the choice of the templates was investigated using PYTHIAB $b\bar{b}+c\bar{c}$ QCD MC and is discussed in section §6.6.3.
In Figure 6.5 the shape comparison for different templates and different background subtraction schemes is also shown.

Table 6.6 and Figure 6.5 show that `AntiDOWithIso` template has the

QCD Fractions without contamination subtraction (%)						
Template	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
AntiD0WithIso	4.1 ± 0.2	14.2 ± 0.3	13.2 ± 0.5	12.8 ± 1.1	11 ± 2	7 ± 2
AntiD0WithIso_D0Window	2.7 ± 0.3	12.7 ± 0.6	11.6 ± 1.1	11 ± 2	8 ± 4	5 ± 4
RestrAntiIso	2.22 ± 0.02	12.03 ± 0.02	10.90 ± 0.05	10.55 ± 0.09	7.7 ± 0.2	5.5 ± 0.2
RestrAntiIsoWithD0	2.23 ± 0.03	11.93 ± 0.05	10.80 ± 0.09	10.4 ± 0.2	7.8 ± 0.3	5.3 ± 0.3
Signal Contamination Fractions (%)						
AntiD0WithIso	34.05 ± 0.14	4.01 ± 0.04	3.59 ± 0.09	2.78 ± 0.15	2.6 ± 0.3	2.0 ± 0.5
AntiD0WithIso_D0Window	12.04 ± 0.12	1.16 ± 0.03	1.07 ± 0.07	0.72 ± 0.11	0.6 ± 0.2	0.4 ± 0.3
RestrAntiIso	8.56 ± 0.03	0.723 ± 0.007	0.719 ± 0.013	0.63 ± 0.02	0.57 ± 0.05	0.50 ± 0.08
RestrAntiIsoWithD0	12.09 ± 0.05	1.118 ± 0.011	1.26 ± 0.02	1.15 ± 0.05	1.08 ± 0.09	1.0 ± 0.2
Total Contamination Fractions (%)						
AntiD0WithIso	43.69 ± 0.15	6.96 ± 0.06	7.63 ± 0.13	9.3 ± 0.3	13.9 ± 0.7	21.4 ± 1.6
AntiD0WithIso_D0Window	23.17 ± 0.17	3.26 ± 0.06	3.65 ± 0.13	4.5 ± 0.3	6.4 ± 0.7	11 ± 2
RestrAntiIso	9.17 ± 0.03	0.833 ± 0.007	1.018 ± 0.016	1.67 ± 0.04	3.20 ± 0.11	6.0 ± 0.3
RestrAntiIsoWithD0	12.88 ± 0.05	1.337 ± 0.011	1.68 ± 0.03	2.52 ± 0.07	4.6 ± 0.2	8.4 ± 0.5
QCD Fractions with contamination subtraction (%)						
Template	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
AntiD0WithIso	2.15 ± 0.14	12.7 ± 0.3	11.7 ± 0.5	11.0 ± 1.0	9 ± 2	6 ± 2
AntiD0WithIso_D0Window	1.4 ± 0.2	12.1 ± 0.6	11.0 ± 1.0	10 ± 2	8 ± 4	5 ± 4
RestrAntiIso	1.85 ± 0.02	11.85 ± 0.03	10.70 ± 0.05	10.25 ± 0.09	7.3 ± 0.2	5.1 ± 0.2
RestrAntiIsoWithD0	1.74 ± 0.03	11.64 ± 0.05	10.48 ± 0.09	9.99 ± 0.17	7.2 ± 0.3	4.8 ± 0.3

Table 6.6: QCD fraction and contamination for the QCD templates investigated. Only the statistical uncertainty are quoted.

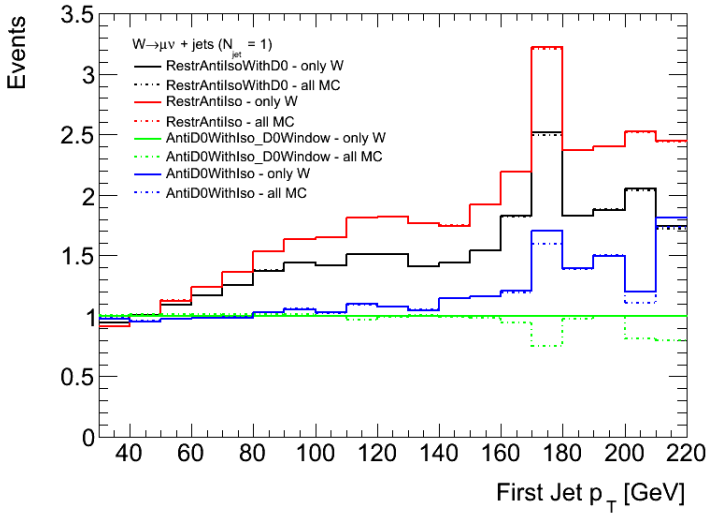
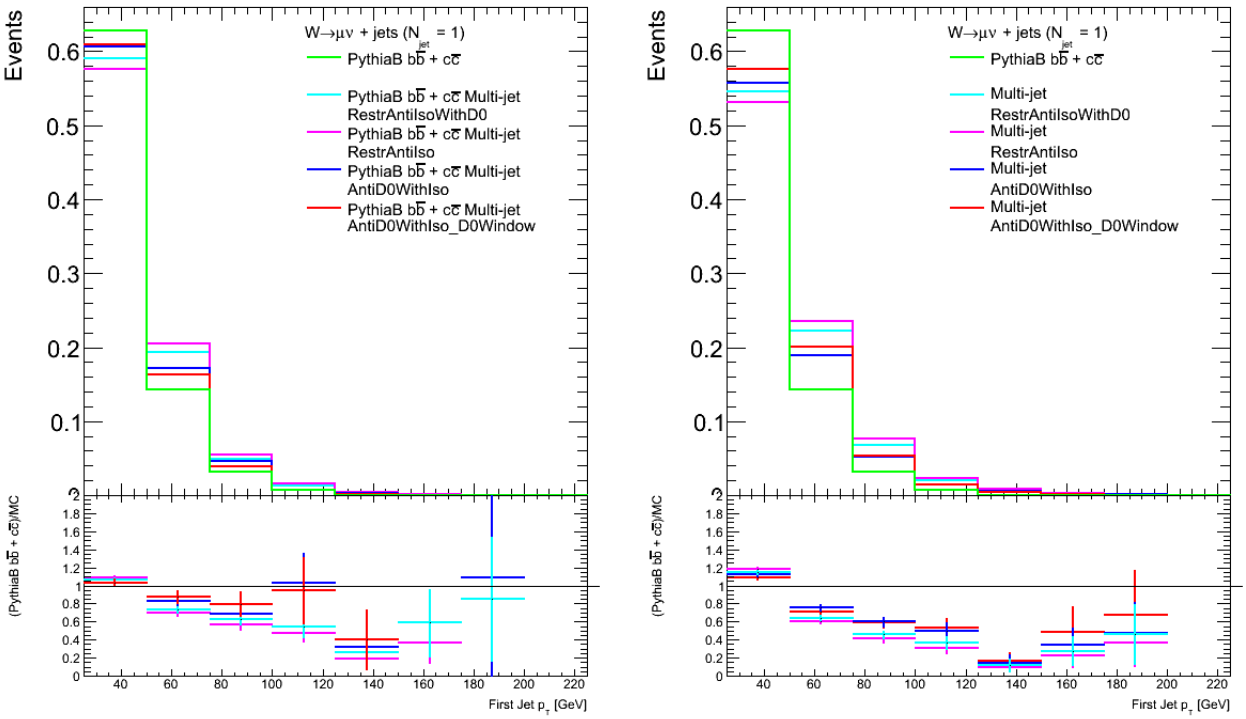


Figure 6.5: Shape comparison of QCD templates for the W channel. Shapes are shown as a ratio to the `AntiD0WithIso_D0Window` template with only signal contamination removed. Dotted lines show the results when full contamination subtraction is carried out whilst the solid lines show the ratio when only signal is subtracted from the QCD templates.

biggest amount of contamination and thus it is dependent on the shape modelling of diboson, $t\bar{t}$ and signal MC used for the subtraction. So, `AntiD0WithIso` is rejected as nominal template. Among the other templates, that show similar signal contamination, `AntiD0WithIso_D0Window` has a significantly higher total contamination including diboson and $t\bar{t}$. For this reason this template was not chosen as the nominal one. Thus the nominal template is chosen to be either `RestrAntiIso` or `RestrAntiIsoWithD0` which are the least contaminated.
The decision on which of these two restricted isolation templates to use was based on the studies shown in Figure 6.6. Here the various templates are compared to the leading jet p_T distribution in PYTHIAB $b\bar{b} + c\bar{c}$ MC. Figure 6.6(a) compares the QCD template derived from QCD MC using the different template definitions to the MC shape derived using the signal selection. Even if the `AntiD0WithIso` template gives the p_T distribution closest to the signal sample it has already been rejected due to its high levels of contamination. Among the other templates, the one closest to the signal sample in the p_T distribution is `RestrAntiIsoWithD0`, having the same cut on d_0/σ_{d_0} as for the signal sample. Figure 6.6(b) shows the same plot but this time comparing the MC signal selection shape to the templates derived from data. Again, the same features are observed, therefore `RestrAntiIsoWithD0` is chosen as



(a) (b)

Figure 6.6: Shape comparison (normalised to area) of different QCD template choices in the muon channel compared to the signal selection applied to PYTHIA $b\bar{b}+c\bar{c}$ MC. In (a) the templates are derived from Pythia MC. In (b) the templates are taken from data and contamination subtracted using W signal, diboson and $t\bar{t}$ MC.

	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$
QCD fractions	$(1.74 \pm 0.03)\%$	$(11.64 \pm 0.05)\%$	$(10.48 \pm 0.09)\%$
	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
QCD fractions	$(9.98 \pm 0.17)\%$	$(7.2 \pm 0.3)\%$	$(4.8 \pm 0.3)\%$

Table 6.7: QCD fraction in muon channel (ratio between the estimated number of QCD events over the total number of events observed in data) versus jet multiplicity. Only the statistical uncertainty are quoted.

the nominal QCD template.
The fit of QCD Background is carried out as explained in §6.6.1 in the E_T^{miss} variable in the fit range [15; 70] GeV. The results of the fits are shown in Table 6.7 and in Figure 6.7.

QCD template bias

This section presents the studies of the bias of the QCD template on the shape of distributions of QCD in various observables, *e.g.* leading jet p_T, H_T (scalar sum of all jets, plus the lepton and E_T^{miss}), S_T (scalar sum of all the jets in events), etc., due to the control sample selection. We want also to understand if the correlation among relative track isolation, d_0 significance and other observables is correctly reproduced by MC or causes a bias on data that is not reproduced by MC.

Correlation studies Let's consider the relative track isolation, the d_0 significance and the leading jet p_T distributions for 1-jet events in data. The following scatter plots have been produced:

- leading jet p_T vs isolation
- leading jet p_T vs d_0 significance

The distributions of leading jet p_T in bins of both relative track isolation and d_0 significance (Figures 6.8(a) and 6.8(b)) have been normalised to unit area for comparison. Then, the ratios of normalised distributions in the different bins have been considered: values close to 1 for each scatter plot are expected for non-correlated variables.
Figures 6.9(a) to 6.9(c) show the ratios of the normalised leading jet p_T distribution in bins of d_0 significance with respect to the first bin distribution, with $0 < |d_0/\sigma_{d_0}| < 1$, using different requirements on isolation. Results without any cuts applied on isolation are shown in Figure 6.9(a) while in Figures 6.9(b) and 6.9(c) isolation is required to be as in the signal region (relative

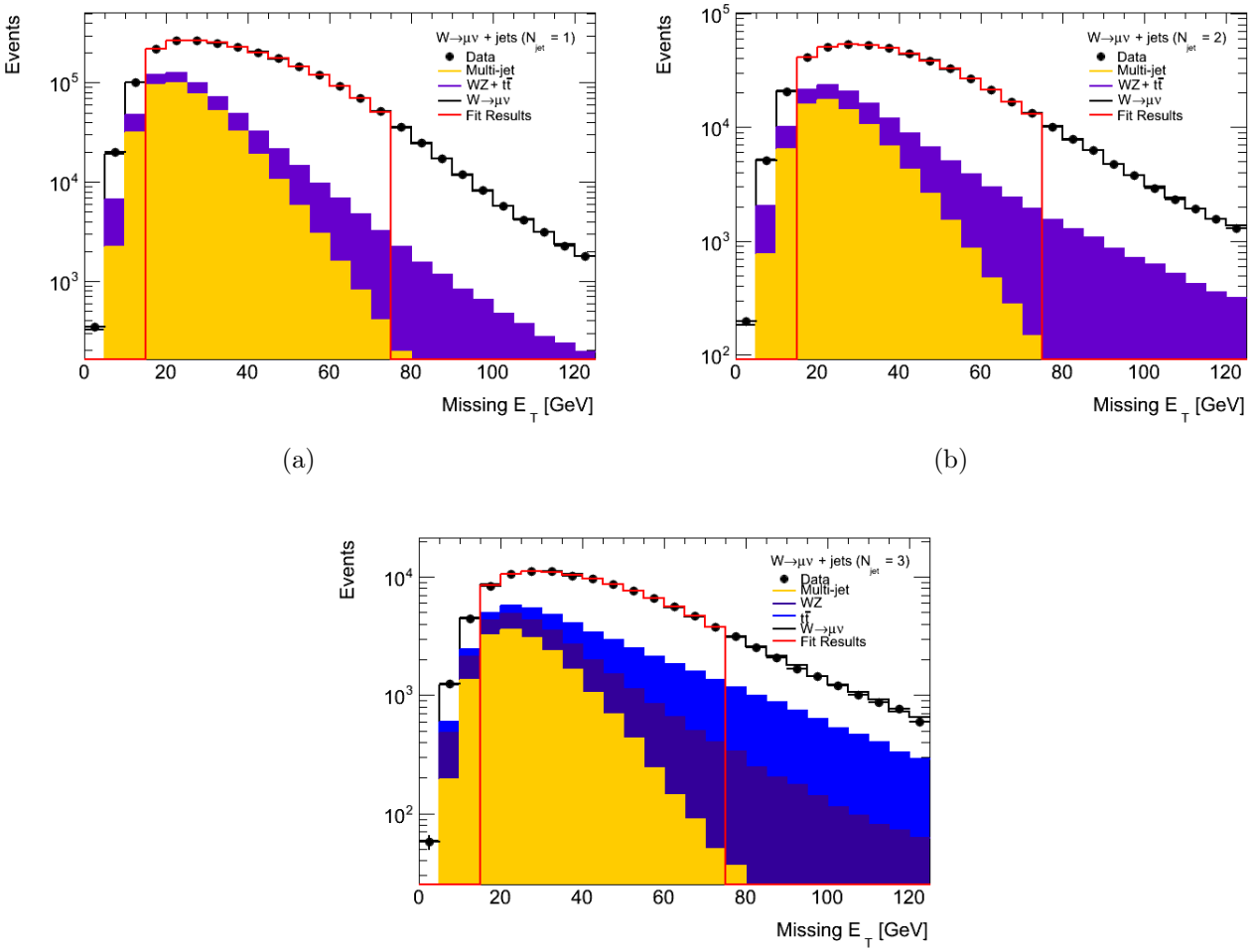
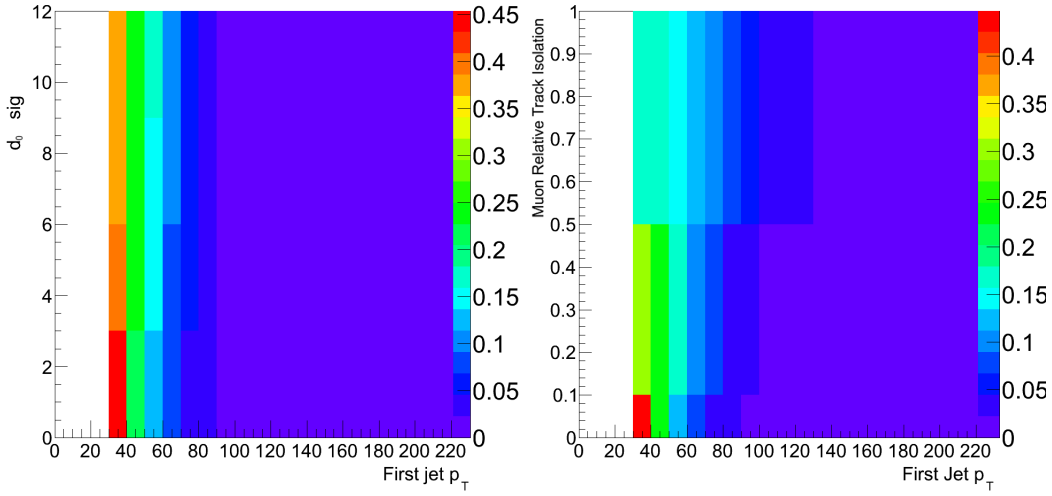


Figure 6.7: Template fits to determine the QCD fraction for the muon channel for 1-, 2- and 3-jets events.Only the statistical uncertainty returned by the fit are quoted.



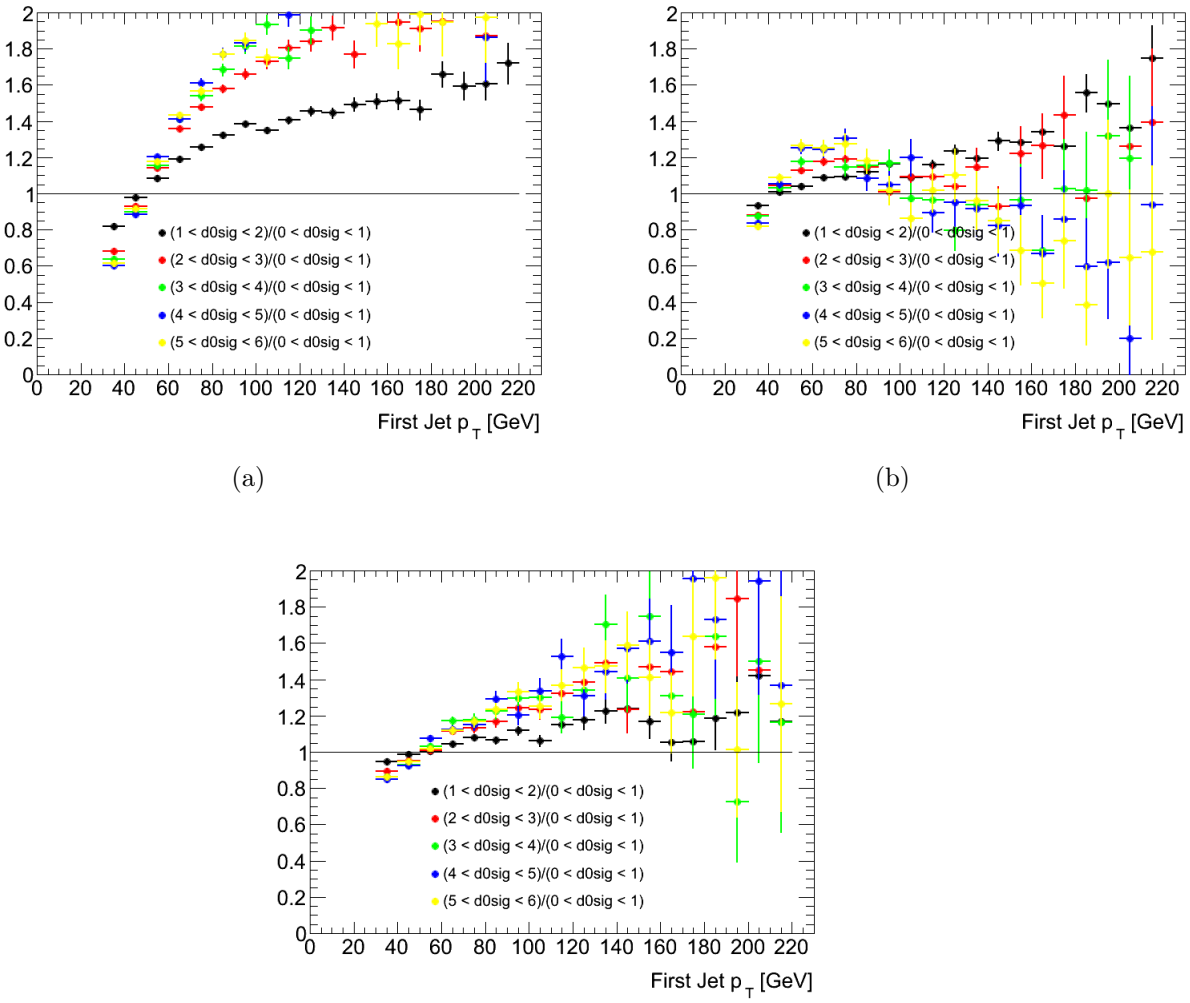
(a) (b)

Figure 6.8: Distributions of the leading jet p_T vs d_0 significance and of the leading jet p_T vs isolation.

track isolation < 0.1) and in the control region (relative track isolation in $]0.1; 0.5[$) respectively. From these plots it is possible to see that the leading jet p_T is correlated to d_0 significance (see Figure 6.9(a)) but the application of cuts on the relative track isolation (see Figures 6.9(b) and 6.9(c)) makes their correlation smaller with small or no dependence on d_0 significance cut.

It is important also to understand what is the impact of the isolation cut on signal and control regions in d_0 significance. Figure 6.10 shows again the ratios of the normalised leading jet p_T distribution in bin of d_0 significance. Now, the ratios are performed having as numerator the leading jet p_T distribution for different ranges of d_0 significance with relative track isolation between 0.1 and 0.5 (*i.e.* as for the control sample selection, labelled as CS) and as denominator the leading jet p_T distribution with d_0 significance and relative track isolation requirements (*i.e.* d_0 significance < 3 and relative track isolation < 0.1) as for the signal region (labelled as SS). It shows that using the same cut on d_0 significance for control region and signal region the bias decreases as already discussed in 6.6.3.

The same procedure has been followed on the normalised leading jet p_T distribution in bin of isolation with different requirements on d_0 significance. Figures 6.11(a), 6.11(b) and 6.11(c) show that the leading jet p_T is strongly correlated to isolation regardless of whether we use $|d_0/\sigma_{d_0}| < 3$ (Figure 6.11(b)) or $|d_0/\sigma_{d_0}| > 3$ (Figure 6.11(c)) for the control region definition.



(a) (b)

(c)

Figure 6.9: Ratios of the normalised leading jet p_T distribution in bin of d_0 significance with respect to the $0 < |d_0/\sigma_{d_0}| < 1$ bin distribution using different requirements on isolation: without any cuts applied on isolation (a), isolation requirement as in the signal region (b), isolation requirement as in the control region (c).

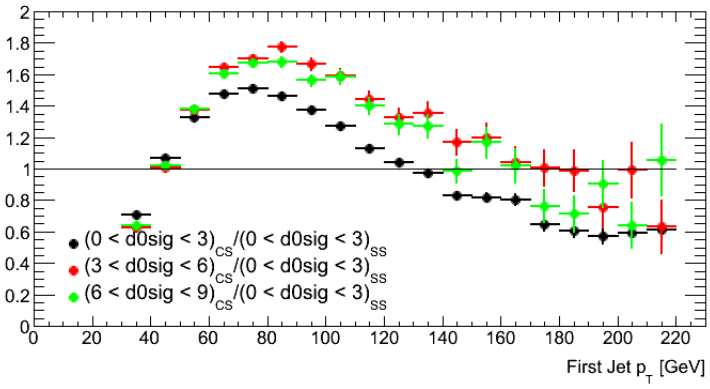


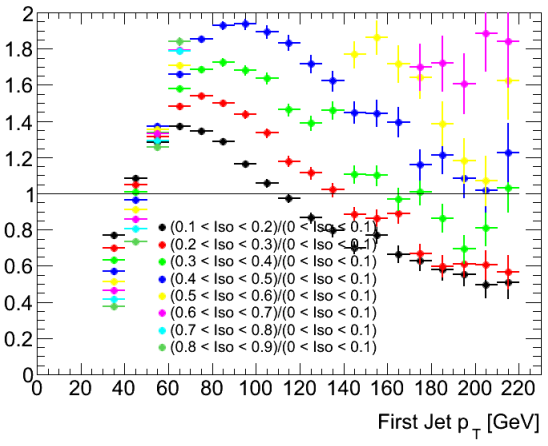
Figure 6.10: Ratios of the normalised leading jet p_T distribution in bins of d_0 significance having as numerator the leading jet p_T distribution for different ranges of d_0 significance and as denominator the leading jet p_T distribution with d_0 significance as for the signal region (*i.e.* d_0 significance < 3). In both numerator and denominator a relative track isolation is applied: the numerator has the relative track isolation in $]0.1; 0.5[$ (*i.e.* as for the control sample selection, labelled as CS) and the denominator has the same relative track isolation requirement of the signal region (*i.e.* relative track isolation < 0.1).

Re-weighting procedure An additional study was carried out to understand if the strong correlation shown above between relative track isolation, d_0 significance and leading jet p_T is reproduced also by MC or causes a bias on the distribution on data which is not reproduced by MC.
To do this we have re-weighted the PYTHIA $b\bar{b} + c\bar{c}$ MC to have the same d_0 significance and relative track isolation distributions as in data; the difference between the re-weighted MC and the default MC is used to estimate the bias. This procedure is done in two steps: re-weighting only QCD PYTHIA MC and re-weighting all MC samples including W signal and EWK backgrounds under two different assumptions, as explained here below:

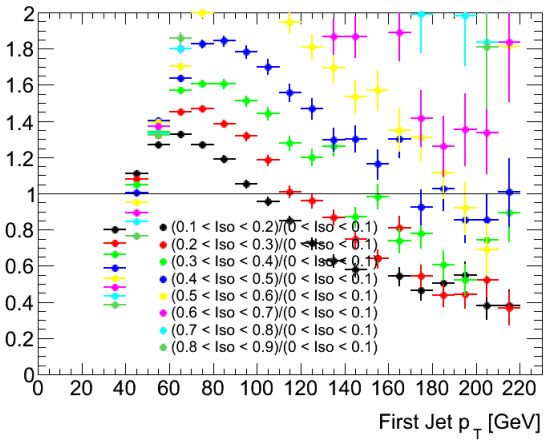
1. Assuming that all MC except the one for QCD correctly model measured observables, we have calculated the event weights $f_{(i,j)}$ to apply to QCD PYTHIA MC:

$$f_{(i,j)} = \frac{\frac{1}{\sum_{(i,j)} N_{(i,j)}^{QCD,Data}} N_{(i,j)}^{QCD,Data}}{\frac{1}{\sum_{(i,j)} N_{(i,j)}^{QCD,MC}} N_{(i,j)}^{QCD,MC}} \tag{6.2}$$

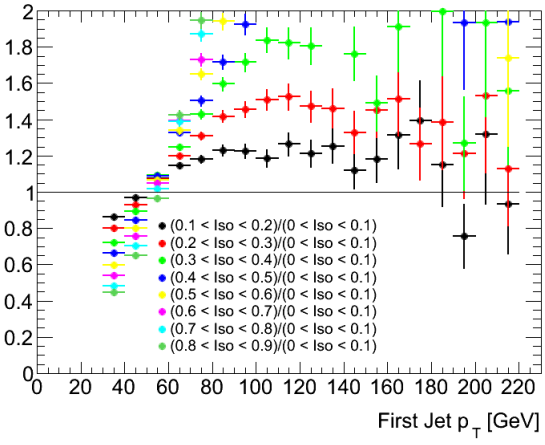
   where (i, j) is a bin in the 2-dimensional distribution of d_0 significance versus the relative track isolation, $N_{(i,j)}^{QCD,Data}$ is the number of events in that bin obtained subtracting EWK background and W signal to data



(a)



(b)



(c)

Figure 6.11: Ratios of the normalised leading jet p_T distribution in bin of isolation with respect to the the relative track isolation bin $[0.0; 0.1[$ using different requirements on d_0 significance: without any cuts applied on d_0 significance (a) and d_0 significance < 3 (b) and d_0 significance > 3 (c).

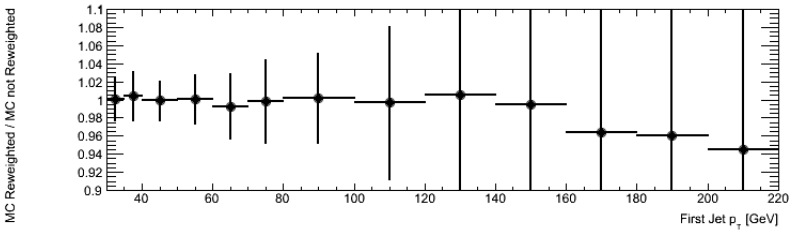


Figure 6.12: Ratio between PYTHIAB $b\bar{b} + c\bar{c}$ MC reweighted sample and not reweighted sample.

and $N_{(i,j)}^{QCD,MC}$ is the number of QCD events of PYTHIA $b\bar{b} + c\bar{c}$ MC in the same bin. Numerator and denominator in Eq. 6.2 are normalised to the integral of the distributions.
Results obtained for the leading jet p_T re-weighting PYTHIA MC are shown in Figure 6.12. The results show that the ratio of the leading jet p_T distribution of the re-weighted PYTHIA MC with respect to the non re-weighted one is compatible with 1. Therefore we conclude that no bias is due to the correlation between the selection cuts applied to extract the QCD control sample and the leading jet p_T. Similar results are obtained for other observables, e.g. leading jet rapidity, muon rapidity, H_T, S_T.

2. A second re-weighting procedure was done applying to all MCs the event weight, $f_{(i,j)}$, calculated as:

$$f_{(i,j)} = \frac{\frac{1}{\sum_{(i,j)} N_{(i,j)}^{Data}} N_{(i,j)}^{Data}}{\frac{1}{\sum_{(i,j)} N_{(i,j)}^{MC}} N_{(i,j)}^{MC}} \tag{6.3}$$

where (i, j) is the particular bin in the 2-dimensional distribution of d_0 significance versus the relative track isolation, $N_{(i,j)}^{Data}$ is the number of events in data in that bin, $N_{(i,j)}^{MC}$ is the number of events obtained adding all MC samples (W signal, EW backgrounds and QCD PYTHIAB $b\bar{b}+c\bar{c}$) in the same bin. Numerator and denominator in Eq. 6.3 are normalised to the integral of the distributions.
Results obtained for the leading jet p_T are shown in Figure 6.13. The results show again that the ratio of the leading jet p_T distribution of the re-weighted MCs with respect to the non re-weighted ones is compatible with 1. Therefore we can conclude no bias is due to the different modelling of the correlation between relative track isolation, d_0 significance and leading jet p_T in data and MCs. Similar results are obtained

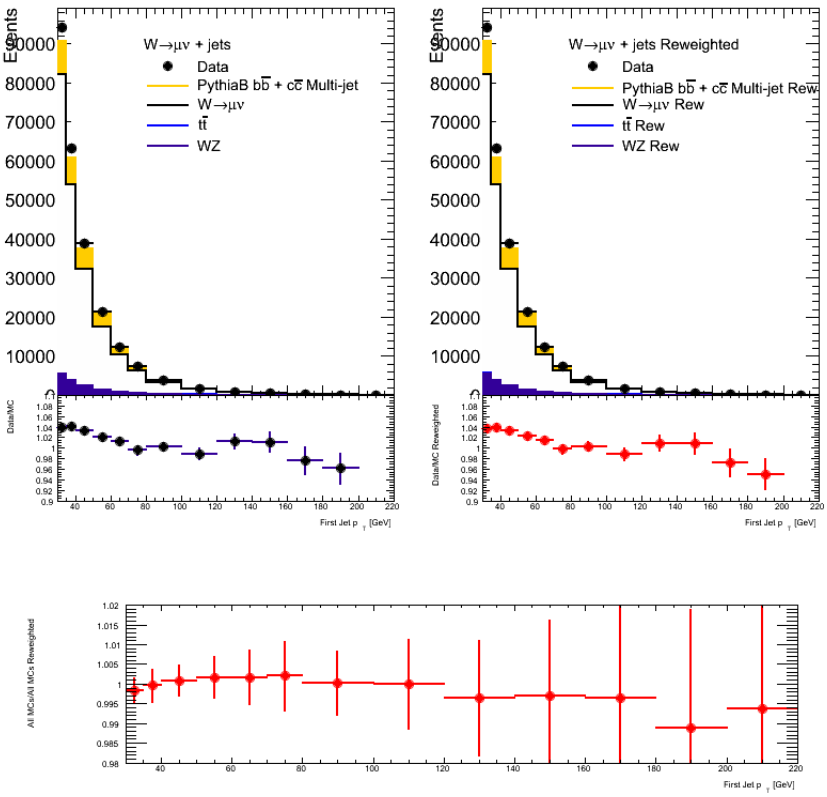


Figure 6.13: Detector-level plots using default MCs and reweighted MCs(top) and their ratio(bottom).

for other observables, e.g. leading jet rapidity, muon rapidity, H_T, S_T.

As no bias is visible in Figures 6.12 and 6.13, we can conclude that the correlation between relative track isolation, d_0 significance and other observables, such as leading jet p_T, leading jet rapidity, muon rapidity, H_T and S_T, is well reproduced by MCs and that the chosen control sample selection doesn't bias the observables' distribution in data.

Effect of pileup on QCD template

As shown in §6.6.2, the prescales of triggers used to select multijet background to $W \to e\nu$ event changed over the course of the 2011 run and pileup conditions changed dramatically across the 2011 run. All these effects influence the QCD background shape and rate in the electron channel due to pileup interactions. In order to take care of the time dependence of this ef-

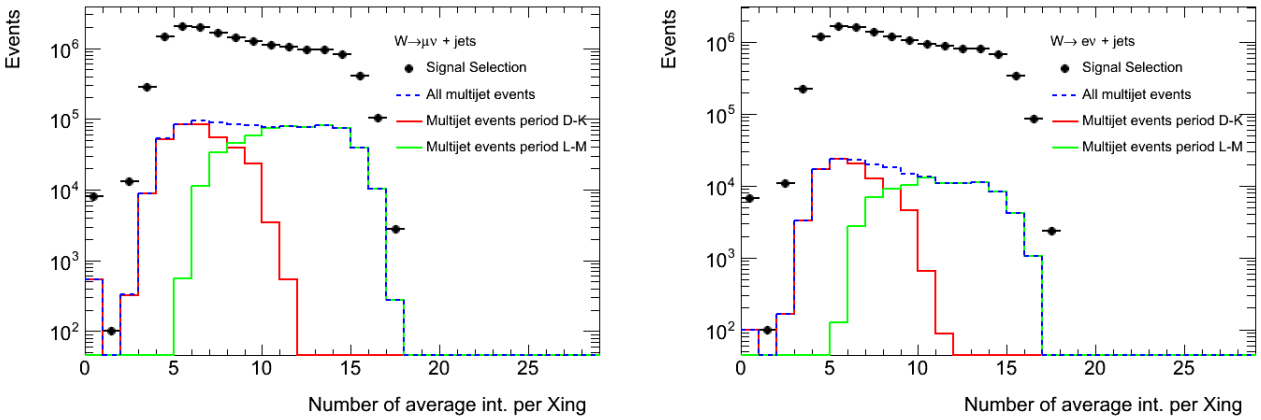
	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
Reference	$(1.83 \pm 0.03)\%$	$(11.64 \pm 0.05)\%$	$(10.48 \pm 0.09)\%$	$(9.98 \pm 0.17)\%$	$(7.2 \pm 0.3)\%$	$(4.8 \pm 0.3)\%$
$0 \leq \mu < 4$	$(1.05 \pm 0.02)\%$	$(5.3 \pm 0.3)\%$	$(5.7 \pm 0.6)\%$	$(3.4 \pm 0.7)\%$	$(5.9 \pm 3.5)\%$	$(2.3 \pm 1.2)\%$
$4 \leq \mu < 8$	$(1.08 \pm 0.03)\%$	$(8.60 \pm 0.07)\%$	$(7.7 \pm 0.1)\%$	$(7.9 \pm 0.3)\%$	$(5.6 \pm 0.5)\%$	$(4.2 \pm 0.6)\%$
$\mu \geq 8$	$(1.75 \pm 0.04)\%$	$(14.26 \pm 0.07)\%$	$(12.8 \pm 0.1)\%$	$(11.7 \pm 0.3)\%$	$(8.9 \pm 0.6)\%$	$(4.8 \pm 0.6)\%$
D-K	$(1.11 \pm 0.03)\%$	$(8.76 \pm 0.07)\%$	$(7.9 \pm 0.1)\%$	$(8.0 \pm 0.3)\%$	$(5.7 \pm 0.5)\%$	$(3.4 \pm 0.5)\%$
L-M	$(1.74 \pm 0.04)\%$	$(14.18 \pm 0.08)\%$	$(12.73 \pm 0.1)\%$	$(11.4 \pm 0.3)\%$	$(9.0 \pm 0.7)\%$	$(6.2 \pm 0.8)\%$

Table 6.8: QCD fractions under different conditions in pileup and in different periods. Only the statistical uncertainty are quoted.

fect, data were divided in two different periods: the early 2011 data (periods D-K) and the later periods (L and M). Thus, the fit of QCD background in the electron channel is performed in the two different data periods templates and then combined into a single QCD estimation.
In the muon channel, since the same trigger is used for signal and background selection, it is not needed a time dependent fit. The background template reflects the same time dependence of the pileup contribution as for the signal template. However, the following tests are performed to prove that the same effect seen in the electron channel is also seen in the muon channel, i.e. a significant increase of QCD multijet background in data periods with high pileup. For these tests the dataset is split into samples with low ($0 \leq \mu < 4$), medium ($4 \leq \mu < 8$) and high ($\mu \geq 8$) pileup conditions and in different data periods D-K and L-M. Table 6.8 shows the QCD multijet fraction for different jet multiplicities in these samples with different pileup conditions. The reference value is the one obtained using the standard procedure described in §6.6.1. The results show that QCD fractions increase when pileup increases. Figures 6.14(a) and 6.14(b) show the distribution of the average number of interaction per bunch crossing ($<\mu>$) in the electron and muon channels for different data periods. For the electron channel the distribution of $<\mu>$ for the signal sample selection, i.e. obtained with the single unprescaled electron trigger, is also shown.

6.6.4 Fit closure tests in the QCD background fit

As seen in §6.6.1, `TFractionFitter` is the class used to perform the QCD background fit and determine the number of multijet events. To test for any possible biases in the fit procedure itself, pseudo-data sets were used for closure tests.
In the toy fits, a functional form was chosen to describe the E_T^{miss} for signal and multijet templates. The functional form was then used to generate templates and thousands of pseudo-data sets. The statistics of the signal



(a) Muon channel (b) Electron channel

Figure 6.14: Distribution of the number of interaction per crossing for multijet events in muon(a) and electron(b) channel for the whole dataset and split in period D-K and L-M. The distributions of the number of interaction per crossing for the signal selection (black dots) are also shown.

and the QCD templates match that from the samples used in the nominal fit procedure on data.
For each pseudo-data set fit, the pull value is calculated as:

$$pull_i = \frac{N_{exp} - N^i_{obs}}{Err_{N^i_{obs}}} \tag{6.4}$$

where N_{exp} is the number of expected multijet events and N^i_{obs} is the number of multijet events obtained by the i-th fit with its error $Err_{N^i_{obs}}$.
Table 6.9 shows the parameters of the pull distributions for electron (divided in periods D-K and L-M) and muon channels from 0 to at least 5-jet multiplicities. The mean of the pull distribution shows any potential bias due to the fit method. The bias is calculated as the difference between the expected and the observed number of events with respect to the expected ones. As seen in the Table 6.9, the bias is compatible with zero for both channels.
Table 6.9 shows also the values of the sigma of the pull distribution, which tests whether or not the statistical uncertainties returned by the fit are indeed statistical. Since it is close to one, the statistical errors returned by `TFractionFitter` can be considered valid.

	Muon Channel		Electron Channel			
			Period D-K		Period L-M	
Jet Multiplicities	bias (%)	pull sigma	bias (%)	pull sigma	bias (%)	pull sigma
$N_{jets} = 0$	0.004 ± 0.009	0.958 ± 0.006	0.003 ± 0.007	1.03 ± 0.01	0.001 ± 0.007	1.048 ± 0.008
$N_{jets} = 1$	0.000 ± 0.0005	1.029 ± 0.007	0.002 ± 0.011	1.025 ± 0.007	0.008 ± 0.010	1.089 ± 0.008
$N_{jets} = 2$	0.007 ± 0.013	1.034 ± 0.007	0.01 ± 0.02	1.036 ± 0.007	0.00 ± 0.02	1.035 ± 0.007
$N_{jets} = 3$	0.01 ± 0.03	1.036 ± 0.007	0.00 ± 0.05	0.997 ± 0.007	0.00 ± 0.04	1.02 ± 0.01
$N_{jets} = 4$	0.03 ± 0.08	1.016 ± 0.007	0.00 ± 0.12	1.015 ± 0.007	0.04 ± 0.09	1.012 ± 0.007
$N_{jets} \geq 5$	0.0 ± 0.02	0.982 ± 0.007	0.0 ± 0.2	1.003 ± 0.007	0.0 ± 0.2	1.044 ± 0.007

Table 6.9: Parameters of the pull distributions for electron and muon channel from 0 to at least 5-jet multiplicities.

6.7 Data-driven top background

6.7.1 Template selection

Understanding the top background for jet multiplicities of four or more jets is an important aspect of this analysis since the background is large. The background from semileptonic top decays is 30% of total number of events with four associated jets and 80% of the total number of events with six associated jets. To determine the top background a data-driven method is used in order to reduce the background uncertainty of MC. Indeed the estimation of the expected top background using MC results in large systematic uncertainty, especially because of uncertainty on the top production cross section and on the theoretical prediction of the kinematic distributions for top events. For instance in the previous ATLAS measurement of W+jets events properties with 35 pb^{-1} of integrated luminosity [92], the uncertainty associated to the top prediction was estimated to be almost 40% of the total uncertainty for events with four associated jets.

To estimate the shape of the top background in data, we can exploit the decay of the top quark. Since the top quark has a probability of 99.8% to decay via $t \to Wb$, events from top pair production are characterize by two b-quarks (see the Feynman diagram in Figure 6.15). By requiring at least one b-jet in the event in data, we can select a background sample that consists largely of top events and models well the expected differential distributions for the top production background. To select b-jets, the MV1 tagger algorithm is used (see §3.6). Several different working points were considered: a tight working point with 60% efficiency, a medium working point with 70% efficiency and a loose working point with 75% efficiency. The medium working point with only a single b-tag gives the best $t\bar{t}$ purity with a tolerable contamination

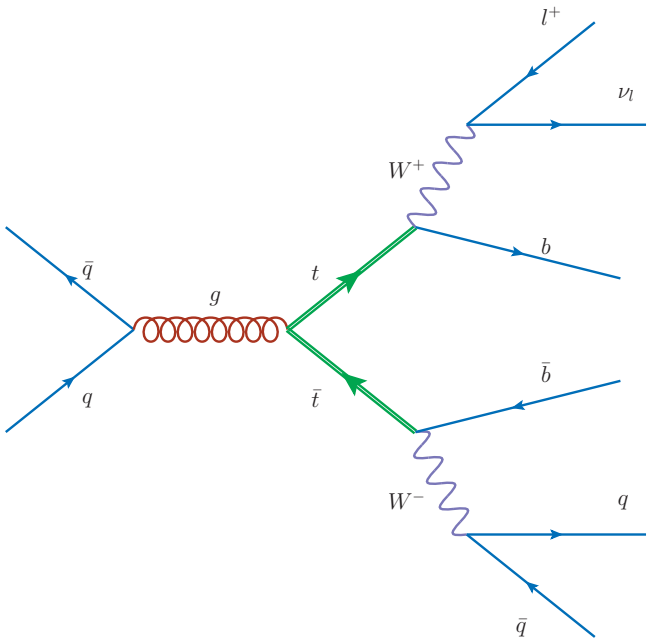


Figure 6.15: Feynman diagram for $t\bar{t}$ production and decay via $t \to Wb$.

from W events. The requirement of at least one b-jet or two b-jets was also considered. Requiring two b-tagged jets in the event resulted in distortions in the kinematic distributions for the control sample, therefore the single b-tag requirement with a medium working point is the most optimal for selecting the control sample. Table 6.10 summaries the purity for the top control sample as well as the contamination in the control sample from W events for the single b-tag requirement.

6.7.2 b-tagging corrections and contamination subtraction

Any potential bias in the b-tagged top sample compared to the inclusive (no b-tagging requirement) top sample was evaluated using two different MCs for $t\bar{t}$: POWHEG and ALPGEN. In Figures 6.16 - 6.19 the ratio of the number of events with no b-tagging requirement applied with respect to the number of events with single b-tagging applied is shown as a function of several observables. The ratios in these figures show the bias of the b-tagging on observables to be measured. Figures 6.16(a) and 6.16(b) show that the bias as a function of the rapidity of the leading jet is roughly 30%. This is due

Loose b-tag	**Number of jets**		
	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$
$t\bar{t}$ fraction	56.7%	75.3%	83.6%
W fraction	16.8%	9.5%	6.4%
Medium b-tag	**Number of jets**		
	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$
$t\bar{t}$ fraction	61.3%	78.6%	85.8%
W fraction	12.8%	7.0%	4.6%
Tight b-tag	**Number of jets**		
	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$
$t\bar{t}$ fraction	66.6%	82.2%	88.2%
W fraction	8.6%	4.6%	3.1%

Table 6.10: Summary of fractions of $t\bar{t}$ and W events in the single b-tagged (b-jet ≥ 1) control sample for different working points.

to the lack of b-tagging beyond the tracker acceptance of $|\eta| < 2.5$. Figures 6.16(c) and 6.16(d) show the ratio as a function of the leading jet p_T for 3 and 4-jet events. The overall bias on the leading jet p_T distribution is $< 20\%$ and considering the ratio as a function of 2^{nd}-4^{th} leading jet it is possible to conclude that the overall bias on the p_T distribution for all jets to be $< 10\%$. In all the distributions shown in Figures 6.16 - 6.19, the two MC samples, POWHEG and ALPGEN show very similar ratios, indicating that the bias is well reproduced by MCs.

For the final kinematic distributions, a correction factor for the b-tagging-induced bias is derived using the ALPGEN $t\bar{t}$ MC. These correction factors are shown in Figures 6.16 - 6.19 by the light blue bands. For the corrections, the numerator and denominator of this inclusive to b-tagged ratio have each been normalized to unity. This is done so that correction factors when applied to the kinematic distributions of the control sample, do not change the normalization of the control sample. The binning for the corrections is chosen so that the statistical uncertainty for the correction term is less than 25%. The uncertainty on the correction factor is taken as the statistical uncertainty from the MC unless the difference in the correction factor between ALPGEN vs POWHEG $t\bar{t}$ samples is statistically significant. In this case, the difference between the two generators is used as the uncertainty.

As seen in Table 6.10 the b-tagged control sample is not purely $t\bar{t}$ events and non-negligible contamination from W signal events is found. This contamination in the control sample is more significant in the lower jet multiplicities ($\sim 12\%$ for 3-jet events, becoming $\sim 5\%$ for 5-jet events) and is dominated

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
QCD b-tagged Fraction	$(6 \pm 3)\%$	$(4 \pm 4)\%$	$(1 \pm 2)\%$

Table 6.11: QCD fractions for fit with b-tagging requirement in electron channel. Only the statistical uncertainty are quoted.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
QCD b-tagged Fraction	$(9.1 \pm 0.3)\%$	$(4.4 \pm 0.4)\%$	$(2.3 \pm 0.3)\%$

Table 6.12: QCD fractions for fit with b-tagging requirement in muon channel. Only the statistical uncertainty are quoted.

by $W + c$, $W + cc$, $W + bb$ and QCD events as seen in Figures 6.20. The contribution from non-top events such as W-light, $W + c$, $W + cc$, $W + bb$, single top as well as other EWK processes are subtracted from the control sample using MC. The normalization for these processes are taken from the MC. The contribution from QCD event is also subtracted from the control sample. For the QCD distribution, the normalization is obtained using the same data-driven method as outlined in §6.6.1 but with b-tagging requirement applied in the fit. The results of the QCD fit with b-tagging for electron and muon channel are summarized in Tables 6.11 and 6.12.

6.7.3 Top background fit

The estimation of the normalization of the top control sample with respect to the non-b-tagged signal selection sample is challenging. As in the fit of the QCD background, it is needed to find a kinematic variable which distinguishes between top events and W events in order to derive the two templates. The template for top is derived using the b-tag control sample while the template for the W signal is taken from MC. These two templates are fit to the data using a maximum likelihood fit. The difficulty lies in the fact that since top events decay to a W, the kinematic distributions between W events and top events are often very similar and the maximum likelihood fit fails to determine the top background normalization well.
Several variables were studied including the lepton η, the top mass, the trans-

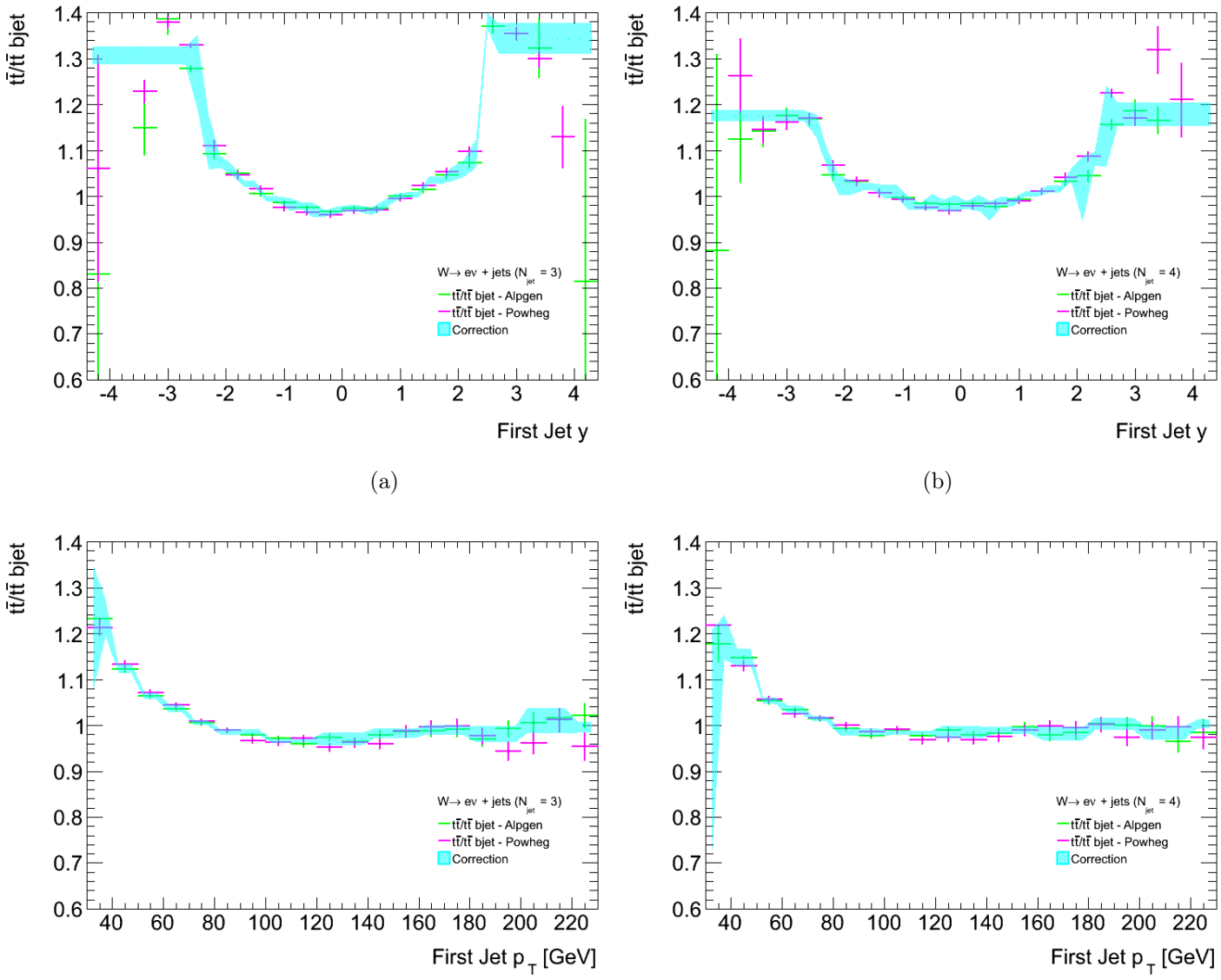


Figure 6.16: Ratio of the events passing the standard jet selection to events passing the b-tagging selection for ALPGEN and PowHeg $t\bar{t}$ MCs as a function of the rapidity of the leading jet for 3-jet(a) and 4-jet(b) events and of the leading jet p_T for 3-jet(c) and 4-jet(d) events. The correction factor for the b-tagging-induced bias is represented by the light blue bands. The uncertainty on the correction factor is taken as the statistical uncertainty from the MC or, if the difference in the correction factor between ALPGEN vs POWHEG $t\bar{t}$ samples is statistically significant, as the difference between the two generators.

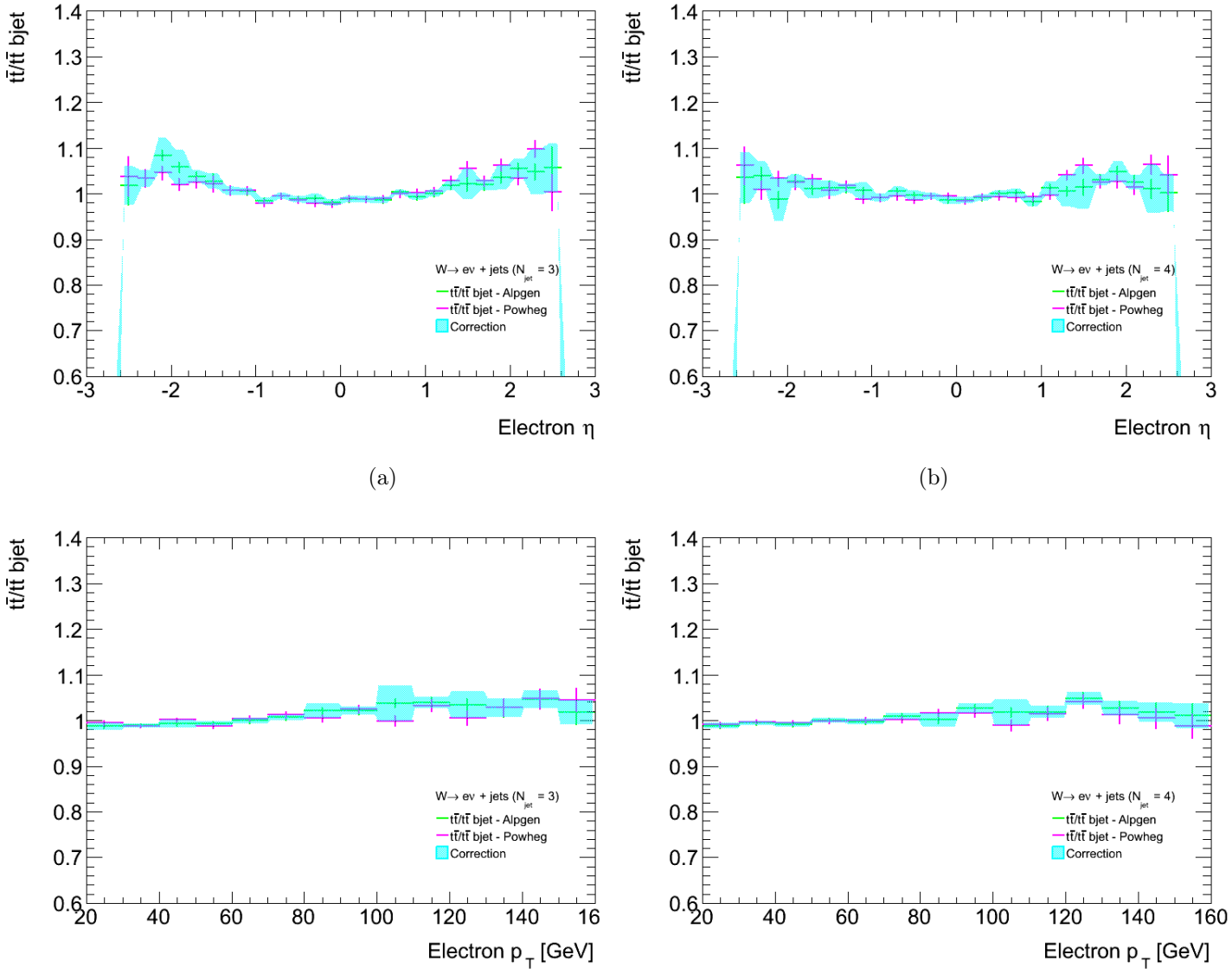


Figure 6.17: Ratio of the events passing the standard jet selection to events passing the b-tagging selection for ALPGEN and PowHeg $t\bar{t}$ MCs as a function of the electron η for 3-jet(a) and 4-jet(b) events and of the electron p_T for 3-jet(c) and 4-jet(d) events. The correction factor for the b-tagging-induced bias is represented by the light blue bands. The uncertainty on the correction factor is taken as the statistical uncertainty from the MC or, if the difference in the correction factor between ALPGEN vs POWHEG $t\bar{t}$ samples is statistically significant, as the difference between the two generators.

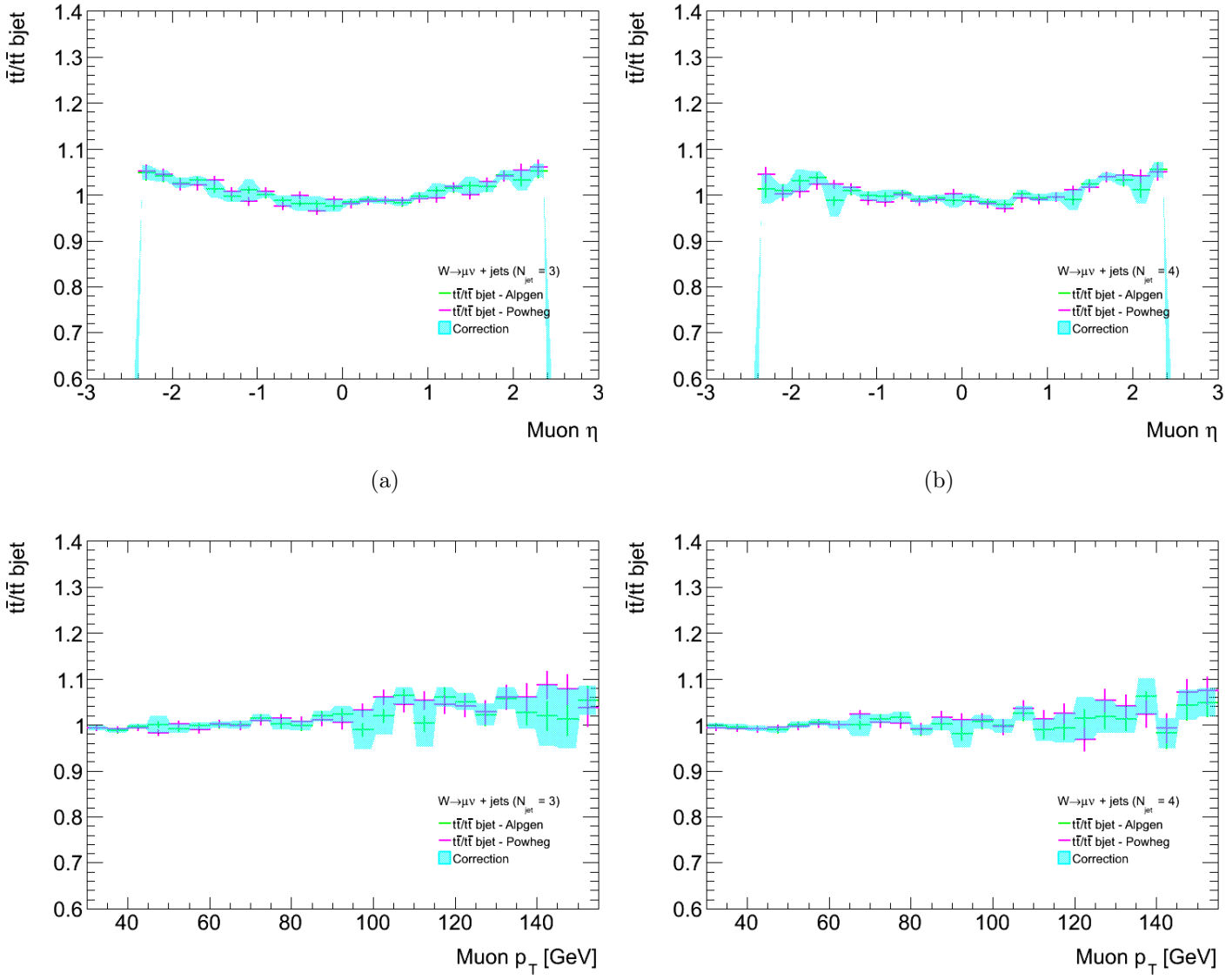
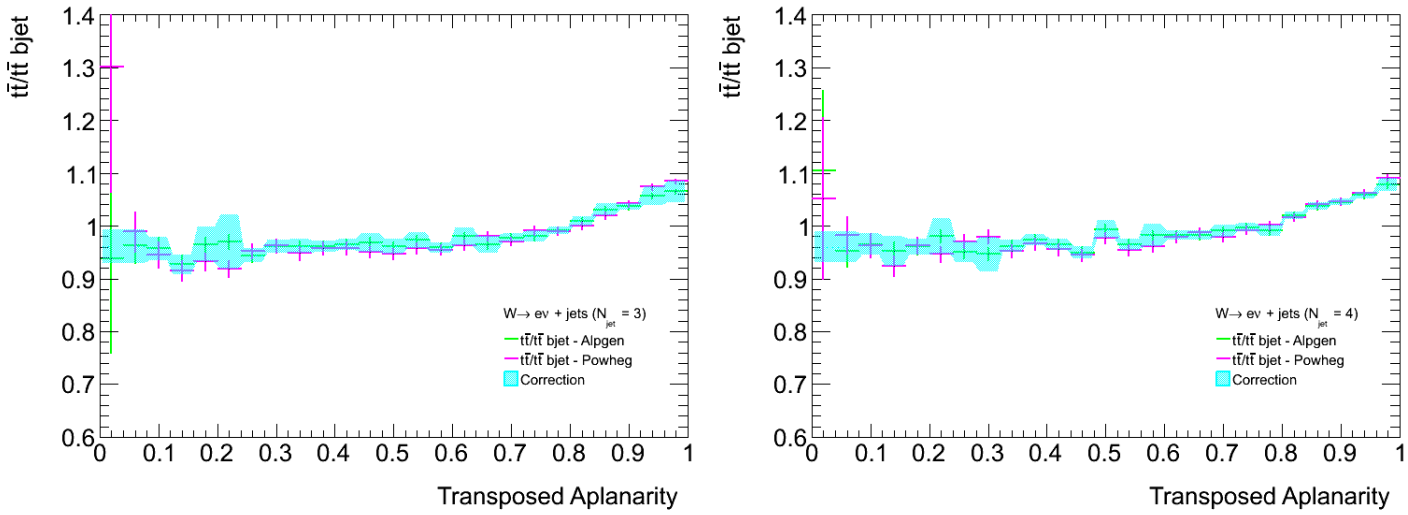
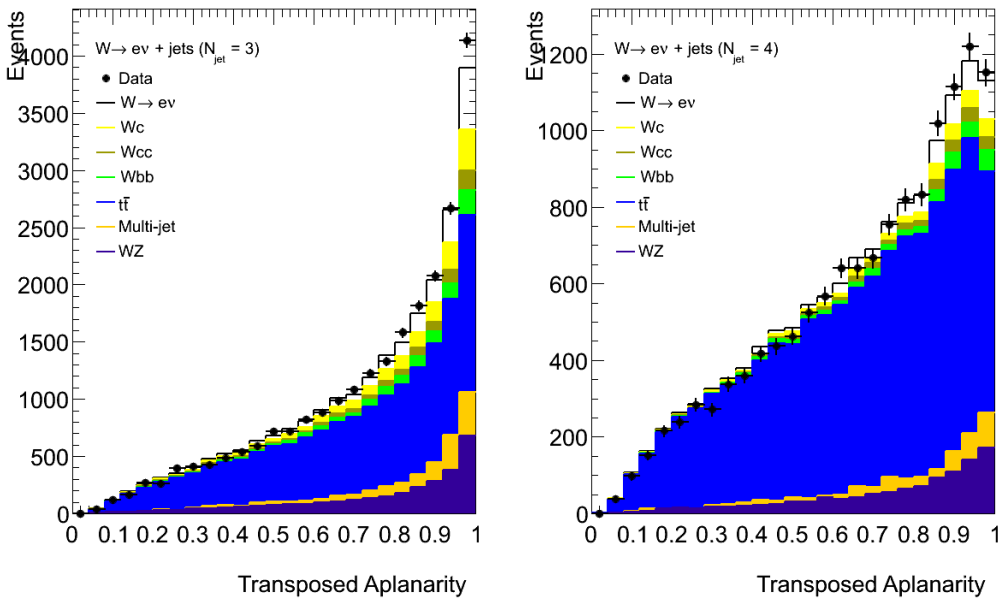


Figure 6.18: Ratio of the events passing the standard jet selection to events passing the b-tagging selection for ALPGEN and PowHeg $t\bar{t}$ MCs as a function of the muon η for 3-jet(a) and 4-jet(b) events and of the muon p_T for 3-jet(c) and 4-jet(d) events. The correction factor for the b-tagging-induced bias is represented by the light blue bands. The uncertainty on the correction factor is taken as the statistical uncertainty from the MC or, if the difference in the correction factor between ALPGEN vs POWHEG $t\bar{t}$ samples is statistically significant, as the difference between the two generators.



(a) (b)

Figure 6.19: Ratio of the events passing the standard jet selection to events passing the b-tagging selection for ALPGEN and PowHeg $t\bar{t}$ MCs as a function of the transposed aplanarity for 3-jet(a) and 4-jet(b) events. The correction factor for the b-tagging-induced bias is represented by the light blue bands. The uncertainty on the correction factor is taken as the statistical uncertainty from the MC or, if the difference in the correction factor between ALPGEN vs POWHEG $t\bar{t}$ samples is statistically significant, as the difference between the two generators.



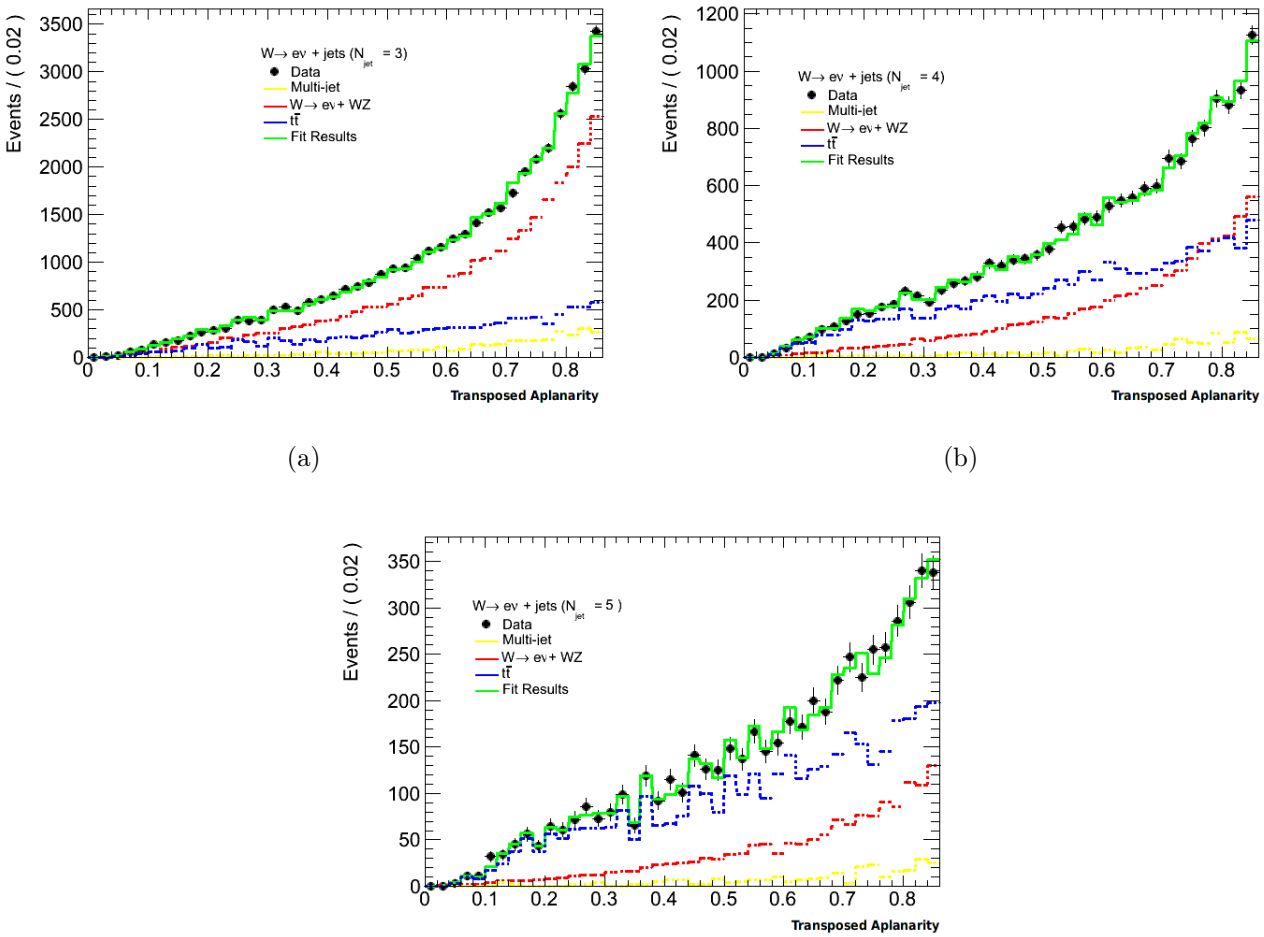
(a) (b)

Figure 6.20: The transposed aplanarity distribution for single medium b-tagged control sample for 3-jet(a) and 4-jet(b) events. The control sample from data is shown by black points and compared to the MC distributions. The MC for $t\bar{t}$ is ALPGEN. The Alpgen W-light, $W + cc$, $W + c$ and $W + bb$ MCs are also included.

verse sphericity[6], the transposed aplanarity[7] and E_T^{miss}, all of which demonstrated good separation between top and W. Some of them, like lepton η, are sensitive to the parton density function of the W signal MC and therefore were not considered further. The variable that has shown to be the most stable to variation of the W template is the transposed aplanarity. Therefore this variable is used for the maximum likelihood fit. For values of the transposed aplanarity greater than 0.9, the effect from the JES uncertainty is large and leads to unstable fits. Therefore the fit range is restricted to 0.0 to 0.85 to remove this poorly modelled region and mitigate the effect of the JES uncertainty on the fit. The b-tagging bias corrections are shown in Figure 6.19 and are applied to the top data-driven template before fitting.
The fit to data is performed using RooFit [93] instead of `TFractionFitter`. The reason why RooFit is preferred to `TFractionFitter` is that during the fit `TFractionFitter` allows the template to have Poissonian fluctuations within each bin and we are interested in decreasing this effect because of the correlation between the $t\bar{t}$ template and the data (discussed in §6.7.4). Since RooFit offers more flexibility from this point of view it is used to carry on the fit to normalize the $t\bar{t}$ background.
The signal template, which is composed of W signal and EWK backgrounds, including single top, is derived from MC. For events with less than 3-jets, the top template is derived from MC because it provides more statistics than the data-driven method. For events with at least 3 jets the top template is derived from data with the contamination from b-tagged W+jets signal and QCD events subtracted (see §6.7.2 for details). The non-b-tagged QCD template is also included in the fit but the normalization is fixed from the fit described in §6.6.1. Figures 6.21 and 6.22 show the fit templates and results for exclusive jet multiplicities for electron and muon channels. The distributions show good agreement to the data. Tables 6.13 and 6.14 show the top fraction for both channels as results of the fit.

[6]The transverse sphericity is a measure of the summed p_T^2 with respect to the event axis. It has values between 0 and 1, where a value of zero corresponds to "pencil-like" events and a value of one corresponds to an isotropic event. For the calculation of transverse sphericity the lepton, E_T^{miss}, and all jets passing the selection are used in the momentum tensor.

[7]The transposed aplanarity is defined as $\exp^{-8 \cdot aplanarity}$ where aplanarity is defined as $A = 3\lambda_3/2$ and λ_3 is the smallest eigenvalue of the normalised momentum tensor. Aplanarity measures the transverse momentum component out of the event plane: a planar event has $A \approx 0$ and an isotropic one $A \approx \frac{1}{2}$. For the aplanarity calculation, the lepton and all jets passing the selection are used in the momentum tensor.



(a) (b)

(c)

Figure 6.21: Templates and results of the data-driven top fits for 3-, 4- and 5-jets events for the electron channel.

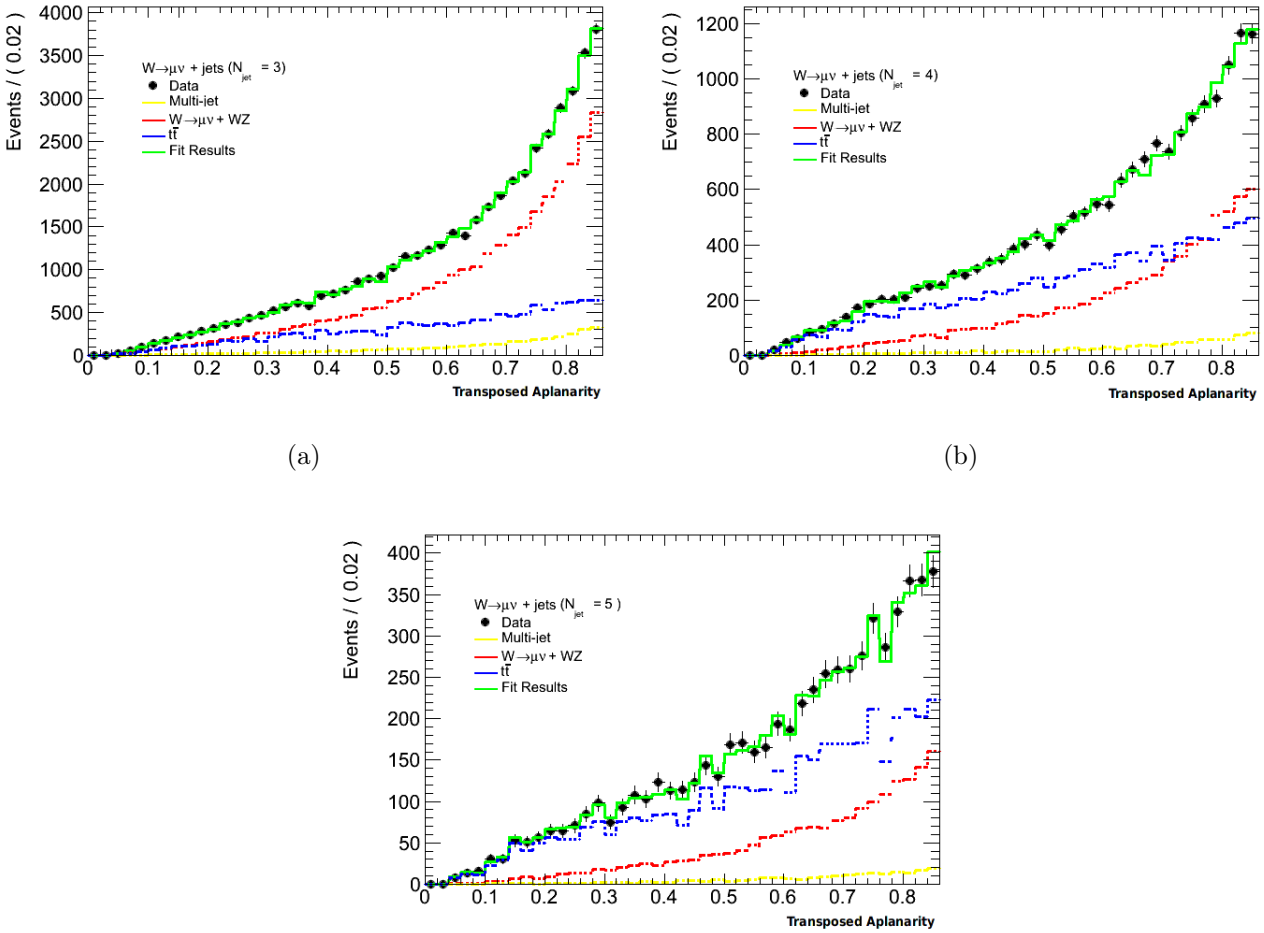


Figure 6.22: Templates and results of the data-driven top fits for 3-, 4- and 5-jets events for the muon channel.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$	$N_{jets} = 6$	$N_{jets} = 7$
$t\bar{t}$ fraction	$(25.4 \pm 1.2)\%$	$(58 \pm 2)\%$	$(71 \pm 4)\%$	$(87 \pm 9)\%$	$(76 \pm 8)\%$

Table 6.13: $t\bar{t}$ fraction (ratio between the estimated number of $t\bar{t}$ events over the total number of events observed in data) versus jet multiplicity in electron channel. Only the statistical uncertainty are quoted.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$	$N_{jets} = 6$	$N_{jets} = 7$
$t\bar{t}$ fraction	$(28.1 \pm 1.2)\%$	$(57 \pm 2)\%$	$(69 \pm 4)\%$	$(73 \pm 7)\%$	$(81 \pm 6)\%$

Table 6.14: $t\bar{t}$ fraction (ratio between the estimated number of $t\bar{t}$ events over the total number of events observed in data) versus jet multiplicity in muon channel. Only the statistical uncertainty are quoted.

6.7.4 Statistical correlation between $t\bar{t}$ template and data

It must be noticed that since the b-tagged sample is a sub-set of the signal sample, top template is correlated to the data in the signal region and therefore the statistical uncertainties returned by RooFit are underestimated. This correlation is partially broken through the application of the b-tagging corrections factors and the contamination subtraction. In addition the top template is only a partial sub-set of the full signal region (b-tagging acceptance $|\eta| < 2.5$). To check the statistical uncertainties, a toy MC procedure is performed to evaluate the correlation between the $t\bar{t}$ template and data and also any additional bias in the fit procedure itself.
In the toy fits, a functional form was chosen to describe the transposed aplanarity for signal and top templates. The functional form was then used to generate templates and thousands of pseudo-data sets. The pseudo-data sets were generated as a mixture of correlated and uncorrelated top samples including one of the template used in the fits. The top template has the b-tagging corrections and contamination subtraction applied. For the signal template, the statistics of the template match that from the W Alpgen samples.
For each pseudo-data set fit, the pull value is calculated as in Eq. 6.4:

$$pull_i = \frac{N_{exp} - N^i_{obs}}{Err_{N^i_{obs}}} \tag{6.5}$$

where N_{exp} is the number of expected $t\bar{t}$ events and N^i_{obs} is the number of $t\bar{t}$ events obtained by the i-th fit with its error $Err_{N^i_{obs}}$.
Table 6.15 shows the parameters of the pull distributions for electron and muon channel from 3 to 7-jet multiplicities. The mean of the pull distribu-

	Muon Channel			Electron Channel		
Jet Multiplicities	bias (%)	pull sigma	stat. correction	bias (%)	pull sigma	stat. correction
$N_{jets} = 3$	0.56 ± 0.04	0.816 ± 0.005	1.15	0.61 ± 0.02	0.831 ± 0.006	1.14
$N_{jets} = 4$	0.25 ± 0.01	0.716 ± 0.005	1.25	0.16 ± 0.02	0.706 ± 0.005	1.24
$N_{jets} = 5$	0.81 ± 0.02	0.633 ± 0.004	1.34	1.02 ± 0.02	0.639 ± 0.004	1.33
$N_{jets} = 6$	1.68 ± 0.04	0.578 ± 0.004	1.40	1.89 ± 0.04	0.588 ± 0.004	1.39
$N_{jets} = 7$	7.00 ± 0.07	0.538 ± 0.004	1.40	5.28 ± 0.04	0.565 ± 0.004	1.40

Table 6.15: Parameters of the bias and pull distributions for electron and muon channel from 3 to 7-jet multiplicities, as well as the total uncertainty including also correlation between top templates and data.

tion shows how significant the bias is. The bias is calculated as the difference between the expected and the observed number of events with respect to the expected ones. The sigma of the pull distribution, that is less than 1, shows that RooFit underestimates the statistical uncertainty of about 20-50% due to the correlation between $t\bar{t}$ template and data. Thus the statistical uncertainty returned by the fit is increased by the same amount. Table 6.15 shows the total uncertainty to use as error for the top fraction.
Table 6.15 also shows that the bias of the data-driven top-background method is $< 2\%$ for all jet multiplicities except for 7-jet events (where it is $\sim 5-7\%$) where the JES and JER uncertainties, respectively $\sim 5\%$ and $\sim 20\%$, are also large. The bias of the data-driven top-background method is added as an additional uncertainty on the top background estimate and it will be propagated to the final cross-section measurement.

6.8 Systematic Uncertainties on $t\bar{t}$ and QCD background estimation

The systematic uncertainties on QCD and $t\bar{t}$ background fraction estimations are estimated by varying the selection requirements and redoing the fits.
For the fit of QCD background alternative templates are defined both in electron and muon channel: in the former changing requirements on calorimeter and track isolation and electron identification and in the latter using the `AntiDOWithIso_DOWindow` and `RestrAntiIso` templates investigated above. The variation of the templates selection changes the shape of the distribution to fit and, obviously, also the QCD fractions returned by the fit.

Because of the use of b-tagging in top background fit, it is needed to take into account model uncertainty for the b-tagged W signal subtracted for the data driven top template and uncertainty on the b-tagging bias correction factors.
Moreover, other common variations are performed in both fits, such as variation of the minimum and the maximum of the fit range, usage of alternative signal models (SHERPA instead of the default ALPGEN), variation of the W signal and EWK templates according to JES, JER, lepton energy scale and resolution, reconstruction and trigger scale factor uncertainties.
A fit is performed separately for each new template and for each source the envelope of the largest deviations from the nominal fit was taken as the systematic uncertainty. These uncertainties are propagated through the analysis chain to the final cross-section measurement.
A global description on the systematic uncertainties is shown in §7.

6.9 Correlation between QCD and top fits

As explained in §6.7.3, the data-driven QCD template is used in the top background fit and its normalization is fixed from the QCD fit method described in §6.6.1. In order to study the correlations between the top and QCD fits, iterative fits of top and QCD backgrounds are carried out in the following way.
The fit of QCD background uses MC as top template and its results are used to normalize the QCD in the top background fit, which uses data-driven top template. In the second iteration, the fit of QCD background is done using data-driven top template whose normalisation is fixed from the previous iteration. A further top background fit is then performed with the QCD background template from the second iteration of the QCD fit.
Tables 6.16 and 6.17 show the fractions of QCD background returned by the fits using MC and data-driven top templates with statistical uncertainties in muon and electron channels. These values are compared to the QCD fraction range variation due to JES and JER systematic uncertainties since they are the largest sources of systematic uncertainty. The Tables also show the QCD fraction obtained using MC as top template and scaling it accordingly to its 10% of theoretical normalisation uncertainty. Tables 6.18 and 6.19 show the fractions of top background returned by the first and the second iterations of top fits respectively for muon and electron channels.
We can see from Tables 6.18 and 6.19 that the top fraction is not correlated to the QCD fit for 4 jets or more. For 3 jets a small difference between the top fractions estimated in the first and the second iteration is seen, how-

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
First QCD Fit (Using MC as top template)	$(9.98 \pm 0.17)\%$	$(7.2 \pm 0.3)\%$	$(4.8 \pm 0.3)\%$
Second QCD Fit (Using data-driven top template)	$(10.30 \pm 0.17)\%$	$(7.5 \pm 0.3)\%$	$(5.0 \pm 0.3)\%$
Scaling top MC +10%	$(10.17 \pm 0.18)\%$	$(7.5 \pm 0.3)\%$	$(5.0 \pm 0.3)\%$
Scaling top MC -10%	$(9.79 \pm 0.17)\%$	$(6.9 \pm 0.3)\%$	$(4.6 \pm 0.3)\%$
JES uncertainty	10.12% – 10.55%	7.3% – 7.5%	4.9% – 5.1%
JER uncertainty	11.15%	8.2%	5.6%

Table 6.16: QCD fractions after the two fit iterations with statistical uncertatinties and JES and JER systematic uncertainties in muon channel. The second and the third rows show the QCD fraction for each jet multiplicity obtained using MC as top template and scaling it accordingly to its 10% of normalisation uncertainty. The last two rows show the range variation of the QCD fractions due to JES and JER uncertainties propagation.

ever it is small compared to others top background uncertainties and can be neglected. From Tables 6.16 and 6.17 we can see that the choice of top template doesn't make a significant difference to the QCD fractions since it is within both the variation of MC top shape uncertainties (JES, JER) and $t\bar{t}$ MC theoretical normalisation.
Thus for top and QCD background fractions the results from the first iteration are used for the rest of the analysis. In particular the data-driven top template provides better description of top distribution shapes in data.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} \geq 5$
First QCD Fit (Using MC as top template)	$(10 \pm 3)\%$	$(7 \pm 5)\%$	$(6 \pm 7)\%$
Second QCD Fit (Using data-driven top template)	$(14 \pm 4)\%$	$(11 \pm 7)\%$	$(9 \pm 10)\%$
Scaling top MC +10%	$(11 \pm 3)\%$	$(8 \pm 5)\%$	$(7 \pm 8)\%$
Scaling top MC -10%	$(10 \pm 3)\%$	$(7 \pm 5)\%$	$(5 \pm 7)\%$
JES uncertainty	13% – 14%	11% – 11%	9% – 9%
JER uncertainty	14%	12%	9%

Table 6.17: QCD fractions after the two fit iterations with statistical uncertatinties and JES and JER systematic uncertainties in electron channel.The second and the third rows show the QCD fraction for each jet multiplicity obtained using MC as top template and scaling it accordingly to its 10% of normalisation uncertainty. The last two rows show the range variation of the QCD fractions due to JES and JER uncertainties propagation.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$	$N_{jets} = 6$	$N_{jets} = 7$
First top Fit (Using top template from MC in QCD fit)	$(28.1 \pm 1.2)\%$	$(57 \pm 2)\%$	$(69 \pm 4)\%$	$(73 \pm 7)\%$	$(81 \pm 6)\%$
Second top Fit (Using data-driven top template in QCD fit)	$(27.7 \pm 1.2)\%$	$(57 \pm 2)\%$	$(69 \pm 4)\%$	$(73 \pm 7)\%$	$(81 \pm 6)\%$

Table 6.18: Fractions of $t\bar{t}$ after the two fit iterations with statistical uncertatinties in muon channel.

	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$	$N_{jets} = 6$	$N_{jets} = 7$
First top Fit (Using top template from MC in QCD fit)	$(25.4 \pm 1.2)\%$	$(58 \pm 2)\%$	$(71 \pm 4)\%$	$(87 \pm 9)\%$	$(76 \pm 8)\%$
Second top Fit (Using data-driven top template in QCD fit)	$(24.7 \pm 1.3)\%$	$(58 \pm 2)\%$	$(71 \pm 4)\%$	$(87 \pm 9)\%$	$(74 \pm 8)\%$

Table 6.19: Fractions of $t\bar{t}$ after the two fit iterations with statistical uncertatinties in electron channel.

7 Detector level systematic uncertainties

In this section an account of all the systematic uncertainties derived at detector level, i.e. without correcting for all the detector effects, is given. First common across channels systematics are detailed, followed by electron and muon specific systematics.

7.1 Introduction

Systematic uncertainties are estimated by applying shifts to MC predictions of signal samples and backgrounds. In the case of the data-driven backgrounds, the fit is performed again after the variation of the MC templates. A summary of the size of all these systematics is provided at the end of this section for the jet multiplicity spectrum at detector level. For Table 7.1, and the detector level plots shown in figures 7.4 and 7.6, all the up and down systematic variations have been symmetrized around the nominal values. Moreover all the systematics are evaluated at detector-level, *i.e* without correcting for all the detector effects (see §6.2).

7.2 Jet Energy Scale

The uncertainties on JES have been already described in §3.5.6. They can be divided into four groups:

- uncertainties coming from the various "in-situ" measurements used for calibration;
- uncertainties on pile-up corrections applied to jets;

- uncertainties due to the presence of jets close to the jet of interest;
- uncertainties accounting for different quark-gluon responses and compositions between the MC samples used for the jet calibration (*e.g.* di-jet and multijet) and the samples used in the analysis (*e.g.* W+jets).

A function of the jet kinematics and pile-up conditions with 14 parameters is used to take all these sources into account. These 14 components are:

- 9 components corresponding to:
  - uncertainty on the η-intercalibration of the jets;
  - uncertainty on single-hadron response at large p_T;
  - uncertainty associated with physics modelling.
- 2 components derived from the uncertainty on the terms making up the pile-up offset correction;
- 1 component derived from the uncertainty due to close-by jets;
- 2 components derived from the uncertainty on the flavour of the jets. These are separated into the uncertainty on the composition of the jets and the uncertainty on the response of different flavours of the jet.

This results in a total 14 parameters characterising the JES uncertainty which are propagated through the analysis chain and then combined in quadrature in the final result. The uncertainty resulting from the JES is the largest source of systematic uncertainty in each channel.

7.3 Jet Energy Resolution

The JER is derived by comparison of data and MC using two distinct "in-situ" methods. Jets are over-smeared 5 times using a different random seed each time and when the errors are combined an average is taken of the absolute uncertainty resulting from the smearing, and then symmetrised to form the final uncertainty band.

7.4 E_T^{miss}

The E_T^{miss} is found by combining the measurements of all objects in the event (muon, electron and jets) and including soft terms for calorimeter depositions which are not included in these physics objects. Any changes in the physics

objects are always propagated to the E_T^{miss} varying each term according to its resolution and scale. The uncertainties on the soft terms are assumed as uncorrelated to the physics object variations and combined in quadrature later on.

7.5 Background uncertainties

Background predictions taken from MC are normalised to the respective (N)NLO cross sections. The uncertainty on the normalisation is assigned as systematic for each background. The cross sections of the electroweak backgrounds have an uncertainty of 5%, stemming from scale and PDF uncertainties in the NNLO prediction. The uncertainty of the cross sections of the diboson processes is taken to be 5% for WW and ZZ and 7% for WZ. The uncertainty on the single top is taken 3.4% for the t-channel and 4% for the s-channel.
To propagate this effect through the measurement, each MC background normalisation is varied independently up and down by the uncertainty on the theoretical cross section. These are then combined quadratically when forming the background normalisation uncertainty band.
The uncertainties on multijet and $t\bar{t}$ have been discussed in §6.8. These are propagated separately through the analysis chain.

7.6 Electron energy and electron efficiency scale factors

Systematics on the electron energy scale, electron energy resolution, electron reconstruction and electron trigger efficiency are evaluated using systematic variations of several terms. The electron energy scaling is separated into terms corresponding to:

- a statistical component from tag-and-probe studies;
- a method component from tag-and-probe studies;
- a generator component from tag-and-probe studies,
- a material-budget component;
- a presampler component.

The electron energy resolution uncertainty is estimated by varying the smearing applied to MC. The total uncertainty is then assigned as the averaged and symmetrised combination of these two variations.
Up and down variations on the reconstruction, identification and trigger scale factors (SFs) are also taken into account. Reconstruction and identification scale factors uncertainties are separated into 7 components, an uncorrelated term and 6 correlated terms. All these uncertainties are propagated through the analysis.

7.7 Muon systematics

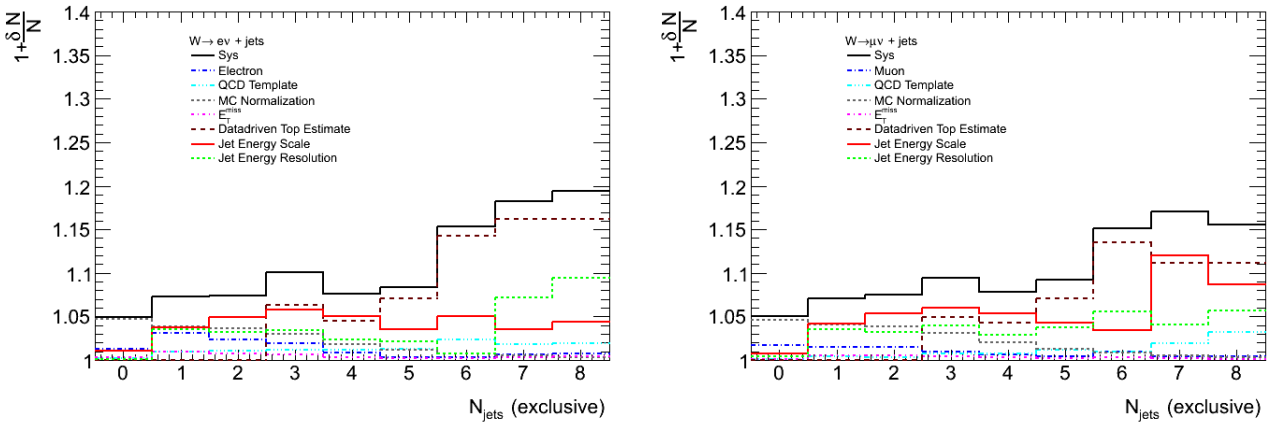
Systematics on the muon reconstruction, muon trigger efficiency, muon momentum scale and resolution are estimated. The following muon systematics are included in the analysis:

- Muon p_T resolution of the Inner Detector (ID) and of the Muon Spectrometer (MS);
- Muon p_T scale;
- Reconstruction efficiency;
- Trigger efficiency.

These systematics are very small compared to other uncertainties in the analysis.

7.8 Summary of systematic uncertainties

The effect of all the systematics considered at detector level are summarised in Table 7.1 which shows the systematic uncertainties for all these components in the exclusive jet multiplicity spectrum for all channels. This table and the plots from Figure 7.1 to 7.10 show the uncertainty on the predictions after the detector simulation: $\Delta N/N$, where N is the sum of MC signal, backgrounds estimated from MC and data-driven backgrounds.
Figures from 7.1 to 7.10 show the breakdowns of the systematics for both channels as a function of the number of jets, leading jet p_T, leading jet rapidity, scalar sum of all the jets in the events (S_T), scalar sum of all the objects in the events (H_T) and angular distance between the two leading jets in events ($\Delta R(\text{jet}_1, \text{jet}_2)$).
From Table 7.1 it is observed that in all bins, except for the 0 jet bin, JES



(a) (b)

Figure 7.1: Breakdown of electron (a) and muon (b) channels detector level systematics for the exclusive jet multiplicity spectrum.

and JER uncertainties are the dominant uncertainties in both channels: indeed they range from 5% for $N_{jet} = 1$ to 10% for $N_{jet} = 7$ and $N_{jet} = 8$. In the 0-jet bin the lepton systematics is dominant being 0.20% bigger than the sum in quadrature of JES and JER uncertainties. The lepton systematics then decrease with increasing jet multiplicity remaining lower that 1% for $N_{jet} \geq 4$.

For $N_{jet} \geq 3$ the data-driven top template contribution becomes more important than the jet and lepton terms. Indeed, when the data-driven top template is used, *i.e.* in more than 3-jet events, systematics which do not change the shape of the fitted distribution (such as lepton SFs) do not affect the fit of top template and as a result the data-driven top estimate is independent of these systematics. At high jet multiplicity, where the top is the dominant part of the detector-level fraction of events, such systematics are therefore very small leading to the observed decrease with increasing jet multiplicity.

Electron channel									
Systematics	$N_{jet} = 0$	$N_{jet} = 1$	$N_{jet} = 2$	$N_{jet} = 3$	$N_{jet} = 4$	$N_{jet} = 5$	$N_{jet} = 6$	$N_{jet} = 7$	$N_{jet} = 8$
Stat. Unc.	0.07%	0.10%	0.18%	0.37%	0.67%	1.16%	2.14%	4.17%	7.70%
JES	1.05%	3.77%	4.93%	5.79%	5.07%	3.50%	5.00%	3.55%	4.41%
JER	0.16%	3.57%	3.18%	3.47%	2.37%	2.15%	0.69%	7.20%	9.45%
JES + JER	1.06%	5.19%	5.87%	6.75%	6.09%	4.11%	5.04%	8.03%	10.43%
Electron	1.26%	3.08%	2.30%	1.98%	0.91%	0.30%	0.32%	0.69%	0.73%
E_T^{miss}	0.31%	0.96%	0.71%	0.64%	0.29%	0.26%	0.26%	0.34%	0.37%
QCD template	0.20%	1.00%	1.10%	1.16%	1.20%	1.26%	2.32%	1.83%	1.95%
$t\bar{t}$ template	0.00%	0.00%	0.00%	6.40%	4.56%	7.11%	14.25%	16.24%	16.35%
MC normalisation	4.68%	3.89%	3.68%	2.96%	1.87%	1.20%	0.70%	0.58%	0.36%
TOT	5.00%	7.36%	7.46%	10.03%	7.64%	8.38%	15.32%	18.23%	19.42%

Muon channel									
Systematics	$N_{jet} = 0$	$N_{jet} = 1$	$N_{jet} = 2$	$N_{jet} = 3$	$N_{jet} = 4$	$N_{jet} = 5$	$N_{jet} = 6$	$N_{jet} = 7$	$N_{jet} = 8$
Stat. Unc.	0.06%	0.09%	0.14%	0.31%	0.56%	1.03%	2.00%	4.17%	7.42%
JES	0.75%	4.19%	5.40%	6.03%	5.31%	4.34%	3.43%	12.04%	8.65%
JER	0.38%	3.57%	3.18%	3.94%	2.93%	3.77%	5.62%	4.11%	5.72%
JES + JER	0.84%	5.50%	6.27%	7.20%	6.06%	5.75%	6.58%	12.72%	10.37%
Muon	1.10%	1.51%	1.52%	1.01%	0.64%	0.44%	0.98%	0.43%	0.41%
E_T^{miss}	0.16%	0.54%	0.51%	0.47%	0.32%	0.26%	0.33%	0.17%	0.12%
QCD template	0.11%	0.41%	0.36%	0.71%	0.73%	1.22%	1.01%	1.96%	3.18%
$t\bar{t}$ template	0.00%	0.00%	0.00%	4.98%	4.31%	7.03%	13.56%	11.19%	11.71%
MC normalisation	4.63%	4.07%	3.90%	3.12%	2.01%	1.28%	0.85%	0.57%	0.29%
TOT	5.08%	6.99%	7.49%	9.34%	7.77%	9.23%	15.17%	17.07%	15.60%

Table 7.1: Detector level systematic uncertainties in $W \to e\nu$ and $W \to \mu\nu$ channels. The uncertainty is defined as the total uncertainty on the prediction using MC and data-driven backgrounds and MC signal.

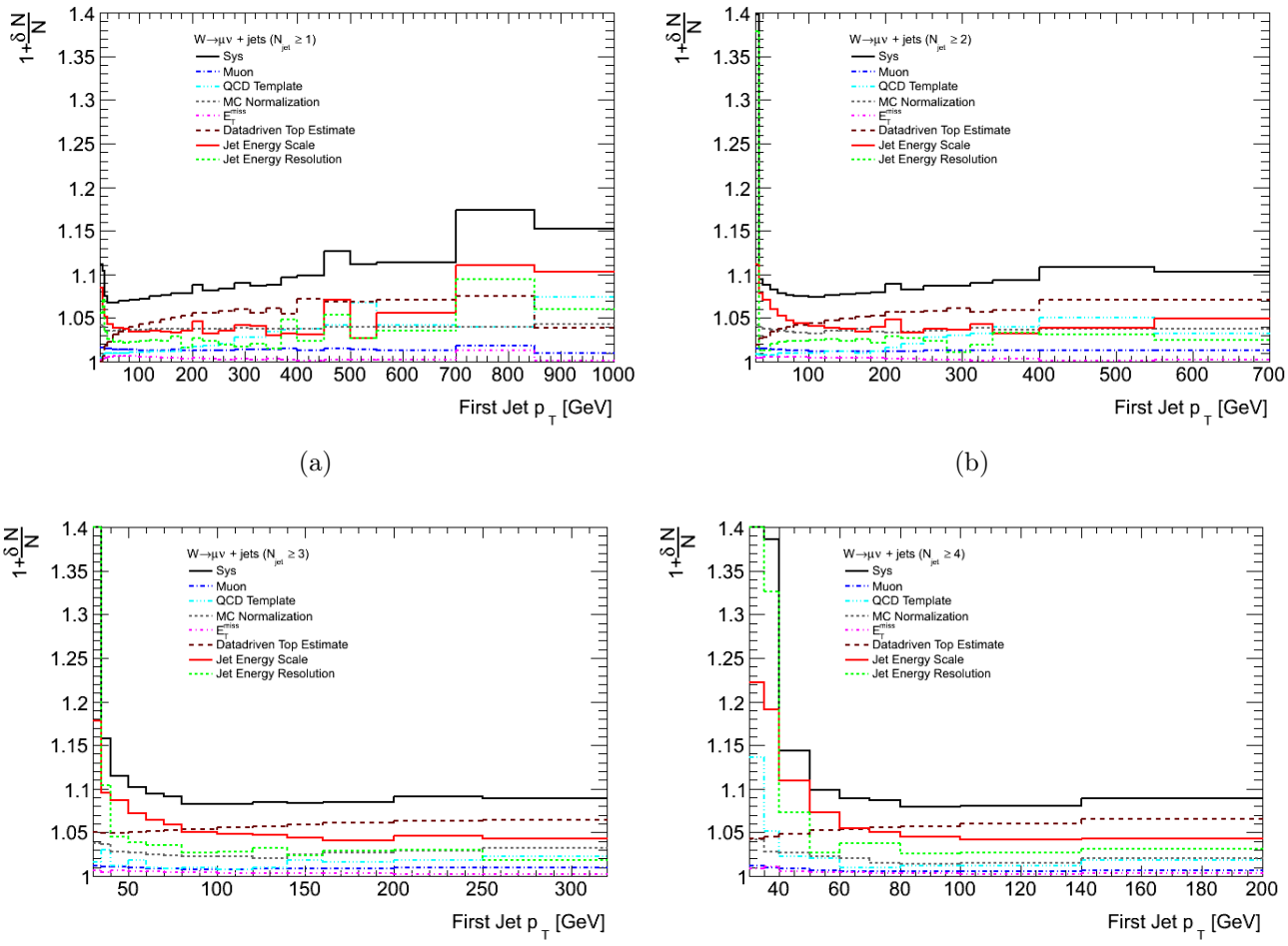


Figure 7.2: Breakdown of the muon channel detector level systematics as a function of the leading jet p_T for events with at least one jets (a) up to at least 4-jets event (d).

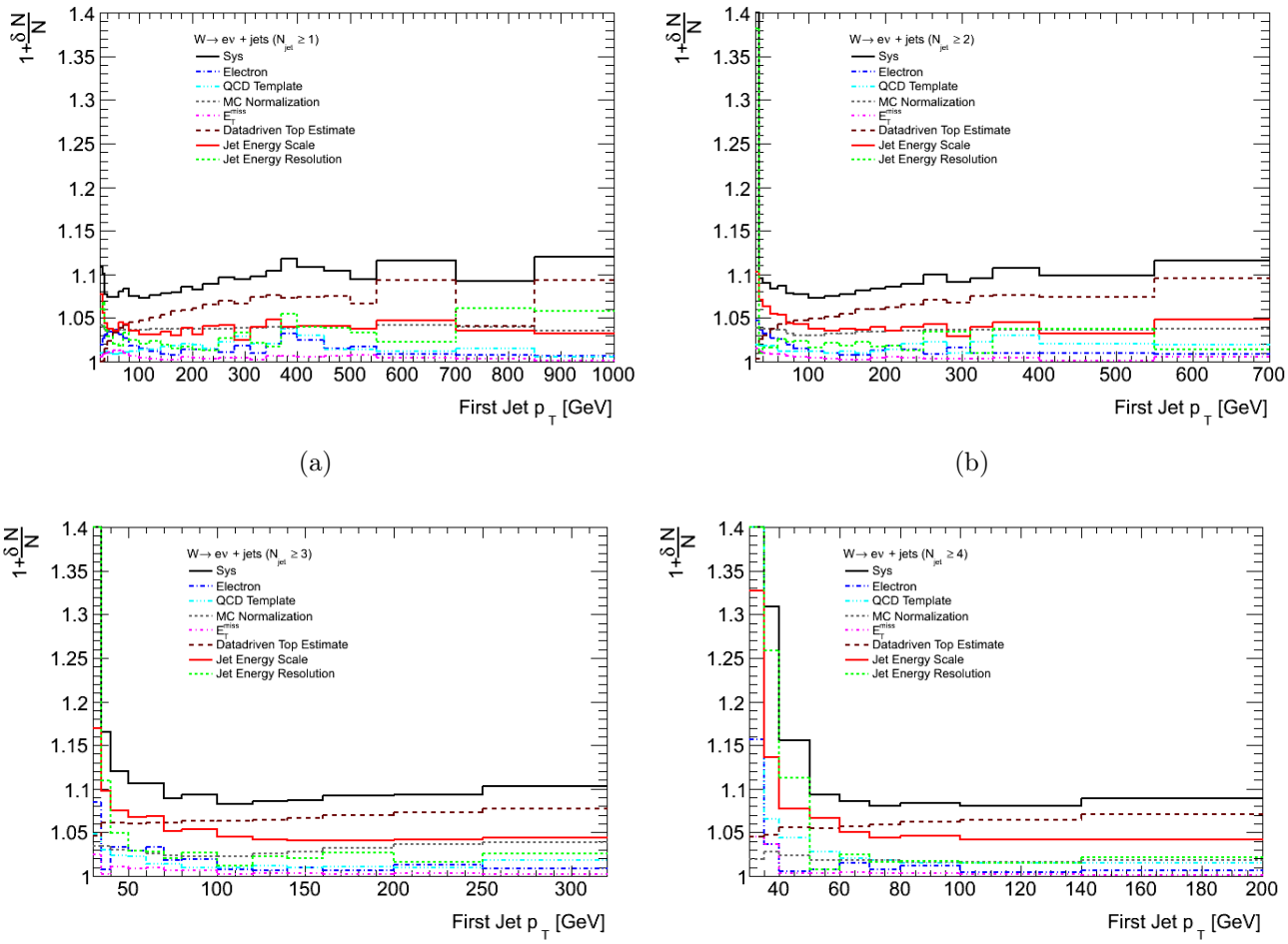


Figure 7.3: Breakdown of the electron channel detector level systematics as a function of the leading jet p_T for events with at least one jets (a) up to at least 4-jets event (d).

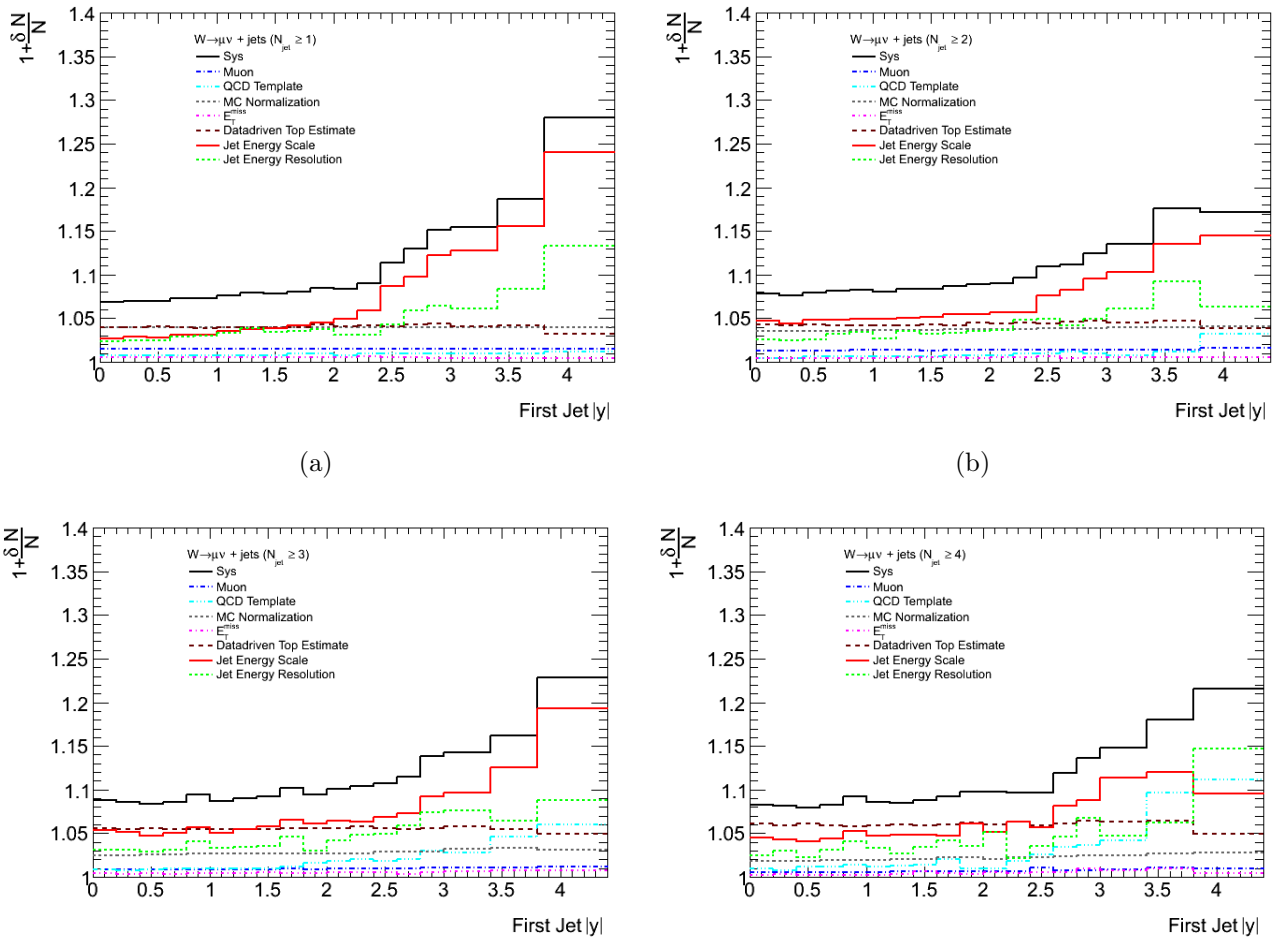


Figure 7.4: Breakdown of the muon channel detector level systematics as a function of the leading jet rapidity for events with at least one jets (a) up to at least 4-jets event (d).

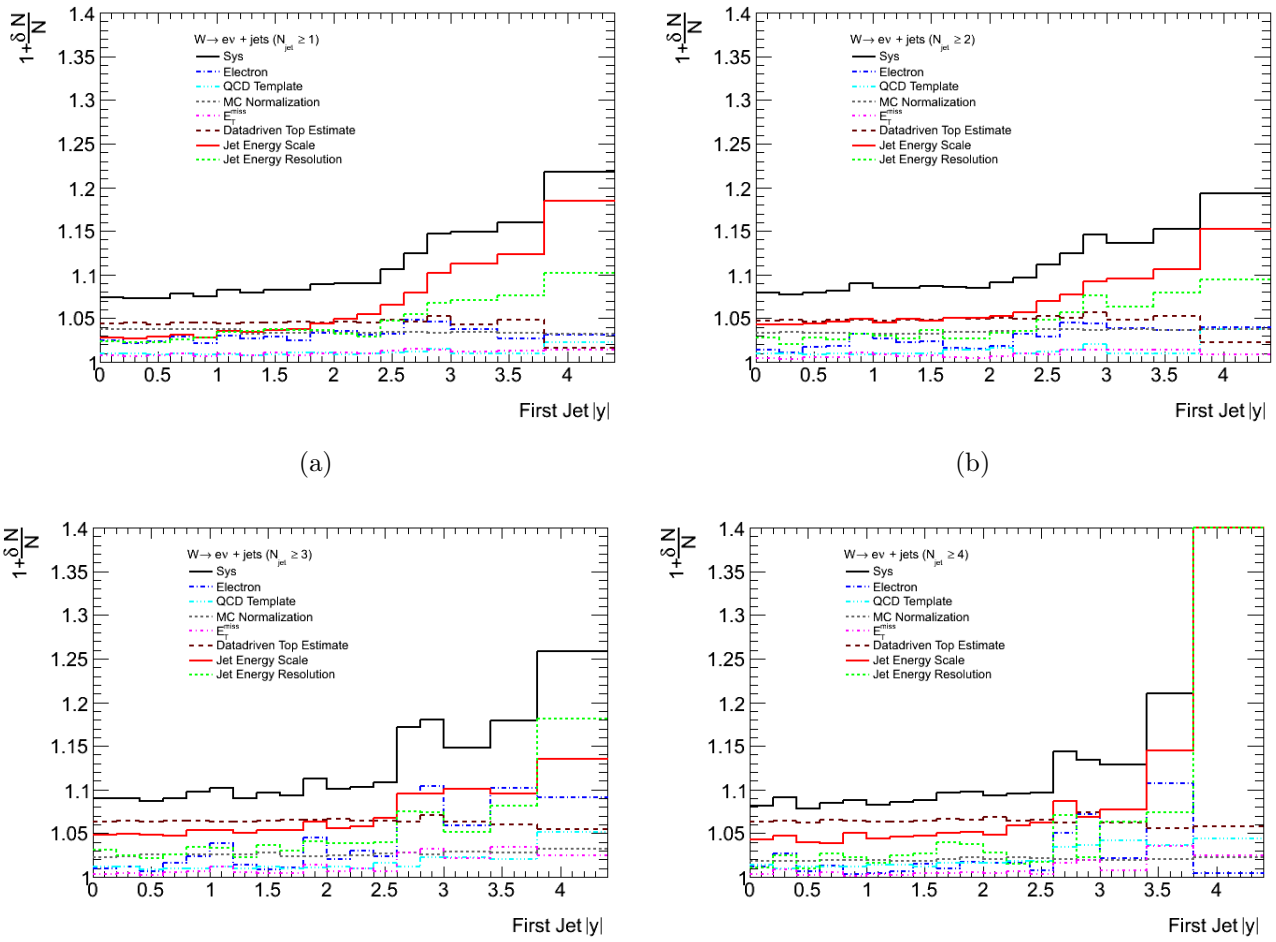


Figure 7.5: Breakdown of the electron channel detector level systematics as a function of the leading jet rapidity for events with at least one jets (a) up to at least 4-jets event (d).

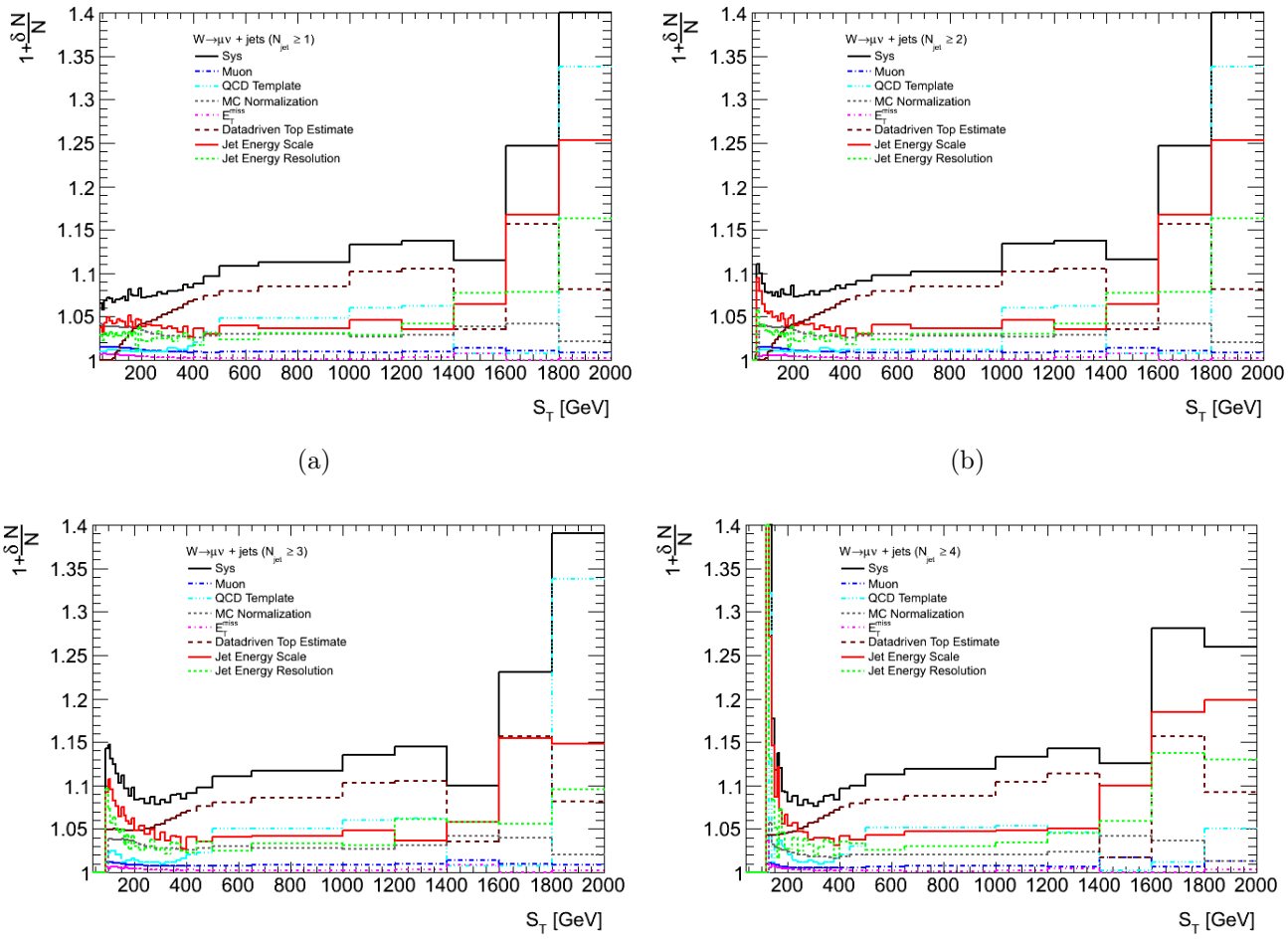


Figure 7.6: Breakdown of the muon channel detector level systematics as a function of the scalar sums of all the jets in events (S_T) for events with at least one jets (a) up to at least 4-jets event (d).

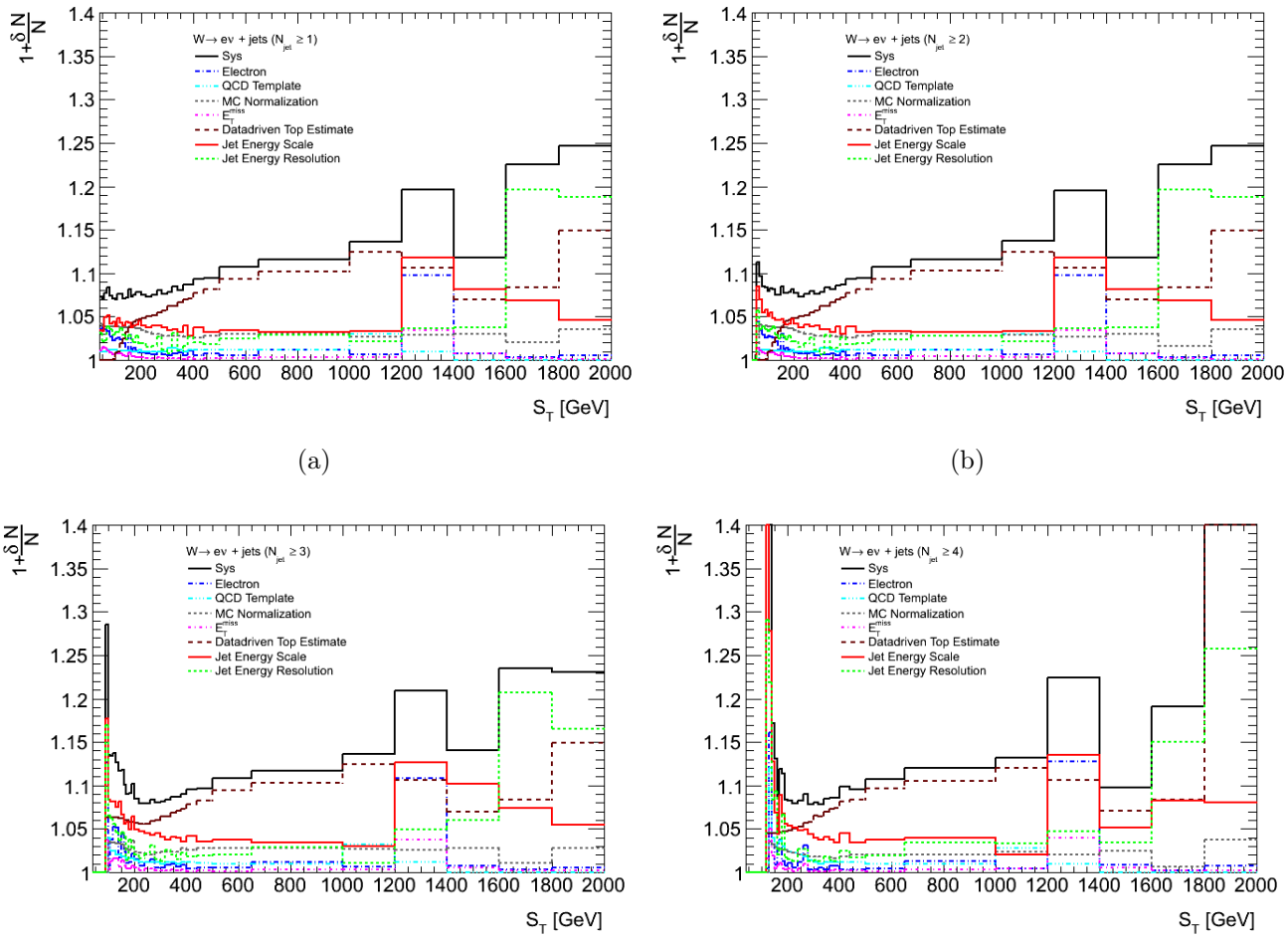


Figure 7.7: Breakdown of the electron channel detector level systematics as a function of the scalar sums of all the jets in events (S_T) for events with at least one jets (a) up to at least 4-jets event (d).

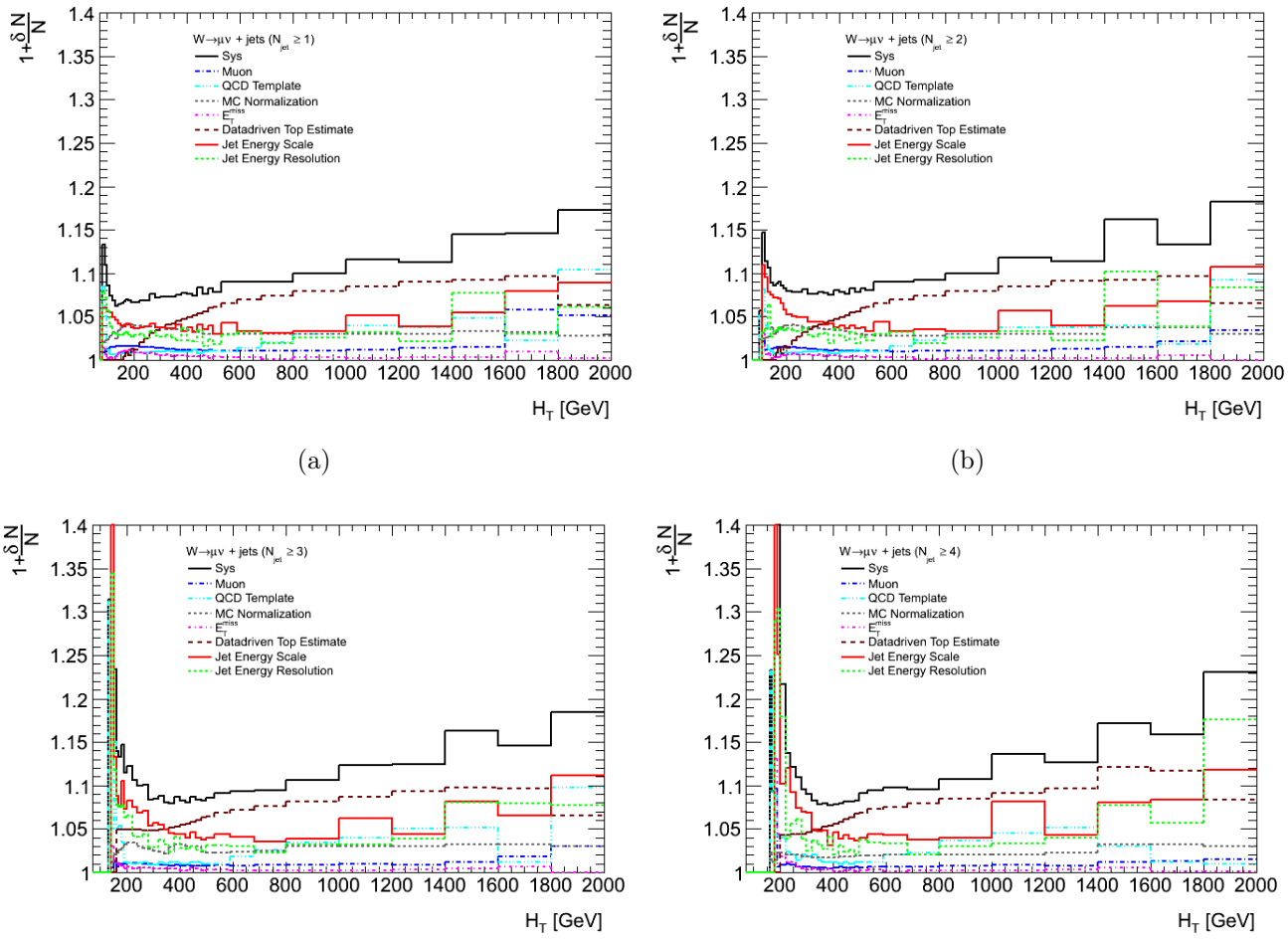


Figure 7.8: Breakdown of the muon channel detector level systematics as a function of the scalar sums of all the objects in events (H_T) for events with at least one jets (a) up to at least 4-jets event (d).

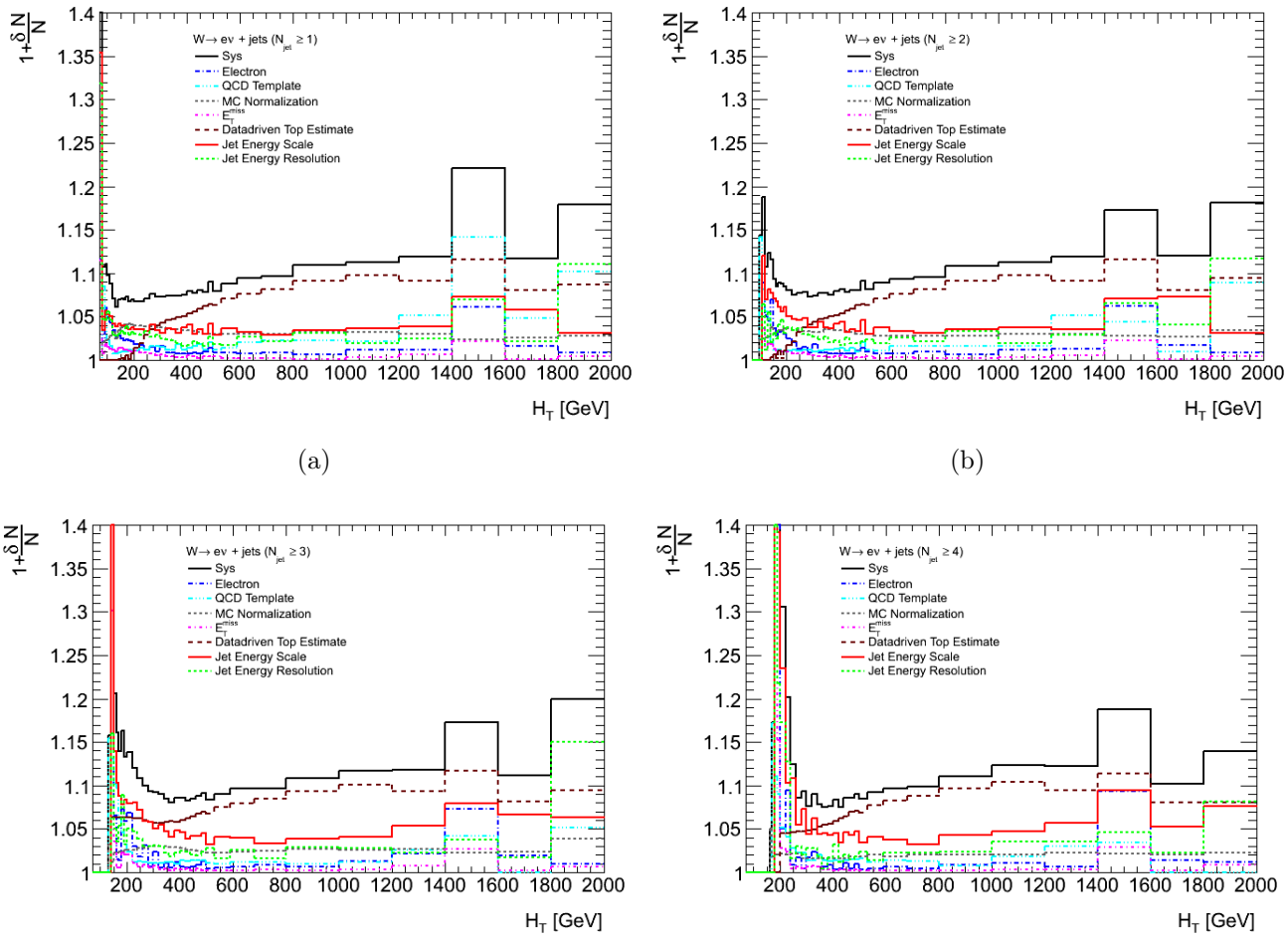
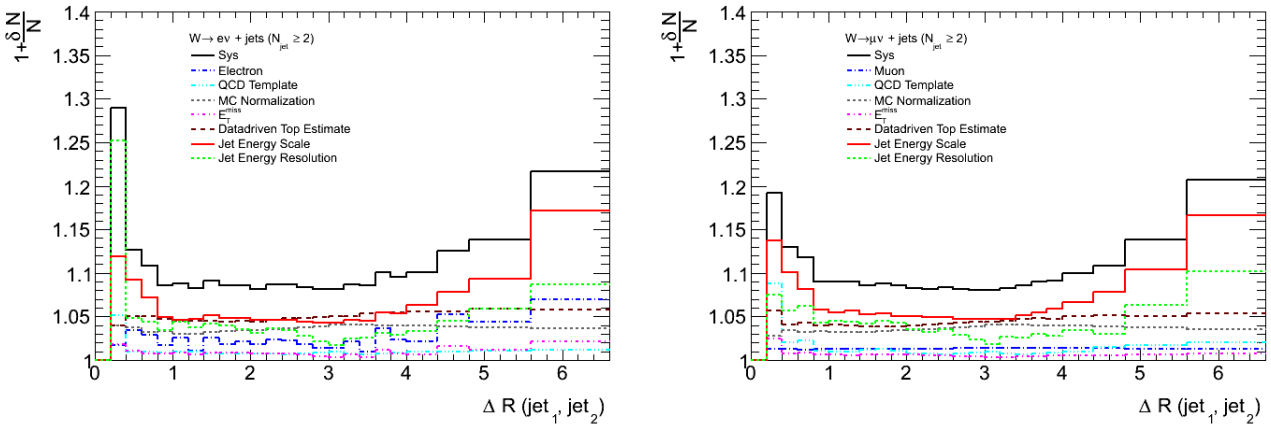


Figure 7.9: Breakdown of the electron channel detector level systematics as a function of the scalar sums of all the objects in events (H_T) for events with at least one jets (a) up to at least 4-jets event (d).



(a) (b)

Figure 7.10: Breakdown of electron (a) and muon (b) channels detector level systematics as a function of the angular distance between the two leading jets in events ($\Delta R(\text{jet}_1, \text{jet}_2)$).

8 Results

This chapter presents and discusses the results obtained comparing ALPGEN *and* SHERPA *theoretical predictions with data.*

8.1 Results

Figures from 8.1 to 8.10 show several detector level distributions: the number of jets, the leading jet p_T, the leading jet rapidity, the scalar sum of all the jets in the events (S_T), the scalar sum of all the objects in the events (H_T) and the angular distance between the two leading jets in events ($\Delta R(\text{jet}_1, \text{jet}_2)$). In the plots the shaded bands represent the uncertainty on the predicted distributions. Statistical uncertainties on the simulated distributions are included into the shaded bands. The observed distributions are shown with statistical uncertainties only. In the bottom part of the plots, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.
Figure 8.1 presents the comparison of the jet multiplicity with predictions at detector level for exclusive bins in the electron and muon channels. For both channels, a good agreement between ALPGEN and SHERPA with data is seen. In particular, while ALPGEN fluctuates within the experimental systematic uncertainty, SHERPA tends to overestimate the number of jets at high multiplicities. This disagreement could be due to the differences between the two MC generators: parton-jet matching, parton shower and multi parton interactions descriptions, PDF and scale choice. The expected and measured numbers of events for signal and backgrounds as a function of jet multiplicity in muon and electron channels are shown in Tables 8.1 and 8.2 and are found

to be in general agreement.
Figures 8.2 and 8.3 show the detector level distributions of the leading jet transverse momentum when there is at least one, two, three and four jets above 30 GeV in the events for muon and electron channels respectively. For both channels, despite some fluctuations, SHERPA and ALPGEN's predictions are in good agreement with the data. Larger differences between the predicted and the observed distributions at high p_T are shown for $N_{jets} \geq 1$ while the agreement improves for higher jet multiplicity, becoming good for $N_{jets} \geq 3$.
Figures 8.4 and 8.5 show the detector-level distributions of the leading jet rapidity. As seen in the figures, ALPGEN describes the data better than SHERPA at high rapidity in both channels. The convolution of the different PDFs in SHERPA and ALPGEN and parton shower effects might explain this, but further investigations would be needed to confirm it.
The scalar sum of all the jets in events (S_T), shown in Figures 8.6 and 8.7, is well modelled by ALPGEN up to 2 TeV while SHERPA presents large variation from the observed distribution in both channels.
Figures 8.8 and 8.9 show the H_T distributions which is the scalar sum of all jets, plus the lepton and E_T^{miss} for muon and electrons channels respectively. Similar to S_T distribution, it is better modelled by ALPGEN than by SHERPA. Both predictions consistently overestimate the data, indicating potential mis-modeling of the W transverse momentum.
For the $\Delta R(\text{jet}_1, \text{jet}_2)$ distribution, shown in Figure 8.10, both SHERPA and ALPGEN's predictions show discrepancies especially at large values of $\Delta R(\text{jet}_1, \text{jet}_2)$. ALPGEN especially underestimates the number of events with very large jet separation while SHERPA models the large rapidity intervals better but doesn't predict well the number of close-by jets.
In Figures 8.11(a) and 8.11(b) the ratio of the number of events of signal for adjacent jet multiplicities, $(N(W + (N_{jet} + 1) \text{ jets})/N(W + N_{jet} \text{ jets}), N_{jet} = 0{-}4)$ is shown. This observable allows to large cancellations of theoretical and experimental uncertainties. Then $(N(W+(N_{jet}+1) \text{ jets})/N(W+N_{jet} \text{ jets})$ is sensitive to α_s: the visible constant ratio (Berends scaling, see §5.1.1) starting from $N_{jet} = 1$ tests the applicability of pQCD. For both the electron and muon channel, there is good agreement between the theoretical predictions and the measured ratios.

	$N_{jet} = 0$	$N_{jet} = 1$	$N_{jet} = 2$	$N_{jet} = 3$
$W \to \mu\nu$	$(121 \pm 6) \cdot 10^5$	$(13.7 \pm 1.0) \cdot 10^5$	$(30 \pm 2) \cdot 10^4$	$(59 \pm 6) \cdot 10^3$
single top, $Z \to \tau\tau$, $W \to \tau\nu$, diboson, $Z \to \mu\mu$	$(72 \pm 4) \cdot 10^4$	$(111 \pm 8) \cdot 10^3$	$(33 \pm 3) \cdot 10^3$	$(93 \pm 9) \cdot 10^2$
$t\bar{t}$	$(38 \pm 2) \cdot 10$	$(34 \pm 2) \cdot 10^2$	$(121 \pm 9) \cdot 10^2$	$(18.9 \pm 1.0) \cdot 10^3$
QCD	$(24.2 \pm 1.2) \cdot 10^4$	$(20.2 \pm 1.4) \cdot 10^4$	$(42 \pm 3) \cdot 10^3$	$(98 \pm 9) \cdot 10^2$
Total Predicted	$(131 \pm 7) \cdot 10^5$	$(16.8 \pm 1.1) \cdot 10^5$	$(38 \pm 3) \cdot 10^4$	$(97 \pm 9) \cdot 10^3$
Data Observed	13251872	1735633	398009	98551

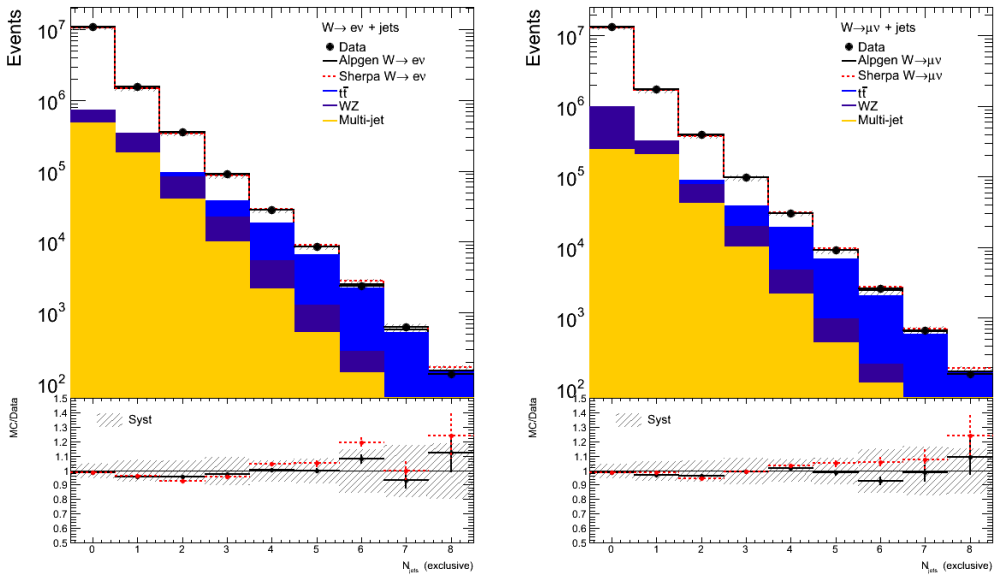
	$N_{jet} = 4$	$N_{jet} = 5$	$N_{jet} = 6$	$N_{jet} = 7$
$W \to \mu\nu$	$(119 \pm 9) \cdot 10^2$	$(23 \pm 2) \cdot 10^2$	$(40 \pm 6) \cdot 10$	73 ± 14
single top, $Z \to \tau\tau$, $W \to \tau\nu$, diboson, $Z \to \mu\mu$	$(25 \pm 2) \cdot 10^2$	$(53 \pm 5) \cdot 10$	99 ± 15	17 ± 3
$t\bar{t}$	$(140 \pm 6) \cdot 10^2$	$(58 \pm 4) \cdot 10^2$	$(18 \pm 2) \cdot 10^2$	$(53 \pm 6) \cdot 10$
QCD	$(22 \pm 2) \cdot 10^2$	$(43 \pm 4) \cdot 10$	117 ± 18	30 ± 5
Total Predicted	$(30 \pm 2) \cdot 10^3$	$(91 \pm 8) \cdot 10^2$	$(24 \pm 4) \cdot 10^2$	$(6.4 \pm 1.1) \cdot 10^2$
Data Observed	30103	9238	2610	658

Table 8.1: Number of events expected from Monte Carlo simulation (*i.e.* $W \to \mu\nu$, single top, $Z \to \tau\tau$, $W \to \tau\nu$, dibosons production, $Z \to \mu\mu$) and data-driven methods ($t\bar{t}$ and QCD) and observed in data for several exclusive jet multiplicities for the $W \to \mu\nu$ selection. The uncertainties are combination of the statistical and the systematic uncertainties.

	$N_{jet} = 0$	$N_{jet} = 1$	$N_{jet} = 2$	$N_{jet} = 3$
$W \to e\nu$	$(100 \pm 5) \cdot 10^5$	$(114 \pm 8) \cdot 10^4$	$(25 \pm 2) \cdot 10^4$	$(51 \pm 5) \cdot 10^3$
single top, $Z \to \tau\tau$, $W \to \tau\nu$, diboson, $Z \to ee$	$(23.9 \pm 1.2) \cdot 10^4$	$(15.4 \pm 1.1) \cdot 10^4$	$(41 \pm 3) \cdot 10^3$	$(12.0 \pm 1.2) \cdot 10^3$
$t\bar{t}$	$(44 \pm 2) \cdot 10$	$(38 \pm 3) \cdot 10^2$	$(120 \pm 9) \cdot 10^2$	$(15.7 \pm 1.0) \cdot 10^3$
QCD	$(47 \pm 2) \cdot 10^4$	$(17.9 \pm 1.3) \cdot 10^4$	$(40 \pm 3) \cdot 10^3$	$(10.0 \pm 1.0) \cdot 10^3$
Total Predicted	$(107 \pm 5) \cdot 10^5$	$(14.7 \pm 1.1) \cdot 10^5$	$(34 \pm 2) \cdot 10^4$	$(88 \pm 9) \cdot 10^3$
Data Observed	10859518	1543553	361116	91001

	$N_{jet} = 4$	$N_{jet} = 5$	$N_{jet} = 6$	$N_{jet} = 7$
$W \to e\nu$	$(100 \pm 8) \cdot 10^2$	$819 \pm 2) \cdot 10^2$	$(33 \pm 5) \cdot 10$	61 ± 12
single top, $Z \to \tau\tau$, $W \to \tau\nu$, diboson, $Z \to ee$	$(32 \pm 2) \cdot 10^2$	$(74 \pm 6) \cdot 10$	$(14 \pm 2) \cdot 10$	26 ± 5
$t\bar{t}$	$(128 \pm 6) \cdot 10^2$	$(53 \pm 4) \cdot 10^2$	$(19 \pm 3) \cdot 10^2$	$(46 \pm 8) \cdot 10$
QCD	$(21 \pm 2) \cdot 10^2$	$(52 \pm 5) \cdot 10$	$(14 \pm 2) \cdot 10$	34 ± 6
Total Predicted	$(28 \pm 2) \cdot 10^3$	$(85 \pm 7) \cdot 10^2$	$(25 \pm 4) \cdot 10^2$	$(5.8 \pm 1.0) \cdot 10^2$
Data Observed	28005	8513	2357	620

Table 8.2: Number of events expected from Monte Carlo simulation (*i.e.* $W \to e\nu$, single top, $Z \to \tau\tau$, $W \to \tau\nu$, dibosons production, $Z \to ee$) and data-driven methods ($t\bar{t}$ and QCD) and observed in data for several exclusive jet multiplicities for the $W \to e\nu$ selection. The uncertainties are combination of the statistical and the systematic uncertainties.



(a) (b)

Figure 8.1: Electron (a) and muon (b) channels detector level plots for the exclusive jet multiplicity spectrum. The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

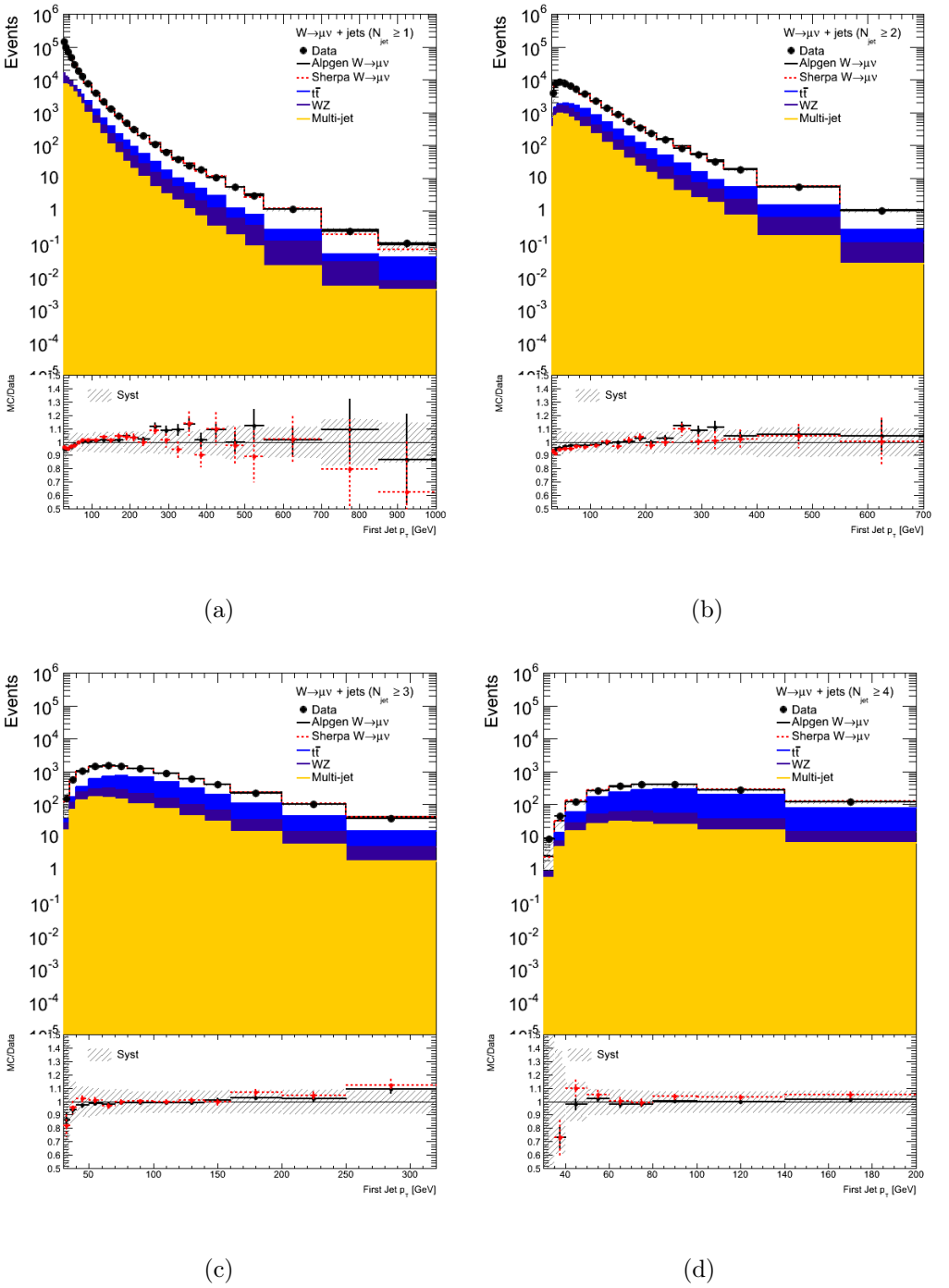


Figure 8.2: Muon channel detector level plots as a function of the leading jet p_T for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

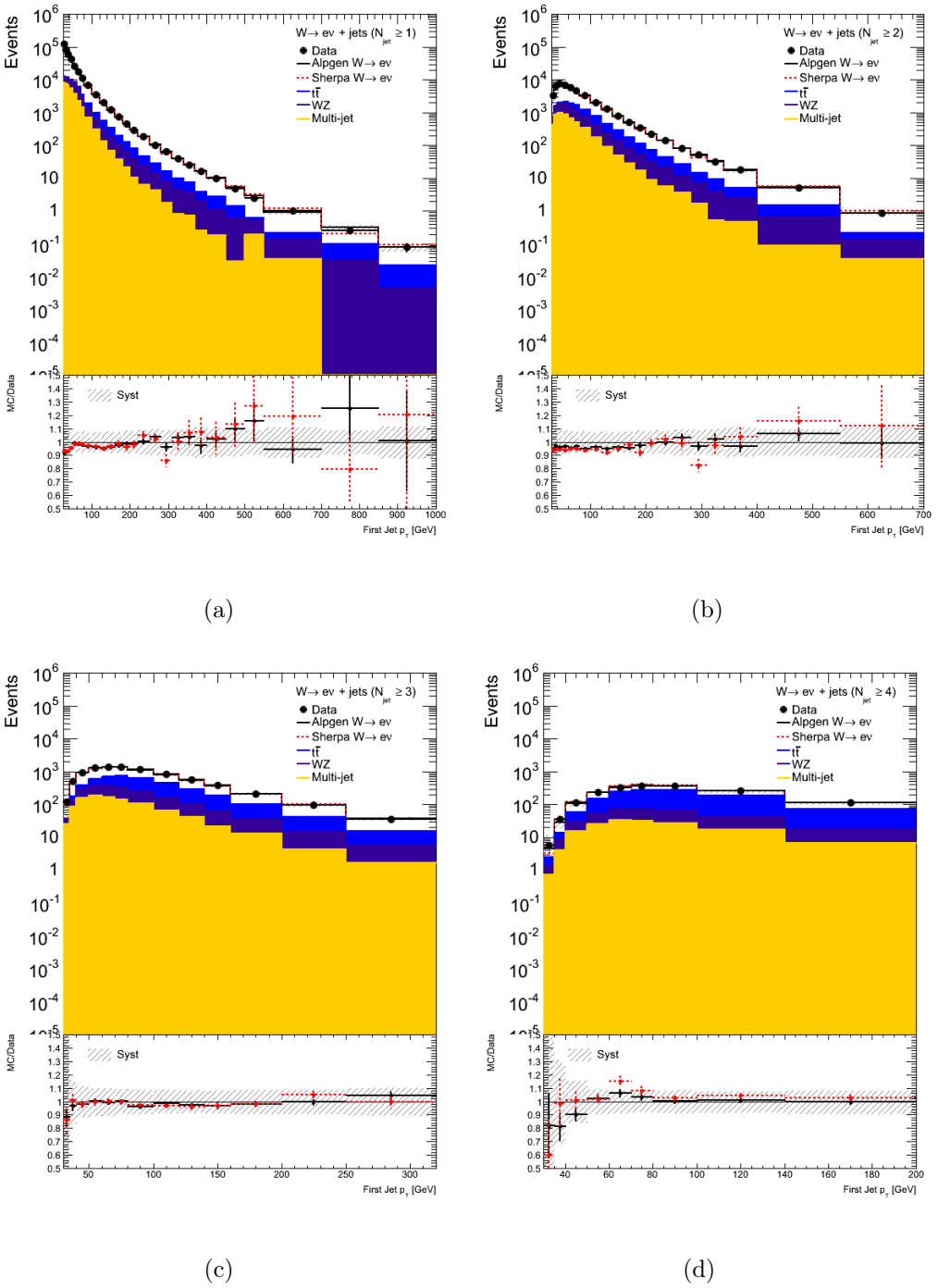


Figure 8.3: Electron channel detector level plots as a function of the leading jet p_T for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

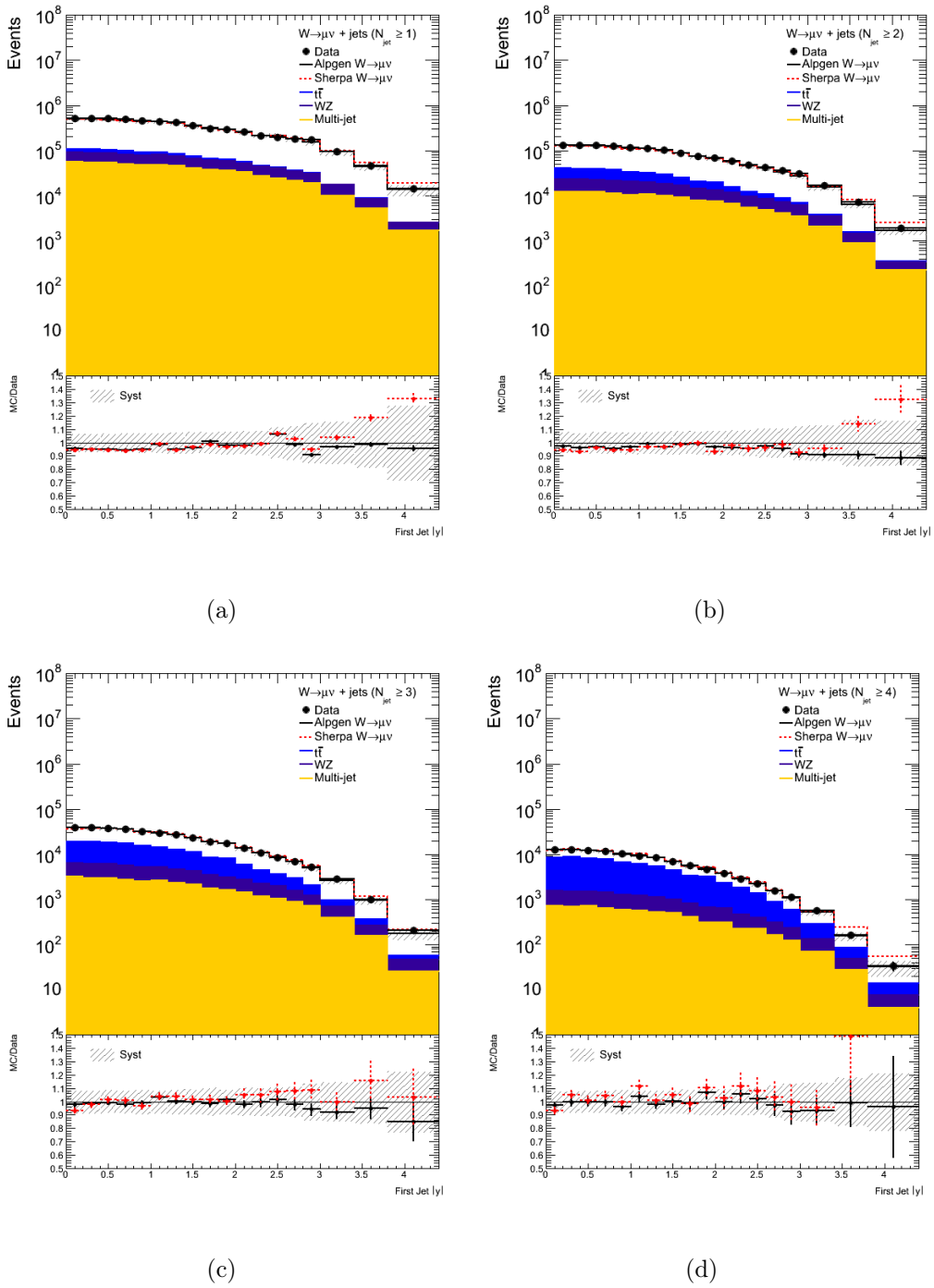


Figure 8.4: Muon channel detector level plots as a function of the leading jet rapidity for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

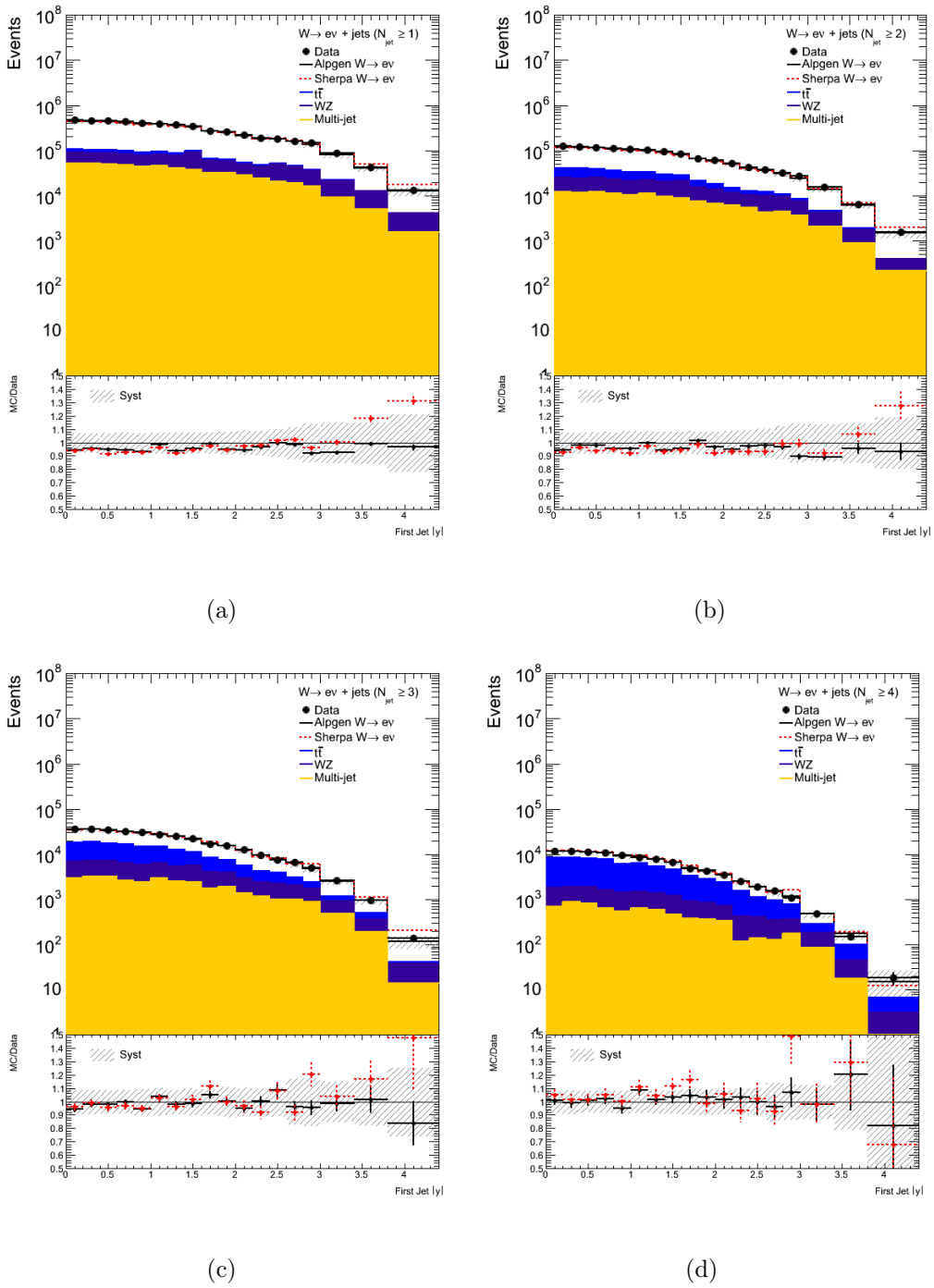


Figure 8.5: Electron channel detector level plots as a function of the leading jet rapidity for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

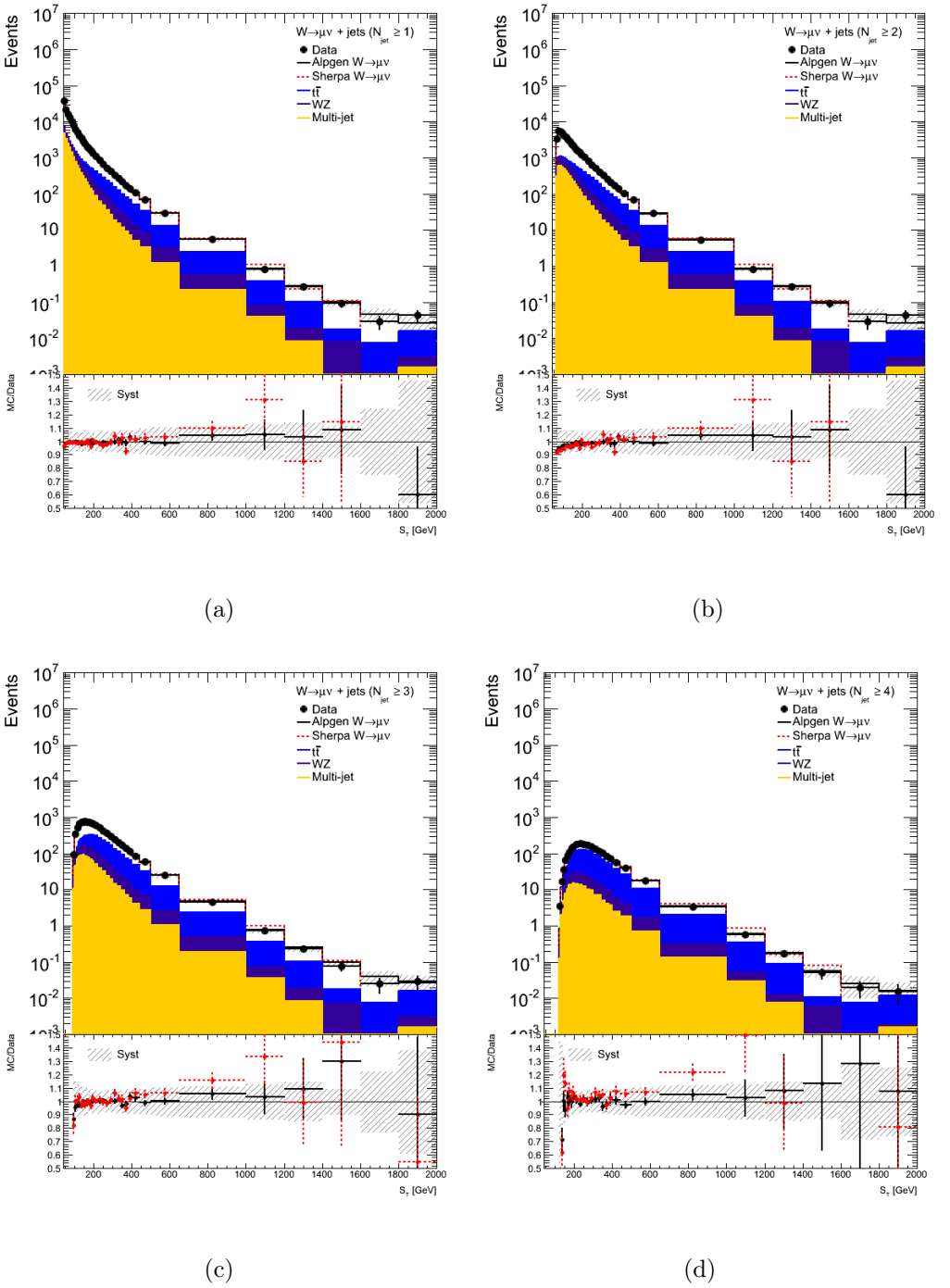


Figure 8.6: Muon channel detector level plots as a function of the scalar sums of all the jets in events (S_T) for events with at least one jets (a) up to at least 4-jets event (b). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

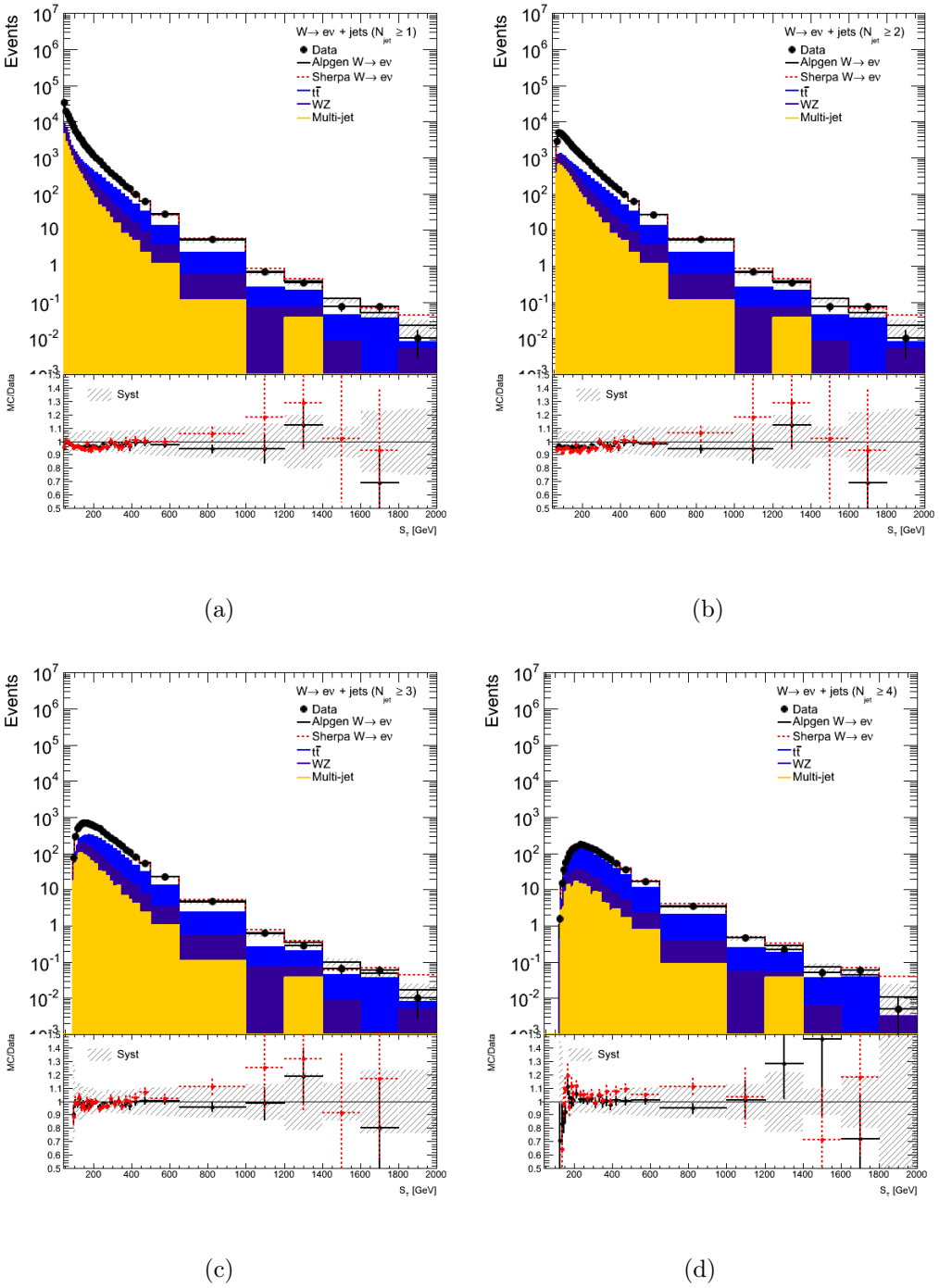
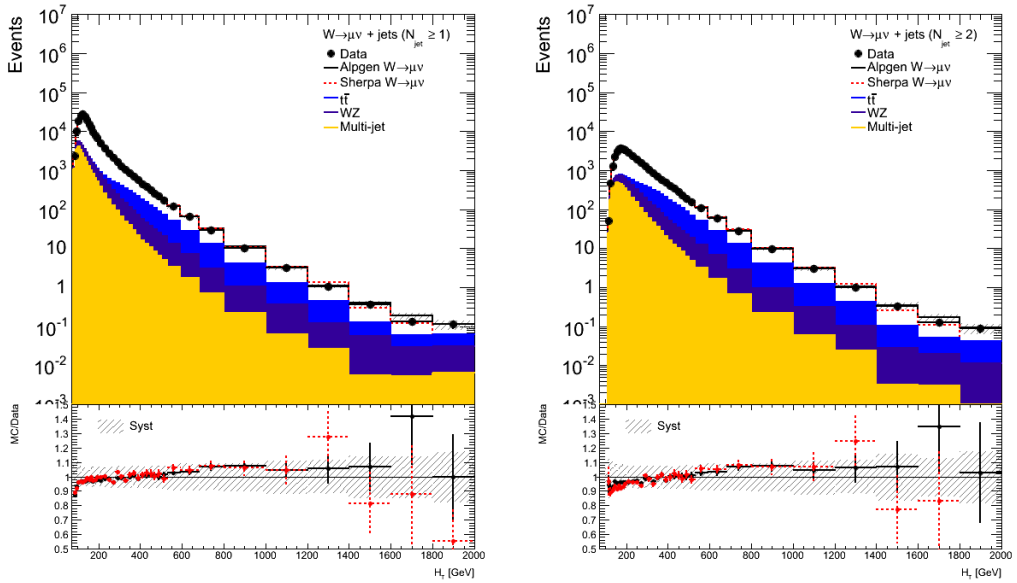
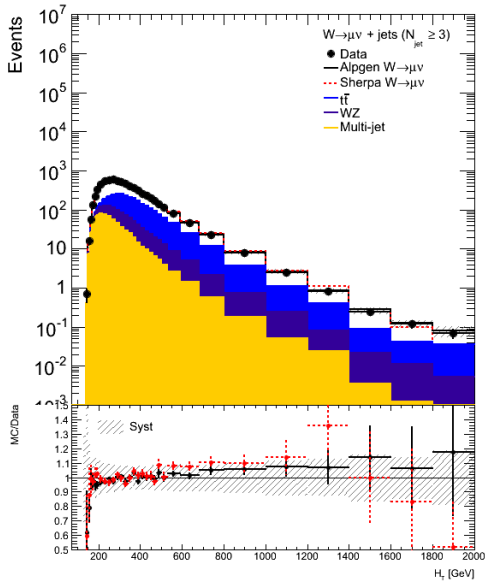


Figure 8.7: Electron channel detector level plots as a function of the scalar sums of all the jets in events (S_T) for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

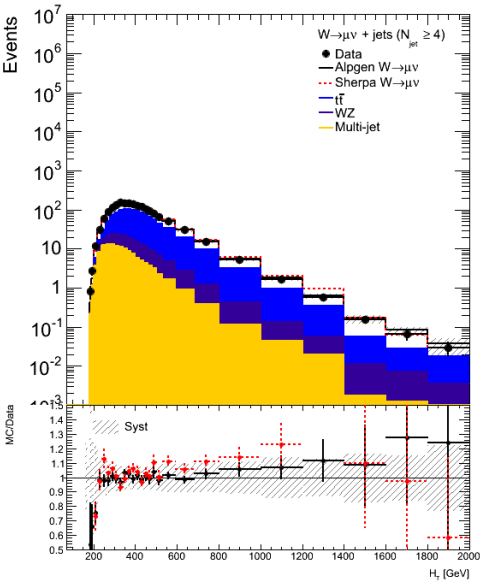


(a)

(b)



(c)



(d)

Figure 8.8: Muon channel detector level plots as a function of the scalar sums of all the objects in events (H_T) for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.

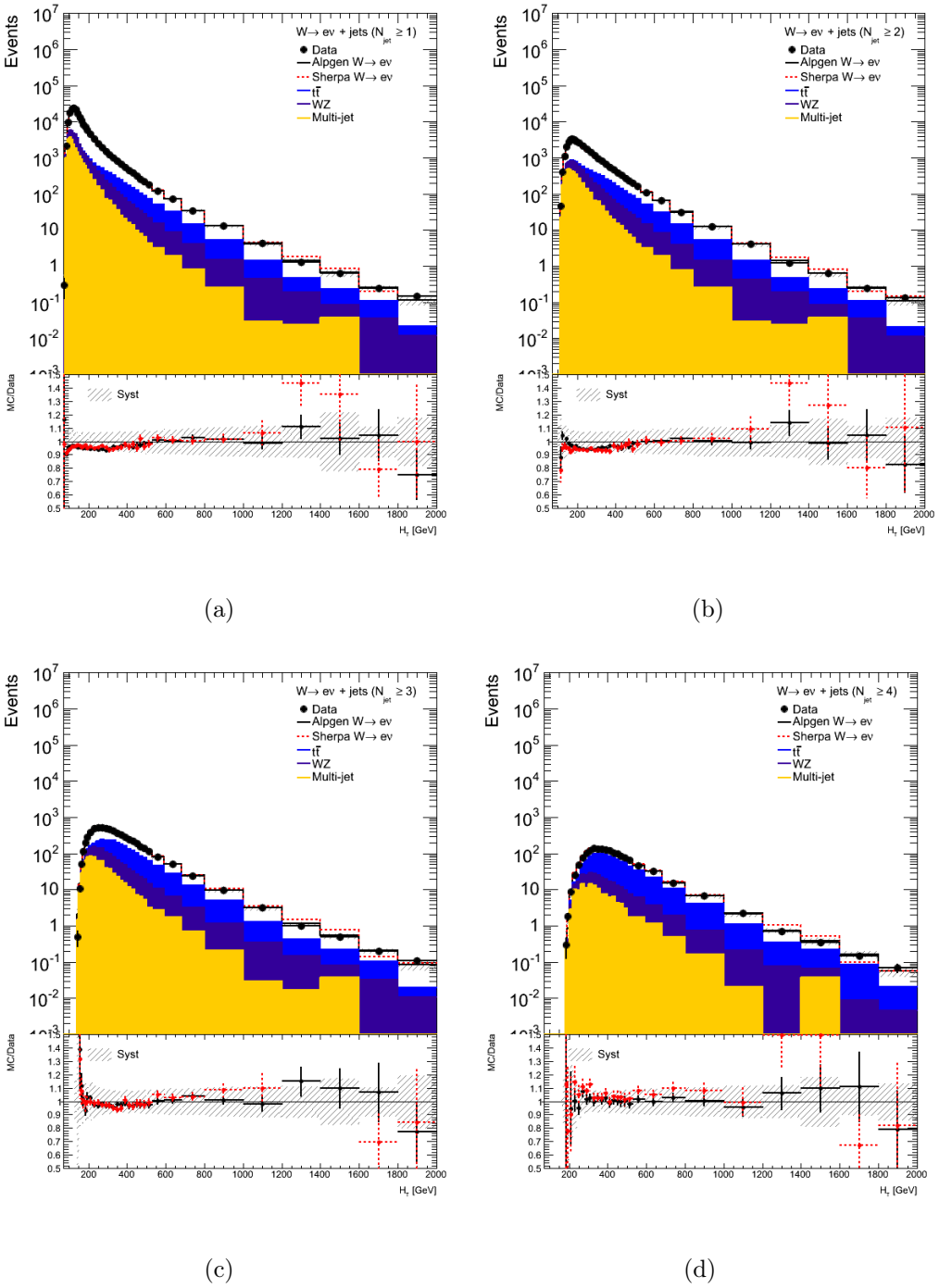
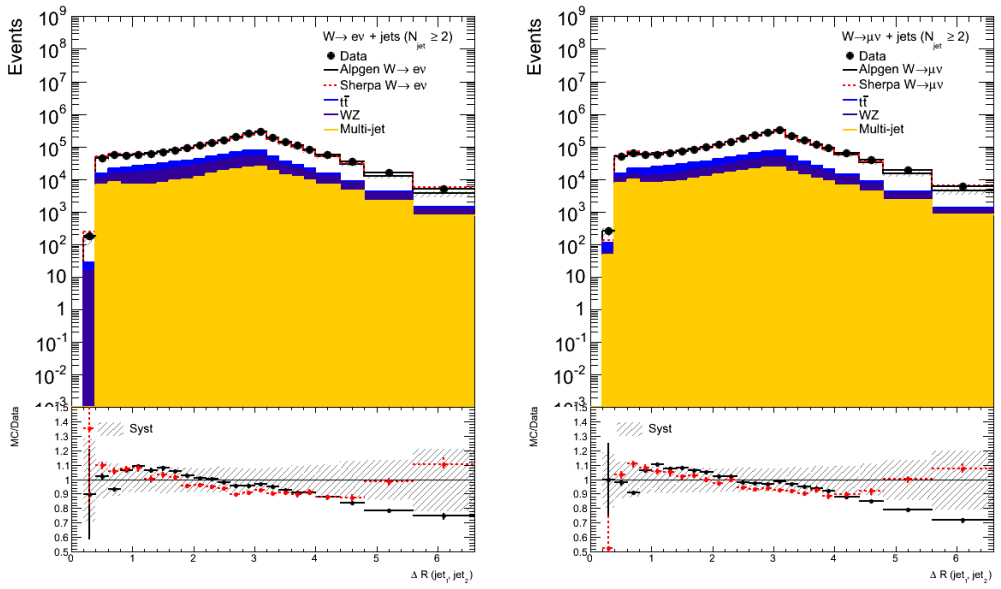
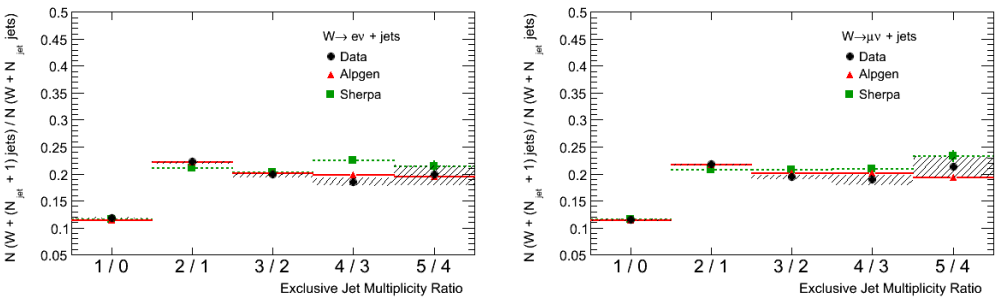


Figure 8.9: Electron channel detector level plots as a function of the scalar sums of all the objects in events (H_T) for events with at least one jets (a) up to at least 4-jets event (d). The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.



(a) (b)

Figure 8.10: Electron (a) and muon (b) channels detector level plots as a function of the angular distance between the two leading jets in events ($\Delta R(\text{jet}_1, \text{jet}_2)$) for events with at least two jets. The shaded bands represent the combination of the statistical and systematic uncertainties on the predicted distributions. The observed distributions are shown with statistical uncertainties only. In the bottom part, the ratios between the predicted and the observed distributions are shown. The systematic uncertainties on the predicted distributions are represented by the dashed bands while the statistical uncertainties are attached to the dots.



(a) (b)

Figure 8.11: Results for the number of events of signal for adjacent jet multiplicities in the electron channel (a) and muon channel (b). For the data the statistical and systematic uncertainties are combined.

Conclusions

In this thesis W boson production in association with jets is studied using $\sqrt{s} = 7$ TeV proton-proton collisions of the 2011 ATLAS dataset with an integrated luminosity of 4.64 fb^{-1}. Particular attention was given to the data-driven background estimations of QCD multijet and $t\bar{t}$ events. The data-driven methods allow to estimate the number of events of these two backgrounds to an unprecedented level for this process. Indeed, they are not subject to the limited statistics in MC samples and to large uncertainties on the predictions for multijet and $t\bar{t}$ total and differential cross-sections.
QCD-enriched samples are selected in electron and muon channels. In the muon channel several QCD control samples are investigated and studies on the template bias are carried on. The observable used to normalize the QCD template is the E_T^{miss}. The accuracy of the fit procedure has been tested using a statistical method of toy MC.
b-tagging is applied on data to select the $t\bar{t}$ control sample. The transposed aplanarity is exploited to normalize $t\bar{t}$ template. A toy MC statistical test is used to validate the fit procedure and to evaluate the correlation between the top templates and data.
Systematic uncertainties on leptons, E_T^{miss}, jets and MC cross-sections are also considered in order to determine the uncertainty on the number of events for signal and backgrounds. Expected and measured numbers of events for signal and background as a function of jet multiplicity in muon and electron channels are compared and are found to be in general agreement. Detector level distributions for several observables, *i.e.* transverse momentum and rapidity of the leading jet, angular distance between the two leading jets and inclusive event quantities (e.g. S_T and H_T), have been measured and compared to predictions. A detailed comparison of experimental results with

ALPGEN and SHERPA's theoretical predictions shows that SHERPA doesn't reproduce perfectly the data spectrum of the observable sensitive to parton-jet matching. ALPGEN shows better agreement with data but not for all the distributions. The agreement can be improved by tuning these generators to ATLAS data.
In light of the results shown and considerations made in §6.2, is possible to determine the W+jets fiducial cross-section for both channels using Eq. 6.1:

$$\sigma^{fid}_{meas}(W + jets) \cdot BR(W \to l\nu) = \frac{N_{obs} - N_{bkg}}{\mathcal{L}_{int} \cdot \epsilon} \tag{8.1}$$

MC and data-driven backgrounds have been subtracted to data and the signal yield has been unfolded to the particle level (see §6.2).
The preliminary fiducial cross-sections measured in the electron and muon channels in the different inclusive jet multiplicities are:

- $\sigma(W \to \mu\nu + \geq 0$ jets) is about 4200 pb with a statistical uncertainty of about 2 pb and a systematic uncertainty of about 120 pb;
- $\sigma(W \to e\nu + \geq 0$ jets) is about 4000 pb with a statistical uncertainty of about 2 pb and a systematic uncertainty of about 60 pb;
- $\sigma(W \to \mu\nu + \geq 1$ jets) is about 370.0 pb with a statistical uncertainty of about 0.5 pb and a systematic uncertainty of about 31.0 pb;
- $\sigma(W \to e\nu + \geq 1$ jets) is about 360.0 pb with a statistical uncertainty of about 0.7 pb and a systematic uncertainty of about 33.0 pb;
- $\sigma(W \to \mu\nu + \geq 2$ jets) is about 87.0 pb with a statistical uncertainty of about 0.2 pb and a systematic uncertainty of about 8.0 pb;
- $\sigma(W \to e\nu + \geq 2$ jets) is about 83.0 pb with a statistical uncertainty of about 0.2 pb and a systematic uncertainty of about 9.0 pb;
- $\sigma(W \to \mu\nu + \geq 3$ jets) is about 17.0 pb with a statistical uncertainty of about 0.1 pb and a systematic uncertainty of about 2.1 pb;
- $\sigma(W \to e\nu + \geq 3$ jets) is about 16.6 pb with a statistical uncertainty of about 0.1 pb and a systematic uncertainty of about 2.4 pb.

All the studies described in this thesis have been used for the measurement of the W+jets cross-section and will contribute to the ATLAS forthcoming publication of the W+jets cross-section measurement.

Bibliography

[1] S. L. Glashow. "Partial symmetries of weak interactions". *Nucl. Phys.*, **22**:579-588, 1961

[2] S. Weinberg. "A model of leptons". *Phys. Rev. Lett.*, **19**:1264–1266, 1967.

[3] A. Salam. "Weak and electromagnetic interactions". *Conf. Proc.*, **C680519**:367–377, 1968.

[4] S. Weinberg. "The Quantum Theory of Fields". Cambridge Univ. Press, Cambridge, 1995.

[5] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys.Lett.*, **B716**:1-29, 2012

[6] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Phys.Lett.*, **B716**:30, 2012

[7] N. Cabibbo. "Unitary symmetry and leptonic decays". *Phys. Rev. Lett.*, **10**:531–533, 1963.

[8] M. Kobayashi and T. Maskawa. "CP violation in the renormalizable theory of weak interaction". *Prog. Theor. Phys.*, **49**:652–657, 1973.

[9] J. Beringer *et al.*(PDG). "Electroweak model and constraints on new physics". *Phys. Rev.*, **D86**:010001, 2012, (`http://pdg.lbl.gov`).

[10] J. Gunion *et al.*. "The Higgs Hunter's Guide". Addison-Wesley, Redwood City, 1990.

[11] I. C. Brock and T. Schorner-Sadenius. "Physics at the terascale". Weinheim: Wiley, 2011.

[12] J. Beringer *et al.*(PDG). "Quantum Chromodynamics". *Phys. Rev.*, **D86**:010001, 2012, (`http://pdg.lbl.gov`).

[13] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *Journal of Instrumentation 3*, **S08003**:437, 2008.

[14] L. Evans and P. Bryant. "LHC Machine". *JINST 3.08*, **S08001**, 2008.

[15] CMS Collaboration. "The CMS experiment at the CERN LHC. The Compact Muon Solenoidexperiment". *JINST 3*, **S08004**: 361, 2008.

[16] ALICE collaboration. "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment". *JINST 3*, **S08002**: 259, 2008.

[17] LHCb Collaboration. "The LHCb Detector at the LHC". *JINST 3*, **S08005**, 2008.

[18] TOTEM Collaboration. "The TOTEM Experiment at the CERN Large Hadron Collider". *JINST 3*, **S08007**, 2008

[19] O. Adriani *et al.*. "LHCf experiment: Technical Design Report". *Tech. rep.* CERN-LHCC-2006-004, Geneva, 2006.

[20] P. Jenni and M. Nessi. "ATLAS Forward Detectors for Luminosity Measurement and Monitoring". *Tech. rep.* CERN-LHCC-2004-010. LHCC-I-014, Geneva: CERN, Mar. 2004.

[21] J. Pinfold *et al.*. "Technical Design Report of the MoEDAL Experiment". *Tech. rep.* CERN-LHCC-2009-006. MoEDAL-TDR-001. Geneva: CERN, June 2009.

[22] ATLAS Collaboration. "ATLAS magnet system: Technical Design Report". *Tech. rep.* CERN-LHCC-97-18, Geneva, 1997.

[23] A. Yamamoto *et al.*. "The ATLAS central solenoid". *Nucl. Instr. and Meth. in Phys. Res.*, Section A **584.1**:53-74, 2008.

[24] ATLAS Collaboration. "ATLAS central solenoid: Technical Design Report". *Tech. rep.* CERN-LHCC-97-21, 1997.

[25] ATLAS Collaboration. "ATLAS barrel toroid: Technical Design Report". *Tech. rep.* CERN-LHCC-97-19, 1997.

[26] ATLAS Collaboration. "ATLAS endcap toroids: Technical Design Report". *Tech. rep.* CERN-LHCC-97-20, 1997.

[27] ATLAS Collaboration. "ATLAS inner detector: Technical Design Report". *Tech. rep.* CERN-LHCC-97-17, Geneva, 1997.

[28] ATLAS Collaboration. "ATLAS pixel detector: Technical Design Report". *Tech. rep.* CERN-LHCC-98-13, Geneva, 1998.

[29] A. Ahmad *et al.*. "The Silicon microstrip sensors of the ATLAS semiconductor tracker". *Nucl. Instr. and Meth. in Phys. Res.*, Section A **578**:98-118, 2007.

[30] The ATLAS TRT collaboration. "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance". *JINST 3.02*, **P02013**, 2008.

[31] The ATLAS collaboration. "Performance of Particle Identification with the ATLAS Transition Radiation Tracker". ATLAS-CONF-2011-128, 2011. `arXiv:1109.5925 [physics]`

[32] ATLAS Collaboration. "ATLAS liquid-argon calorimeter: Technical Design Report". *Tech. rep.* CERN-LHCC-96-41, Geneva, 1996.

[33] ATLAS Collaboration. "ATLAS tile calorimeter: Technical Design Report". *Tech. rep.* CERN-LHCC-96-42, Geneva, 1996.

[34] ATLAS Collaboration. "ATLAS muon spectrometer: Technical Design Report". *Tech. rep.* CERN-LHCC-97-22, ATLAS-TDR-10, Geneva, 1997.

[35] P. Jenni, M. Nessi and M. Nordberg. "Zero Degree Calorimeters for ATLAS". *Tech. rep.* LHCC-I-016. CERN-LHCC-2007-001, Geneva: CERN, Jan. 2007.

[36] ATLAS Collaboration. "ATLAS level-1 trigger: Technical Design Report". *Tech. rep.* CERN-LHCC-98-14, ATLAS-TDR-12, Geneva, 1998.

[37] P. Jenni *et al.*. "ATLAS high-level trigger, data-acquisition and controls: Technical Design Report". *Tech. rep.* CERN-LHCC-2003-022, ATLAS-TRD-016, Geneva, 2003.

[38] W. Lampl et al. "Calorimeter Clustering Algorithms: Description and Performance". `https://cdsweb.cern.ch/record/1099735/`, 2008. ATL-LARG-PUB-2008-002.

[39] ATLAS Collaboration. "Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data". *Eur. Phys. J.*, **C72**, 2012, arXiv:1110.3174 [hep-ex].

[40] Electron efficiency measurements, ATLAS public results. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/EGAMMA/PublicPlots/20120611/ElectronEfficiency2012/ATL-COM-PHYS-2011-783/index.html, 2012.

[41] ATLAS Collaboration. "Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV". ATLAS-CONF-2010-005, 2010.

[42] ATLAS Collaboration. "Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung". ATLAS-CONF-2012-047, 2012.

[43] ATLAS Collaboration, "Expected electron performance in the ATLAS experiment". ATL-PHYS-PUB-2011-006, 2011, http://cds.cern.ch/record/1345327.

[44] ATLAS Collaboration, "Muon Reconstruction and Identification: Studies with Simulated Monte Carlo Samples". ATLAS-PHYS-PUB-2009-008, 2009.

[45] ATLAS Collaboration. "Determination of the muon reconstruction efficiency in ATLAS at the Z resonance in proton-proton collisons at $\sqrt{s} = 7$ TeV". ATLAS-CONF-2011-008, 2011.

[46] ATLAS Collaboration. "ATLAS Muon Momentum Resolution in the First Pass Reconstruction of the 2010 pp Collision Data at $\sqrt{s} = 7$ TeV". ATLAS-CONF-2011-046.

[47] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPerformancePublicPlots.

[48] M. Cacciari, G. Salam, and G. Soyez. "The $anti - k_t$ jet clustering algorithm". JHEP, **04**:063, 2008.

[49] ATLAS Collaboration. "Jet energy scale and its systematic uncertainty in ATLAS for jets produced in proton-proton collisions at 7 TeV". https://cdsweb.cern.ch/record/1281329. ATLAS-CONF-2010-056, 2010

[50] ATLAS Collaboration. "Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS Detector". *Eur. Phys. J.*, **C71**, 2010.

[51] ATLAS Collaboration. "Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS 2011 data". ATLAS-CONF-2013-004, 2013.

[52] ATLAS Collaboration. "Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector". *Eur. Phys. J.*, **C73**:2306, 2013.

[53] ATLAS EXPERIMENT. Public Results. JetEtmissApproved2013Jer2011. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013Jer2011`

[54] ATLAS Collaboration. "Identification and Tagging of Double b-hadron jets with the ATLAS Detector". ATLAS-CONF-2012-100, 2012.

[55] ATLAS Collaboration. "Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector". ATLAS-CONF-2012-043

[56] ATLAS Collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV with ATLAS". *Eur. Phys. J.*, **C72**:1844, 2011.

[57] ATLAS Collaboration. "Reconstruction and calibration of missing transverse energy and performance in Z and W events in ATLAS proton-proton collisions at 7 TeV". *Tech. rep.*, ATLAS-CONF-2011-080.

[58] ATLAS Collaboration. "ATLAS computing: Technical Design Report". *Technical Design Report ATLAS*. CERN, Geneva, 2005.

[59] Frank Krauss, Sherpa.

[60] T. Sjeostrand, S. Mrenna, and P. Z. Skands. "PYTHIA 6.4 Physics and Manual". JHEP, **05**:026, 2006.

[61] A. Bialas and M. Gyulassy. "Lund model and an outside-inside aspect of the inside-outside cascade". *Nucl. Phys. B*, **291(0)**:793-812, 1987. `http://www.sciencedirect.com/science/article/pii/0550321387904962, doi: 10.1016/0550-3213(87)90496-2.`

[62] G. Corcella et al. "HERWIG 6.5: an Event Generator for Hadron Emission Reactions With Interfering Gluons (Including Supersymmetric Processes)". JHEP, **01**:010, 2001.

[63] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour. "Multiparton interactions in photoproduction at HERA". *Z. Phys.* **C72**:637-646, 1996, `arXiv:9601371 [hep-ph]`.

[64] M. Bahr et al.. "HERWIG++ Physics and Manual". *Eur. Phys. J.* **C58**:639-707,2008, `arXiv:0803.0883 [hep-ph]`.

[65] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert, J. Winter. "Event generation with SHERPA 1.1". JHEP, **02**:007, 2009.

[66] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa. "ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions". JHEP, **07**:001, 2003.

[67] J. Alwall, S. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, F. Maltoni, M.L. Mangano, M. Moretti, C.G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, M. Worek. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". *Eur. Phys. J.*, **C53**:473-500, 2008. `arXiv:0706.2569 [hep-ph]`.

[68] S. Frixione and B. R. Webber. "Matching NLO QCD Computations and Parton Shower Simulations". JHEP, **06**:029, 2002.

[69] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", JHEP, **0411**:040, 2004.

[70] S. Hoeche, F. Krauss, M. Schonherr, F. Siegert. "QCD matrix elements + parton showers: The NLO case". `arXiv:1207.5030 [hep-ph]`. JHEP, **04**:027, 2013. `arXiv:1207.5030 [hep-ph]`.

[71] The GEANT4 Collaboration. "GEANT4: A simulation toolkit". *Nucl. Instr. Meth.*, **A506**:250, 2003.

[72] R. K. Ellis, W. J. Stirling, B. R. Webber. "QCD and collider physics". Cambridge University Press, 2003.

[73] J. C. Collins, D. E. Soper. "The Theorems of Perturbative QCD". *Ann. Rev. Nucl. Part. Sci.*, **37**:383-409, 1987.

[74] J. C. Collins, D. E. Soper, and G. Sterman. "Heavy Particle Production in High Energy Hadron Collisions". *Nucl. Phys.*, **B263**:37, 1986. doi: 10.1016/0550-3213(86) 90026-X.

[75] Martin-Stirling-Thorne-Watt Parton Distribution Functions. `http://mstwpdf.hepforge.org/.`

[76] V.N. Gribov and L.N. Lipatov. *Sov. J. Nucl. Phys.*, **15**:438, 1972.
G. Altarelli and G. Parisi, *Nucl. Phys.*, **B126**:298, 1977.
Yu.L. Dokshitzer, *Sov. Phys.* JETP, **46**:641, 1977.

[77] F.A. Berends, H. Kuijf, B. Tausk, W.T. Giele. "On the production of a W and jets at hadron colliders". *Nucl. Phys. B*, **357**(1):32-64, 1991.

[78] CDF and D0 Collaborations. "Measurements of vector boson plus jets at the Tevatron". Contribution to the proceedings of the Top 2010 Conference. `arXiv:1008.2291 [hep-ex]`.

[79] C. Berger et al.. "Precise predictions for W + 4-jet production at the Large Hadron Collider". *Phys. Rev. Lett.*, **106**: 092001, 2011.

[80] H. Ita and K. Ozeren. "Colour decompositions of multi-quark one-loop QCD amplitudes". JHEP, **1202**:118, 2012.

[81] H. Ita et al.. "Precise predictions for Z + 4 jets at hadron colliders". *Phys.Rev.*, **D85**:031501, 2012.

[82] Jeppe R. Andersen, Tuomas Hapola, Jennifer M. Smillie. "W plus Multiple Jets at the LHC with High Energy Jets". JHEP, **09**:047, 2012. `arXiv:1206.6763 [hep-ph]`.

[83] M. Rubin, G. P. Salam, S. Sapeta. "Giant QCD K-factors beyond NLO". JHEP, **09**:084, 2010. `arXiv:1006.2144 [hep-ph]`.

[84] ATLAS Collaboration. "Study of jets produced in association with a W boson in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector". *Phys. Rev.*, **D85**:092002, 2012. `arXiv:1201.1276 [hep-ex]`.

[85] ATLAS Collaboration. "Measurement of the production cross section of jets in association with a Z boson in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector". JHEP, **07**:032, 2013. `arXiv:1304.7098 [hep-ex]`.

[86] E. Abouzaid and H. J. Frisch. "The Ratio of W + N jets to Z^0/γ^* + N jets versus N as a precision test of the standard model". *Phys. Rev.*, **D68**:033014, 2003. `arXiv:0303088 [hep-ph]`.

[87] ATLAS Collaboration. "Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC". *Eur. Phys. J.*, **C73**:2518, 2013. `arXiv:1302.4393 [hep-ex]`.

[88] G. D'Agostini, "A Multidimensional Unfolding Method Based On Bayes' Theorem". *Nucl. Istr. and Meth. in Phys. Res.*, Section A **362**:487-498, 1995.

[89] C. Anastasiou, L. J. Dixon, K. Melnikov, F. Petriello. "High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO". *Phys. Rev.*, **D69**:094008, 2004

[90] A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt. "Parton distributions for the LHC". *Eur. Phys. J.*, **C63**:189-285, 2009

[91] R. J. Barlow, C. Beeston. "Fitting using finite Monte Carlo samples". *Comput. Phys. Commun.*, **77**:219-228, 1993

[92] ATLAS Collaboration. "Measurement of the production cross section for W-bosons in association with jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". `arXiv:1012.5382 [hep-ex]`, 2010.

[93] `http://root.cern.ch/drupal/content/roofit`

Acknowledgements

The last 12 months spent at CERN have been among the most the most intense and unforgettable of my life, full of work as well as great satisfaction, expectations, new friends, new languages. I would like to thanks all the people for helping me and took part of this time.
My deepest gratitude goes first and foremost to my supervisor Alessandro Tricoli for encouraging me, answering all my doubts and questions and helping me when I needed. Alessandro thank you because you taught me to *pistolare*.
I am specially grateful with Paolo Iengo for accepting to be the referees of this document and for supporting my work with confidence and criticism.
My gratitude goes also to my professor Mariagrazia Alviggi who, step by step, guided me in the conclusive sprint of my university development.
I have to give a special grateful to Monica Dunford for being a colleague and a friend, for her guide and help: I leave after having learned many things from you!!
I would like to thank the ATLAS R-jet group members, in particular Evelin Meoni, for their comments, suggestions and physics discussions when needed.
Thanks to my neighbors, Sergio and Elisa, for creating together a small corner of Italy.
Thanks to Ludovica, who shared with me the office: I cannot forget our laughs and friendly chats, especially about french people and bitten pieces of chocolate.
Thanks to Nicola, Felice, Mirko e Federica who became the best friend ever to share lunches, dinners, tears and laughs.
Thanks to Ylenia, the best friend I have ever met: our friendship is among

the most wonderful things the last year returned me. Thank you for sharing *struffoli*, shopping and zumba (obviously to digest *struffoli*).
To my old high-school friends for keeping close in distance and always being there during the re-encounters that I organized: thank you Elena and Anna!
Thanks to Francesco for his love and bravery. Thank you because you always make me laugh when I have *saudade*. Yes, we are building our family now.
Finally, I would like to thank my family, for all the encouragement provided during this year despite the distance. Especially to my parents, who with their work, efforts and sacrifices have supported all my studies. Thank you also for being on Skype almost all the days and giving me a wonderful nativity scene last Christmas ("Ma a te...te piace 'o presepe??" "No. Nun me piace. Voglio 'a zuppa 'e latte!")!!!
Thank you very much to all of you!

Ringraziamenti

Gli ultimi 12 mesi trascorsi al CERN sono stati i più intensi e indimenticabili della mia vita, pieni di lavoro ma anche di soddisfazioni, di aspettative, di nuove amicizie, di nuove lingue. Vorrei ringraziare tutte le persone che mi hanno aiutato e hanno fatto parte di questo periodo.
La mia più profonda gratitudine va al mio supervisor Alessandro Tricoli per avermi incoraggiata, aver risposto ai miei dubbi e alle mie domande, avermi aiutato quando ne avevo bisogno. Alessandro, grazie perché mi hai insegnato a *pistolare*.
Sono particolarmente grata a Paolo Iengo per aver accettato di essere il relatore di questa tesi e per aver supportato il mio lavoro con fiducia e criticità.
La mia gratitudine va anche alla Prof.ssa Mariagrazia Alviggi che, passo dopo passo, mi ha guidato nello sprint finale del mio iter universitario.
Un grazie speciale va a Monica Dunford, per essere stata una collega e un'amica, per la sua guida e il suo aiuto: vado via dopo aver imparato tante cose da te!! Vorrei ringraziare i membri dell'*ATLAS R-jet group*, in particolare Evelin Meoni, per i loro commenti, suggerimenti e discussioni di fisica quando ce ne è stata la necessità.
Grazie ai miei vicini, Sergio ed Elisa, perché insieme abbiamo creato un piccolo angolo di Italia.
Grazie a Ludovica, che ha condiviso con me l'uffico: non potrò dimenticare le nostre risate e le nostre chiaccherate, soprattutto su francesi e pezzi morsicati di cioccolato.
Grazie a Nicola, Felice, Mirko e Federica che sono diventati gli amici di sempre, per condividere pranzi, cene, lacrime e risate.
Grazie ad Ylenia, la migliore amica che abbia mai incontrato: la nostra ami-

cizia è tra le cose più belle che ho ricevuto nell'ultimo anno. Grazie per aver condiviso struffoli, shopping e lezioni di zumba (ovviamente per smaltire gli struffoli).
Grazie alle mie vecchie amiche del liceo per essermi state vicine nonostante la distanza e sono state sempre presenti ai nostri ritrovi: grazie Elena e Anna!
Grazie a Francesco per il suo amore e il suo coraggio. Grazie perché mi porti un sorriso anche nei momenti di "saudade". Sì, adesso siamo noi a costruire la nostra famiglia.
Infine, ringrazio la mia famiglia, per l'incoraggiamento che ho ricevuto durante quest'anno nonostante la distanza. Un ringraziamento speciale ai miei genitori che con il loro lavoro, i loro sforzi e i loro sacrifici hanno sempre supportato i miei studi. Grazie per essere su Skype quasi tutti i giorni e per avermi costruito un bellissimo presepe ("Ma a te...te piace 'o presepe??" "No. Nun me piace. Voglio 'a zuppa 'e latte!")!!!
Grazie mille a tutti voi!!