School of Physics and Astronomy

Particle Physics Research Center

Measurement of the production cross section of a W boson with a single charm quark using the ATLAS detector at the Large Hadron Collider

Giacomo Snidero

Supervisor: Dr. Lucio Cerrito

Thesis submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

2014

Statement of Orginality

I, Giacomo Snidero, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Giacomo Snidero

London, July 2014

Abstract

This thesis presents the measurement of the production cross section of a W boson with a single charm quark in 7 TeV proton-proton collisions using the ATLAS detector at the Large Hadron Collider. The data analysed correspond to an integrated luminosity of 4.6 fb^{-1} and were collected during 2011. This is the first time that ATLAS data has been used for this particular measurement.

This cross section is of particular interest as a probe of the strange quark density of the proton. Typically, the strange quark density is considered to be suppressed relative to that of the other light-quarks in the proton sea. However, some analyses suggest a more symmetric composition of light-quarks in the proton sea. The results of this study aim to improve the precision of the determination of the strange quark density.

The analysis uses events where the W boson decays to a muon and a neutrino. In such events, the charm quark is identified by its semileptonic decay to a soft muon within a hadronic jet. The charge correlation between the W boson and the soft muon is exploited to reduce the backgrounds substantially. The analysis results are combined with those obtained using additional decay channels.

The measured cross section provides further constraint for the determination of the strange quark density, advancing the knowledge of the fundamental structure of the proton. The results are compared with predictions of next-to-leading-order QCD calculations obtained using various parton distribution function parameterisations. Additionally, the ratio of the strange-to-down sea-quark distributions is determined to be $0.96^{+0.26}_{-0.30}$ at $Q^2 = 1.9$ GeV^2. This supports the hypothesis of a symmetric composition of light-quarks in the proton sea. The cross section ratio $\sigma(W^+ + \bar{c})/\sigma(W^- + c)$ is also determined and compared with different predictions for the asymmetry of the strange and anti-strange quark distribution functions.

Preface

The Large Hadron Collider (LHC) at the CERN laboratory is the world's most powerful particle accelerator. This machine has been built to advance the human knowledge in the discipline of particle physics – the study of the elementary constituents of the Universe and their interactions. The current paradigm of particle physics is the theoretical framework developed over the 1960s and 70s which is known as the standard model (SM).

The ATLAS detector is one of the two major experiments operating at the LHC. The ATLAS research programme can be subdivided into three main areas: Higgs boson physics, searches for phenomena beyond the SM and precision measurements of the SM. The latter constitute the benchmark studies with the LHC data. The measurement of the production cross section of a W boson with a single charm quark (W+c) described in this thesis is a precision measurement of the SM.

This thesis is structured as follows. The scientific research context is introduced in chapter 1 with an overview of the standard model, the LHC and the ATLAS detector. Chapter 2 presents an analysis of the time stability of one of the ATLAS subdetectors, namely the L1Calo trigger preprocessor. This study is the author's contribution to the operation and performance of the ATLAS detector.

The measurement of the W+c production cross section is presented in chapter 3 onwards. The theoretical description, the motivations and the strategy of the W+c production analysis are given in chapter 3. The LHC collision data sample, the selections for $W \to \mu\nu$ decays and c-quark jets, and the Monte Carlo simulated samples for the modelling of the signal and background processes are explained in chapter 4.

Chapter 5 describes the determination of the yield of W+c events in the data.

Both data-driven methods and simulations are used for the background estimation. Dedicated selection criteria are applied to reduce the Z+jets and QCD multijet backgrounds and enhance the signal purity of the selected data. In chapter 6 the number of W+c events is measured as a function of the charge and pseudorapidity of the muon from the W-decay.

In chapter 7 the integrated and differential W+c cross sections are determined from the measured yields. The integrated cross section is combined with that resulting from a similar analysis which uses $W \rightarrow e\nu$ events. The combined result is compared to theoretical predictions of next-to-leading-order QCD calculations obtained using different parton distribution function parameterisations. In addition, the combined results are reported for the cross section ratio $\sigma(W^{+}+\bar{c})/\sigma(W^{-}+c)$ and for the differential cross sections, as a function of i) the pseudorapidity of the W-decay muon; ii) the jet multiplicity. Finally, the ratio of strange-to-down sea-quark densities is shown as determined from the combination of this analysis with those using the $W \rightarrow e\nu$ events and additional charm quark decay channels.

Furthermore, appendix A describes the multijet background estimation, carried out by the author, which was used in an ATLAS measurement of the top quark pair production cross section. The contribution of the background process is evaluated by extending the method used for the $W+c$ analysis.

Details of collaboration and publications

The $W+c$ analysis described in this thesis is part of the measurement by the ATLAS Collaboration published in reference [1]. Within this collaboration, which includes almost 3,000 researchers, the W+c analysis was carried out by a team of 9 researchers, including the author of this thesis. The W+c analysis in this thesis is also detailed in the ATLAS internal analysis report [2].

The operation and maintenance of the ATLAS L1Calo trigger preprocessor is the work of around 100 ATLAS researchers. Specifically, the analysis described in this thesis was performed by 2 researchers, including the author of this thesis, and is included in the summary of analyses and results published in reference [3].

The QCD multijet estimation in appendix A has been used in a measurement of the top quark pair production cross section by the ATLAS Collaboration, which is reported in [4]. This cross section measurement was carried out by 10 ATLAS researchers, including the author of this thesis. Moreover, this analysis team calibrated the soft muon tagging algorithm used in both the W+c and the top quark pair analyses. These studies of the soft muon tagger are also documented in the ATLAS internal analysis report [5].

Due to the complexity of modern particle physics experiments, a single researcher cannot carry out the work required for all aspects of an ATLAS analysis. This thesis specifically describes the contributions which are the own and original work of the author. A summary of these contributions is listed below. Any aspect of analysis or plot which was not produced by the author is clearly stated as such and referenced throughout the thesis.

Chapter 2: The author developed the analysis strategy, implemented the algorithm in a package within the L1Calo software framework, analysed the data and presented the results.

Chapter 3: The author evaluated from simulation the expected contributions from the different scattering processes to the $W + c$ production, and the expected fraction of opposite sign signal events.

Chapter 4: The author wrote his own software for the W+c event selection and implemented the soft muon tagging algorithm; analysed datasets of LHC collisions and Monte Carlo simulated collisions using the CERN's *World LHC Computing Grid* system[1]. The author also produced datasets used for the calibration of soft muons mistag rate.

Chapter 5: The author estimated the multijet, W+light and Z+jets backgrounds using data-driven methods; evaluated the other analysis backgrounds with simulation. The author also studied dedicated selection requirements on the charm-jet and dimuon invariant mass to suppress the Z+jets background and optimised the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W} selections to reduce the multijet and Z+jets

---

[1] http://wlcg.web.cern.ch.

backgrounds. The author also obtained the results for the yield of $W+c$ events and compared kinematical distributions of the selected data with predictions.

Chapter 6: The author evaluated the backgrounds as a function of the W-decay muon's charge and pseudorapidity, using both data-driven methods and simulated samples. The author also obtained the results for the charge divided yields of W+c events; measured the pseudorapidity distribution for the W+c production and compared it to prediction.

Chapter 7: The author calculated the combination of integrated cross sections using the profile likelihood method, using dedicated statistical software packages and generating Monte Carlo pseudo-experiments to validate the result of the maximum likelihood fit. The author also compared the combined result to the theoretical predictions taken from reference.

Appendix A: The author obtained the results for the QCD multijet background estimation in the muon channel analysis, which was used to measure the top quark pair production cross section.

Appendix B: The author obtained the results for the predicted yields in the pretag data sample and compared kinematical distributions with predictions.

Contents

List of Figures

List of Tables

Chapter 1

The Large Hadron Collider and the ATLAS experiment

The work presented in this thesis is contextualised in this chapter. A concise scheme of particle physics together with the goals and early results from the Large Hadron Collider (LHC) are summarised. The LHC accelerator at the CERN laboratory is briefly described and the structure of one its the two major experiments, the ATLAS detector, is summarised.

1.1 Physics at the LHC

The LHC is a particle accelerator constructed with the aim of furthering knowledge in the field of particle physics. Ultimately, this discipline intends to recognise the fundamental constituents of the Universe, or elementary particles, and to understand their mutual interactions.

The current paradigm of particle physics is the theoretical framework known as the *standard model* (SM). Originally, this name was given to the theory of electroweak interactions ideated in the 1960's [6–9]. More commonly, the SM also incorporates the theory of the strong interactions formulated in the 1970's, which is known as quantum chromodynamics (QCD) [10–12]. In this thesis "standard model" refers both to the electroweak and the QCD sectors.

The SM is a quantum field theory: in the equations of motion, the elementary particles are represented by quantum fields. Such particles are classified into fermions and bosons. In the SM there are twelve fermions represented by spin-1/2 fields, which are divided into three families of leptons and three families of quarks. There are also five gauge bosons represented by spin-1 fields: the W^+, W^-, Z and photon (γ) mediate the electroweak interaction, and the gluon[1] is the carrier of the strong interaction. One more particle, know as the Higgs boson, is incorporated in the SM as a result of the spontaneous symmetry breaking mechanism. The gravitational interaction is not included in the SM framework. Figure 1.1 is a scheme of the elementary particles of the SM.

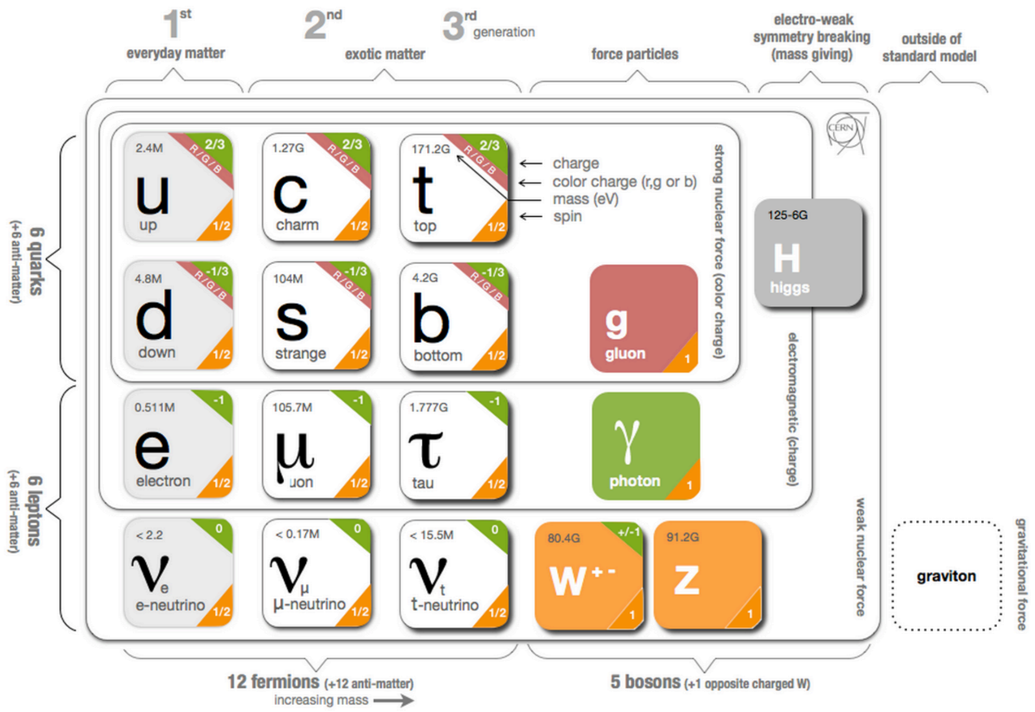


Figure 1.1: Scheme of the elementary particles of the standard model. Taken from [13].

The first generation of quarks and leptons contains the lightest and hence stable elementary particles. The quarks and leptons of the other two generations are heavier and eventually decay into the lightest particles of the first generation.

While leptons experience only the electroweak interactions, quarks are also subject to the strong interaction. This leads to a substantial difference in the observations of the electroweak and the QCD sectors. Quarks and gluons have never

---

[1] In the SM there are eight gluons, each of them with different quantum numbers.

been observed as free particles, as they are “confined” into composite particles known as hadrons. The latter are further classified into baryons, e.g. the proton, and mesons, e.g. the pion.

The SM is a gauge theory formulated in the Lagrangian formalism. The interaction between the fermions is introduced by postulating the invariance or symmetry of the Lagrangian with respect to an arbitrary local gauge transformation of the fields. Initially, the requirement of the Lagrangian’s invariance forces all the gauge bosons to be massless. This implication is in contradiction with the observation of massive W and Z bosons.

Mass generation is restored through the spontaneous symmetry breaking mechanism, i.e. when the set of minimum energy states of the Lagrangian, or vacuum, does not share the gauge symmetry of the Lagrangian. As one such state must be chosen for the quantisation of the Lagrangian, the symmetry is broken, or hidden, by this specific choice. This mechanism is included in the SM by adding a neutral, spin-0 boson field with a non-zero vacuum expectation value. This minimalistic model for the symmetry breaking originates the particle known as the Higgs boson [14–16].

Since its formulation in the late 1960’s the standard model has been capable of predicting an extraordinary amount of experimental results. Two remarkable and relatively recent achievements are the predictions of the top quark, observed at the Tevatron collider [17, 18], and of the W boson mass, measured with sub-percent precision from the combination of Tevatron and LEP collider data [19].

Nonetheless, the standard model has not been exhaustively tested and does not provide a complete picture of the observed Universe. For this reason, the LHC has been built to test the validity of the SM and to search for physics phenomena which may lie beyond it, at an unprecedented energy scale. The research conducted at the LHC can be divided into the three broad categories outlined below.

Higgs boson and spontaneous symmetry breaking

The existence of the standard model Higgs boson was an unconfirmed theoretical model prior to the LHC. Previous experiments attempting to observe this particle were unsuccessful.

On 4 July 2012 the ATLAS and CMS experiments at the LHC announced the direct observation of a new particle compatible with the Higgs boson [20, 21]. Figure 1.2 shows the invariant mass distribution of the di-photon candidates measured by the ATLAS experiment, which was presented in the observation of the new boson in [20]. To date, all the studies concerning the properties of the newly discovered particle are consistent with expectations for the standard model Higgs boson. For example, the ATLAS measurements of the spin and coupling strength are reported in references [22] and [23], respectively.

However, different models to that of the standard model exist to describe the symmetry breaking mechanism. For instance, non-minimal theories encompass the existence of multiple "Higgs-like" bosons or lead the boson associated with the symmetry breaking mechanism to be a composite state. A review of such theories is found in [24].

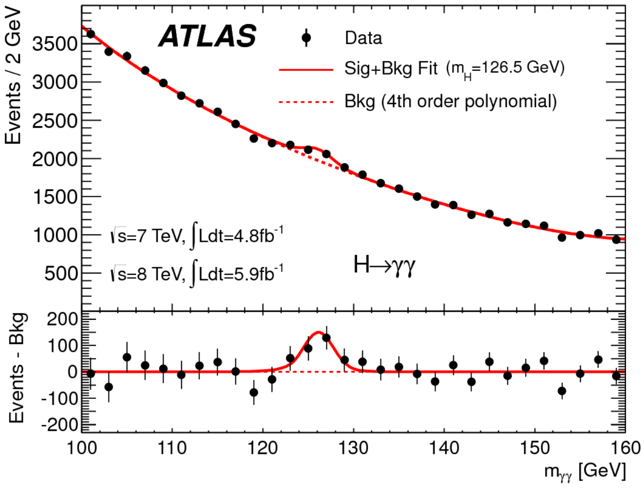


Figure 1.2: Invariant mass distribution of di-photon candidates measured by the ATLAS experiment. The data are fitted to the sum of a signal component with mass 126.5 GeV and the background component. This result, combined with data of other Higgs boson decays, led to the discovery of a new particle compatible with the Higgs boson. Taken from [20].

Searches for phenomena beyond the standard model

Both observations and theoretical arguments suggest that the SM framework is not complete. A prominent example is the dark matter/dark energy problem: cosmological observations indicate that the Universe is composed of dark matter for the 27%, of dark energy for the 68% and of ordinary matter for 5%, where only the latter is included in the SM framework [25].

Searches for phenomena beyond the SM can be classified into two broad sectors: those involving supersymmetric particles and those concerning all other hypothetical particles; the latter are generally referred to as "exotica". Supersymmetry is an extension of the SM which is well motivated on theoretical grounds (hierarchy problem). It introduces a large number of new particles with the same quantum numbers as their SM partners, but differing in spin by half a unit and larger mass. Typical examples of exotica are the existence of heavy gauge bosons Z'/W' or excited quark states. A review of supersymmetric and exotica models is found, for example, in reference [26].

Searches for phenomena beyond the SM with collider data are led by the LHC, owing to its highest energy collisions. To date, no indications of such phenomena have been found. However, searches for phenomena beyond the SM will reach their full exploratory potential in the next decade of LHC operation.

Precision measurements of the standard model

The measurement of W+c production cross section presented in thesis is a precision measurement of the standard model at the LHC.

The LHC also allows the scrutiny of the SM predictions to an higher energy scale than previous experiments. An overview of the SM measurements at the LHC is described, for instance, in reference [27]. By means of precision measurements of the SM processes it is possible to: (i) constrain the parton distribution functions, which describe the proton's structure; (ii) assess methods to calculate high-order theoretical predictions for the electroweak and QCD interactions; (iii) test and "tune" Monte Carlo generators which are adopted for the calculation of the SM predictions. Moreover, as the SM processes constitute the backgrounds for Higgs measurements and for beyond SM searches, their precise knowledge is crucial for the success of the LHC research programme.

1.2 The LHC accelerator

The Large Hadron Collider is built on the site of the European Organisation for Nuclear Research, known as CERN, which operates the world's largest particle physics laboratory in Geneva.

The LHC is described in reference [28] and briefly summarised in this section. This machine has been build primarily to collide two proton beams with nominal values for the centre of mass energy and instantaneous luminosity of $\sqrt{s} = 14$ TeV and $L \sim 10^{34}$ cm^{-2} s^{-1}, respectively, making it the most powerful particle accelerator ever built. The LHC also collides lead ion beams at $\sqrt{s} = 2.7$ TeV per nucleon and $L \sim 10^{27}$ cm^{-2} s^{-1}. Table 1.1 compares the main parameters of the LHC and with those of the previous leading hadron collider, the Tevatron. The latter was a proton-antiproton collider in operation until September 2011 at the Fermi National Accelerator Laboratory in Chicago.

	LHC	**Tevatron**
Centre of mass energy $\sqrt{s}$ (TeV)	14	1.96
Tunnel ring length (km)	26.7	6.3
Instantaneous luminosity L (cm^{-2} s^{-1})	$\sim 10^{34}$	4×10^{32}
Beams hadrons	proton - proton	proton - antiproton
Bunches per beam	2808	36
Distance between bunches	25 ns $\approx$ 7.5 m	396 ns $\approx$ 120 m
Bunch crossing frequency (MHz)	40	2.5

Table 1.1: Comparison of the main parameters of the LHC and Tevatron colliders. The LHC centre of mass energy and luminosity refer to the nominal maximum values. The particles inside the two colliding beams are grouped into "packets" known as bunches.

The LHC is built in a circular tunnel of 27 km circumference at an average depth of 80 m. Before being injected into the LHC itself, the protons undergo several acceleration stages in an accelerator complex whose scheme is displayed in figure 1.3.

The pre-LHC accelerator chain is made up of a linear accelerator (LINAC2), the Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). This chain accelerates proton beams up to 450 GeV. Inside the LHC, the beams are accelerated by 8 radio frequency cavities; the latter determine the structure of the beams as packets of particles which are referred to as *bunches*. The nominal time interval between two bunches is 25 ns which corresponds to a *bunch crossing* frequency of 40 MHz.

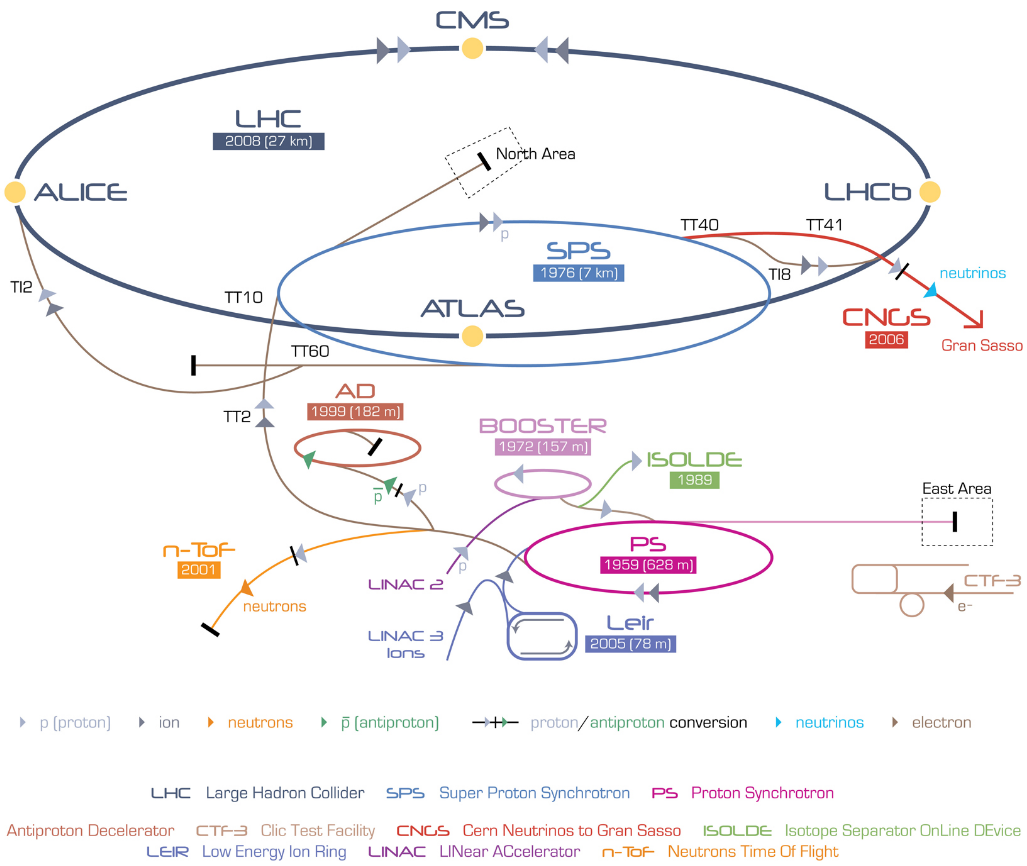


Figure 1.3: Scheme of the LHC accelerator complex. The accelerator chain pre-LHC includes the LINAC2, Booster, PS and SPS. The displacement of the four main LHC experiments is also shown: ATLAS and CMS are the two major multi-purpose detectors, while LHC*b* and ALICE are dedicated to the study of CP-violation and heavy-ion collisions, respectively. Taken from [29].

The particle beams are deflected to maintain a circular trajectory by 1232 superconducting magnetic dipoles, cooled with superfluid helium at a temperature of 1.9 K. The magnets are designed to produce a magnetic field of up to 8.36 T. Before a bunch crossing, the beams are collimated using a total of about 7000 magnetic quadrupoles. This process reduce the dimensions of the bunches to maximise the collision probability.

Four main experiments are displaced in underground caverns around the LHC ring, as shown in figure 1.3. ATLAS (A Toroidal LHC Apparatus) [30] and CMS (Compact Muon Solenoid) [31] are the two major detectors. They are referred to as multi-purpose experiments, being designed for a broad range of different studies. Their research programme and capabilities are similar and they differ in the construction design and detector technologies adopted.

The other two main LHC experiments are LHC*b* (LHC beauty) [32], for CP-violation studies in the b-quark sector, and ALICE (A Large Ion Collider Experiment) [33], dedicated to the study of heavy ion collisions.

The LHC began to deliver sufficiently stable collisions for physics measurements in December 2009. However, the centre of mass energy and instantaneous luminosity so far achieved are lower than the nominal design values. A center of mass energy $\sqrt{s} = 7$ TeV was reached in 2010 and continued into 2011. In 2012, the $\sqrt{s}$ was increased to 8 TeV and ATLAS recorded a maximum instantaneous luminosity of $L = 7.7 \times 10^{33}$ cm^{-2} s^{-1}. The maximum nominal values of energy and instantaneous luminosity are foreseen after 2015. During 2013 and 2014 a LHC shutdown is necessary for machine components upgrade.

To date, the integrated luminosity of the LHC collision data recorded by ATLAS is 5.08 fb^{-1} at $\sqrt{s} = 7$ TeV and 21.3 fb^{-1} at $\sqrt{s} = 8$ TeV. Figure 1.4 shows the integrated luminosity at $\sqrt{s} = 7$ TeV as a function of time. The integrated luminosity recorded by ATLAS is lower than that delivered by the LHC due to data acquisition inefficiencies. Data quality criteria are also applied to the recorded data, thereby further reducing the size of the dataset used for physics analyses.

The full dataset at $\sqrt{s} = 7$ TeV, after data quality criteria are applied, is used in the $W{+}c$ production measurement presented in this thesis. This sample corre-

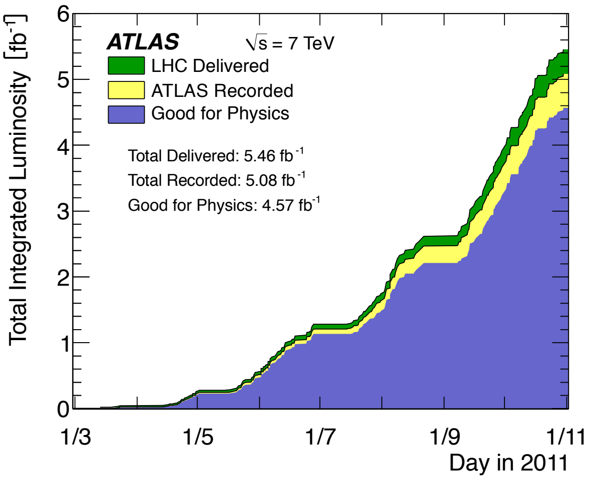


Figure 1.4: Integrated luminosity versus time delivered by LHC (green), recorded by ATLAS (yellow), and passing data quality requirements (blue) for collisions at $\sqrt{s} = 7$ TeV. The latter dataset is used in the $W+c$ measurement presented in this thesis. Taken from [34].

sponds to an integrated luminosity $\int Ldt = 4.57$ fb^{-1}, which has been measured with an uncertainty of 1.8% [34].

1.3 The ATLAS experiment

The ATLAS experiment is the largest particle detector ever built: it has a length of 44 m, a diameter of 25 m and weights approximately 7000 tonnes. The description of the detector is found in reference [30]; a short summary is presented here.

ATLAS is composed of sub-systems as represented in figure 1.5. Three main sub-systems can be identified: the inner detector (ID) for tracking of charged particles, the calorimeters for the energy measurement of electrons, photons and hadrons, and the muon spectrometer (MS) for muon identification and tracking. All the sub-detectors are built with a cylindrical central part known as *barrel*, enclosed at each end by an *end-cap*.

The design of ATLAS has been shaped by the magnet system, which is composed of a 2 T central solenoid enclosing the inner detector and three toroids serving the muon spectrometer. Each of the toroids is build with eight air-core coils symmetrically positioned around the z-axis to form a barrel plus end-caps structure. The toroidal field is 0.5 T and 1 T in the barrel and the end-caps, respectively. The magnet system is built with superconducting wires made of a mixture of aluminium, niobium, titanium and copper, and requires an operating temperature of 4.5 K.

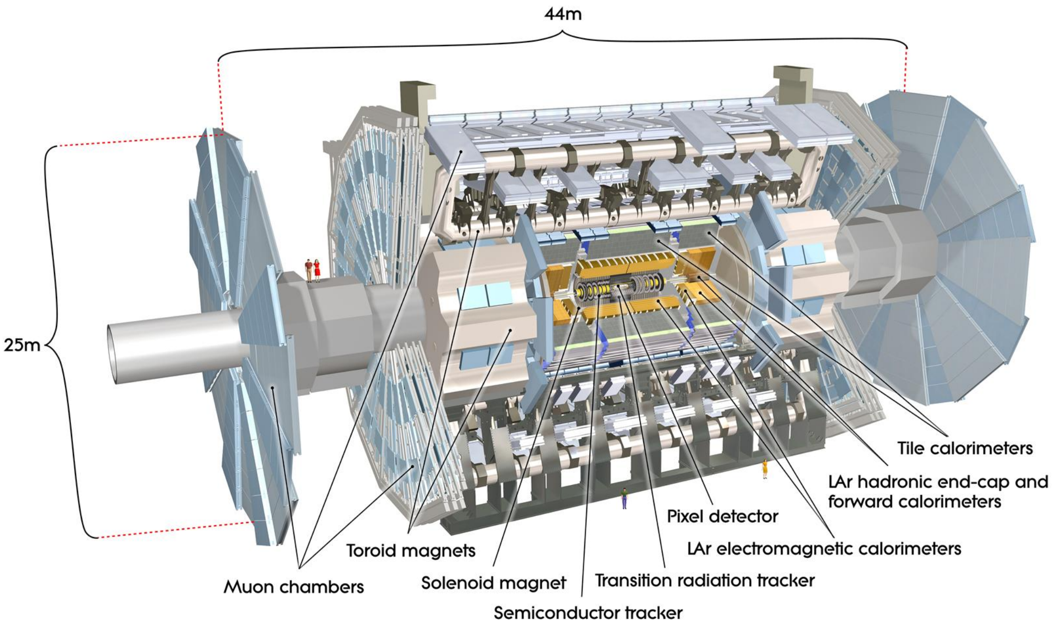


Figure 1.5: Schematic view of the ATLAS detector displaying its sub-system structure. Three main sub-detectors are identifiable: the inner detector (pixel detector, semiconductor detectors and transition detector tracker) is located at the core, surrounded by the calorimeters (electromagnetic, hadronic and forward) which is enclosed by the muon spectrometer. The magnet system is composed by a central solenoid enclosing the inner detector and three toroids serving the muon spectrometer. Taken from [30].

1.3.1 Coordinate system and kinematic variables definitions

The origin of the ATLAS coordinate system coincides with the nominal point where the LHC beams undergo a bunch-crossing, which is referred to as the nominal *interaction point* (IP). The coordinate system is right-handed, with the x-axis pointing towards the centre of the LHC ring and the y-axis upwards, as shown in figure 1.6 (a). ATLAS is nominally symmetric in the positive and negative

z-direction.

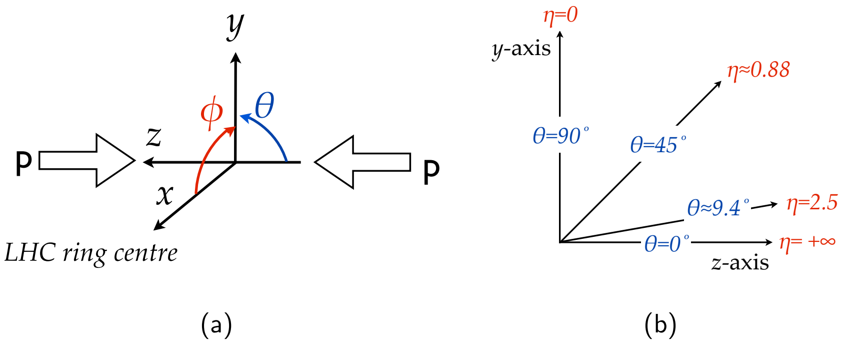


Figure 1.6: Coordinate reference system adopted by the ATLAS detector: (a) is the cartesian reference system and (b) illustrates the relation between the pseudorapidity η and the polar angle θ.

In the coordinate system, the azimuthal direction around the z-axis is denoted by the angle ϕ, which ranges from $-\pi$ to π the x-axis being $\phi = 0$. The variable referred to as *pseudorapidity*,

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{1.3.1}$$

is used to identify the direction with respect to the z-axis. In equation 1.3.1, θ is the polar angle, which lies in the range $]0, \pi]$ starting from the positive z-axis. The relation between η and θ is illustrated in figure 1.6 (b). The pseudorapidity and polar angle are used to define the distance in the η–ϕ space:

$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}\,. \tag{1.3.2}$$

In an inelastic hadron-hadron collision, two of the partons undergo the scattering interaction. Although the longitudinal momentum, p_z, of these partons in the initial state is unknown, their *transverse momentum*, p_{T}, perpendicular to the z-axis, is approximately null. Hence the conservation of energy and momentum can be exploited only in the transverse plane of the detector and the transverse momentum and *transverse energy*, E_{T}, are extensively used in ATLAS. These two variables are defined as

$$\begin{aligned} p_{\mathrm{T}} &= |\mathbf{p}_{\mathrm{T}}| = \sqrt{p_x^2 + p_y^2} \\ E_{\mathrm{T}} &= E\sin\theta\,. \end{aligned} \tag{1.3.3}$$

Accordingly the *missing transverse energy*, $E_{\mathrm{T}}^{\mathrm{miss}}$, and its azimuthal angle, ϕ^{miss}, are defined from the transverse momentum imbalance as

$$\begin{aligned} E_{\mathrm{T}}^{\mathrm{miss}} &= |\mathbf{E}_{\mathbf{T}}^{\mathbf{miss}}| = \left| -\sum_i \mathbf{p_T}^i \right| \\ \phi^{\mathrm{miss}} &= \arctan(E_{\mathrm{x}}^{\mathrm{miss}} / E_{\mathrm{y}}^{\mathrm{miss}})\,, \end{aligned} \tag{1.3.4}$$

where $\mathbf{E}_{\mathbf{T}}^{\mathbf{miss}}$ is vector sum of the transverse momenta of all detected particles a collision.

1.3.2 Inner detector

The ID is located at the core of ATLAS surrounding the IP. It is surrounded by the solenoidal magnet, has a diameter of 2.1 m and a length of 6.2 m which corresponds to a coverage of $|\eta| < 2.5$. The ID is composed of three sub-detectors which successively envelope each other as shown in figure 1.7. They are known as the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT).

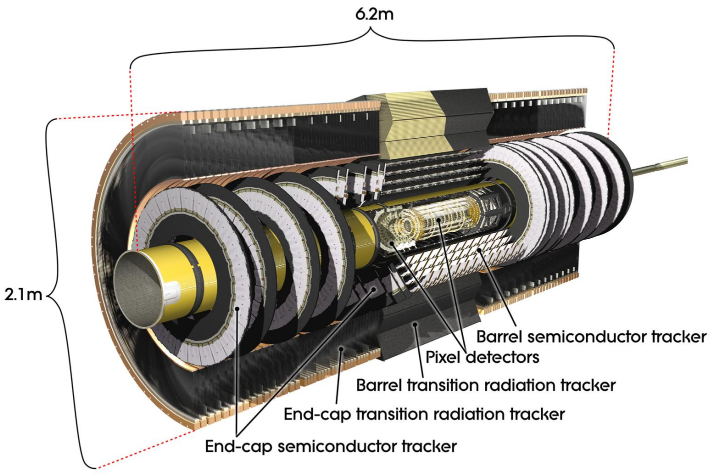


Figure 1.7: Scheme of the ATLAS inner detector displaying its three sub-detectors: the pixel detector, the semiconductor tracker and the transition radiation tracker. Taken from [30].

The pixel detector has the highest granularity[2] and its innermost part lies at 5 cm from the IP. Pixels are silicon sensors which allow a precision on the determi-

---

[2] The granularity is the number of readout channel per unit of solid angle.

nation of the hit position of 10 μm in the r-ϕ plane and 115 μm in the z direction. The pixel detector has a total of 80 million read-out channels.

The SCT is composed of silicon micro-strip detectors which provide a resolution for the measurement of the hit position of 17 μm in the r-ϕ plane and 580 μm in the z direction.

The TRT is a gaseous detector which, besides tracking, aids with the identification of electrons. The TRT is made of straw-tubes filled with a Xenon-based gas mixture and the gaps between these straws are filled by radiator material. A particle crossing the detector emits transition radiation that contributes to the identification of the electrons. The TRT coverage in pseudorapidity is $|\eta| < 2.0$ and the measurement of the hit position, available only in the r-ϕ plane, has a resolution of 130 μm.

The tracks are reconstructed from the ID measurements by combining the hits from its three subsystems. The particles' momenta are measured with a resolution of approximately $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\% \cdot p_\mathrm{T} \oplus 1\%$, where p_T is in GeV. In the formula, the first term represents the uncertainty on the track curvature while the second term includes all other effects, which are dominated by multiple scattering within the detector material.

1.3.3 Calorimeters

The ATLAS calorimetric system encloses the inner detector and covers a region of $|\eta| < 4.9$. The calorimeters have a twofold function: (i) to contain and measure the energy of all the particles interacting electromagnetically and/or strongly except for muons and neutrinos; (ii) to serve as trigger detectors.

ATLAS employs three calorimeters: electromagnetic (EM), hadronic (HAD) and forward (FCal), whose scheme is shown in section 1.3.3. All these sub-detectors are sampling calorimeters, i.e. composed of absorber material interlaid with active material. The absorber is a dense material with high stopping power, while the active part measures the energy deposited by the interacting particles.

The EM calorimeter is designed to measure the energy of electrons and photons. It is composed of lead and stainless steel as the absorbing materials and liquid

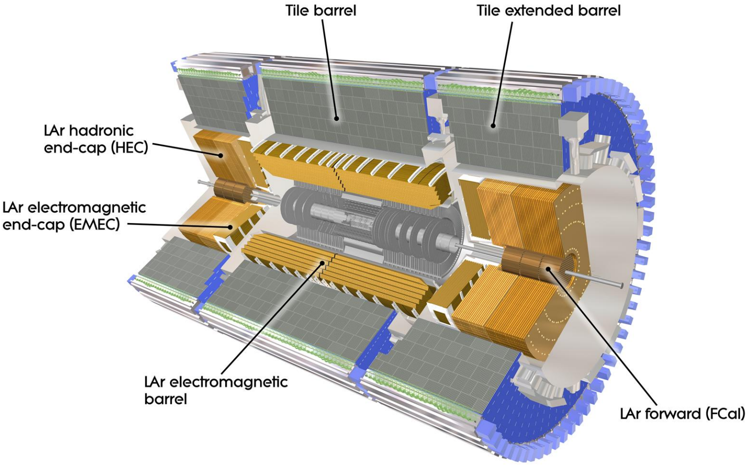


Figure 1.8: Scheme of the ATLAS calorimetric systems composed of the electromagnetic, hadronic and forward calorimeters. Taken from [30].

argon (LAr) as the active medium. It extends up to $|\eta| = 3.2$ with several active layers in depth: three in the central region $0 < |\eta| < 2.5$ and two in the high pseudorapidity region $2.5 < |\eta| < 3.2$. In the central region, a more precise position measurement is obtained by finely segmenting the first layer, closest to the interaction point. In the region $0 < |\eta| < 1.8$ the EM calorimeter is complemented by a presampler consisting of a LAr layer which measures the energy lost due to the material upstream of the calorimeter.

The energy resolution provided by the EM calorimeter is approximately $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$, where E is in GeV. In the formula, the first term is the stochastic uncertainty on the energy measurement and the second term accounts for all other effects, which are dominated by calibration and detector non-uniformity.

The HAD calorimeter is designed to measure the energy of hadrons. It surrounds the EM calorimeter and extends up to $|\eta| < 3.2$. Iron absorber plates with plastic scintillating tiles as active medium are used in the central region $|\eta| < 1.7$, which is known as *TileCal*. The forward region, $1.7 < |\eta| < 3.2$, is instrumented with copper plates and LAr. The energy resolution provided by the HAD calorimeter is approximately $\sigma_E/E \approx 50\%/\sqrt{E} \oplus 3\%$. The same considerations as per EM calorimeter are valid for the terms in the relation.

The calorimetry system is completed by a radiation-resistant forward calorimeter covering the region $3.1 < |\eta| < 4.9$. This detector is divided into three parts positioned consecutively along the beam pipe. The first part serves as an electromagnetic calorimeter with a copper absorber and the other two parts are hadronic calorimeters with a tungsten absorber. All parts use LAr as the active medium. The energy resolution of the FCal is lower than that of the HAD calorimeter. Nonetheless the coverage in η provided by FCal improves the precision of the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement.

1.3.4 Muon spectrometer

The MS completes the ATLAS detector. It is designed to track charged particles exiting the calorimeters, which are nominally only muons. The spectrometer covers a pseudorapidity range $|\eta| < 2.7$. The MS is also used to trigger on muons in the region $|\eta| < 2.4$. At $\eta \approx 0$ there is no MS coverage due to the passage of the "service cables" of the ID, solenoid magnet and calorimeters, which provide, for example, sensors read-out and voltage supplies. The bending power for the spectrometer is provided by the toroid magnets which deflect the muon trajectories mostly in the z-direction.

The MS, whose scheme is shown in figure 1.9, includes four types of gaseous detector chambers with different performance in terms of precision and response speed, which are accordingly employed for tracking or triggering.

Precision tracking chambers are located in the magnet coils. The pseudorapidity range $|\eta| < 2.7$ is instrumented with monitored drift tube chambers (MDT). An exception to this is in the innermost layer of the forward region, $2 < |\eta| < 2.7$, where cathode strip chambers (CSC) are used due to their higher radiation resistance. The resolution of the measurements provided by the MDT is 80 μm, along the z-axis for the barrel and along the radial direction in the end-caps. The CSC provides a position resolution of 40 μm along the z-axis and approximately 5 mm in the transverse plane. The spectrometer operating in stand-alone mode, i.e. without the inner detector information, allows transverse momentum measurements ranging, from 3 GeV up to 3 TeV and with a resolution of about 2% for p_{T} of 50 GeV.

The resistive plate chambers (RPC) in the barrel $|\eta| < 1.05$ and the thin gap chambers (TGC) in the end-caps $1.05 < |\eta| < 2.4$ serve as trigger detectors. These are designed for fast response: the RPCs have a time resolution of 1.5 ns and the TGCs of 4 ns. Moreover, the RPC and TGC provide the measurement of the muon coordinate in the direction orthogonal to that determined by the precision tracking chambers.

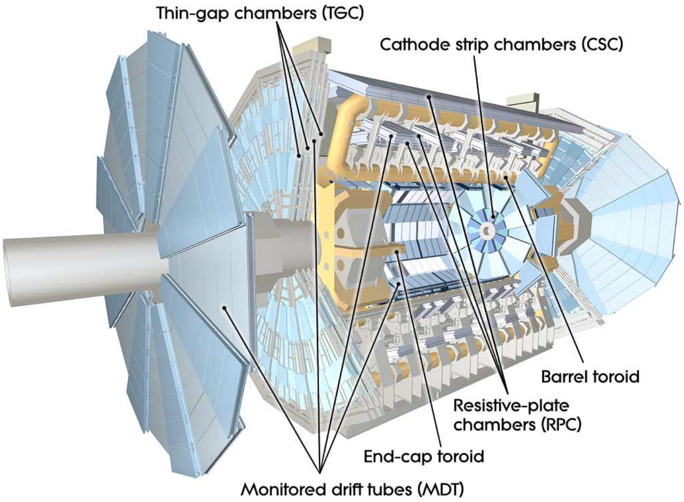


Figure 1.9: Scheme of the ATLAS muon spectrometer. The MDT and CSC chambers are used to measure the muons' momentum. The TGC and RPC chambers are used as muon triggers. Taken from [30].

1.3.5 Trigger and data acquisition systems

The ATLAS detector employs a three-level system trigger to select collision events according to specific pre-defined criteria. The system is made up of the level-1 (L1), level-2 (L2) and event filter (EF) triggers. Each subsequent level refines the decisions made at the previous stage and, where necessary, applies additional selection criteria. The ATLAS trigger system reduces the LHC nominal collision rate of 40 MHz down to a rate of 200 Hz, which is the maximum rate at which the data can written to tape.

The L1 trigger consists of custom-made hardware. It selects events relying on reduced-granularity information from the calorimeters and the muon trigger chambers. The L1-trigger has a maximum latency[3] of 2.5 μs and a maximum accept rate of 75 kHz, the latter being limited by the ATLAS sub-systems read-out electronics. The data from the read-out electronics of each of the ATLAS sub-detectors are initially stored in buffers at the L1 trigger accept rate. If an event is selected by the L1 trigger, the collision data recorded by ATLAS are transferred off the detector to the data acquisition system. The L1 trigger also identifies the regions of interest (RoIs) of the accepted events. The RoIs are the η–ϕ regions in ATLAS where the trigger selection process has identified the event features matching the selection criteria.

Chapter 2 presents a study of the time stability of a subsystem of the L1 trigger. This analysis has been performed by the author of this thesis as a direct contribution to the operation and maintenance of the ATLAS detector.

The L2 trigger is based on software algorithms which analyse only data associated with the RoIs. At this stage, the trigger system employs the full granularity information from all the ATLAS sub-detectors. The L2 trigger has an average latency of 40 ms and further reduces the event accept rate below 5 kHz.

When a collision event is accepted by the L2 trigger, the complete read-out information from all ATLAS sub-detectors is passed to the event builder module. Here, the detector information is analysed by the software algorithms of the EF trigger, which have an average output rate of 200 Hz. The events passing the EF trigger selection are finally stored to disk.

---

[3] The latency is the time the trigger requires to make a decision.

Chapter 2

Time stability study of the L1Calo trigger calibration

This chapter presents a study of the calibration of the ATLAS level-1 calorimeter (L1Calo) trigger. Specifically, a set of quantities used for the calibration of the L1Calo trigger preprocessor is analysed for their systematic variations over time. Systematic trends over time can highlight problems in the calibration or the functioning of the trigger system. This analysis is part of the studies of the L1Calo trigger preprocessor performances whose summary is published in [3].

The author of this thesis implemented a procedure to “clean” data samples and to evaluate trends over time. This software package has been included in the L1Calo trigger software framework. The results of the analysis indicate that systematic trends of the calibration variables considered are generally at a level of 0.1% or smaller.

2.1 ATLAS L1 trigger

The ATLAS three-level trigger system is briefly described in section 1.3.5. The first level, known as the L1 trigger, is responsible for the initial event selection based on information from the calorimeters and muon trigger system. The L1 trigger consists of custom-made electronic modules which are able to withstand the LHC bunch crossing rate of 40 MHz.

A schematic representation of the L1 trigger and its main components, the L1 calorimeter trigger, L1 muon (L1Muon) trigger and central trigger processor, is shown in figure 2.1. The L1Calo and L1Muon triggers receive and process the analog signals from the respective detectors. The central trigger processor collects the processed information and, if predefined selection criteria are met, issues the L1 trigger accept signal.

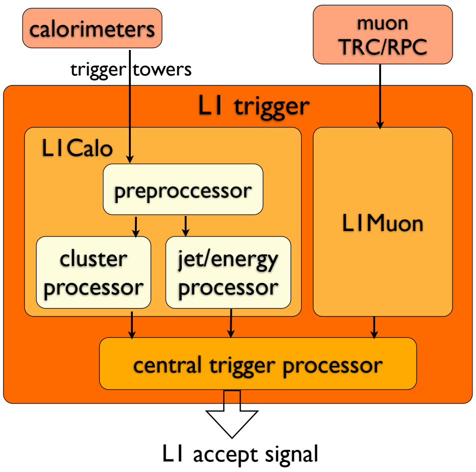


Figure 2.1: Schematic representation of the ATLAS L1 trigger and its three main subsystems, the L1 calorimeter trigger (L1Calo), the L1 muon trigger (L1Muon) and the central trigger processor. The main components of the L1Calo trigger, the preprocessor, the cluster processor and the jet/energy processor, are also shown.

The ATLAS calorimeters, which are briefly described in section 1.3.3, have a total of about 200,000 cells. The L1Calo trigger cannot processes all their information simultaneously due to the stringent requirements on the L1 trigger latency. Therefore the analog signals from multiple calorimeter cells are summed within the detector's front-end electronics to form 7168 analog signals, which are referred to as *trigger towers* (TTs) or *channels*. One TT can be the sum of up to 60 single calorimetric cells, and has typical dimensions $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$ which increases in the $|\eta| > 2.5$ region. There are independent TTs for the EM and the HAD layers of the calorimetric system.

The L1Calo trigger is designed to identify high-E_{T} objects, such as electrons, photons, τ-leptons and jets, and events with large magnitude of E_{T} or $E_{\mathrm{T}}^{\mathrm{miss}}$. This task is divided amongst its three main components: the *preprocessor* (PP), the cluster processor and the jet/energy processor. The TTs' analog signals are passed to the PP to be processed, digitised and converted into GeV units. The results of the PP are sent in parallel to the cluster and jet/energy processors, which run the algorithms to identify high-E_T objects or values, whose η–ϕ coordinates constitute the regions of interest and are transmitted to the L2 trigger.

2.2 Calibration of the L1Calo trigger preprocessor

The pedestals, energy conversion and timing of the PP input analog signals need to be calibrated. Dedicated data taking runs for this calibration occur on a regular basis when ATLAS is not operating in physics data acquisition mode. With these datasets the PP is calibrated and its functioning is monitored.

There are five types of such PP calibration runs, whose description is found in reference [3]. Three of them, which are referred to as the DAC scans, pedestal scans and energy scans, respectively, are used in the analysis presented here and are summarised below. The remaining types of PP calibration runs concern the timing of the input signals with respect to the LHC bunch crossing.

DAC scans

Digital-to-analog converters (DACs) within the PP produce a voltage level which is added to the input analog signals of each TT. Such an offset is necessary to ensure that the analog signals are properly digitised by falling within the linear digitisation window of the analog-to-digital converters (ADCs) present in the PP. Hence, the *pedestal* measured by each ADC, i.e. the value it records when there are no input signals, has a linear dependence from the voltage set in the corresponding DAC.

In a scan, for each of the TTs, the DAC voltage is varied and the ADC pedestal is measured. From this, the slope and offset of each of the ADC-DAC linear relationships are evaluated.

These results are used to recalibrate the DAC voltage in each of the TTs so that the pedestal corresponds to the default value of 32 ADC units [1]. In addition,

---

[1] In the L1Calo trigger, an input signal of 1 ADC unit corresponds to an energy of 0.25 GeV.

the DAC scans can be used to identify problems, e.g. non-linearity in the ADCs or DACs response. The DAC scans are performed typically 6 times per month. Since signals of the TTs are not required in these calibration runs, the inputs of the PP are disabled to suppress any noise coming from the upstream electronics.

Pedestal scans

The pedestal scans measure the distribution of the pedestal for each of the TTs (at the default DAC voltage). These calibration runs allow the identification of problematic channels, which show multi-peaked or broad pedestal distributions.

A pedestal scan is run typically after a DAC scan. Similarly to the DAC scans, the inputs to the PP are disabled, and a pedestal scan is typically run after the DAC scan.

Energy scans

The energy scans allow the comparison of the transverse energy measured in the PP ($E_{\mathrm{T}}^{\mathrm{L1Calo}}$) with that reconstructed from the calorimeters' full information ($E_{\mathrm{T}}^{\mathrm{calo}}$). The former is used in the online trigger algorithms and the latter, which is more accurate, in the offline analysis.

Pulse test signals at multiple fixed energies are produced in the calorimeters for the energy scans. At each of these energy points, measurements are taken in the PP and the linear relationship between $E_{\mathrm{T}}^{\mathrm{L1Calo}}$ and $E_{\mathrm{T}}^{\mathrm{calo}}$ is extracted. Consequently, gain factors are set in the PP modules to adjust $E_{\mathrm{T}}^{\mathrm{L1Calo}}$ to $E_{\mathrm{T}}^{\mathrm{calo}}$ for each of the TTs.

2.3 Study of the time stability

The time stability of the PP calibration is studied because it can highlight problems in the calibration or functioning of this trigger system. The systematic variation over time of five variables, which are referred to as *attributes*, is studied using data of calibration runs described above. The five attributes are: (i) the slope and the offset of the linear relationship between pedestal and DAC voltage extracted from the DAC scans; (ii) the mean of the pedestal distribution measured in the pedestal scans; (iii) the slope and the offset of the $E_{\mathrm{T}}^{\mathrm{L1Calo}}$–$E_{\mathrm{T}}^{\mathrm{calo}}$ linear relationship evaluated in the energy scans. The L1Calo PP calibration data taken in 2012 are considered.

A software package, which is referred to as *calibtools*, has been developed by the author within the L1Calo software framework. The calibtools algorithm takes as input a set of attributes and a time interval, retrieves the selected dataset from the L1Calo calibration database and executes the analysis described below, which is described for the case of the pedestal mean data.

2.3.1 Time series and residuals

The *time series* of a channel c, $x_c(t)$, is an attribute's value measured at the trigger tower c as a function of time. Each point in the time series corresponds to the calibration run at time t. The *time average* of the channel c is calculated as

$$\overline{x}_c = \frac{1}{N} \sum_{t=t_1}^{t_N} x_c(t) \, , \tag{2.3.1}$$

where N is the number of points in the time series. The *residual* of the channel c at the time t,

$$res_c(t) = \frac{x_c(t) - \overline{x}_c}{\overline{x}_c} , \tag{2.3.2}$$

evaluates the variation of the attribute's value relative to the time average for the channel considered. A check of the data consistency is executed by calibtools and "dead" channels registering $\overline{x}_c = 0$ are excluded.

An example of a time series is shown in figure 2.2. The values of $x_c(t)$ at all time steps are close to the expected value of 32 ADC counts; the time average, $\overline{x}_c$, results 32.3 ADC counts.

The distribution across all channels of $\overline{x}_c$ and of the time series' root mean square (RMS), maximum and minimum values are evaluated by calibtools. Examples are shown in figure 2.3 for the pedestal mean analysis. The x-axis range of each histogram corresponds exactly to the measured range. For example, in figure 2.3 (a) the channels with the smallest and the largest $\overline{x}_c$ measure approximately 17 and 34 ADC counts, respectively.

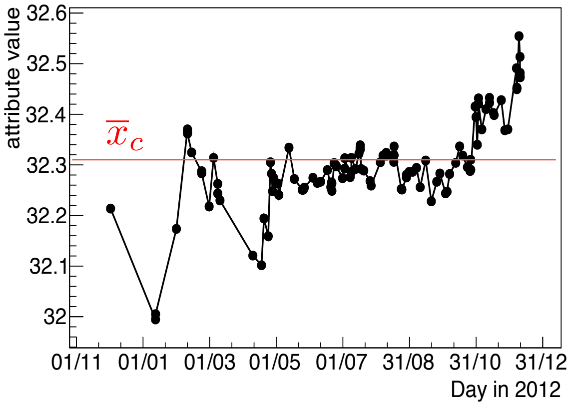


Figure 2.2: Example of a channel time series for the analysis of the pedestal mean data. The time average, $\overline{x}_c$, is displayed by the line.

The same distributions are also plotted in η–ϕ maps of the TT coordinates in the calorimeter. Two sets of maps are produced, for the calorimeter's EM and the HAD layers, respectively. An example of the EM layer maps for the pedestal mean analysis is shown in figure 2.4. The $\overline{x}_c$ values and those of the time series RMS, maximum and minimum values are generally uniformly distributed in η–ϕ and only a few outlier values are present.

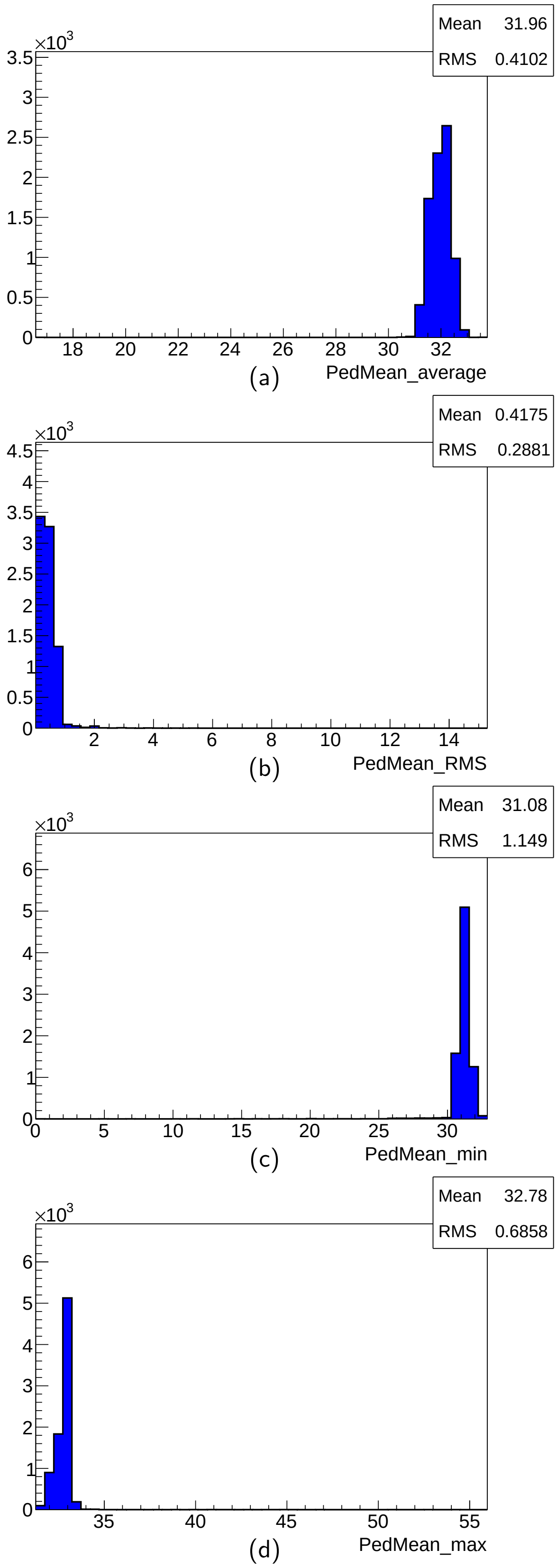
The varying granularity of the TTs is also visible in the η–ϕ maps. The channels' $\Delta\phi \times \Delta\eta$ size in the $|\eta| < 2.5$ and the $2.5 < |\eta| < 3.2$ regions is 0.1×0.1 and 0.2×0.2, respectively. In the forward region $3.2 < |\eta| < 4.9$, where FCal operates as described in section 1.3.3, the dimensions of a TT are $\Delta\phi \times \Delta\eta = 0.4 \times 0.4$.

2.3.2 Average residuals and data cleaning

The *average residuals* across all channels,

$$\overline{res}(t) = \frac{1}{N_c(t)} \sum_{c=1}^{N_c(t)} res_c(t)\,, \tag{2.3.3}$$

evaluate the average variation at time t of each channel value relative to its $\overline{x}_c$. The number of channels which have a measurement at t, $N_c(t)$, is a time dependent variable. In fact, temporary failures might occur in the measurement of an attribute, due, for instance, to malfunctioning of the electronic modules of a



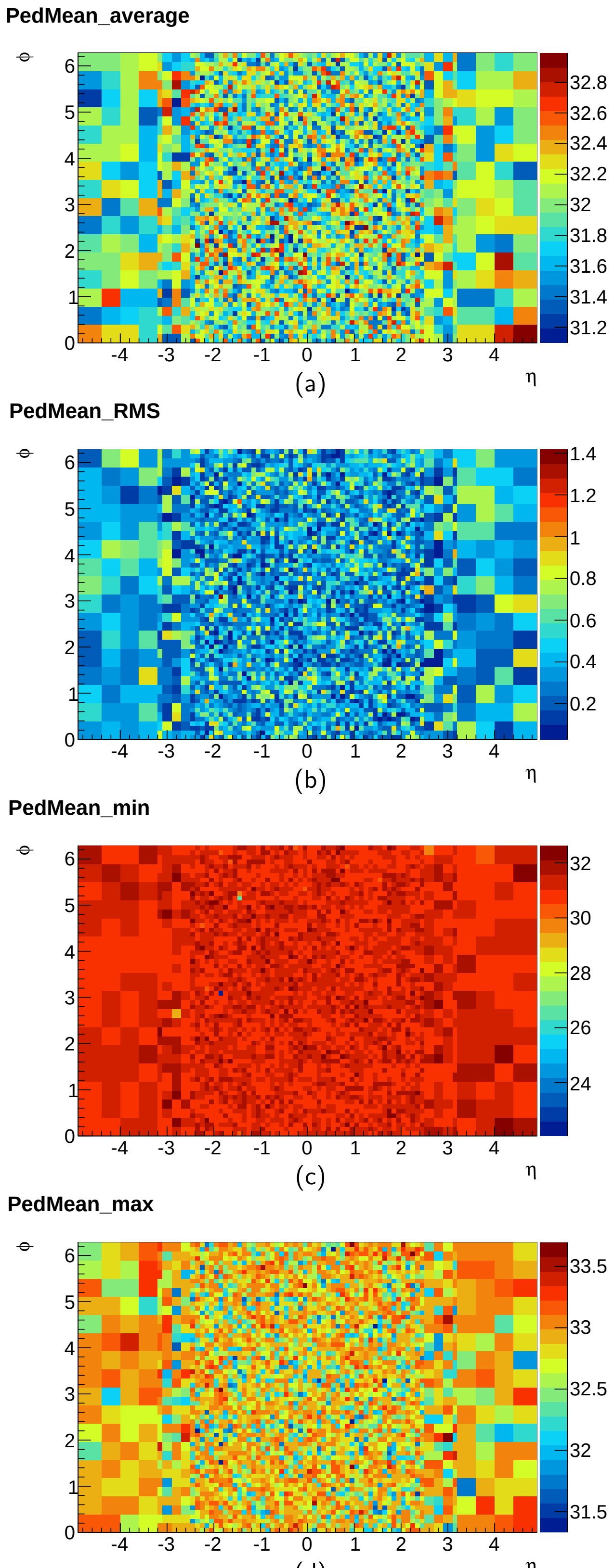
(a)

(b)

(c)

(d)

Figure 2.3: Distributions across all channels of (a) $\overline{x}_c$, and of the time series (b) RMS, (c) minimum and (d) maximum values, for the analysis of the pedestal mean data. The x-axis range in each histogram covers the range of the measured quantity for all channels.



(a)

(b)

(c)

(d)

Figure 2.4: Maps in the channel η–ϕ coordinates for (a) $\overline{x}_c$, and the time series (b) RMS, (c) minimum and (d) maximum values. Channels of the EM layer of the calorimeter for the pedestal mean data have shown here.

channel.

The average residuals are used to identify possible systematic shifts of the attribute value over time. Figure 2.5 shows the $\overline{res}(t)$ as a function of time which is obtained from the pedestal mean analysis. A linear regression using the line

$$\overline{res}(t) = a_0 + a_1 \frac{t}{t_N - t_1} \tag{2.3.4}$$

is calculated, where $t_N - t_1$ is the interval corresponding to the analysed time series. [2] The slope of the regression line is taken as the magnitude of the attribute's systematic variation over the time interval considered. For instance, the systematic variation of the pedestal mean value shown in figure 2.5 is 0.22%.

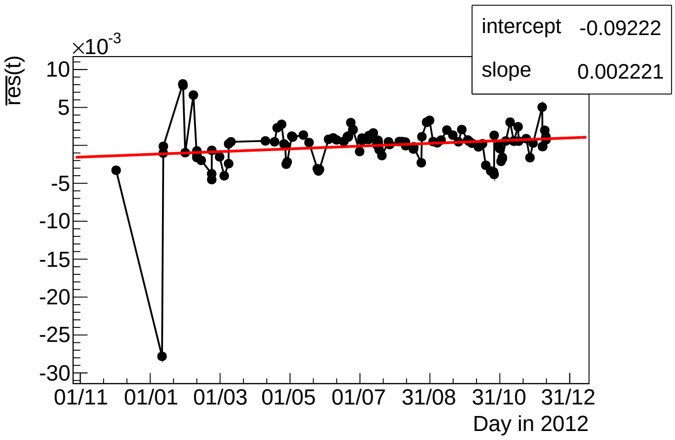


Figure 2.5: Average residuals as a function of time for the pedestal mean analysis. The slope of the regression line indicates the magnitude of the attribute's systematic variation over the time interval analysed. All points before the data cleaning procedure explained in the text are considered.

A single calibration run might measure a large average residual due to only few single points of the times series. The example in figure 2.5 shows few points with $\overline{res}(t)$ at percent level, while the majority of points have $\overline{res}(t) \sim 10^{-3}$. Single measurements with large residuals might occur, for example, due to malfunctioning or human mistake in the calibration runs.

---

[2] The regression line coefficients calculated with least squares are $a_1 = Cov\left(\overline{res}(t), t\right) / Var(t)$ and $a_0 = \langle \overline{res}(t) \rangle - a_1 \langle t \rangle$, where the operators $\langle\ \rangle$, $Cov()$ and $Var()$ indicate the sample mean, covariance and variance, respectively.

A *data cleaning* procedure is applied in the analysis to reject the outlier points of the $\overline{res}(t)$ distribution,

$$\overline{res}(t) \notin [m \pm \delta] \ , \tag{2.3.5}$$

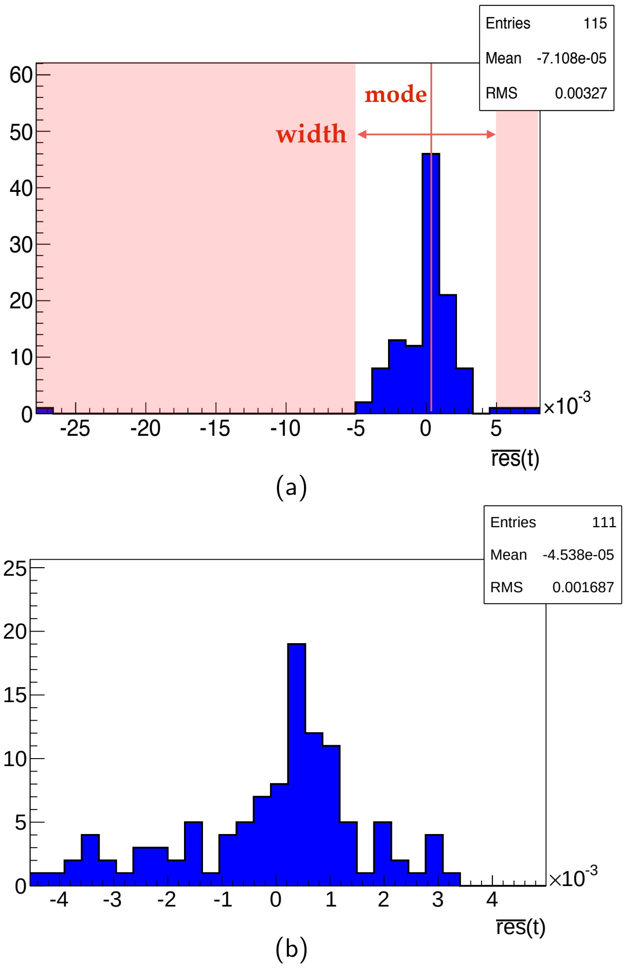
where m and δ measure the distribution's "peak" and width, respectively.

The value m is taken to be the distribution's mode, i.e. the central value of the bin with the largest number of entries in the $\overline{res}(t)$ histogram. The width is calculated as $\delta = u \cdot w_{\text{bin}}$, where w_{bin} is the width of the bins in $\overline{res}(t)$ histogram and u is a defined number of such bins. The values of u and w_{bin} are user defined inputs in calibtools, which can be used to control sensitivity and magnitude of the data cleaning executed by the program. A different evaluation of the outlier points was attempted, with m and δ set to the mean and the RMS of the $\overline{res}(t)$ distribution, respectively. This alternative method generally does not work because the mean and RMS values are sensitive and hence "pulled" to the outlier points.

As an example, in figure 2.6 the $\overline{res}(t)$ distribution is shown (a) before and (b) after the data cleaning, for the pedestal mean analysis. The values of m and δ are displayed with lines, and the shaded area encompasses the resulting 4 outlier points.

Calibtools repeats the entire calculation of the average residuals after executing the data cleaning procedure, and the resulting regression line is taken as the analysis final result. Figure 2.7 shows the values of $\overline{res}(t)$ as a function of time for the analysis of the pedestal mean, after the exclusion of the outlier points in figure 2.5. The regression line after data cleaning gives a systematic variation of the pedestal mean value of 0.19%.

The final results of the analysis are considered after data cleaning because this procedure can change the evaluated systematic variation by many orders of magnitude. For instance, in the analysis of the slope measured from the DAC scans, the linear regression slope coefficient before and after this procedure is -0.067 and 1.5×10^{-8}, respectively, as shown in figure 2.8.



(a)

(b)

Figure 2.6: Distribution of the $\overline{res}(t)$, for the pedestal mean analysis, (a) before and (b) after the data cleaning procedure. The red area encompasses 4 outlier points, which are excluded by applying equation 2.3.5.

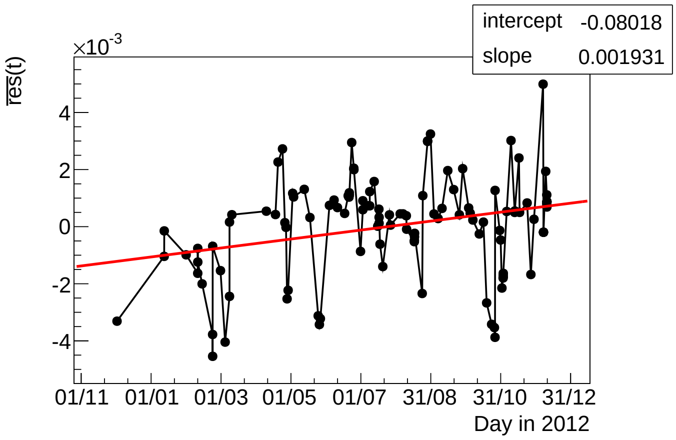
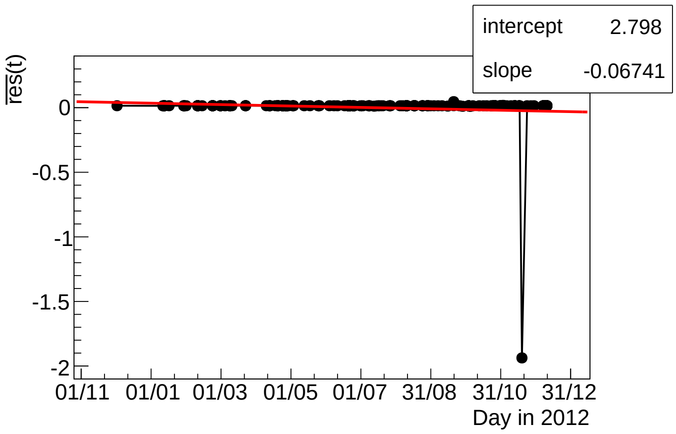
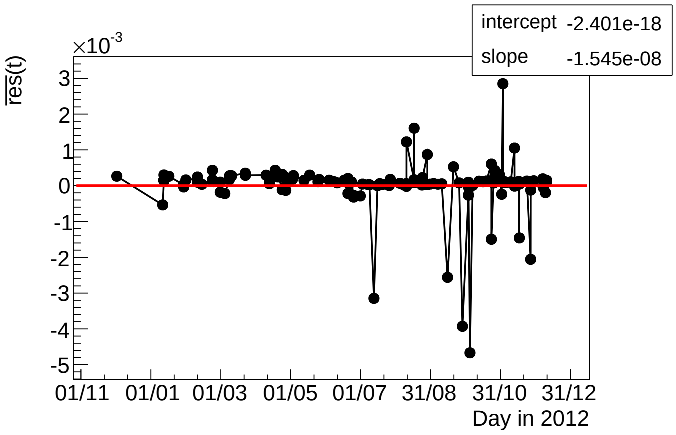


Figure 2.7: Average residuals as a function of time for the pedestal mean analysis, after the data cleaning procedure. The slope of the regression line is taken as the magnitude of the attribute's systematic variation over the time interval analysed.



(a)



(b)

Figure 2.8: Average residuals as a function of time for analysis of the DAC scans slope, (a) before and (b) after the data cleaning procedure. This procedure results in a change of the evaluated linear regression coefficient from -0.067 to 1.5×10^{-8}.

2.3.3 Results

The final results obtained for the pedestal mean, and for the slope and the offset measured in the DAC scans, are presented in figures 2.7, 2.8 (b) and 2.9, respectively. The systematic variation of these attributes evaluated from the 2012 PP calibration runs data is at the order 10^{-3} or less. This magnitude of systematic trend has a negligible effect on the performance of the ATLAS trigger system and can be associated, for example, to changes in the environmental conditions of the ATLAS cavern or to "ageing" effects in hardware modules.

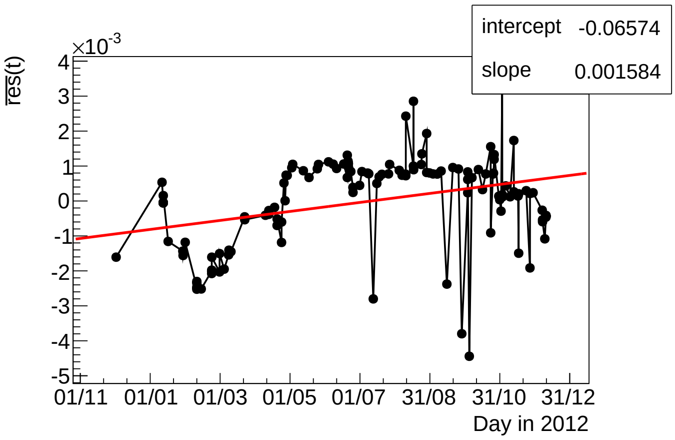
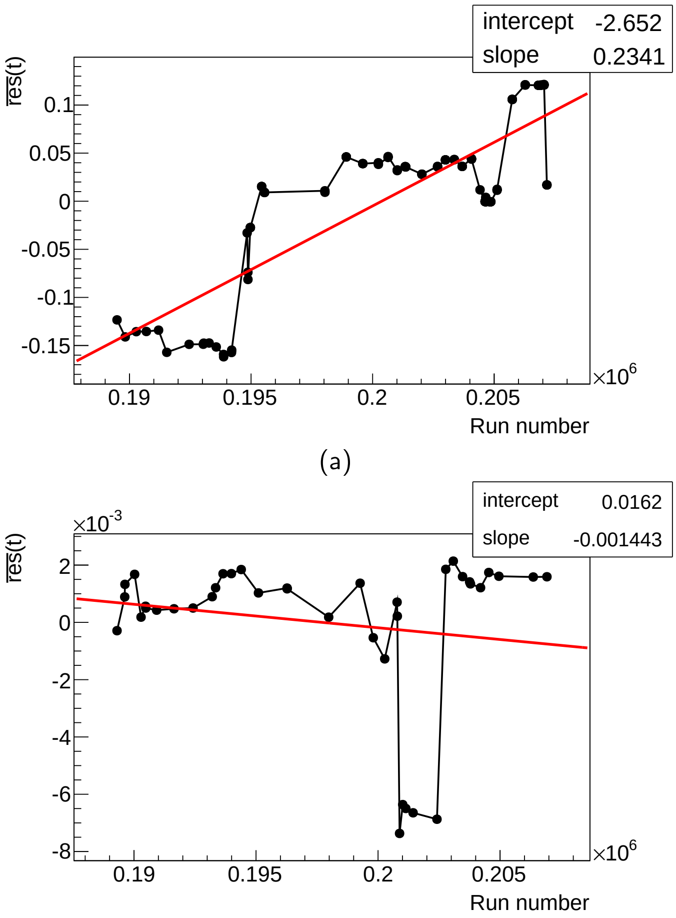


Figure 2.9: Average residuals as a function of time for the DAC scans offset data.

The results obtained for the slope of the $E_{\mathrm{T}}^{\mathrm{L1Calo}}$–$E_{\mathrm{T}}^{\mathrm{calo}}$ linear relationship are shown in figures 2.10 (a) and (b), for TileCal and the LAr instrumented calorimeters, respectively. As described in section 1.3.3, the hadronic calorimeter in the region $|\eta| < 1.7$ is instrumented with plastic scintillator tiles while the rest of the calorimetric systems with liquid argon. The time values are measured by run number in this case, as the energy runs occur during LHC luminosity fills, i.e. when the proton beams are injected into the accelerator. A luminosity fill in the LHC is associated with run number in the ATLAS data acquisition system.

A systematic increase of the attribute's value of approximately 20% can be seen in figure 2.10 (a). The $\overline{res}(t)$ values can be approximately subdivided into three plateau regions, which differ by approximately 10–15% from one to another. The origin of such systematic shift has been traced to the periodical recalibration of the



(a)

(b)

Figure 2.10: Average residuals as a function of time for the slope measured in the energy scans. Result for the TileCal and the LAr instrumented calorimeters are shown in (a) and (b), respectively, with the regression lines obtained. In (a) a systematic trend of approximately 20% is visible.

TileCal subdetector, which is required due to the deterioration of the scintillator material caused by the LHC collisions' radiation [35]. The L1Calo trigger was not synchronised with these TileCal detector re-calibrations hence causing the systematically different measurements of $E_{\mathrm{T}}^{\mathrm{L1Calo}}$ and $E_{\mathrm{T}}^{\mathrm{calo}}$. This misfunctioning was estimated to have an impact of the order of the percent on the performance of the ATLAS trigger and corrected for the future running (i.e. L1Calo trigger and TileCal re-calibrations are now synchronised).

Systematic variations of the energy scans slope of the LAr instrumented calorimeters, shown in figure 2.10 (b), are instead at the level of 10^{-3}. Data were also analysed for the offset measured in the energy scans. The evaluated $\overline{res}(t)$ and their variation over time are large, of the order of 100%. However, since the expected value for this attribute is zero, this magnitude of the relative variations is expected.

Chapter 3

W+c production and measurement strategy

This chapter introduces the measurement of the production of a W boson with a single charm quark. First, the general calculation of the cross section for a process produced in proton-proton (pp) collisions is outlined and then the W+c production processes are explained.

The main motivation for the measurement of the $W + c$ production is the sensitivity to the parton density function of the strange quark. This distribution is poorly known from existing data and its accurate determination will improve the description of the proton structure. This process is also a relevant background to studies at the LHC, such as top quark measurements and beyond the SM searches; furthermore, it provides a test of the current theoretical predictions in perturbative QCD.

Finally, the strategy adopted in the measurement of this thesis is explained. The analysis is based on the charm quark identification technique known as soft muon tagging. To select the signal with high-purity from data, the anti-correlation of the W boson and charm quark charges is exploited. A list of the background processes to the signal selection concludes the chapter.

3.1 Production cross sections in pp collisions

The proton is described in the standard model as a bound state of three *valence quarks*, two up quarks and one down quark, held together by the strong interaction. Gluons and other quarks, which are collectively known as *sea partons*, are produced in the proton as a result of the valence quarks' interaction.

Each of the valence quarks and sea partons carries a fraction of the proton momentum. The probability density of finding a parton of type i, carrying a fraction of the proton's longitudinal momentum x, in interactions at an energy scale, or momentum transfer, Q, is known as the *parton density function* (PDF), $f_i(x, Q^2)$. The PDFs represent the knowledge of the composite structure of the proton and are therefore a crucial element in the theoretical description of the *pp* collisions.

Most of the studies with *pp* collision data rely on *hard scattering events*, i.e. those in which at least one parton-parton interaction between the two incoming protons occurs with a high momentum transfer, for instance $Q \sim 10$ GeV or larger. In such events, the hard interaction can be considered and evaluated as independent from the remaining partons in the two protons, which are referred to as *spectator partons*.

The hard scattering partons may radiate gluons or photons, both before and after the interaction, phenomena which are known as initial and final state radiation (ISR, FSR), respectively. The spectators may also produce additional radiation, or undergo further interactions. However, such processes are at a lower, or soft, energy scale and are collectively referred to as the *underlying event*.

The production cross section measures the likelihood of a process to occur in the scattering events considered. The cross section of the collision of two protons A and B, in which a hard scattering process produces the final state c, is calculated as

$$\sigma_{A+B\to c+X} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b \, \left[f_{a/A}(x_a, Q^2) \, f_{b/B}(x_b, Q^2) \cdot \sigma_{a+b\to c} \right] , \quad (3.1.1)$$

where X indicates the other products of the collision. In equation 3.1.1, the PDFs of the partons a and b belonging to the proton A and B, respectively, are convolved with the cross section of the partonic processes $a + b \to c$, $\sigma_{a+b\to c}$, according to the QCD factorisation theorem [36].

The partonic cross section is evaluated in perturbative QCD as an expansion in terms of the QCD running coupling $\alpha_S(Q^2)$:

$$\sigma_{a+b\to\ c} = [\,\sigma_0\ +\ \alpha_S(Q^2)\cdot\sigma_1\ +\ \alpha_S^2(Q^2)\cdot\sigma_2 +\ ...\,]_{a+b\to c}\ . \tag{3.1.2}$$

The first term of the expansion, σ_0, is known as the leading order (LO) cross section; the first two terms collectively as the next-to-leading order (NLO) cross section, and so forth.

The mathematical computation of $\sigma_{A+B\to c+X}$ requires the introduction of two parameters which are known as the *factorisation scale*, μ_F, and the *renormalisation scale*, μ_R. The former indicates the energy scale which separates the parton radiation accounted for either in the PDFs or in the partonic cross section; the latter fixes the value of the QCD running coupling [36].

If calculated to all orders in $\alpha_S(Q^2)$, the cross section $\sigma_{A+B\to c+X}$ is invariant with respect to the choice of values for μ_F and μ_R. However, only the computation of the first few of terms in equation 3.1.2 is possible[1]. Hence, the theoretical prediction for $\sigma_{A+B\to c+X}$ is calculated after truncating the series, e.g. at NLO, and therefore depends on the particular choice of μ_F and μ_R. The chosen values of the QCD scales are of the order of the hard scattering momentum transfer, $\mu_F \sim \mu_R \sim Q$, [36].

The PDFs cannot be calculated directly from QCD and their parametrisation is determined from measured cross sections and from the DGLAP equations [37–39]. The latter describe the evolution of a PDF with respect to Q.

The PDF for a given parton, or a linear combination thereof, is parametrised as a function of x at a starting value of the energy scale, $Q_0 \sim 1$ GeV. This PDF is first evolved via the DGLAP equations to the Q of a measured cross section and then convolved with the associated partonic cross sections as per equation 3.1.1.

---

[1] The inclusive NLO cross section is available for most of the processes studied at the LHC. The full calculation at next-to-next-to-leading order, i.e. up to the $\alpha_S^2(Q^2)\cdot\sigma_2$ term of equation 3.1.2, is achieved for a few specific processes, such as the Drell-Yan lepton pair production.

From the comparison of the predicted and the measured cross sections, information can be extracted about the parameters of the PDF considered.

The analysis is expanded to include multiple measurements and various PDFs, and PDF set is obtained from a global fit to the data. Various PDF sets are available, which generally differ in the choice of the PDFs' parametrisation and in the data included in the global fit.

3.2 W+c production in pp collisions

The production of a W boson in association with a single charm quark in pp collisions is the process $p + p \rightarrow W + c + X$ [2], where X indicates all the other particles in the final state.

The partonic process is described at LO in perturbative QCD by the scattering of a gluon off a down-type quark (u-, s- and b-quark). The LO diagrams for the t-channel and the s-channel are drawn in figures 3.1 (a) and (c) for the $W^- + c$ production, and in figures 3.1 (b) and (d) for the $W^+ + \overline{c}$ production. The contribution of each of the three down-type quarks to the initial state is determined by the relative PDF and by the elements of the CKM matrix, $|V_{cd}|$, $|V_{cs}|$ and $|V_{cb}|$.

The *Cabibbo-Kobayashi-Maskawa (CKM) matrix* arises in the standard model because the weak interaction allows the *quark-mixing*, that is the decay of one quark i to another quark j belonging to a different generation. In quantum field theory this is included with quarks' weak interaction eigenstates which are not the same as the quarks' mass eigenstates. Through the CKM matrix, the weak interaction eigenstates of the down-type quarks are expressed as a linear combination of mass eigenstates. The magnitude of CKM matrix elements, $|V_{ij}|$, describes the probability of the weak decay $i \rightarrow j$. The partonic cross section of the decay is proportional to $|V_{ij}|^2$.

The s-quark initiated production, $g + s \rightarrow W + c$, is the dominant process and has been shown to account for approximately 90% of the total LO $W + c$

---

[2] When the W boson charge is not specified both the $W^- + c$ and $W^+ + \overline{c}$ processes are considered.

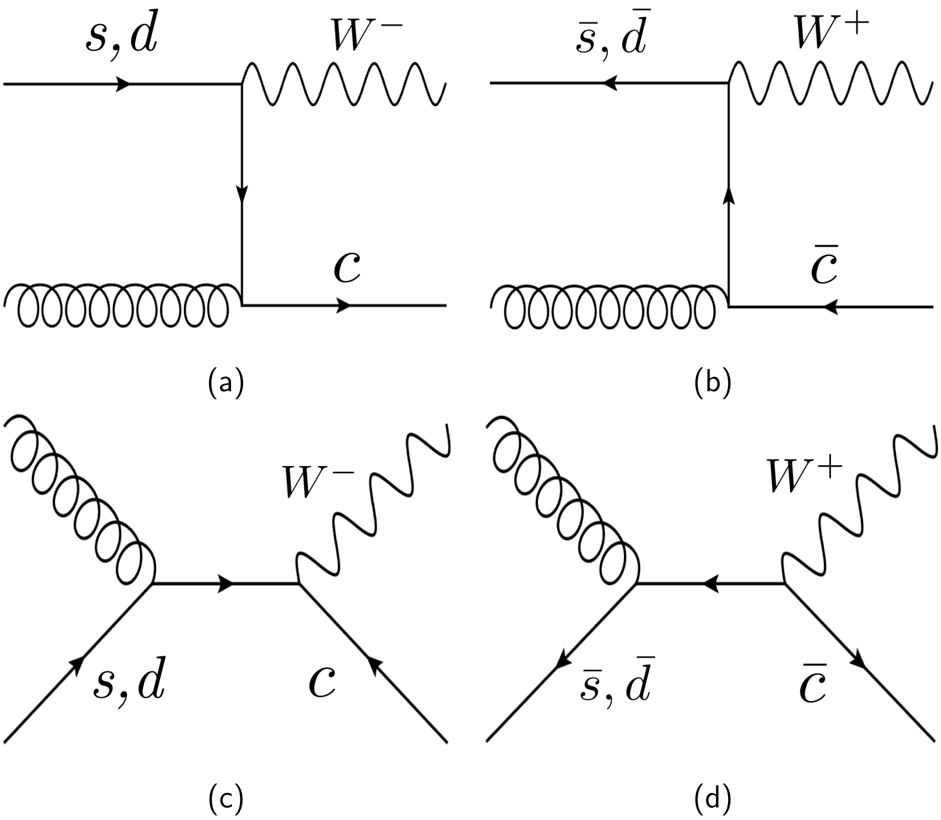


Figure 3.1: Diagrams for the W+c production at LO. The t-channel diagrams for the W^-+c and W^++$\bar{c}$ production are shown in (a) and (b), respectively; similarly the s-channel diagrams are illustrated in (c) and (d).

production at the LHC [40]. This is due to large value of the corresponding matrix element, $|V_{cs}|^2 = 1.012$.

The remaining 10% is due to the d-quark initiated contribution, $g + d \to W + c$, which despite being enhanced by the larger d-quark PDF, is suppressed by $|V_{cd}|^2 = 0.053$. In fact, in the kinematical regime of the measurement, $x \sim 0.01$ and $Q^2 \sim 10^4$ GeV^2, the ratio of the strange and down PDFs is about 0.5. The b-quark contribution is negligible because of both the suppression from the small b-quark PDF and $|V_{cb}|^2 = 0.002$.

In perturbative QCD, a NLO correction of about 30% is calculated in [41] to the W+c production cross section at LO. The QCD NLO diagrams can be subdivided into three categories. The first category, whose examples are illustrated in figures 3.2 (a) and (b), encompasses the emissions of virtual and real gluons in the LO processes. The second category includes the processes initiated by two gluons, $g + g \to (s/d) + W + c$, for which a diagram is displayed in figure 3.2 (c). The last category, for which an example is shown in figure 3.2 (d), accounts for the processes initiated by two quarks, $q + q^{(\prime)} \to q + W + c$, where q indicates either a valence or a sea (anti)-quark in the proton.

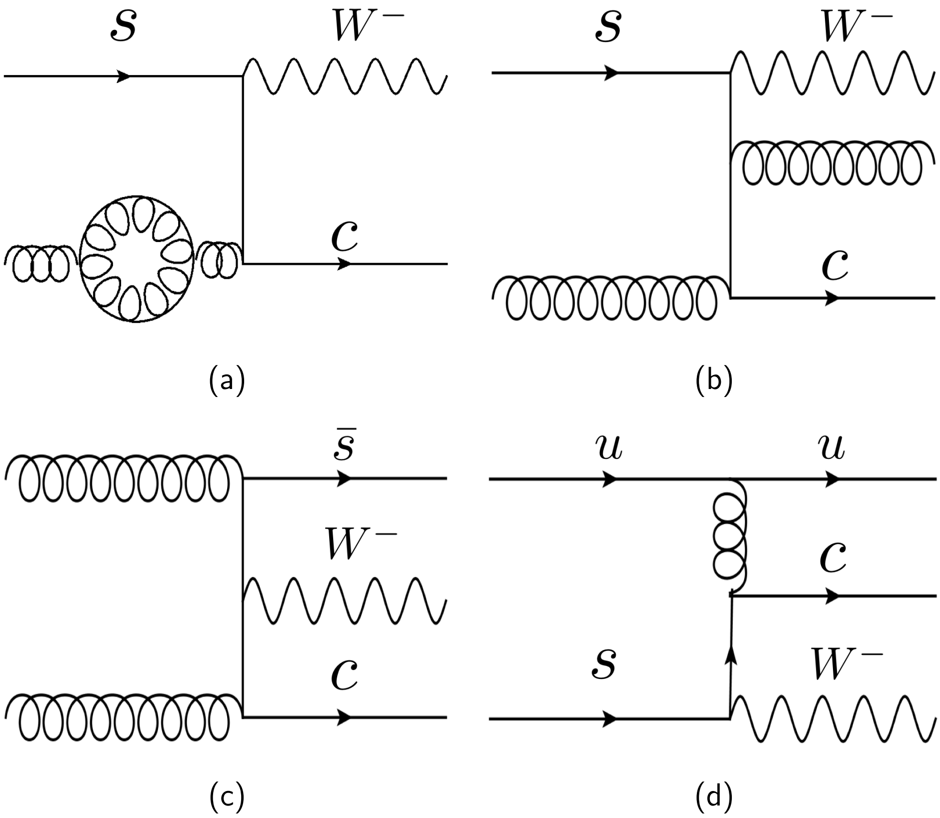


Figure 3.2: Examples of QCD NLO diagrams for the $W + c$ production: emission of a virtual (a) or real (b) gluon in the LO processes; (c) initiated by two gluons, $g + g \to (s/d) + W + c$; (d) initiated by two quarks, $q + q' \to q + W + c$.

Table 3.1 shows the contributions to the W+c production subdivided by the types of the two partons initiating the scattering. The fractions of events are evaluated with the W+c simulated sample and the selection criteria described in section 4.3 and section 4.2, respectively. The partonic processes sensitive to the s-quark PDFs are the subprocesses $g + (s/\bar{s})$ and $q + (s/\bar{s})$, which correspond to approximately 80% of the total W+c production at the LHC with $\sqrt{s} = 7$ TeV.

It should be stressed that "W+c production" in this analysis refers exclusively

Contribution to the $W+c$ production from the processes with initiating partons [%]				
$g+(s/\bar{s})$	$g+(d/\bar{d})$	$g+g$	$q+(s/\bar{s})$	other $q+q^{(\prime)}$
71.2 ± 0.9	10.63 ± 0.34	7.96 ± 0.30	9.07 ± 0.32	1.12 ± 0.15

Table 3.1: Predicted contribution to the $pp \to W+c$ production at $\sqrt{s} = 7$ TeV subdivided by the type of the two initiating partons. Values are shown in percent. The fractions of events are evaluated with the $W+c$ simulated sample and the selection criteria described in section 4.3 and section 4.2, respectively. The different partonic processes are explained in the text. The uncertainties are statistical.

to events with a single c-quark produced in the partonic scattering. Partonic processes generating a W boson and a c-quark pair can also occur, but these are not measured with the strategy adopted in this thesis, as explained in section 3.5.

3.3 Motivations for the W+c production measurement

The main motivation to study the W+c production is its sensitivity to the s-quark PDFs, $s(x)$ and $\bar{s}(x)$. PDFs play a fundamental role at hadron colliders as they relate theoretically calculated partonic cross sections with measured cross sections. In fact, the understanding of the PDFs is a potential source of uncertainty for the vast majority of the studies at the LHC, such as Higgs measurements.

The s-quark PDFs suffer from large uncertainties because of the lack of measurements to constrain them in the PDF fits. A review of the strange content of the proton can be found in e.g. [42]; a brief summary is presented here.

Initially, the $s(x) = \bar{s}(x) = \kappa \left[\bar{u}(x) + \bar{d}(x)\right]/2$ parametrisation was the common choice in PDF analyses, where $\bar{u}(x)$ and $\bar{d}(x)$ indicate the $\bar{u}$- and $\bar{d}$-quark PDF, respectively. This relation reflected the lack of measurements directly sensitive to the strange content of the proton. The functions $s(x)$ and $\bar{s}(x)$ were evaluated from the better-determined $\bar{u}(x)$ and $\bar{d}(x)$ and no asymmetry between them was considered. A value $\kappa \approx 0.5$ at $Q_0^2 \sim 1$ GeV^2 was set to suppress the strange content of the proton. This was theoretically motivated owing to the larger mass of the s-quark with respect to that of the u- and d-quarks.

Recent PDF analyses typically use parametrisations of the s-quark PDFs which do not depend on $\overline{u}(x)$ and $\overline{d}(x)$, and also might consider that $s(x) \neq \bar{s}(x)$. This was made possible after the measurements sensitive to both $s(x)$ and $\bar{s}(x)$ performed by the NuTeV [43, 44] and CCRF [45] experiments. These data provide important constraints to the s-quark PDFs in the region with $Q^2 \sim 10$ GeV^2 and $x \sim 0.1$.

The NuTeV and CCRF experiments measured di-muon production from the deep inelastic scattering (DIS) process $(\nu/\bar{\nu}) + N \to \mu^+ + \mu^- + X$, where N indicates a nucleon (proton or neutron). Neutrino beams were directed at an iron target to produce the scattering. The diagram of this neutrino-nucleon DIS processes is shown in figure 3.3. For the reasons explained in section 3.2, this reaction proceeds primarily through the $W^+s \to c$ and $W^-\bar{s} \to \bar{c}$ subprocesses and, hence, allows probing individually $s(x)$ and $\bar{s}(x)$. One muon originates from the weak interaction of the neutrino with the nucleon; the second muon, which carries opposite charge, stems from the semileptonic decay of the charm quark.

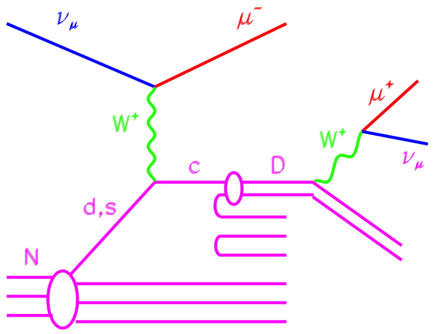


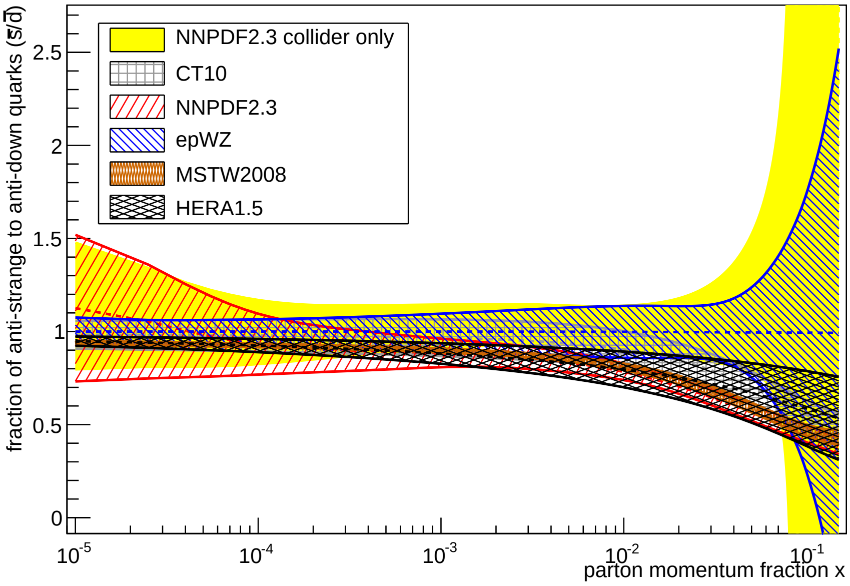
Figure 3.3: Diagram of the neutrino-nucleon deep inelastic scattering processes measured by the NuTeV and CCRF experiments. Di-muon events are produced: one muon originates from the neutrino's weak interaction with the nucleon; a second muon, with opposite charge, is produced in the semileptonic decay of a charm quark. Taken from [44].

However, recent PDF analyses have shown sizeable differences in the size and the asymmetry of the s-quark PDF, indicating that the neutrino-nucleon DIS data do not provide enough constraint on the strange content of the proton. This can be seen in figure 3.4, which compares the s-quark PDF obtained in six PDF sets: MSTW2008 [46], HERAPDF1.5 [47], CT10 [48], NNPDF2.3 and NNPDF2.3coll [49], and ATLAS-epWZ12 [50]. The uncertainty bands have been calculated according to the prescriptions from each PDF analysis.

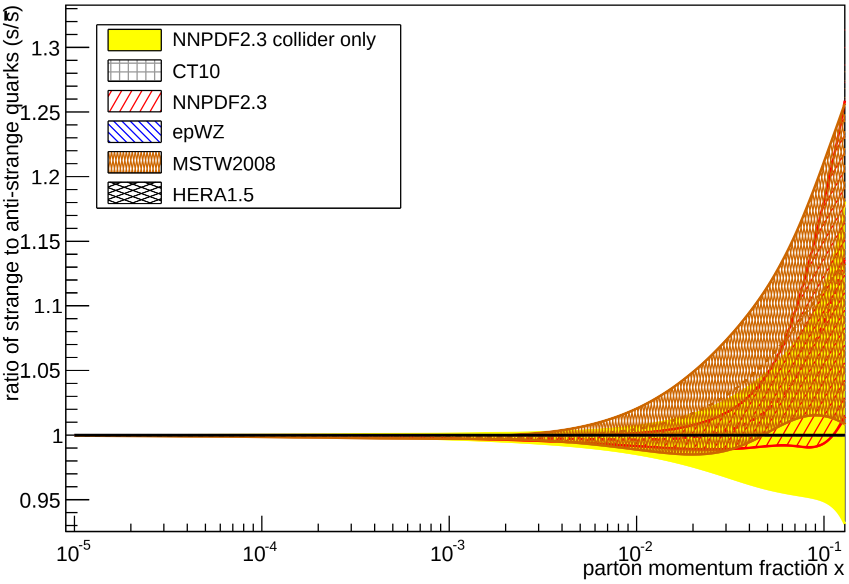
The different size of the s-quark PDF can be seen in figure 3.4 (a), where the PDF ratio $\bar{s}(x)/\overline{d}(x)$ is shown as a function of x. In the region between 10^{-3} and 10^{-1}, which is relevant for the $W+c$ measurement presented in this thesis, the MSTW2008 (brown) and NNPDF2.3 (red) analyses obtain a suppressed s-quark sea with respect to the d-quark sea. A similar but smaller suppression is found in the HERAPDF1.5 (black) and CT10 (grey) results. The ATLAS-epWZ12 and NNPDF2.3coll analyses use exclusively collider data, so that the neutrino-nucleon DIS measurements are excluded. These two PDF sets have more symmetric composition of the light-quark sea, which is suggested by flavour symmetry of the strong interaction. The ATLAS-epWZ12 analysis (blue), which combines an ATLAS measurement of the W and Z bosons' cross section with electron-proton DIS data from the HERA collider, fixes $\bar{s}(x) \equiv \overline{d}(x)$. The NNPDF2.3coll analysis (yellow), which includes data from the HERA, the Tevatron and the LHC, obtains an s-quark component that can be slightly larger than that of the d-quark.

Figure 3.4 (b) show the ratio $s(x)$ over $\bar{s}(x)$ as obtained from the six PDF analyses. In the CT10, HERAPDF1.5 and ATLAS-epWZ12 analyses, the ratio is fixed to unity, $s(x) \equiv \bar{s}(x)$. The MSTW2008, NNPDF2.3and NNPDF2.3coll PDF sets allow a small strange PDF asymmetry, with $s(x)$ few percent larger than $\bar{s}(x)$ for $x \sim 10^{-2}$. Such an asymmetry is suggested by results of the NuTeV and CCFR experiments; however, the hypothesis of symmetric $s(x)$ and $\bar{s}(x)$ is not ruled out, as explained, for example, in [51].

The possibility of using measurements of the $W+c$ production to further constrain the strange content of the proton has long been discussed for the Tevatron [52] and for the LHC [40]. The W+c production allows to probe the s-quark PDFs in a region with momentum transfer $Q^2 \sim 10^4$ GeV^2 and momentum frac-



(a)



(b)

Figure 3.4: Comparison of (a) the $\bar{s}(x)/\bar{d}(x)$ ratio and (b) the $s(x)/\bar{s}(x)$ ratio for six PDF sets: MSTW2008 (brown), NNPDF2.3 (red), HERAPDF1.5 (black), CT10 (gray), ATLAS-epWZ12 (blue) and NNPDF2.3coll (yellow). The uncertainty bands are calculated according to the prescriptions from each PDF analysis. The momentum scale is $Q^2 = m_W^2$. Taken from [2].

tion $x \sim 0.01$. This kinematical regime is different from the neutrino-nucleon DIS data; hence, by means of including the $W+c$ production in the PDF fits, a more accurate determination of the s-quark PDFs can be achieved.

The W+c production was measured with a precision of 20–30% at the Tevatron collider [53–55], but recent PDF analyses typically do not include these results. Due to the sizeable measurement uncertainties, the Tevatron W+c measurements do not add a significant amount of constraint to the global PDF fits. The larger W+c production rates available at the LHC offer the first opportunity to measure this process with enough precision to constrain the s-quark PDFs. Measurements of the $W + c$ production cross section at the LHC were performed recently by CMS [56] and ATLAS [1], the latter containing the analysis of this thesis.

Two other motivations support the study of the $W + c$ production. Firstly, this process can be a sizeable background for LHC measurements which involve heavy quarks (charm, bottom and top) and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state. Particularly relevant cases are top quark studies and beyond the SM searches of, for example, squarks of the third-generation. Secondly, the comparison of the $W + c$ cross section measured at the LHC with the current NLO predictions offers a unique test of the perturbative QCD calculations performed by the latest MC generators.

3.4 Strategy for the W+c production measurement

In the measurement of the W+c production, the W bosons are identified via the leptonic decay into muons or electrons: $W \to l\,\nu$, where $l = \mu, e$. The c-quark can be identified either by the semileptonic decay $c \to l\,\nu\,q$, where q indicates a down-type quark, or via the reconstruction of a c-hadron decay, e.g. $D \to K\pi\pi$.

The analysis presented in this thesis uses both the W boson and the c-quark decays into muons, whose branching ratios are $BR(W \to \mu\nu) = 0.1057 \pm 0.0015$ and $BR(c \to q\mu\nu) = 0.096 \pm 0.004$ [57]. The corresponding ATLAS analyses using the electron channel of the W boson decay and that exploiting the reconstruction of c-hadrons decays are reported in [1]. Electron channel decays were also used to tag the c-quark at the Tevatron collider [54].

The muons originating from the $W \to \mu\,\nu$ and $c \to \mu\,\nu\,q$ decays are referred to as *W-decay muon* and *soft muon*, respectively. The latter owes its name to the softer energy spectrum with respect to that of a W boson decay. In the $W{+}c$ production, the W boson and the c-quark carry opposite sign charge, $W^-{+}c$ and $W^+{+}\bar{c}$; consequently the W-decay and soft muons have opposite charge, as highlighted in figure 3.5.

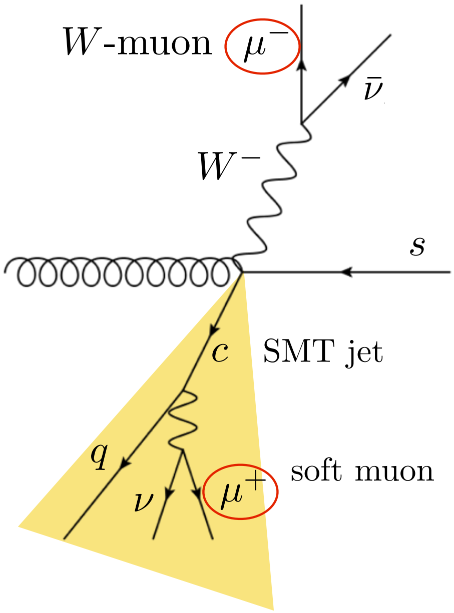


Figure 3.5: Diagram of the decays used to identify the $W{+}c$ events in the analysis of this thesis. The W-decay and the soft muons always carry opposite charge. The c-quark is identified with the soft muon tagging (SMT) method, i.e. by the presence of a soft muon within the jet, which is reconstructed from the quark's hadronisation.

The identification of a soft muon within a jet is used as the experimental signature of a c-quark. Jets are collimated bunches of particles reconstructed in the detectors which result from the hadronisation of quarks and gluons. The presence of a soft muon is one of the methods used to identify the *heavy-flavour jets*, which originate from c- or b-quarks. This method is known as *soft muon tagging (SMT)* and so an identified jet is referred to as an *SMT-jet*. The definition of reconstructed jets and the description of the SMT algorithm used for this analysis are given in section 4.2.2.

By exploiting the charge anti-correlation between the W boson and the soft muon, the $W+c$ production signal can be extracted with high efficiency from collision events data. The fraction of opposite sign (OS) events predicted at detector level for the $W{+}c$ production is found to be 0.924 ± 0.004. The detector level event selections are described in section 4.2. The remaining 7.6% of $W+c$ events is reconstructed as same sign (SS) events and is mainly due to the misidentifications of the soft muon. At particle level, in the fiducial region defined in section 7.1, the fraction of SS events has been found to be smaller than 1% and is due to c-quarks produced by gluon splittings or the underlying event. Therefore, to a first approximation, the yield of the $W+c$ production can be estimated as the number of OS events selected from data, $N^{\text{OS}}_{\text{data}}$.

Furthermore, the relative sign of the W boson and the soft muon charges is exploited to extract the $W+c$ signal with high purity. The background processes produce events in large part symmetrically distributed between the OS and SS categories, as explained in section 3.5. Consequently, the largest part of the background contribution to $N^{\text{OS}}_{\text{data}}$ can be estimated by the number of SS events selected from data, $N^{\text{SS}}_{\text{data}}$. To a second approximation, the $W+c$ production yield can be evaluated from the selected data event sample after the *OS–SS subtraction*,

$$N^{\text{OS}-\text{SS}}_{\text{data}} \equiv N^{\text{OS}}_{\text{data}} - N^{\text{SS}}_{\text{data}}. \tag{3.4.1}$$

Finally, the OS/SS asymmetric background contribution, $N^{\text{OS}-\text{SS}}_{\text{bkg}}$, is subtracted to obtain the final yield for the $W+c$ production, $N^{\text{OS}-\text{SS}}_{\text{data}} - N^{\text{OS}-\text{SS}}_{\text{bkg}}$.

The integrated $W+c$ production cross section is calculated as

$$\sigma_{W{+}c} \times BR(W \to \mu\nu) = \frac{N^{\text{OS}-\text{SS}}_{\text{data}} - N^{\text{OS}-\text{SS}}_{\text{bkg}}}{U \cdot L}, \tag{3.4.2}$$

where L and U are the integrated luminosity of the data sample and the unfolding factor correction, respectively, which are described in section 4.1 and section 7.1.

In addition, the cross section is also measured differentially as a function of the W-decay muon pseudorapidity. The integrated and differential measurements are performed separately for events with a positively and a negatively charged W boson, and the ratio

$$R^{\pm}_{W{+}c} \equiv \sigma(W^{+}{+}\overline{c})/\sigma(W^{-}{+}c) \tag{3.4.3}$$

is also measured. All measurements are compared to predictions of NLO QCD calculations obtained with various PDF sets and the sensitivity to the choice of PDFs is presented in chapter 7.

3.5 Backgrounds to the W+c production measurement

The background processes are those that contribute to the sample of signal candidate events which is selected from data. In the case of the $W{+}c$ production, the background processes are those that can produce either "real" or "fake" $W \to \mu\nu$ decays and SMT-jets. However, such processes are only relevant if they present OS–SS asymmetry and, hence, contribute to the $N_{\mathrm{bkg}}^{\mathrm{OS-SS}}$ term in equation 3.4.2.

The potential background processes to the $W + c$ production are described below:

$\mathbf{W + c\bar{c}}$ and $\mathbf{W + b\bar{b}}$. These processes refer to the production a W boson in association with a pair of c- or b-quarks, for which an example diagram is shown in figure 3.6 (a). Similarly to c-quarks, b-quarks can decay semileptonically producing a soft muon. The $W{+}c\bar{c}$ and $W{+}b\bar{b}$ productions have a similar signature to that of the $W{+}c$ production; however, they are expected to be OS–SS symmetric. In fact, there is an equal probability that the jet originating from either the quark or the antiquark is identified by the SMT algorithm. Therefore there is no expected charge correlation between the W boson and the soft muon.

W+light. This refers to the production of a W boson in association with *light-flavour jets*, i.e. those originating from the hadronisation of a light quark (up, down and strange) or a gluon. The soft muon from a c- or b-quark is faked by muons which might be produced, for example, in the decays of pions and kaons. The W+light production is expected to present some OS/SS asymmetry due to the processes such as, for instance, $g + u \to W^{+}d$, whose diagram is displayed in figure 3.6 (b). In this example, the W boson and the d-quark carry opposite sign charges. Owing to charge conservation among the quarks' hadronisation products, there is a higher probability that the misidentified soft muon carries same charge sign of the d-quark and, hence, an OS event is produced.

Z+jets. This refers to the production of a Z/γ^* decay into a pair of electrons, muons or τ-leptons, which can occur in association with jets. For this thesis, the most relevant Z+jets background source is $Z \to \mu^+\mu^-$ [3]. The muons from the Z decay can (i) be falsely reconstructed as a W boson decay if missing energy is also present; (ii) fake the SMT signature if they emit photons leading to a mistakenly reconstructed jet or they are accidentally associated with a jet. Owing to the production of two oppositely charged muons, the Z+jets production is expected to be OS/SS asymmetric.

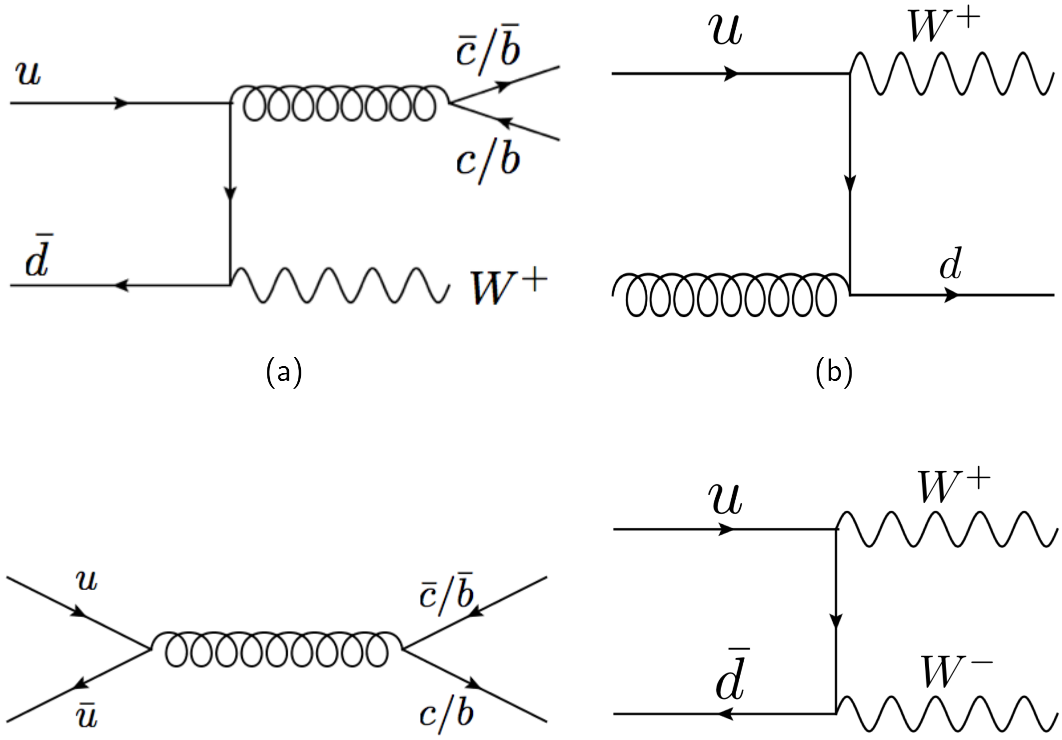
Multijet. This refers to the QCD mediated scatterings producing jets in the final state. For the analysis in this thesis, the most relevant multijet background source is the production of c- and of b-quark pairs, $c\bar{c}$ and $b\bar{b}$, for which an example diagram is shown in figure 3.6 (c). The W boson decay might be misreconstructed from a muon and a neutrino produced in a semileptonic decay of the c- or b-quarks. The $c\bar{c}$ and $b\bar{b}$ production are expected to present some OS/SS asymmetry, as the quark and the antiquark carry opposite charges. Due to charge conservation in the quarks' hadronisation processes, the misidentified W-decay muon is more likely to carry same charge sign of the original (anti)-quark, and, hence, give an OS event.

Top quark. This refers collectively to the production of t-quark pairs, $t\bar{t}$, and of single t-quarks. Since a t-quark decays into a W boson and a b-quark, these processes can produce a similar signature to that of the W+c production. Owing to the charge correlation between the W boson and b-quark, the top quark production is expected to present some OS/SS asymmetry.

---

[3] When referring to the Z boson production or decay, both the Z boson and the γ^* processes are included.

Diboson. This refers to the production of the ZZ, ZW and W^+W^- boson pairs, which can occur in association with jets. An example diagram for the W^+W^- production is shown in figure 3.6 (d). This process is expected to be OS/SS asymmetric if, for instance, the final state contains a W boson with an oppositely charged c-quark, $W^+W^- \rightarrow W^+\bar{c}s$.



(a) (b)

(c) (d)

Figure 3.6: Example of diagrams for the (a) W+$c\bar{c}$ and W+$b\bar{b}$ production, (b) W+light production, (c) multijet $c\bar{c}$ and $b\bar{b}$ production, and (d) diboson production.

Chapter 4

Datasets and analysis selections

This chapter describes the datasets of collision events used for the analysis of this thesis. First, the data sample of LHC collision events is described. Then, the event selection criteria and physics objects used for the signal events are specified. Finally, the datasets obtained from Monte Carlo simulations of the W+c production and of the background processes are presented.

The author of this thesis has implemented his own algorithms to select the event samples for the analysis. These algorithms were interfaced with software packages used to calibrate the physics objects and correct the MC simulations. These packages were typically provided by ATLAS performance analysis groups.

Specifically, the SMT algorithm was implemented in a software package entirely by the author of this thesis, who has also collaborated in the study of the SMT calibration reported in [5].

4.1 LHC data sample

The analysis of this thesis is based on the LHC collision data collected by the ATLAS detector during 2011. The centre of mass energy of the pp collisions was $\sqrt{s} = 7$ TeV. The dataset's integrated luminosity is $\int Ldt = 4.58$ fb^{-1}, measured with an uncertainty of 1.8% [34]. The data sample corresponds to the order of 10^{14} collision events in the LHC.

The data have been verified to be of good quality for physics measurements by applying requirements provided by the ATLAS data quality group. Such requirements ensure a stable conditions and the correct operating of the relevant ATLAS components during data taking.

4.2 Event selections and physics object definitions

The identification of a $W \to \mu\nu$ decay requires exactly one isolated, high-p_{T} muon and large $E_{\mathrm{T}}^{\mathrm{miss}}$ stemming from the neutrino. For the selection of the $W+c$ production process, events with exactly one or exactly two jets are kept, and exactly one of these jets is identified as originating from a c-quark, with the SMT algorithm.

The event selection criteria adopted are listed in table 4.1 and explained, with the physics object definitions, in sections 4.2.1 to 4.2.2. The number of events counted in the data sample and the $W+c$ Monte Carlo (MC) simulated sample after applying each of the detector level selection steps is shown in the table.

In the data, the number of events selected by the trigger is approximately 170 million, which is reduced to 17,366 after applying all the selection criteria. The simulated datasets for the $W+c$ signal and background processes, whose descriptions are found in section 4.3, are obtained with the same event selection requirements applied to the data. In the $W+c$ sample, the number of simulated events[1] passing the trigger requirement is approximately 2.2 million, which is reduced to 13,897 after applying all the selection criteria.

4.2.1 Trigger and interaction vertex selections

The collision events are selected with an EF trigger requiring the identification of a muon with $p_{\mathrm{T}} > 18$ GeV. Using $Z \to \mu\mu$ data events, the trigger efficiency has been evaluated as function of the muon p_{T} and separately for the barrel and endcap regions [58]. The trigger efficiency has been found to be approximately 70% and 90% in the barrel and endcap, respectively, approximately constant with

---

[1] This refers to the number of simulated events available, not normalised to the luminosity of the data sample.

Event selection		Event counts	
		Data	$W+c$ (MC)
Trigger and interaction vertex			
Trigger	a muon with $p_{\mathrm{T}} > 18$ GeV	169,620,064	2,154,175
Primary vertex	at least five tracks	168,728,704	2,152,118
$W \to \mu\nu$ selection			
Muon	exactly one muon ($p_{\mathrm{T}} > 20$ GeV, $\|\eta\| < 2.5$)	32,596,362	1,910,502
	no electrons (event veto)	32,586,458	1,910,463
Neutrino	$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV	23,162,674	1,624,119
m_{T}^{W}	$m_{\mathrm{T}}^{W} > 60$ GeV	15,073,392	1,240,435
c-quark selection			
Jets	at most two jets ($p_{\mathrm{T}} > 25$ GeV, $\|\eta\| < 2.5$)	2,236,401	501,944
c-jet	exactly one jet tagged by SMT	17,366	13,897

Table 4.1: Event selection criteria for the measurement of the $W+c$ production cross section. The $W \to \mu\nu$ decay and the SMT algorithm are used for the W boson reconstruction and the c-quark identification, respectively. The requirements and the physics objects' definitions are described in sections 4.2.1 to 4.2.2. The number of events selected in data sample and in the $W+c$ Monte Carlo simulated sample after applying each selection requirement is shown. After applying all the criteria, the data and the $W+c$ simulation count 17,366 and 13,897 events, respectively.

respect to the muon p_{T}. The trigger is available in the simulated events, where the trigger efficiency is corrected to match that measured in data.

The rejection of events not associated with pp collisions, such as those resulting from background radiation in the ATLAS cavern or from cosmic rays, is achieved by requiring at least one interaction vertex, reconstructed with at least five tracks each with $p_{\mathrm{T}} > 0.4$ GeV. If an event has multiple reconstructed vertices, the one with the largest $\sum p_{\mathrm{T}}^2$, where the sum runs over all the tracks associated with the vertex, is defined as the *primary vertex* originating from the hard scattering process.

4.2.2 W $\to \mu\nu$+jets selections

Muon. The W-decay muon candidates are selected with the following criteria.

- Reconstructed as a *combined muon*, i.e. by matching a track measured by the MS with a track identified in the ID [59].
- Compliance with the track quality criteria, requiring at least 6 SCT hits and a hit in the innermost layer of the pixel detector unless the track passes through a region in which no hit is expected. Tracks with more than one missing pixel or SCT hit are discarded.
- Matching the muon identified by the trigger within $\Delta R < 0.15$.
- Reconstructed as an *isolated muon*. The *track isolation* variable is defined as the sum of the p_{T} of the tracks reconstructed within a $\Delta R = 0.3$ cone around the muon: $I_{\mathrm{track}} = \sum_{\Delta R<0.3} p_{\mathrm{T}}^{\mathrm{tracks}}$. Similarly, the energy of the calorimeter cells not associated with muon radiation within a $\Delta R = 0.2$ cone is defined as *calorimeter isolation*, $I_{\mathrm{calo}} = \sum_{\Delta R<0.2} E_{\mathrm{T}}^{\mathrm{cells}}$. An isolated muon is required to measure $I_{\mathrm{track}} < 2.5$ GeV and $I_{\mathrm{calo}} < 4$ GeV. Furthermore, the W-decay muon candidate must be separated by $\Delta R > 0.4$ from any selected jet.
- Reconstructed with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$.

In the simulated samples, the efficiencies of the reconstruction and the isolation requirement, and the energy scale and resolution of the W-decay muon candidates, are corrected to match those measured in data by the ATLAS muon performance group.

Electrons. Events with a W-decay muon candidate are discarded if they contain high-p_{T}, isolated electrons, which are defined with the following criteria.

- Reconstructed by matching a calorimeter cluster with a track [60]. *Calorimeter clusters* are collections of neighbouring calorimeter cells measuring an energy significantly above the expected detector noise [61].
- Compliance with the “tight” quality criteria recommended by the ATLAS e/γ performance group, which are based on: the calorimeter cluster shape, track quality, track-to-calorimeter-cluster matching, particle identification information from the TRT and photon conversion veto.

- Reconstructed as isolated electrons with both track and calorimeter isolation required to be smaller than 3 GeV.

- Reconstructed with $E_{\rm T} > 25$ GeV and $|\eta| < 2.47$ excluding the region $1.37 < |\eta| < 1.52$. This region is not instrumented due to the transition between the EM calorimeter barrel and endcap. The electron energy scale is calibrated from $Z \to ee$ data [62]. In the simulated samples, the energy resolution of the veto electron candidates is corrected to match those measured in data by the ATLAS e/γ performance group.

Jets. The jets are reconstructed with the anti-k_t algorithm [63] with radius parameter $R = 0.4$. The algorithm is an iterative procedure starting with the identification of "pseudo-jets", which correspond to calorimeter clusters and to particle four-momenta for the data and the simulated samples, respectively.

First, the distances $d_i = 1/k_{t,i}^2$ and $d_{i,j} = \min\left(1/k_{t,i}^2\,,\ 1/k_{t,j}^2\right) \cdot (\Delta R_{i,j}/R)^2$ are calculated, where i, j run over the pseudo-jets and k_t is their transverse momentum. If the smallest distance is d_i, the i-th pseudo-jet is identified as a jet and removed from the list of pseudo-jets; conversely, if the smallest distance is $d_{i,j}$, the i-th and j-th pseudo-jets are removed and recombined into a single pseudo-jet. The iterative procedure ends when no pseudo-jets are left.

The energy of the reconstructed jets is calibrated with the scheme known as *EM+JES scale* [64]. The *EM scale* calibrates the calorimeter response to the energy deposited by electromagnetic showers, which are measured from $Z \to ee$ data. Consequently, the *jet energy scale* (JES) is applied to correct the measured jets' energy to be as close as possible to that of the particle-level jet. This correction accounts for effects such as the calorimeter response to hadronic interactions of the jets' particles, pileup contamination and the jets' particles leaking from the calorimeters.

The JES is mainly estimated from simulations, by means of comparing the energy of calorimeter jets with that of the corresponding *truth-jets*. The latter are reconstructed from the *stable particles*, i.e. are those with an expected lifetime longer than 10 ps, present in the final state of a simulated event.

Jet which are reconstructed within $\Delta R < 0.2$ from an electron are removed. A further requirement is applied to remove jets stemming from pileup. The tracks reconstructed within a jet are used to calculate the *jet vertex fraction* (JVF). This variable is defined as the p_{T} sum of the jets' tracks associated with the primary vertex, divided by the p_{T} sum of all the jets' tracks. The JVF of the selected jets is required to be larger than 0.75.

Finally, the event selection of this analysis requires at most two jets reconstructed with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$.

Missing Transverse Energy. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated from calorimeter cells and reconstructed muons as described in reference [65]:

$$\begin{aligned} -E_{\mathrm{x,y}}^{\mathrm{miss}} &= E_{\mathrm{x,y}}^{\mathrm{e}/\gamma} + E_{\mathrm{x,y}}^{\mathrm{muons}} + E_{\mathrm{x,y}}^{\tau\text{-leptons}} + E_{\mathrm{x,y}}^{\mathrm{jets}} + E_{\mathrm{x,y}}^{\mathrm{soft\text{-}jets}} + E_{\mathrm{x,y}}^{\mathrm{cells\text{-}out}} \\ E_{\mathrm{T}}^{\mathrm{miss}} &= |\mathbf{E}_{\mathbf{T}}^{\mathbf{miss}}| = \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2}. \end{aligned} \tag{4.2.1}$$

The $E_{\mathrm{x,y}}^{\mathrm{e}/\gamma}$ term accounts for the energy of reconstructed electron and photons. The muon term includes possible energy deposits in the calorimeter due to reconstructed muons. The $E_{\mathrm{x,y}}^{\mathrm{jets}}$ and $E_{\mathrm{x,y}}^{\mathrm{soft\text{-}jets}}$ terms include jets reconstructed with $p_{\mathrm{T}} > 20$ GeV and 7 GeV $< p_{\mathrm{T}} < 20$ GeV, respectively. The $E_{\mathrm{x,y}}^{\mathrm{cells\text{-}out}}$ term encompasses all the remaining calorimeter cells, which are not associated with any physics object.

Two quality criteria from the ATLAS jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ performance group are applied to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. An event containing a jet identified from the non-collision background processes is rejected [66]. Such processes include cosmic muons, calorimeter detector noise, interactions of the colliding protons with gas molecules or components of the LHC beam pipes. In addition, events with a jet reconstructed in the η–ϕ region referred to as the *LAr hole*, $0 < \eta < 1.5$ and $-0.8 < \phi < -0.6$, are removed. During part of the 2011 data taking, this region of the ATLAS calorimeter was affected by damaged readout electronics.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution depends on the precision of the $E_{\mathrm{x,y}}$ terms in equation 4.2.1. Hence, it depends on total transverse energy deposited in the detector by collision event, ΣE_{T}[2]. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution has been evaluated as a function of

---

[2] $\Sigma E_{\mathrm{T}}$is defined the sum of the transverse energy of the calorimeter clusters within $|\eta| < 4.5$

ΣE_{T} using $Z \to \mu\mu$ data events, and has been found to be 4–9 GeV for $\Sigma E_{\mathrm{T}}=$ 100–400 GeV [65].

W boson transverse mass. The transverse mass of the W boson, m_{T}^{W}, is the magnitude of the W boson four-momentum projected onto the plane transverse to the beam axis. In the case of the $W \to \mu\nu$ decay, it is calculated as:

$$m_{\mathrm{T}}^{W} = \sqrt{2\, p_{\mathrm{T}}^{\mu}\, p_{\mathrm{T}}^{\nu} \cdot (1 - \cos(\phi^{\mu} - \phi^{\nu}))}\,, \qquad (4.2.2)$$

where the neutrino variables are inferred from the missing energy calculation, $p_{\mathrm{T}}^{\nu} = E_{\mathrm{T}}^{\mathrm{miss}}$ and $\phi^{\nu} = \phi^{E_{\mathrm{T}}^{\mathrm{miss}}}$.

4.2.3 Soft muon tagging algorithm

As introduced in chapter 3, the SMT is an algorithm used to discriminate between heavy-flavour and light-flavour jets. It relies on the semileptonic decays of c- and b-quarks in the muon channel. Essentially, the SMT associates muons with reconstructed jets.

The complete description of the SMT algorithm used in this analysis is found in [5] and summarised here. The soft muon candidates are identified with selection criteria listed in table 4.2 and explained below.

Soft muon selection criteria
Combined muon and track quality requirements
$\chi^2_{\mathrm{match}} < 3.2$
$\lvert d_0\rvert < 3$ mm ; $\lvert z_0 \cdot \sin\theta\rvert < 3$ mm
$p_{\mathrm{T}} > 4$ GeV ; $\lvert\eta\rvert < 2.5$
ΔR(soft μ, jet)< 0.5

Table 4.2: List of the soft muon candidates' selection criteria, which are explained in the text. A selected jet is defined as soft muon tagged if exactly one soft muon candidate is found within a distance of $\Delta R = 0.5$.

The combined muon and track quality requirements refer to those used for the W-muon selection in section 4.2.2. The χ^2_{match} variable measures the quality of the fit for the muon track of the combined muon. The variables $|d_0|$ and $|z_0 \cdot \sin\theta|$, which are known as the *transverse* and the *longitudinal impact parameters* respectively, measure the distance of the soft muon track with respect to the primary vertex. Such requirements collectively aim to reject muons originating mainly from pions' and kaons' decays which can occur within a jet.

A selected jet is defined as soft muon tagged, or *SMT-jet*, if exactly one soft muon candidate is found within a distance of $\Delta R = 0.5$ from its axis. The fraction of events with SMT-jets containing or more soft muons, which is at the subpercent level, is discarded.

Two additional selection criteria are applied to suppress the contribution of the Z+jets background, which is described in section 3.5. The motivation of such selection requirements is presented in section 5.1.3.

- The SMT-jet is required to have at least three associated tracks or an *electromagnetic fraction* (EMF) smaller than 0.8. The EMF is the fraction of the jets' total energy recorded in the EM calorimeter.
- An event is rejected if the invariant mass calculated from the soft-muon and the W-muon, $m_{\mu\mu}$, falls in a range around the Z boson mass (80–100 GeV) or the Υ meson mass (8–11 GeV).

The *tagging efficiency*, which is here defined as the probability that a c-jet is identified with the SMT algorithm, is 0.02919 ± 0.00032 with the event selection criteria adopted for the $W+c$ analysis. This value is driven mainly from the low soft muon branching ratio $BR(c \to \mu X) = 0.096 \pm 0.004$.

The calibration of the SMT algorithm is reported in references [4,5]. The author of this thesis has collaborated to this calibration by providing software packages implementing the SMT algorithm and the mistag rate analysis' event selections.

The efficiency of the soft muon selection criteria and the mistag rate have been measured with data-driven methods. The *mistag rate*, or fake rate, is the probability that a light-flavour jet is mistakenly identified by the SMT algorithm. The simulated samples used in the $W+c$ analysis are corrected to match the soft muon efficiency and mistag rate measured from data.

The efficiency of the soft muon selection criteria has been measured from $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ data samples. The SMT efficiency can be evaluated in events without any reconstructed jet because this tagging algorithm relies on the reconstruction of a muon. Since the soft muon coming from a semileptonic b- and c-quark decay is surrounded by tracks, while muons from J/ψ and Z bosons are isolated, the calibration results have been validated by checking that they do not depend on the muon isolation variables. The SMT efficiency *scale factor*, that is the ratio of the value measured from data and that obtained from simulation, has been found to range from 0.95 to 1.03 depending on the soft muon p_{T} and η, with an uncertainty of $\approx 1\%$.

The mistag rate has been measured using two data samples of dijet events containing light-flavour jets with high purity. The fraction of heavy-flavour jets in these samples has been reduced by using a lifetime-based jet tagger (i.e. based on the information from secondary vertices related to b- and c-hadron decays within jets). For each of the two samples, the number of SMT-tagged jets can be related with the mistag rate and the tagging efficiency. The SMT mistag rate has been extracted by solving the system of two equations and found to range, depending on the jet p_{T} and η, from 0.2% to 0.5%. The scale factor has been found to be 1.44 ± 0.20, independent from the jet p_{T} and η.

4.3 Monte Carlo simulation samples for the W+c and background processes

Event samples obtained from Monte Carlo (MC) simulations of the W+c production and the background processes are used in this thesis for multiple purposes, such as estimation of background contributions or efficiencies.

The theoretical description of pp collision events, which was introduced in section 3.1, involves the simulation of several steps: (i) the hard scattering process according to the evaluated *matrix element* (ME); (ii) the propagation of the hard scattering particles emitting ISR and FSR, which are known as the *parton shower* (PS); (iii) the underlying event; (iv) the formation of hadrons from the partons in the final state, which is known as *fragmentation*, and the consequent hadrons' decay into stable particles. A description of these simulation steps is found, for example, in [67].

Algorithms simulating particle scatterings are known as *generators*. Many generators exist, which differ, for example, in the simulation methods adopted or the scattering processes provided.

The final state particles of simulated collision events are passed as input to the simulation of the response of the ATLAS detector, which is described in [68]. The reconstruction of the physics objects from datasets of simulated events in ATLAS is achieved with the same algorithms used for the LHC collision data. The samples of simulated events used in the $W+c$ analysis are selected with the same criteria described for the data in section 4.2. Samples before the event selection are criteria applied are generated by the ATLAS MC production group.

The list of the MC simulations adopted in this thesis for the $W+c$ production and the background processes, which where described in section 3.5 is found in table 4.3, and described below. The generators used and the cross section multiplied by the corresponding branching ratio are specified. The generators cross sections are corrected, by means of multiplicative *k-factors*, to the more accurate NLO or NNLO theoretical calculations.

The $W+c$, $W+b\bar{b}$, $W+c\bar{c}$ and $W+$light production processes, which are collectively referred to as *W+jets production*, are simulated with Alpgen [70] and, for the PS and fragmentation, with Herwig [72] or Pythia [71]. The latter generator is used for $W+c$ production because its modelling of the c-quark fragmentation and semileptonic decays is expected to provide a more representative description of the data with respect to Herwig.

Additional ad hoc corrections to the simulation of the $W+c$ production are applied in this analysis for the c-quark fragmentation and c-hadrons decays. The fragmentation corrections match: (i) the type and relative population of resulting c-hadrons to the data in [80]; (ii) the fraction of c-hadron energy carried by the c-jet energy to that obtained from Herwig++ [81]. The decay corrections match (i) the branching ratio of the decay into muons of the c-hadrons species to the world

Process	Generator(s)	σ·BR (pb)
Total $W(\to l\nu)$+jets		31×10^3, NNLO [69]
$W(\to l\nu)+c$	Alpgen [70]+Pythia [71]	1.1×10^3
$W(\to l\nu)+c\bar{c}$	Alpgen+Herwig [72]	0.37×10^3
$W(\to l\nu)+b\bar{b}$	Alpgen+Herwig	0.13×10^3
$W(\to l\nu)$+light	Alpgen+Herwig	30×10^3
$Z(\to l\,l)$+jets	Alpgen+Herwig	3.2×10^3, NNLO [69]
$t\bar{t}$	MC@NLO [73]+Herwig	91, NNLO [74]
Single-top (t-channel)	AcerMC [75]+Pythia	21, NNLO [76]
Single-top (s-channel)	MC@NLO+Herwig	1.5, NNLO [77]
Single-top (Wt)	MC@NLO+Herwig	16, NNLO [78]
WW	Herwig	17, NLO [79]
WZ	Herwig	5.7, NLO [79]
ZZ	Herwig	1.3, NLO [79]

Table 4.3: Simulated samples used in this thesis. The generators and the cross section multiplied by the corresponding branching ratio is listed for each sample. Cross sections are corrected to NLO or NNLO calculations where specified. The W+c, W+$c\bar{c}$, W+$b\bar{b}$ and W+light production processes are simulated independently as detailed in the text.

average results [82]; (ii) the distribution of the momentum of outgoing muons in the c-hadron rest frame to that of obtained from EvtGen [83].

In the W+jets simulated samples, the W boson is allowed to decay exclusively in the leptonic channels. The total W+jets production cross section is corrected to the NNLO calculation in [69] and the fraction of each different W+jets process is that predicted with Alpgen.

In the simulation of the Z+jets production, the Z boson is required to decay into an electron, muon, or τ-lepton pair. In the simulation of the $t\bar{t}$, single-top (s- and t-channels) and diboson production processes, at least one W or Z boson decaying in the leptonic channel is required.

Chapter 5

Determination of the W+c production yield

This chapter presents the determination of the number of $W+c$ events, which is used to calculate the production cross section. The measured yield is determined from the data sample selected after estimating and subtracting the contribution from the background processes.

The event selection and the background processes are described in sections 3.5 and 4.2, respectively. To attain reliable estimations, the most relevant backgrounds are evaluated with data-driven methods. In addition, distributions of several kinematical quantities for the events selected from data are compared to predictions.

The contents of this chapter are the own work of the author[1] and have also been published in reference [1].

5.1 Estimation of the background processes

The most relevant background processes in the selection of the $W+c$ signal stem from the QCD multijet, $W+$light and $Z+$jets productions. Their contributions are estimated with data-driven methods, in which event samples independent from that used to extract the $W+c$ yield are selected from the data. The $t\bar{t}$, single-top and

---

[1] In two cases specified in the text, results of the ATLAS top quark analysis group are used as *partial* input of the background estimation method adopted.

diboson backgrounds are estimated from the simulations described in section 4.3. The contribution of the W+$c\bar{c}$ and the W+$b\bar{b}$ productions cancel out in the OS–SS subtraction.

The following definitions of events samples and quantities are used in the process of determining the $W+c$ and the background yields.

- The data event sample used to determine the $W+c$ production yield is referred to as the *signal sample*. Other data event samples employed for background estimation are referred to as *control samples*.

- The *1-jet, 2-jets* and *1,2-jets samples* are selected according to the number of reconstructed jets per event. The latter includes events with either one or two jets, the others include events with exactly one and exactly two jets, respectively. The $W+c$ production yield is evaluated separately in each of these samples.

- Event samples selected without the requirement of exactly one SMT-jet is referred to as *pretag samples*. When the SMT-jet requirement is applied samples are referred to as *OS+SS samples (or tagged samples)*.

- The *SMT tagging rate*, that is the probability that a pretag event contains exactly one SMT-jet, is defined as

  $$R^{\mathrm{SMT}} = \frac{N^{\mathrm{OS+SS}}}{N^{\mathrm{pretag}}} , \tag{5.1.1}$$

  where N^{pretag} and $N^{\mathrm{OS+SS}}$ are the yields of pretag and of OS+SS events, respectively.

- The fraction of tagged events remaining after the OS–SS subtraction can be evaluated from the *OS/SS asymmetry*, which is defined as

  $$A = \frac{N^{\mathrm{OS}} - N^{\mathrm{SS}}}{N^{\mathrm{OS}} + N^{\mathrm{SS}}} , \tag{5.1.2}$$

  where N^{OS} (N^{SS}) is the yield of the opposite (same) sign events.

5.1.1 Multijet background

The contribution of the multijet background to the signal sample is evaluated as

$$N^{\mathrm{OS-SS}}_{\mathrm{multijet}} = N^{\mathrm{pretag}}_{\mathrm{multijet}} \cdot R^{\mathrm{SMT}}_{\mathrm{multijet}} \cdot A_{\mathrm{multijet}} \,, \tag{5.1.3}$$

where $N^{\mathrm{pretag}}_{\mathrm{multijet}}$, $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ and A_{multijet} are the pretag sample yield, the SMT tagging rate and the OS/SS asymmetry of the multijet events, respectively. All the terms of equation 5.1.3 are derived using data-driven methods.

$\mathbf{N^{pretag}_{multijet}}$. The estimation of $N^{\mathrm{pretag}}_{\mathrm{multijet}}$ is derived from data with the method used in [84], which is known as the *matrix method*. A "looser" category for the W-muon is defined by removing the "standard" isolation requirements, $I_{\mathrm{track}} < 2.5$ GeV and $I_{\mathrm{calo}} < 4$ GeV. Accordingly, a *loose sample* is selected from data with the pretag event selection criteria. The events of the loose sample whose W-muon pass the isolation requirements define the *standard (std) sample*, which corresponds to the pretag signal sample. [2]

In the matrix method, the W-muon is classified as *real* if it originates from a W boson decay (or a Z boson decay), and as *fake* otherwise. The yields of the loose and the standard samples, N^{loose} and N^{std} respectively, can be expressed as:

$$\begin{aligned} N^{\mathrm{loose}} &= N^{\mathrm{loose}}_{\mathrm{real}} + N^{\mathrm{loose}}_{\mathrm{fake}} \,, \\ N^{\mathrm{std}} &= N^{\mathrm{std}}_{\mathrm{real}} + N^{\mathrm{std}}_{\mathrm{fake}} = r \cdot N^{\mathrm{loose}}_{\mathrm{real}} + f \cdot N^{\mathrm{loose}}_{\mathrm{fake}} \,. \end{aligned} \tag{5.1.4}$$

The quantities r and f are referred to as the *efficiencies*, and represent the fraction of muons passing the isolation requirements for the real and the fake categories, respectively.

---

[2] The standard and the pretag signal samples are approximately identical. Events containing two muons, one of which is classified as loose and the other as standard, are not included in the loose sample, owing to the requirement of exactly one W-muon. Hence, events of this type are included in the standard sample; however, they are selected in the pretag signal sample. The difference of the yields of the standard and pretag signal samples is 0.2%, which is negligible as the uncertainty on $N^{\mathrm{pretag}}_{\mathrm{multijet}}$ is 11%.

If the efficiencies are known, $N^{\rm std}_{\rm fake}$ can be computed from $N^{\rm loose}$ and $N^{\rm std}$ by solving equations 5.1.4:

$$N^{\rm std}_{\rm fake} = f \cdot N^{\rm loose}_{\rm fake} = \frac{f}{f-r}(N^{\rm std} - r \cdot N^{\rm loose}). \tag{5.1.5}$$

Since muons of the fake category stem from the multijet production processes, this quantity corresponds to $N^{\rm pretag}_{\rm multijet}$. Muons produced by all the other processes (W+jets, Z+jets, top and diboson productions) originate from W (or Z) boson decays, and therefore are of real type.

The efficiencies can be measured in data samples with a high-purity of muons of either the real or the fake type. The values of r and f used in this thesis are provided by the ATLAS top quark analysis group and were measured in two different analyses, which are outlined below.

In the first analysis, which is referred to as *low* $m_{\rm T}^{W}$, f was evaluated in a control sample selected after substituting the requirement $m_{\rm T}^{W} > 60$ GeV with $m_{\rm T}^{W} < 20$ GeV and, in addition, by requiring $m_{\rm T}^{W} + E_{\rm T}^{\rm miss} < 60$ GeV. In the second analysis, which is referred to as d_0 *significance*, f was computed by fitting the distribution of the significance of the transverse impact parameter of the muon track, d_0/σ_{d_0}, in a control sample selected by relaxing the requirement $m_{\rm T}^{W} > 60$ GeV to $m_{\rm T}^{W} + E_{\rm T}^{\rm miss} > 60$ GeV. For both these control data samples, all the other event selection criteria were identical to those adopted for the $W+c$ signal selection. Both analyses measured r in data samples of $Z \to \mu\mu$ events.

To improve the accuracy of the matrix method, the efficiencies were evaluated as a function of the muon pseudorapidity, η^{μ}. Additionally, r and f were also measured as a function of the $p_{\rm T}$ of the leading jet[3], $p_{\rm T}^{\rm jet}$, or of the number of jets, $N_{\rm jets}$, in the low $m_{\rm T}^{W}$ and the d_0 significance analyses, respectively.

The matrix method is generalised to include these parametric efficiencies by defining the event weight

$$w = \frac{f(\eta^{\mu}, v)}{f(\eta^{\mu}, v) - r(\eta^{\mu}, v)}(w^{\rm std} - r(\eta^{\mu}, v)), \tag{5.1.6}$$

where v represents $p_{\rm T}^{\rm jet}$ or $N_{\rm jets}$, and $w^{\rm std}$ is equal to 1 for events of the standard

---

[3] The leading jet is that reconstructed with the largest $p_{\rm T}$ in the event.

sample and to 0 otherwise. The event weight applied to the events of the loose sample gives $\sum_{i\in\text{loose}} w_i = N^{\text{std}}_{\text{fake}}$, which is the equivalent result of equation 5.1.5.

The results for $N^{\text{pretag}}_{\text{multijet}}$ estimated with the matrix method are presented in table 5.1. The yields and statistical uncertainties deriving from both the low m^W_{T} and the d_0 significance efficiencies are shown. The unweighted average of the two results is taken as the best estimation of $N^{\text{pretag}}_{\text{multijet}}$. The systematic uncertainty on the average result is evaluated as half the difference of the two measurements, and results in a value of 11% for the 1,2-jets sample.

	$N^{\text{pretag}}_{\text{multijet}}$		
Matrix method	1-jet	2-jets	1,2-jets
Low m^W_{T}	24080 ± 120	7328 ± 53	31410 ± 130
d_0 significance	19826 ± 83	5180 ± 38	25005 ± 91
Average	22000 ± 2100	6300 ± 1100	28200 ± 3200

Table 5.1: Estimation of the multijet background yield in the pretag sample. The low m^W_{T} and d_0 significance yields are determined from the matrix method with two different measurements of the efficiencies, and their uncertainty is statistical. The average result and half the difference are taken as the best estimation and the systematic uncertainty of $N^{\text{pretag}}_{\text{multijet}}$.

$\mathbf{R^{SMT}_{multijet}}$. Two control samples with high purity of multijet events are selected in data to determine $R^{\text{SMT}}_{\text{multijet}}$ and A_{multijet}. The selection requirements on either the W-muon isolation variables or m^W_{T} are "inverted", while all the others requirement are identical to those used for the $W+c$ signal selection:

- **inverted isolation sample**: $I_{\text{track}} > 3.5$ GeV and $I_{\text{calo}} > 5$ GeV,
- **inverted $\mathbf{m^W_T}$ sample**: $m^W_{\text{T}} < 30$ GeV.

The distributions of the events as a function of the isolation variables and m^W_{T} are presented in figures 5.1 and 5.2, respectively. The pretag and tagged events are shown separately. The values of I_{track}, I_{calo} and m^W_{T} which define the signal and the control samples are displayed. The number of events originating from non-multijet production processes (W+jets, Z+jets, top quark and diboson productions) is determined from simulation. A systematic uncertainty of 25% is assigned from the studies in [84] to each of the W+jets and the Z+jets predictions, and represented by the error bands.

The inverted isolation and inverted m_{T}^{W} samples contain predominantly multijet events. The fractions of W+jets and Z+jets events in the control samples are 7–22%, as shown in table 5.2. The uncertainties are the quadrature sum of the statistical and systematic contributions. The fraction of top quark and diboson events in each control region is collectively at the subpercent level and therefore neglected.

		Fraction of events [%]		
Control sample		W+jets	Z+jets	Total
Inverted isolation	Pretag	21 ± 5	0.78 ± 0.20	22 ± 5
	Tagged	6.6 ± 1.8	0.38 ± 0.12	7.0 ± 1.8
Inverted m_{T}^{W}	Pretag	15 ± 4	5.5 ± 1.4	21 ± 4
	Tagged	10.1 ± 2.5	10.8 ± 2.7	21 ± 4

Table 5.2: Fractions of W/Z+jets events evaluated from simulation in the inverted isolation and inverted m_{T}^{W} samples. Events with one or two reconstructed jets are considered separately for the pretag and the tagged samples.

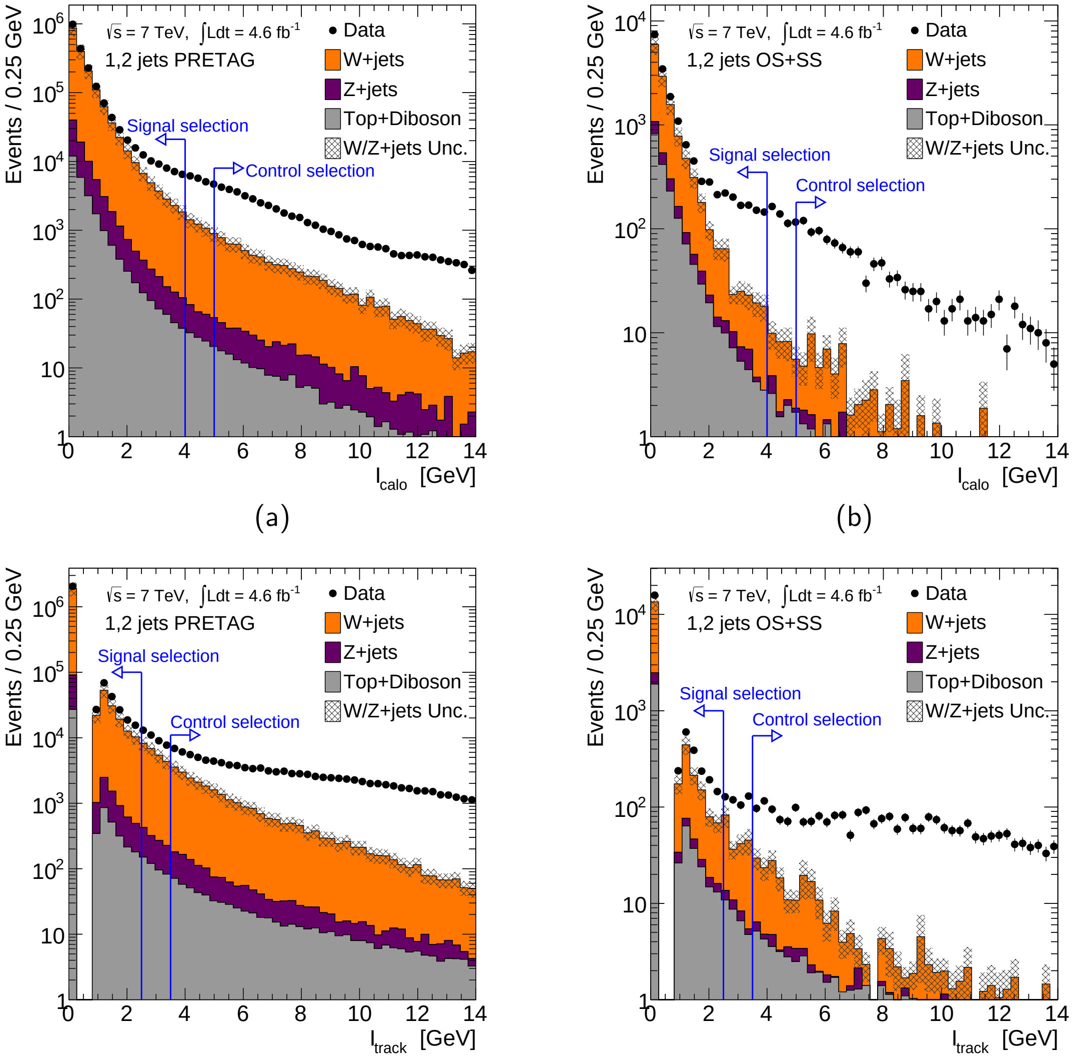
The inverted isolation sample is used to estimate $R^{\mathrm{SMT}}_{\mathrm{multijet}}$. The tagging rate is determined in the control sample as a function of the W-muon's *total isolation*, which is defined as the sum of the track- and the calorimeter-based isolations:

$$I_{\mathrm{tot}} = I_{\mathrm{track}} + I_{\mathrm{calo}}. \tag{5.1.7}$$

The tagging rate is calculated in seven bins of total isolation as

$$R^{\mathrm{SMT}}_{\mathrm{multijet}} = \frac{(N_{\mathrm{data}} - N_{W/Z+\mathrm{jets}})^{\mathrm{OS+SS}}}{(N_{\mathrm{data}} - N_{W/Z+\mathrm{jets}})^{\mathrm{pretag}}}, \tag{5.1.8}$$

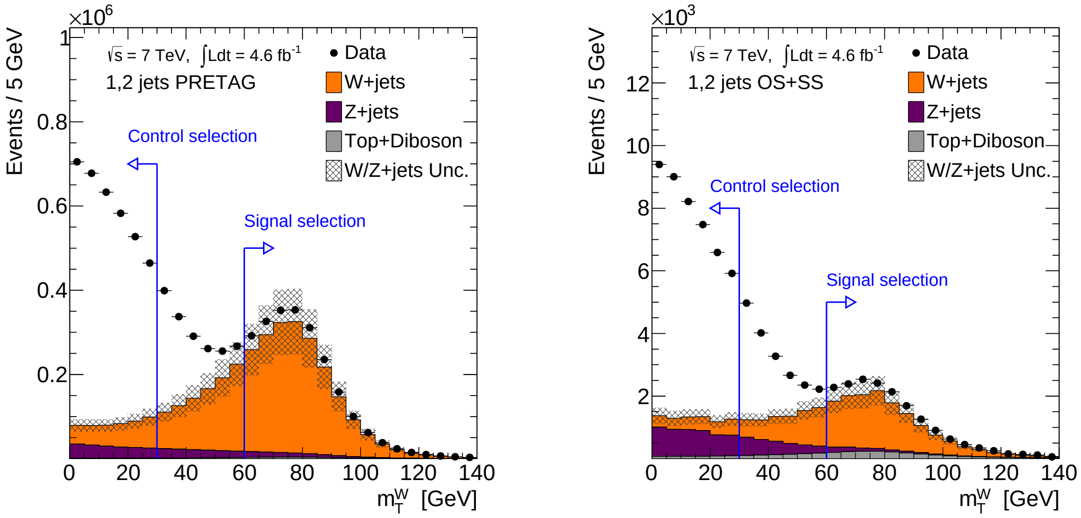
where N_{data} and $N_{W/Z+\mathrm{jets}}$ are the data and W/Z+jets event yields, respectively.



(a) (b)

(c) (d)

Figure 5.1: Distributions of the data events as a function of the of the W-muon isolation variables: (a,b) I_{calo} and (c,d) I_{track}. Events with one or two reconstructed jets are considered for the pretag (a,c) and the tagged (b,d) samples. The values which define the signal and the inverted isolation (control) samples are displayed. Events with large values of the isolation variables are predominantly multijet events. The contributions of non-multijet production processes (W/Z+jets, top quark and diboson) are evaluated from simulation.

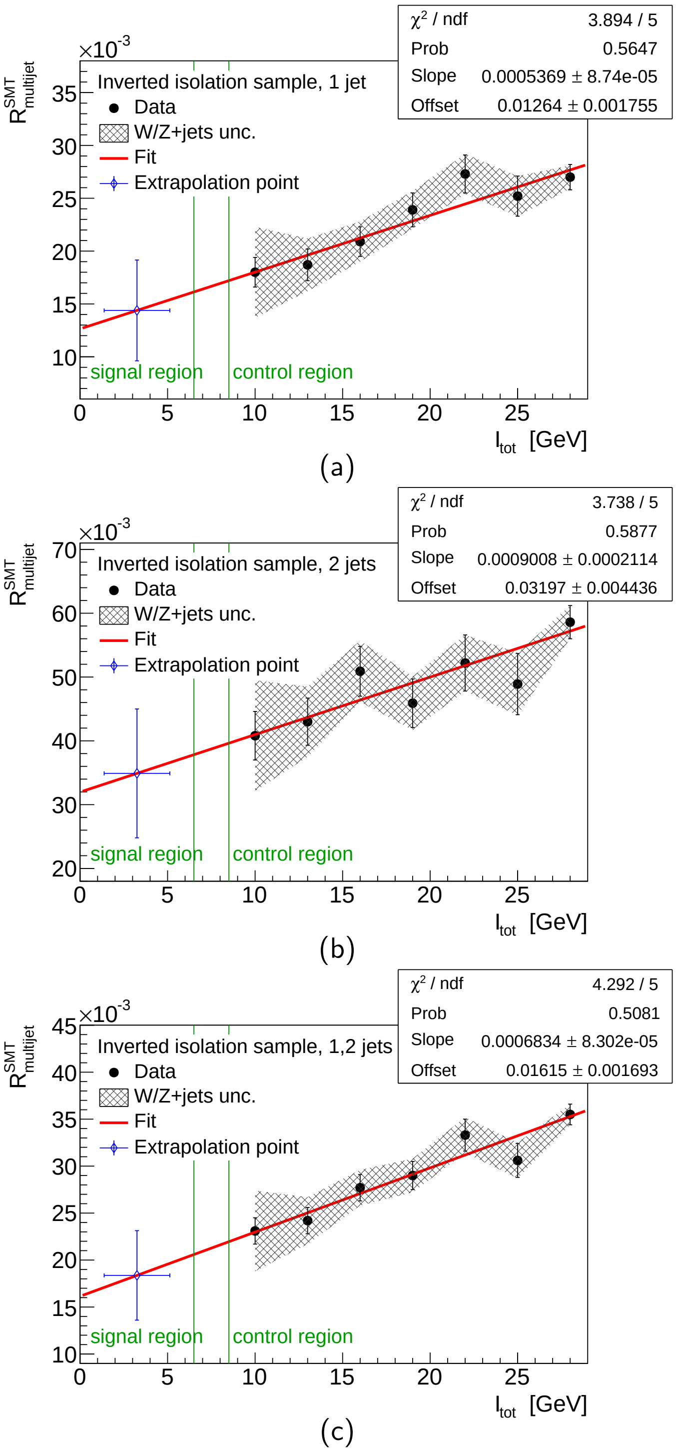


(a) (b)

Figure 5.2: Distributions of the data events as a function of m_{T}^{W}. Events with one or two jets are considered for the pretag (a) and the tagged (b) samples. The values which define the signal and the inverted m_{T}^{W} control samples are displayed. Events with low transverse mass are predominantly multijet events. The contributions of non-multijet production processes (W/Z+jets, top quark and diboson) are evaluated from simulation.

In figure 5.3 the data points show the values of $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ as a function of I_{tot} obtained separately in the 1-jet, 2-jets and 1,2-jets inverted isolation samples. The statistical uncertainties associated with the size of the control data samples are represented with error bars. The error bands show the systematic uncertainties on the W/Z+jets terms of equation 5.1.8, which are evaluated by (i) assigning a 25% uncertainty to the yields predicted for each of the W+jets and the Z+jets productions; (ii) considering uncorrelated uncertainties on $N^{\mathrm{pretag}}_{W/Z+\mathrm{jets}}$ and $N^{\mathrm{OS+SS}}_{W/Z+\mathrm{jets}}$.

The tagging rate decreases as the W-muon is more isolated, i.e. it is reconstructed with a low value of I_{tot}. A χ^2-fit to the data of the control samples, which correspond to the control region $I_{\mathrm{tot}} > 8.5$ GeV, is found to model the correlation between $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ and I_{tot}. Only the statistical uncertainties on the data points are considered in the fitting procedure, while the systematic uncertainties on the method are described below. The choice of the bin-size of I_{tot} is a trade-off



(a)

(b)

(c)

Figure 5.3: Tagging rate of the multijet production as function of the total isolation in the inverted isolation (a) 1-jet, (b) 2-jets and (c) 1,2-jets samples. The error bars and the error bands on the data points represent the statistical and the $W/Z+$jets yields' systematic uncertainties, respectively. The linear χ^2-fit to the control sample data and the point to which the fitted line is extrapolated are shown in red and blue, respectively. The tagging rate at the extrapolation point is taken as $R^{\text{SMT}}_{\text{multijet}}$, whose total systematic uncertainty is shown by the vertical error bar.

between the statistical uncertainties and the number of data points to fit.

The fitted line is extrapolated to the signal region ($I_{\mathrm{tot}} < 6.5$ GeV) assuming that the same $R^{\mathrm{SMT}}_{\mathrm{multijet}}$–$I_{\mathrm{tot}}$ linear relationship holds in the signal and control regions. The value of $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ in the signal region is calculated from the fitted line with $I^{\mathrm{extr}}_{\mathrm{tot}} = 3.25$ GeV. The extrapolation point, which is shown with blue points, correspond to the central value of I_{tot} in the signal selection.

Three different sources of systematic uncertainty are taken into account in the extrapolation of the tagging rate in the signal region. First, the statistical uncertainty on the fit parameters is propagated to the calculation of $R^{\mathrm{SMT}}_{\mathrm{multijet}}$. Second, the fit is repeated after varying the W/Z+jets contributions by their uncertainty, i.e. with the data points shifted by the error bands; the difference in the resulting $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ is taken as a systematic uncertainty. Third, an uncertainty of $6.5/\sqrt{12}$ GeV is assigned to $I^{\mathrm{extr}}_{\mathrm{tot}}$ assuming uniformly distributed total isolation and it is propagated to the calculation of $R^{\mathrm{SMT}}_{\mathrm{multijet}}$. This uncertainty is shown with horizontal error bar on the extrapolation point. The total systematic uncertainty on the signal region's tagging rate is taken as the quadrature sum of the three contributions and it is represented by the extrapolation point's vertical error bar.

Table 5.3 presents $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ and the total uncertainties separately obtained from the fit to the data of the 1-jet, 2-jets and 1,2-jets inverted isolation samples. The relative value of the total uncertainty is 28% in the 1,2-jets sample. The contribution of the three sources of systematic uncertainty is presented and the estimation of the W/Z+jets yields is found to be the dominant one.

As a cross-check of the results in table 5.3, $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ is also computed from equation 5.1.8 using all the events of either the inverted isolation or the inverted m^{W}_{T} samples. Table 5.4 shows the tagging rates so obtained, whose uncertainties are the quadrature sum of the statistical and the systematic components. The latter stems from the estimation of the W/Z+jets yields in the control samples. The unweighted average of the two measurements is also computed, and half the difference is used as the systematic uncertainty[4].

---

[4] The propagation of the uncertainties on the two single measurements to the average is negligible.

	1-jet	2-jets	1,2-jets
$R^{\text{SMT}}_{\text{multijet}}$ [%]	1.4±0.5	3.5±1.0	1.8±0.5
Source	Uncertainty [%]		
Fit parameters	0.18	0.45	0.17
W/Z+jets contributions	0.43	0.89	0.43
Extrapolation point	0.10	0.17	0.13

Table 5.3: Tagging rate of the multijet production evaluated from the fit to data of the inverted isolation sample. The results and the total uncertainties separately determined in the 1-jet, 2-jets and 1,2-jets samples are shown. The contributions from the three sources of systematic uncertainty are individually presented.

As can be seen in table 5.4, the values of $R^{\text{SMT}}_{\text{multijet}}$ obtained from the two control samples are in some tension with each other when considering the uncertainties. This is explained by the different content of heavy-flavour jets, produced in the multijet events, selected in the two control samples. The rate of tagged events depends strongly on the content of heavy-flavour jets. From the results of the fit extrapolation in figure 5.3, the inverted isolation sample is expected to be heavy-flavour-jets-enriched with respect to the multijet events of the signal region. Conversely, due to the low missing energy which is reconstructed in events selected with low transverse mass, the inverted m_{T}^{W} sample is expected to be heavy-flavour-jets-depleted. Hence, the values of $R^{\text{SMT}}_{\text{multijet}}$ obtained from such control samples are expected to "bracket" that of the signal sample.

The estimation of $R^{\text{SMT}}_{\text{multijet}}$ from the fitting procedure and as the average of the control sample measurements leads to compatible results. The fit method is preferred for the final estimation of $R^{\text{SMT}}_{\text{multijet}}$ owing to the smaller uncertainties and the more reliable extrapolation to the signal region. Appendix A describes an extension of this procedure to the estimation of the multijet background for the ATLAS $t\bar{t}$ cross section measurement reported in [4]. This background estimation is the own work of the author of this thesis and exploits a data-driven method similar to that described above.

	$R^{\text{SMT}}_{\text{multijet}}$ [%]		
Control sample	1-jet	2-jets	1,2-jets
Inverted isolation	2.30 ± 0.19	5.01 ± 0.36	2.95 ± 0.33
Inverted m^W_{T}	1.17 ± 0.08	2.34 ± 0.23	1.37 ± 0.22
Average	1.7 ± 0.6	3.7 ± 1.3	2.2 ± 0.8

Table 5.4: Tagging rate of the multijet production evaluated using the events of either the inverted isolation or the inverted m^W_{T} samples. The results for the 1-jet, 2-jets and 1,2-jets samples are reported separately. Their uncertainties include the statistical and systematic components. The average of the two measurements is computed and half the difference is used as the systematic uncertainty. These results are used as a cross-check of the tagging rate obtained from the fit method.

Furthermore, an evaluation of $R^{\text{SMT}}_{\text{multijet}}$ as a function of m^W_{T} was also attempted. This method was unsuccessful because the inverted m^W_{T} sample presents larger contributions from the W/Z+jets productions and the linear dependence hypothesis is not a good description of the data.

$\mathbf{A}_{\text{multijet}}$. The asymmetry of the multijet production processes is evaluated in the inverted isolation and the inverted m^W_{T} samples as

$$A_{\text{multijet}} = \frac{(N_{\text{data}} - N_{W/Z+\text{jets}})^{\text{OS}} - (N_{\text{data}} - N_{W/Z+\text{jets}})^{\text{SS}}}{(N_{\text{data}} - N_{W/Z+\text{jets}})^{\text{OS+SS}}}, \tag{5.1.9}$$

where N^{OS} (N^{SS}) is the number of OS (SS) events, respectively. As shown in table 5.5, the results obtained from the two control samples are compatible. The uncertainties are the quadrature sum of the statistical and systematic components. The latter are associated with the estimation of W/Z+jets yields and they are calculated as described for equation 5.1.8. The average of the asymmetry obtained from the two control regions is taken as the best estimation of A_{multijet}. The uncertainty is set as the largest of those evaluated in the two measurements.

As a cross-check of the results, A_{multijet} is also evaluated after varying the selection requirements of the inverted isolation and the inverted m^W_{T} samples by 1 GeV and 10 GeV, respectively. The asymmetries measured from the modified

	A_{multijet} [%]		
Control sample	1-jet	2-jets	1,2-jets
Inverted isolation	21.7 ± 3.5	25.9 ± 3.9	23.4 ± 2.6
Inverted m_{T}^{W}	23.5 ± 3.8	21.4 ± 4.4	22.9 ± 3.0
Average	22.6 ± 3.8	23.7 ± 4.4	23.2 ± 3.0

Table 5.5: Asymmetry of the multijet production evaluated from the inverted isolation and the inverted m_{T}^{W} samples. The results for the 1-jet, 2-jets and 1,2-jets samples are shown separately. Their uncertainties include the statistical and systematic components. The average value and the largest uncertainty of the two measurements are taken as the best estimation and the uncertainty of A_{multijet}, respectively.

control samples differ within one standard deviation from the results in table 5.5.

Table 5.6 presents the results of the multijet background estimation from equation 5.1.3. The uncertainties on $N_{\text{multijet}}^{\text{pretag}}$, $R_{\text{multijet}}^{\text{SMT}}$ and A_{multijet} include the statistical and systematic components, and they are propagated to the calculation of $N_{\text{multijet}}^{\text{OS}-\text{SS}}$. The relative total uncertainty on the multijet OS–SS yield is 30% in the 1,2-jets sample.

	Multijet background estimation		
	1-jet	2-jets	1,2-jets
$N_{\text{multijet}}^{\text{pretag}}$	22000 ± 2100	6300 ± 1100	28200 ± 3200
$R_{\text{multijet}}^{\text{SMT}}$ [%]	1.4 ± 0.5	3.5 ± 1.0	1.8 ± 0.5
$N_{\text{multijet}}^{\text{OS+SS}}$	320 ± 110	218 ± 82	520 ± 160
A_{multijet} [%]	22.6 ± 3.8	23.7 ± 4.4	23.2 ± 3.0
$N_{\text{multijet}}^{\text{OS}-\text{SS}}$	71 ± 28	52 ± 22	120 ± 39

Table 5.6: Results of the multijet background estimation. The values of $N_{\text{multijet}}^{\text{pretag}}$, $R_{\text{multijet}}^{\text{SMT}}$ and A_{multijet} which are used to evaluate $N_{\text{multijet}}^{\text{OS}-\text{SS}}$ are listed. The uncertainties include the statistical and systematic contributions.

The validity of the multijet background estimation is checked by comparing the predicted m_{T}^{W} distribution with that obtained from data. Figure 5.4 shows the m_{T}^{W} distribution obtained from data after removing the selection requirement $m_{\mathrm{T}}^{W} > 60$ GeV. Events with one or two jets after OS-SS subtraction are considered. The predicted distribution is evaluated as the sum of signal and background distributions. These are obtained from simulation for all processes except for the multijet background. The distributions of Z+jets and W+light events and of $W + c$ events are scaled with the yields estimated and measured in the signal region, respectively (see table 5.16).

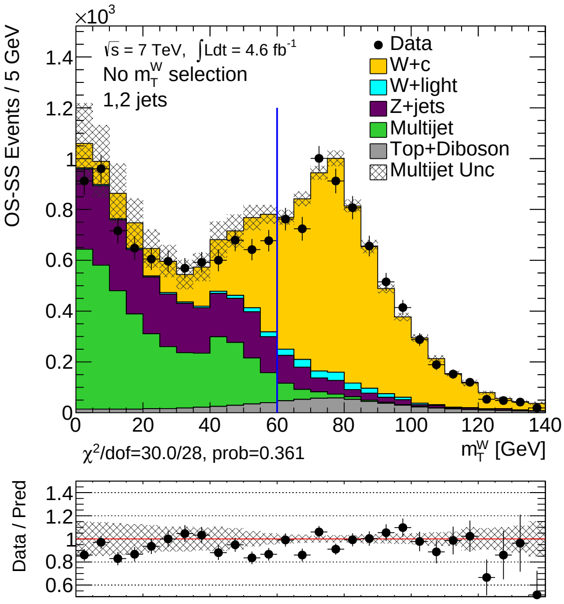


Figure 5.4: Distribution of m_{T}^{W} in data events selected without the $m_{\mathrm{T}}^{W} < 60$ GeV requirement. OS–SS events of the 1,2-jets sample are used. The prediction is obtained by summing the signal and background distributions. The lower panel shows the data-to-prediction ratio and the result of the χ^2-test. The prediction is found to be in good agreement with the data.

The shape of the multijet distribution is taken from the inverted isolation sample. The small contribution of non-multijet events to this control data sample, which arises from the W+jets and Z+jets productions, is neglected (0.086 ± 0.024 after the OS–SS subtraction). The yield of multijet events in the region $m_{\mathrm{T}}^{W} > 60$ GeV is taken from table 5.6. In the region $m_{\mathrm{T}}^{W} < 60$ GeV, the yield of multijet events is calculated as done in equation 5.1.3 where: i) the multijet pretag yield is evaluated with the matrix method applied to the data selected with $m_{\mathrm{T}}^{W} < 60$ GeV;

ii) the multijet SMT tagging rate and asymmetry are taken from the inverted m_{T}^{W} control sample, table 5.4 and table 5.5, respectively.

The uncertainty on the data distribution is statistical. The uncertainty on the predicted distribution in the region $m_{\mathrm{T}}^{W} < 60$ GeV is dominated by that on the multijet yield. The data-to-prediction ratio is shown in the lower panel. A χ^2-test[5] is performed and the predicted m_{T}^{W} distribution is found to be in a good agreement with the data (χ^2/dof=30.0/28).

5.1.2 W+light background

The contribution of the W+light background to the signal sample is estimated as

$$N_{W+\mathrm{light}}^{\mathrm{OS-SS}} = N_{W+\mathrm{jets}}^{\mathrm{pretag}} \cdot f_{\mathrm{light}} \cdot R_{W+\mathrm{light}}^{\mathrm{SMT}} \cdot A_{W+\mathrm{light}}\,, \tag{5.1.10}$$

where $N_{W+\mathrm{jets}}^{\mathrm{pretag}}$ is the yield of W+jets events in the pretag sample, and f_{light} is the fraction of W+light events in $N_{W+\mathrm{jets}}^{\mathrm{pretag}}$. The quantities $R_{W+\mathrm{light}}^{\mathrm{SMT}}$ and $A_{W+\mathrm{light}}$ are the tagging rate and the OS/SS asymmetry of the W+light production, respectively. All the terms of equation 5.1.10 are derived using data-driven methods.

$\mathbf{N}_{\mathbf{W+jets}}^{\mathbf{pretag}}$. The yield of pretag events of the W+jets production in the data sample is evaluated after subtracting the non-W+jets contributions:

$$N_{W+\mathrm{jets}}^{\mathrm{pretag}} = N_{\mathrm{data}}^{\mathrm{pretag}} - \sum_{i \notin W+\mathrm{jets}} N_{i}^{\mathrm{pretag}}. \tag{5.1.11}$$

The sum index i in the equation includes: the multijet production, which is estimated with the matrix method described in section 5.1.1, and the Z+jets, top quark and diboson productions, which are evaluated from simulation. The systematic uncertainty assigned to the Z+jets simulation prediction is 25%, as explained in section 5.1.1. The uncertainties on the $t\bar{t}$, single-top and diboson yields are evaluated as explained in section 5.1.4. The pretag yields used in equation 5.1.11

---

[5] The statistical test used to compare the data and predicted shapes is $\chi^2 = \sum_{i=1}^{N_{\mathrm{bins}}} \frac{(n_i - m_i)^2}{(\delta n_i)^2 + (\delta m_i)^2}$, where n_i (m_i) and δn_i (δm_i) are the event yield and the uncertainty in the i-th bin, respectively.

and the resulting $N^{\mathrm{pretag}}_{W+\mathrm{jets}}$ are listed in table 5.7. In the 1,2-jets signal sample, 96% of the pretag events selected from data stem from the W+jets production.

	N^{pretag}		
	1-jet	2-jets	1,2-jets
Data	1,839,910	403,853	2,243,763
Multijet	$22{,}000 \pm 2{,}100$	$6{,}300 \pm 1{,}100$	$28{,}200 \pm 3{,}200$
Z+jets	$59{,}000 \pm 15{,}000$	$15{,}000 \pm 4{,}000$	$74{,}000 \pm 19{,}000$
$t\bar{t}$	$2{,}240 \pm 320$	$8{,}200 \pm 800$	$10{,}000 \pm 1{,}000$
Single top	$4{,}200 \pm 500$	$5{,}100 \pm 500$	$9{,}200 \pm 900$
Diboson	$6{,}400 \pm 1{,}700$	$4{,}400 \pm 1{,}200$	$11{,}000 \pm 2{,}700$
W+jets	$1{,}746{,}000 \pm 15{,}000$	$365{,}000 \pm 4{,}000$	$2{,}111{,}000 \pm 19{,}000$

Table 5.7: Pretag events yields observed in data and estimated for the non-W+jets production processes (multijet, Z+jets, $t\bar{t}$, single-top and diboson), which are used to evaluate $N^{\mathrm{pretag}}_{W+\mathrm{jets}}$. The uncertainties include the statistical and systematic components.

f_{light}. To obtain the fraction of W+light events in $N^{\mathrm{pretag}}_{W+\mathrm{light}}$, f_{light}, a scale factor correction provided by the ATLAS top quark analysis group is applied to the value obtained from simulation, $f_{\mathrm{light,MC}}$. The correction's derivation, which exploits a lifetime-based technique to tag the heavy-flavour jets, is found in reference [85] and summarised below.

The yields of pretag and tagged events for the W+jets production are evaluated from data as per equation 5.1.11. The tagged events are those containing at least one heavy-flavour jet, identified by the tagging algorithm known as MV1 [86].

The *MV1 tagging algorithm* is based on the information from the vertices related to b- and c-hadron decays within jets. The algorithm inputs are: the number, the masses, track multiplicities, the track energy fraction of the reconstructed vertices as well as the transverse and longitudinal impact parameter significances of each track within the jet. This information is combined, using a neural network, into a single variable, which allows discriminating heavy- and light-flavour jets.

The tagger operating point used in this analysis corresponds to a 70% tagging efficiency for b-jets in simulated $t\bar{t}$ events.

Additionally, the tagged and pretag events are divided into two samples according to the charge of the W boson. The system of four equations relates the yield of pretag ($N^{\mathrm{pretag}}_{W+\mathrm{jets}}$) and tagged ($N^{\mathrm{MV1}}_{W+\mathrm{jets}}$) events for the W+jets production:

$$\begin{aligned} N^{\mathrm{MV1},\pm}_{W+\mathrm{jets}} &= N^{\mathrm{pretag},\pm}_{W+\mathrm{jets}} \left(\sum_{i=bb,cc,c,light} k_i \cdot f^{\pm}_{i,\mathrm{MC}} \cdot R^{\mathrm{MV1}}_i \right) , \\ 1 &= \sum_{i=bb,cc,c,light} k_i \cdot f_{i,\mathrm{MC}} , \\ k_{cc} &= k_{bb} . \end{aligned} \tag{5.1.12}$$

In the relations above, the (±) sign refers to the two independent equations of the W^+ and W^- samples, and the sum index i includes the four different processes of the W+jets production ($W+b\bar{b}$, $W+c\bar{c}$, $W+c$ and W+light). The quantities $f_{i,\mathrm{MC}}$ are the fractions of pretag events of the W+jets production which are associated with each of the four processes, as evaluated from simulation. The values k_i, which are unknown quantities of the equations system, represent the scale factors of $f_{i,\mathrm{MC}}$. The rates of MV1-tagged events, R^{MV1}_i, are evaluated from simulation. Finally, to obtain four unknown scale factors in four equations, the ratio of pretag events from the $W+c\bar{c}$ and $W+b\bar{b}$ productions is fixed to the value obtained from simulation, i.e. $k_{cc} = k_{bb}$.

The system of equations 5.1.12 is solved to find the scale factors k_i. The value of f_{light} is calculated as $f_{\mathrm{light}} = k_{\mathrm{light}} \cdot f_{\mathrm{light,MC}}$. Since the $W+b\bar{b}$ and $W+c\bar{c}$ backgrounds cancel out with the OS–SS subtraction, the respective scale factors are not used.

The results for the W+light pretag fractions are shown in table 5.8, separately for events with exactly one or exactly two jets. The fractions obtained from simulation $f_{\mathrm{light,MC}}$ are shown with the statistical uncertainty. The scale factors, which are provided by the ATLAS top quark analysis group, are compatible with unity and the systematic uncertainty associated with the method above is 8%. As a check of this results, the scale factors have been recalculated with the method above by the thesis author and the results found are compatible with those adopted within one sigma. The fraction of W+light events in the pretag W+jets sample is 0.85 (0.77) in the 1-jet (2-jets) sample. The yield of pretag W+light events is calculated as $N^{\text{pretag}}_{W+\text{light}} = N^{\text{pretag}}_{W+\text{jets}} \cdot f_{\text{light}}$, using the results in table 5.7. The yield of $W{+}b\bar{b}$, $W{+}c\bar{c}$ and $W{+}c$ events, which are collectively referred to as W+heavy production, is computed as $N^{\text{pretag}}_{W+\text{heavy}} = N^{\text{pretag}}_{W+\text{jets}} \cdot (1 - f_{\text{light}})$.

	1-jet	2-jets
$f_{\text{light,MC}}$	0.8461 ± 0.0013	0.7711 ± 0.0020
k_{light}	1.01 ± 0.08	1.00 ± 0.08
f_{light}	0.85 ± 0.07	0.77 ± 0.06
$N^{\text{pretag}}_{W+\text{light}}$	$1{,}490{,}000 \pm 120{,}000$	$280{,}000 \pm 23{,}000$
$N^{\text{pretag}}_{W+\text{heavy}}$	$257{,}000 \pm 120{,}000$	$85{,}000 \pm 23{,}000$

Table 5.8: Fraction of pretag events for the W+light production, f_{light} obtained by applying the scale factor k_{light} to the value obtained from simulation $f_{\text{light,MC}}$. The correction was evaluated with the data-driven method described in the text. The resulting yields for the W+light and W+heavy ($W{+}b\bar{b}$, $W{+}c\bar{c}$ and $W{+}c$) productions in the pretag sample is also shown. The uncertainties include the statistical and systematic components, expect for $f_{\text{light,MC}}$ for which uncertainties are statistical.

As a check of the data selected with the pretag selection criteria and of the W+light pretag yields, distributions of several kinematical quantities are presented in appendix B. The distributions observed in data are compared to the predictions from simulation, normalised with the event yields estimated in tables 5.7 and 5.8. The predicted distributions are found to be representative of the data, with differences of the order of the percent.

$\mathbf{R^{SMT}_{W+light}}$. The probability that a jet in a W+light event is mistakenly tagged as an SMT-jet, $R^{\text{SMT}}_{W+\text{light}}$, is evaluated from simulation. The latter, as explained in section 4.2.3, is corrected for the rate of mistagging SMT-jets measured in data (scale factor). The rate $R^{\text{SMT}}_{W+\text{light}}$ is recalculated after varying the mistag scale factor within the respective uncertainty (14%) and the difference is taken as the systematic uncertainty. The results for $R^{\text{SMT}}_{W+\text{light}}$ are shown in table 5.9. The estimated rate of mistagged events is 0.27% in the 1,2-jets sample. The yield of OS+SS events is calculated from the results in table 5.8 as $N^{\text{OS+SS}}_{W+\text{light}} = N^{\text{pretag}}_{W+\text{light}} \cdot R^{\text{SMT}}_{W+\text{light}}$.

	1-jet	2-jets	1,2-jets
$R^{\text{SMT}}_{W+\text{light}}$ [%]	0.211 ± 0.030	0.58 ± 0.08	0.27 ± 0.04
$N^{\text{OS+SS}}_{W+\text{light}}$	$3{,}100 \pm 500$	$1{,}610 \pm 260$	$4{,}700 \pm 800$

Table 5.9: Rate of SMT tagged events for the W+light production. The values of $R^{\text{SMT}}_{W+\text{light}}$ are shown in percent and evaluated from the simulation corrected for the mistag rate measured in data. The resulting yield of W+light events in the OS+SS signal sample is shown. The uncertainties include the statistical and systematic components.

$A_{W+\text{light}}$. The OS/SS asymmetry of the W+light production, $A_{W+\text{light}}$, is estimated in two steps. First, the OS/SS *tracks* asymmetry of the W+light production is evaluated in the pretag events selected from data ($A^{\text{tracks}}_{W+\text{light,data}}$). The *OS/SS tracks asymmetry* is defined as

$$A^{\text{tracks}} = \frac{N^{\text{OS,tracks}} - N^{\text{SS,tracks}}}{N^{\text{OS,tracks}} + N^{\text{SS,tracks}}}, \tag{5.1.13}$$

where OS and SS yields are counted using the charge of the W boson and of each generic track selected with the soft-muon kinematic requirements (table 4.2). Then, $A^{\text{tracks}}_{W+\text{light,data}}$ is extrapolated to the signal region according to

$$A_{W+\text{light}} = k_A \cdot A^{\text{tracks}}_{W+\text{light,data}}, \tag{5.1.14}$$

where the extrapolation factor k_A is obtained from simulation.

In the data pretag sample, the OS/SS tracks asymmetry measured can be expressed as

$$A^{\text{tracks}}_{\text{data}} = \sum_i F_i \cdot A^{\text{tracks}}_i, \tag{5.1.15}$$

where the sum index i includes all contributing processess: W+light, W+heavy, Z+jets, multijet, top quark and diboson productions. The quantity F_i is the fraction of pretag events in the data originating from the process i. Hence, the

OS/SS tracks asymmetry for the W+light production can be extracted from that measured in data using the equation:

$$A^{\text{tracks}}_{W+\text{light,data}} = \frac{1}{F_{W+\text{light}}} \left(A^{\text{tracks}}_{\text{data}} - \sum_{i \neq W+\text{light}} F_i \cdot A^{\text{tracks}}_i \right). \tag{5.1.16}$$

Table 5.10 lists the event fractions and OS/SS tracks asymmetries used in equation 5.1.16. The fractions F_i are evaluated from the yields calculated in tables 5.7 and 5.8, and the uncertainties include the statistical and systematic components. The asymmetries A^{tracks}_i are evaluated from simulation for all the processes except the multijet production, for which the matrix method is used, and the uncertainties are statistical, except for that of the W+heavy events. Since the W+heavy production gives the largest non-W+light contribution ($F_{W+\text{heavy}}$ is 16% in the 1,2-jets sample), systematic uncertainties on the simulation prediction of $A^{\text{tracks}}_{W+\text{heavy}}$ are evaluated as described below.

	F [%]	A^{tracks} [%]	F [%]	A^{tracks} [%]	F [%]	A^{tracks} [%]
	1-jet		2-jets		1,2-jets	
Data	-	9.54 ± 0.07	-	5.89 ± 0.10	-	8.29 ± 0.06
W+light	81 ± 7	-	69 ± 6	-	79 ± 6	-
W+heavy	15 ± 7	6.4 ± 0.4	22 ± 6	5.21 ± 0.28	16 ± 6	5.90 ± 0.32
Z+jets	3.0 ± 0.8	0.83 ± 0.28	3.5 ± 0.9	0.28 ± 0.23	3.1 ± 0.8	0.62 ± 0.19
Multijet	1.19 ± 0.19	1.4 ± 0.4	1.6 ± 0.4	1.8 ± 0.5	1.26 ± 0.20	1.52 ± 0.31
Singletop	0.224 ± 0.026	8.89 ± 0.26	1.24 ± 0.13	7.28 ± 0.18	0.41 ± 0.04	7.75 ± 0.15
$t\bar{t}$	0.114 ± 0.020	3.08 ± 0.34	1.88 ± 0.27	3.79 ± 0.13	0.43 ± 0.06	3.71 ± 0.12
Diboson	0.34 ± 0.09	12.54 ± 0.20	1.08 ± 0.28	10.64 ± 0.17	0.48 ± 0.12	11.43 ± 0.13

Table 5.10: Fraction of events and OS/SS tracks asymmetry for the processes of the pretag data sample, which are used for the data-driven estimation of $A^{\text{tracks}}_{W+\text{light,data}}$ in-equation 5.1.16. The uncertainties on F_i and $A^{\text{tracks}}_{W+\text{heavy}}$ include statistical and systematic contribution, the other uncertainties are statistical.

The OS/SS tracks asymmetry obtained for the $W{+}b\bar{b}$, $W{+}c\bar{c}$ and $W+c$ productions are different, as shown in table 5.11. Hence, $A^{\text{tracks}}_{W+\text{heavy}}$ depends on the relative contributions of the three processess to the W+heavy sample and the associated systematic uncertainty is derived as follows. The asymmetry is expressed as $A^{\text{tracks}}_{W+\text{heavy}} = \sum_{j=bb,cc,c} f'_j \cdot A^{\text{tracks}}_j$, where f'_j is the fraction of events of the process j. The event fractions are allowed to vary independently by 50%[6] with respect to the values determined from the simulated samples, while their sum is constrained to unity. The value of $A^{\text{tracks}}_{W+\text{heavy}}$ is minimised and maximised with respect to the fractions f'_j. The difference of the maximum and the minimum value divided by $\sqrt{12}$ is taken as the systematic uncertainty.

	f' [%]	A^{tracks} [%]	f' [%]	A^{tracks} [%]	f' [%]	A^{tracks} [%]
	1-jet		2-jets		1,2-jets	
$W+b\bar{b}$	8 ± 4	8.5 ± 0.6	13 ± 6	5.1 ± 0.5	9 ± 4	6.6 ± 0.4
$W+c\bar{c}$	22 ± 11	10.13 ± 0.31	32 ± 16	6.64 ± 0.29	25 ± 12	8.35 ± 0.21
$W+c$	70 ± 30	4.99 ± 0.15	55 ± 28	4.39 ± 0.20	70 ± 30	4.78 ± 0.12
$W+$heavy	-	6.4 ± 0.4	-	5.20 ± 0.28	-	5.81 ± 0.32

Table 5.11: Fraction of events and OS/SS tracks asymmetry for the $W+b\bar{b}$, $W+c\bar{c}$ and $W+c$ processes, which compose the $W+$heavy sample. The uncertainties on the event fractions and $A^{\text{tracks}}_{W+\text{heavy}}$ include statistical and systematic contributions, the other uncertainties are statistical. The systematic uncertainty on $A^{\text{tracks}}_{W+\text{heavy}}$ is evaluated after varying independently the event fractions.

The results for the $W+$light background estimation of equation 5.1.10 are presented in table 5.12. The OS/SS tracks asymmetry measured in data is 8.7% for the 1,2-jets sample and the uncertainties include the statistical and systematic components. The extrapolation factor is obtained as

$$k_A = \frac{A_{W+\text{light,MC}}}{A^{\text{tracks}}_{W+\text{light,MC}}}, \tag{5.1.17}$$

where $A_{W+\text{light,MC}}$ and $A^{\text{tracks}}_{W+\text{light,MC}}$ are the OS/SS asymmetries obtained from simulation with the soft muon and with the tracks, respectively. The uncertainty on k_A is statistical and is found to be 31% in the 1,2-jets sample. Such an uncertainty derives from the small number of simulated $W+$light events which are reconstructed with a SMT-jet. The final OS/SS asymmetry estimated for the $W+$light background in the 1,2-jets sample is 5.7%, whose uncertainty is dominated by

---

[6] This is an arbitrary but conservative value of the uncertainty on the on the fractions f'_j.

that on the extrapolation factor k_A. The yield of OS–SS events is calculated as $N^{\mathrm{OS-SS}}_{W+\mathrm{light}} = N^{\mathrm{OS+SS}}_{W+\mathrm{light}} \cdot A_{W+\mathrm{light}}$ from the results in table 5.9.

W+light background estimation	1-jet	2-jets	1,2-jets
$A^{\mathrm{tracks}}_{W+\mathrm{light,data}}$ [%]	10.0 ± 0.8	5.8 ± 0.6	8.7 ± 0.7
$A_{W+\mathrm{light,MC}}$ [%]	9.6 ± 3.0	3.0 ± 3.2	7.3 ± 2.2
$A^{\mathrm{tracks}}_{W+\mathrm{light,MC}}$ [%]	13.8 ± 0.11	7.58 ± 0.12	11.83 ± 0.08
k_A	0.70 ± 0.22	0.4 ± 0.4	0.61 ± 0.19
$A_{W+\mathrm{light}}$ [%]	7.0 ± 2.2	2.3 ± 2.5	5.3 ± 1.7
$N^{\mathrm{OS-SS}}_{W+\mathrm{light}}$	220 ± 80	40 ± 40	250 ± 90

Table 5.12: Results of the estimation of the W+light background. The estimation of $A^{\mathrm{tracks}}_{W+\mathrm{light,data}}$, which include statistical and systematic uncertainties, is extrapolated the to signal sample with the multiplicative factor k_A. This is evaluated from simulation and the uncertainties are statistical. The OS/SS asymmetry of the W+light background, $A_{W+\mathrm{light}}$, is used to evaluate $N^{\mathrm{OS-SS}}_{W+\mathrm{light}}$ from equation 5.1.10 and the results in tables 5.7 to 5.9.

5.1.3 Z+jets background

The Z+jets production is a prominent background to the $W+c$ analysis of this thesis. [7] Since events with two muons are selected, one for the W-boson reconstruction and the other for the c-jet tagging, the $Z \to \mu^+\mu^-$ and signal events are similar. Moreover, owing to the opposite charge of the two muons, only a small part of this background is cancelled with OS–SS subtraction.

Two selection criteria were introduced in the W+c analysis from a study by the author in order to suppress the Z+jets background. The selection criteria, one of which on the SMT-jet and the other on the invariant mass of the soft and W-decay muons, were introduced in section 4.2.3 and are explained below. In addition, the

---

[7] In the similar analysis using $W \to e\nu$ decays, the Z+jets background is smaller than 1%. When referring the Z boson production, the contribution of γ^* is also included.

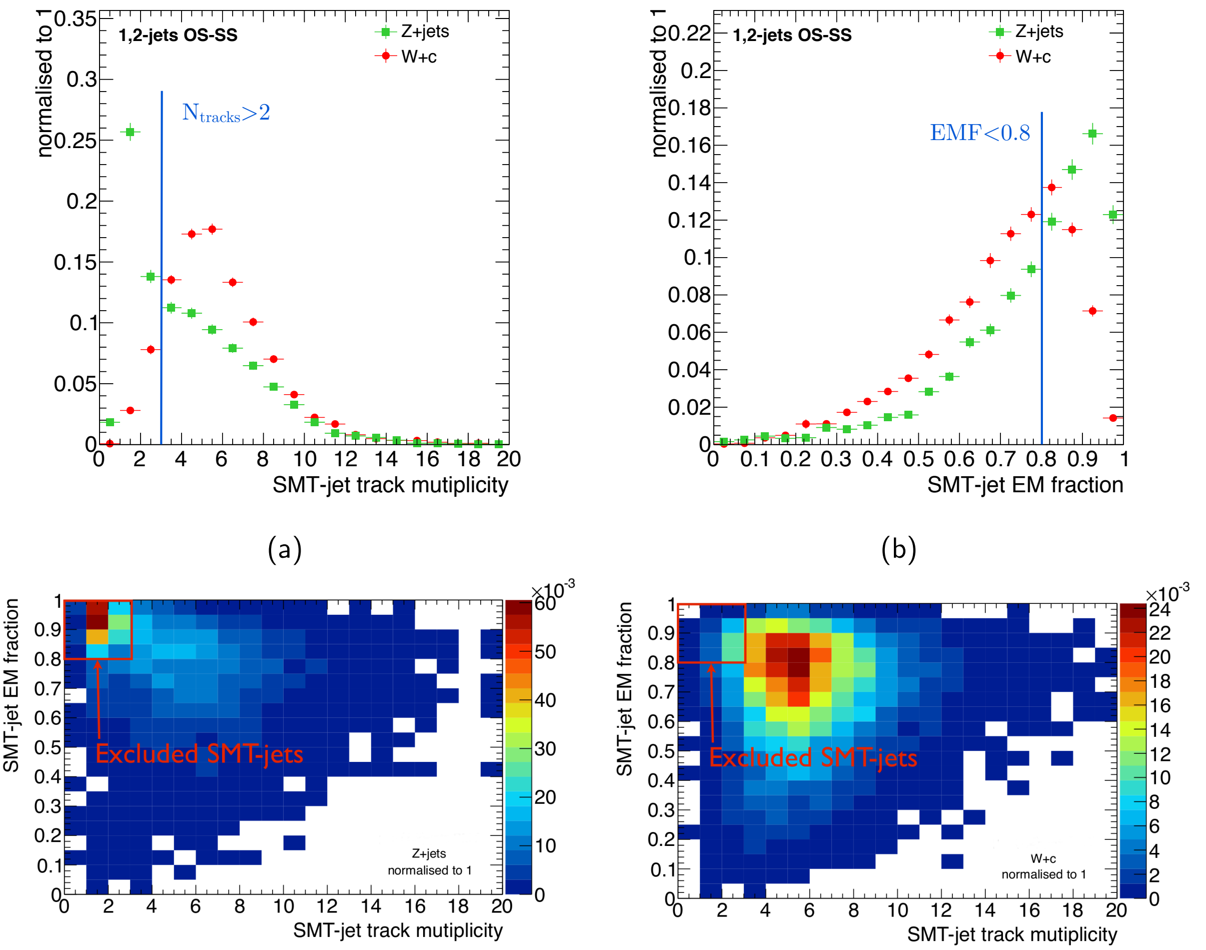
yield of the remaining Z+jets background is estimated from a control sample of Z-resonance events selected from data, as described below.

Selection requirements to suppress the Z+jets background

Muons from the $Z \to \mu^+\mu^-$ decays can radiate photons by interacting with the detector material and be falsely reconstructed as jets. In addition, the Z-decay muons can be accidentally associated with jets originating in the collision event. Both such processes lead to the reconstruction of jets containing a muon which can be mistakenly identified as c-jets by the SMT.

The jets reconstructed from photon radiation have a larger electromagnetic fraction (EMF) and a smaller number of associated tracks (N_{tracks}) with respect to those produced by the c-quark hadronisation. This is shown in figure 5.5, which compares the distribution of (a) EMF and of (b) N_{tracks} for the SMT-jet reconstructed in simulated events of the Z+jets and $W+c$ productions. The events are selected with all requirements of the signal sample except for those on EMF/N_{tracks} and the dimuon invariant mass, which has been removed. The OS–SS events with one or two jets are used and the distributions are normalised to unity. To reduce the Z+jets background in the $W+c$ analysis, the SMT-jets are required to be reconstructed with $N_{\text{tracks}} > 2$ or EMF<0.8. The distribution of EMF versus N_{tracks} is presented in figures 5.5 (c) and (d) for the Z+jets and W+c simulations, respectively. The fraction of SMT-jets which is rejected by the selection requirement is highlighted with a red box, and corresponds to 32.2% and 4.59% for the Z+jets and W+c productions, respectively.

In the majority of $Z \to \mu\mu$ events passing the $W+c$ selection, the invariant mass of the soft and W-decay muons, $m_{\mu\mu}$, is close to that of the Z boson. Conversely, in $W+c$ events, $m_{\mu\mu}$ is expected to be small with respect the Z boson mass. This is shown in figure 5.6 (a), which compares the $m_{\mu\mu}$ distribution obtained from the Z+jets and $W+c$ simulations. Events of the 1,2-jets sample after the OS–SS subtraction are considered, the EMF/N_{tracks} selection requirement is applied and the two distributions are normalised to unity. To suppress the Z+jets background in the $W+c$ analysis, events are vetoed if $m_{\mu\mu}$ is found within the Z



(a)

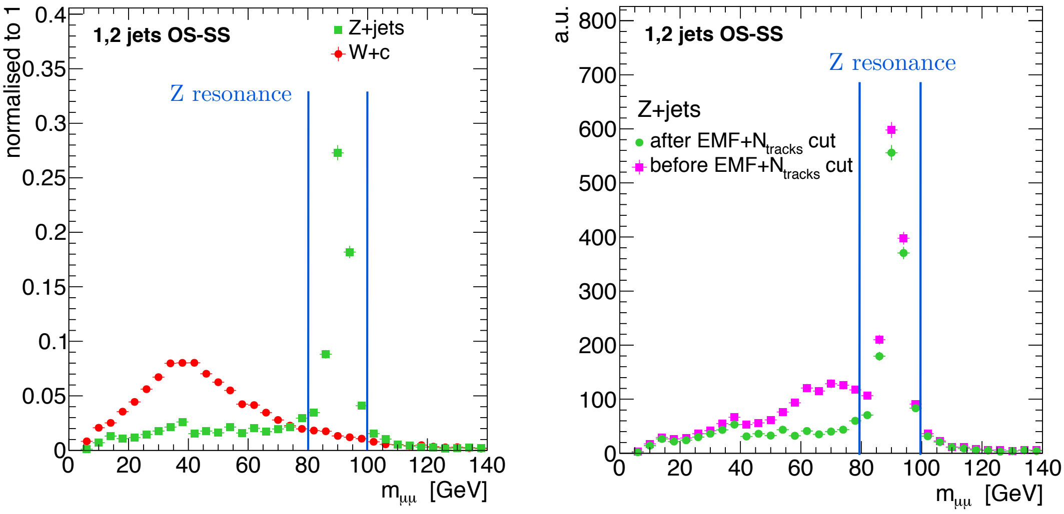
(b)

(c) Z+jets

(d) W+c

Figure 5.5: Distributions of (a) track multiplicity, N_{tracks}, and of (b) electromagnetic fraction, EMF, for the reconstructed SMT-jet in Z+jets (green squares) and W+c (red cirlces) events. In (c) and (d) the distribution of EMF versus N_{tracks} are shown for Z+jets and W+c events, respectively. The distributions are obtained from simulation and normalised to unity. Events of the 1,2-jets sample are considered after the OS–SS subtraction. One of the two selection requirements used to suppress the Z+jets background in the W+c analysis is displayed: the SMT-jets are required to be reconstructed with $N_{\text{tracks}} > 2$ or EMF<0.8.

boson resonance defined as 80–100 GeV. The fraction of events vetoed with this selection requirement is 78.2% and 7.69% for the Z+jets and $W+c$ productions, respectively.



(a) (b)

Figure 5.6: Distribution of the invariant mass of the soft and W-decay muons, $m_{\mu\mu}$. In (a) the distributions of simulated Z+jets (green squares) and $W+c$ (red circles) events are compared. In (b) the effect of the EMF/N_{tracks} selection requirement on the distribution obtained for the simulated Z+jets events is shown. The magenta circles (green squares) points represent events before (after) applying the selection requirement. The region corresponding to the Z boson resonance, 80–100 GeV, is excluded from the signal sample selected from data. All distributions are for the 1,2-jets sample after OS–SS subtraction.

The rejection of the EMF/N_{tracks} selection requirement as a function of $m_{\mu\mu}$ can be seen in figure 5.6 (b). Simulated Z+jets events are shown before (magenta) and after (green) applying the selection requirement. The majority of the events vetoed has $m_{\mu\mu} < 80$ GeV and hence is included in the signal region. In the latter, the fraction of events rejected by the EMF/N_{tracks} selection requirement results 55.2% and 4.45% for the Z+jets and $W+c$ productions, respectively.

In conclusion, two selection requirements are adopted in the $W+c$ analysis to suppress the Z+jets background: (i) the SMT-jet is selected if EMF<0.8 or $N_{\text{tracks}} > 2$; (ii) the events are selected if $m_{\mu\mu}$<80 GeV or $m_{\mu\mu}$>100 GeV. The fraction of events rejected by adding the two selection criteria is listed in table 5.13, separately for the 1-jet, 2-jets and 1,2-jets samples. Events after the OS–SS subtraction are used and the uncertainties are statistical. In the 1,2-jets sample, the two selection requirements reject 82% of the background events and 12% of the signal events. The sample selected from data is reduced by 26%.

	Fraction of rejected events [%]		
	1-jet	2-jets	1,2-jets
Z+jets	83.1 ± 0.9	79.3 ± 1.2	81.5 ± 0.7
$W+c$	12.1 ± 0.5	10.1 ± 0.8	11.5 ± 0.4
Data	24.2 ± 0.4	28.8 ± 0.5	25.53 ± 0.30

Table 5.13: Fraction of events rejected by the selection requirements adopted to suppress the Z+jets background: (i) SMT-jet with EMF<0.8 or $N_{\text{tracks}} > 2$; (ii) events with $m_{\mu\mu}$<80 GeV or $m_{\mu\mu}$>100 GeV. The fractions are obtained from simulation for the Z+jets background and $W+c$ signal. The reduction of the data sample events is also shown. The uncertainties are statistical.

Data-driven estimation of the Z+jets background

The remaining Z+jets background is evaluated from simulation and corrected for the Z+jets yield measured in a data control sample. The latter is selected with the criteria of the signal sample except for the requirement on $m_{\mu\mu}$ which is inverted to obtain events within the Z boson resonance 80–100 GeV. In addition, to minimise the non-Z+jets contribution to the control sample, only events containing exactly one jet are considered.

The Z resonance control region is used to determine the correction scale factor

$$k_{Z+\text{jets}} = \frac{N^{\text{OS}-\text{SS,CR}}_{Z+\text{jets,data}}}{N^{\text{OS}-\text{SS,CR}}_{Z+\text{jets,MC}}}, \tag{5.1.18}$$

where the numerator (denominator) is the OS–SS yield of Z+jets events obtained from data (simulation). The scale factor is extrapolated to the signal region, in which the Z+jets background is estimated as $N^{\text{OS}-\text{SS}}_{Z+\text{jets}} = k_{Z+\text{jets}} \cdot N^{\text{OS}-\text{SS}}_{Z+\text{jets,MC}}$.

In the control region, the yield of Z+jets events is estimated from data using the equation

$$N^{\mathrm{OS-SS,CR}}_{Z+\mathrm{jets,data}} = N^{\mathrm{OS-SS,CR}}_{\mathrm{data}} - \sum_{i \neq Z+\mathrm{jets}} N^{\mathrm{OS-SS,CR}}_{i} \,, \tag{5.1.19}$$

where the sum index i includes $W+c$, W+light and diboson productions. The collective contribution of the multijet and top quark production is of the order of 0.1% and therefore neglected. The yields used in equation 5.1.19, which are listed in table 5.14, are obtained from simulation for the W+light and diboson productions, and in a data-driven way explained below for the $W+c$ production.

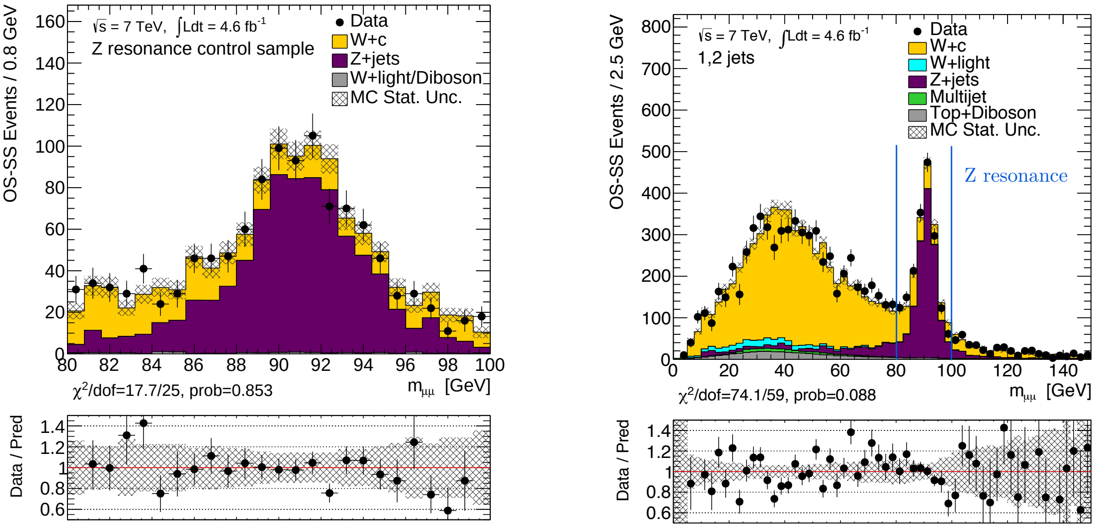
	$N^{\mathrm{OS-SS,CR}}$
Data	1173 (38)
$W+c$	366 ± 21
W+light	8 ± 4
diboson	5.7 ± 1.5
Z+jets - data	790 ± 40
Z+jets - MC	751 ± 18
$k_{Z+\mathrm{jets}}$	1.06 ± 0.06

Table 5.14: Event yields in the Z boson resonance control sample. The Z+jets yield obtained from data and from simulation is used to evaluate the correction scale factor $k_{Z+\mathrm{jets}}$. The statistical uncertainty on the yield observed in data is indicated in parentheses. The uncertainty on the Z+jets MC simulation prediction is statistical. All other uncertainties include the statistical and systematic contributions.

The yield of $W+c$ events in the Z boson control sample is estimated using an iterative data-driven method. Initially, the number of $W+c$ events is determined from simulation and used to compute a preliminary value for the Z+jets scale factor. As done in section 5.2, the latter is used in the signal sample to extract a preliminary yield for the $W+c$ production, which results in being 30% larger than the simulation prediction. Consequently the yield of $W+c$ events in the control sample is increased by 30% and the Z+jets scale factor is recomputed. These two iterations are sufficient for the scale factor value to converge to within 1% and the resulting $W+c$ yield and Z+jets scale factor are shown in table 5.14. The scale factor obtained is $k_{Z+\text{jets}} = 1.06 \pm 0.06$, whose uncertainty includes the statistical and systematic components.

The systematic uncertainties added to the simulation predictions for W+light and diboson productions are 50% (which is compatible with the data-driven estimation in table 5.12) and 25% (section 5.1.4), respectively. The Z+jets yields obtained from data and simulation in the control region are shown with the statistical uncertainty.

The validity of the extrapolation of $k_{Z+\text{jets}}$ from the control region to the signal region is checked by comparing the data and predicted distributions of $m_{\mu\mu}$. Figures 5.7 (a) and (b) show the distributions obtained from the Z boson control sample and from the signal sample extended to include the Z boson resonance region, respectively. The sum of signal and background distributions (predictions) are



(a) (b)

Figure 5.7: Distributions of $m_{\mu\mu}$ in the (a) Z boson control sample and (b) signal sample extended to include the Z boson resonance. The events of the Z boson control sample are vetoed and used to estimate the Z+jets background. OS–SS events selected from the data or simulated samples are used. The lower panels show the ratio of data to the predicted distribution and the result χ^2-test. Predictions are found in good agreement with data.

obtained from simulation for all processes except the multijet production, for which the inverted isolation data sample is used (see section 5.1.1). In figure (a), the prediction is normalised according to table 5.14 using the Z+jets yield measured in data. In figure (b), the predicted distributions for W+c, Z+jets and W+light production are scaled with the yields measured in the signal region (table 5.16). The uncertainties on the data and simulated distributions are statistical and the ratio of the data to prediction is shown in the lower panels. A χ^2-test (footnote 5) is performed and the predicted $m_{\mu\mu}$ distribution is found to be in a good agreement with data in both cases (χ^2/dof=17.7/25 and χ^2/dof=74.1/59 for the control sample and the extended signal samples, respectively).

5.1.4 Other backgrounds

The yields of $t\bar{t}$, single-top and diboson events in the signal sample selected from data are estimated from simulation. The results are shown in table 5.16 and the uncertainty include the statistical and systematic components. The latter are the sum of uncertainties associated with the cross sections (10% and 25% for the top and diboson productions, respectively, from the cross sections in table 4.3) and with the physics objects reconstruction (see section 7.2) in the simulated samples.

The contribution of W+$c\bar{c}$ and W+$b\bar{b}$ events is expected to cancel out with the OS–SS subtraction. The charm (bottom) or anti-charm (anti-bottom) quarks have the same probability to be tagged by the SMT algorithm, hence, the expected yields of OS and SS events are equal. This has been verified in the W+$c\bar{c}$ and W+$b\bar{b}$ simulated samples, whose yields of OS, SS and OS–SS events and statistical uncertainties are shown in table 5.15. The OS–SS yields obtained are compatible with zero (the largest deviation corresponding to 1.7 standard deviations).

5.2 Results for the W+c production yields

The number of OS (SS) events selected from data is 7736 (2775) and 4376 (2479) for the 1-jet and 2-jets samples, respectively. The corresponding number of OS–SS events and the estimated backgrounds are listed in table 5.16. The yields

	N^{OS}	N^{SS}	$N^{\mathrm{OS-SS}}$
1-jet			
W+$b\bar{b}$	233 ± 15	268 ± 15	-35 ± 21
W+$c\bar{c}$	220 ± 14	243 ± 15	-23 ± 21
2-jets			
W+$b\bar{b}$	268 ± 15	238 ± 13	30 ± 20
W+$c\bar{c}$	245 ± 13	256 ± 13	-11 ± 18
1,2-jets			
W+$b\bar{b}$	501 ± 21	506 ± 20	-5 ± 29
W+$c\bar{c}$	465 ± 19	498 ± 20	-33 ± 28

Table 5.15: Yields of OS, SS and OS–SS events for the W+$c\bar{c}$ and W+$b\bar{b}$ productions. The number of events and their statistical uncertainties are evaluated from the simulated samples. The yields of OS–SS events are found to be compatible with zero.

measured for the W+c production, are calculated as $N^{\mathrm{OS-SS}}_{\mathrm{data}} - N^{\mathrm{OS-SS}}_{\mathrm{bkg}}$, where the latter is the total background yield (equation 3.4.2). The statistical uncertainties associated with the observed yields are shown in the table and the uncertainties on the background estimation include the statistical and systematic contributions. The uncertainties on the measured W+c yields are the quadrature sum of those on the data and background yields.

The purity[8] of the signal sample selected from data is 86.9%, 71% and 82.7% for the 1-jet, 2-jets and 1,2-jets samples, respectively. In the 1,2-jets sample, the first two largest backgrounds are the Z+jets and W+light productions, which account for 6.5% and 4.5% of the data events. The size of the multijet, $t\bar{t}$ and single-top backgrounds are similar and each of them amounts to 2% of the data sample. The fraction of data events which originate from diboson production is 1%.

The sources of uncertainty on the measured yield of W+c events are listed

---

[8] The purity of a sample is defined as S/(S+B), where S and B are the number of events of the signal and the background, respectively. In the W+c analysis, the event yield of the signal is $N^{\mathrm{OS-SS}}_{\mathrm{data}} - N^{\mathrm{OS-SS}}_{\mathrm{bkg}}$ and therefore the purity is $(N^{\mathrm{OS-SS}}_{\mathrm{data}} - N^{\mathrm{OS-SS}}_{\mathrm{bkg}})/N^{\mathrm{OS-SS}}_{\mathrm{data}}$.

in table 5.17. The total uncertainty results 3.2%, 7.4% and 3.0% for the 1-jet, 2-jets and 1,2-jets samples, respectively. In the 1,2-jets sample, the data statistical uncertainty (2.3%) is the dominant contribution. The systematic uncertainty arising from the background estimation is 1.9%, whose largest contribution stems from the W+light yield. The evaluation of the number of multijet and Z+jets events gives an uncertainty of 0.7% and 0.6% respectively. Top quark and diboson backgrounds give a small uncertainty (0.1%).

	$N^{\mathrm{OS-SS}}$		
	1-jet	2-jets	1,2-jets
Data	4961 (103)	1897 (83)	6858 (132)
Multijet	71 ± 27	52 ± 20	120 ± 40
W+light	220 ± 80	40 ± 40	250 ± 90
Z+jets	237 ± 22	207 ± 16	445 ± 34
$t\bar{t}$	24 ± 21	129 ± 19	154 ± 21
Singletop	58 ± 18	82 ± 21	140 ± 23
Diboson	37 ± 10	39 ± 13	76 ± 20
Backgrounds	650 ± 90	550 ± 60	1190 ± 110
Measured W+c	4300 ± 140	1350 ± 100	5670 ± 170

Table 5.16: Number of OS–SS events observed in the signal sample selected from data and estimated for the background processes. Data and background yields are subtracted to measure that of the W+c production. The statistical uncertainties associated with the observed yields are shown. The uncertainties on the backgrounds and W+c yields include the statistical and systematic components.

	$\delta(N^{\mathrm{OS-SS}}_{\mathrm{data}} - N^{\mathrm{OS-SS}}_{\mathrm{bkg}})$ [%]		
Uncertainty	1-jet	2-jets	1,2-jets
Statistical	2.4	6.1	2.3
Multijet	0.6	1.5	0.7
W+light	1.8	3.0	1.6
Z+jets	0.5	1.2	0.6
Top & diboson	0.2	0.5	0.1
Backgrounds	2.1	4.2	1.9
Total	3.2	7.4	3.0

Table 5.17: Sources of uncertainty on the yields measured for the W+c production. The statistical uncertainty of the data dominates. In the background estimation, the largest uncertainty arises from the W+light production.

5.3 Optimisation of the $\mathbf{E_{T}^{miss}/m_{T}^{W}}$ selections for the W+c analysis

The author of this thesis performed a study of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W} selection requirements to improve the measurement of the $W+c$ production yields. Specifically, the baseline $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection criteria of the ATLAS standard model analysis group for the selection of W boson events have been optimised specifically for this $W{+}c$ analysis. By modifying these two selection criteria, a drastic suppression of the multijet and Z+jets backgrounds has been achieved.

The $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection requirements adopted in the $W{+}c$ analysis and those of the standard model group's baseline are presented in table 5.18. The $W+c$ production yields have also been measured by selecting the data sample with all the event selection criteria listed in section 4.2 except for those on $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W}, for which the baseline selections were used. Accordingly the complete background estimation has been performed: the multijet, W+light and Z+jets backgrounds have been evaluated with data-driven methods, as done in sections 5.1.1 to 5.1.3, and the top quark and diboson backgrounds have been obtained from simulation.

	Selection requirement	
	$W{+}c$ analysis	Baseline analysis
$E_{\mathrm{T}}^{\mathrm{miss}}$	$>$20 GeV	$>$25 GeV
m_{T}^{W}	$>$60 GeV	$>$40 GeV

Table 5.18: Selection requirements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W} adopted in this $W{+}c$ analysis, compared with the baseline of the ATLAS standard model analysis group for W boson event selection.

The results of the analysis with the baseline $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection criteria are reported in table 5.19 for the 1,2-jets sample. These are to be compared with the results obtained with the $W+c$ selection criteria in tables 5.16 and 5.17. The data yield is shown with the associated statistical uncertainty; the uncertainties on the backgrounds and the $W+c$ production include statistical and systematic contributions. The rightmost column shows the uncertainties on the measured $W+c$ yield.

	$N^{\mathrm{OS-SS}}$ 1,2-jets	Uncertainty	$\delta(N^{\mathrm{OS-SS}}_{\mathrm{data}} - N^{\mathrm{OS-SS}}_{\mathrm{bkg}})$ [%] 1,2-jets
Data	8050 (146)	Statistical	2.5
Multijet	590 ± 170	Multijet	2.9
W+light	210 ± 90	W+light	1.4
Z+jets	940 ± 70	Z+jets	1.3
Top & diboson	420 ± 40	Top & diboson	0.6
Backgrounds	2160 ± 210	Backgrounds	3.6
Measured $W+c$	5890 ± 260	Total	4.3

Table 5.19: Results of the analysis with the baseline $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection criteria: (left) yield of OS–SS events observed in data and estimated for the background processes; (right) uncertainties on the yields measured for the $W+c$ production. These results are to be compared with those obtained with the $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selections optimised for the W+c analysis (tables 5.16 and 5.17). The yield measured for the W+c production is calculated by subtracting the background contribution from data. The statistical uncertainties associated with the data yield are shown in parentheses and the uncertainties on the backgrounds and $W+c$ yields include the statistical and systematic components.

With the baseline selections, the total uncertainty on the yield measured for the $W+c$ production is found to be 4.3%. The total uncertainty achieved in the analysis with the $W+c$ selection criteria is 3.0%. The data sample selected with the baseline selection criteria has purity 73.1%, which is lower than that obtained with the $W+c$ selection criteria (82.7%). This is due to the larger Z+jets and multijet backgrounds, which are responsible for 11.7% and 7% of the events selected from data, respectively. The W+c selection criteria reduce of a factor 5 the multijet background and halve the Z+jets background. In analysis with the baseline selection criteria the backgrounds' systematic uncertainty (3.6%) is found to be larger than the data statistical uncertainty (2.5%). The backgrounds's systematic uncertainty is driven by that assigned to multijet yield estimation, 2.9%. The

In the analysis with the $W+c$ selection requirements, the systematic uncertainty of the background estimation is found to be smaller the statistical uncertainty of the data.

The improvement in the background rejection achieved by changing the $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection criteria can also be seen in figure 5.8. This shows the m_{T}^{W} distribution obtained from OS–SS events of the 1,2-jets sample selected in data with baseline criteria and compared to the prediction (sum of the signal and all backgrounds). The distributions are obtained from simulation for all processess except the multijet production, for which the inverted isolation sample is used (section 5.1.1). The distributions of multijet, W+light and Z+jets events are normalised to the data-driven yields in table 5.19. The $W+c$ distribution is scaled to the measured yield and therefore the prediction is normalised to data. The uncertainties on data and simulated distributions are statistical. The ratio of data to the prediction is displayed in the lower panel.

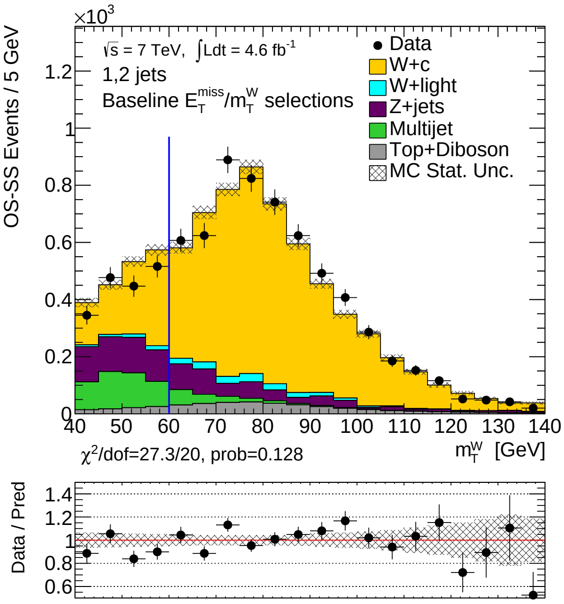


Figure 5.8: Distribution of m_{T}^{W} in data events selected with the baseline $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ selection criteria. OS–SS events of the 1,2-jets sample are used. The prediction obtained by summing signal and backgrounds is normalised to data. The lower panel shows the data-to-prediction ratio and the result of the χ^2-test. The prediction is found to be in good agreement with the data.

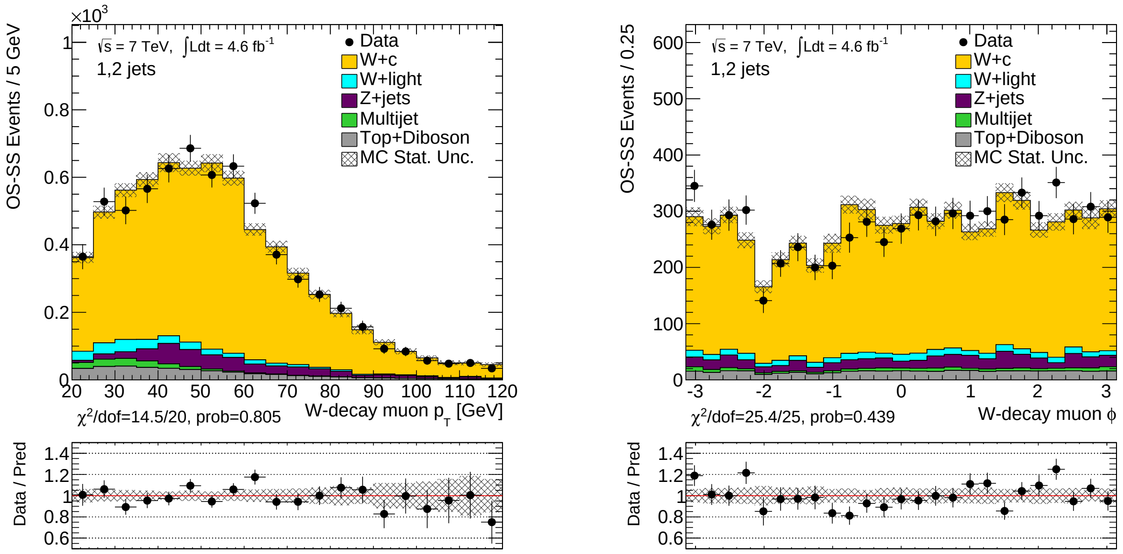
The majority of multijet (green) and Z+jets (purple) events are reconstructed with low values of m_{T}^{W} and the selection $m_{\mathrm{T}}^{W} > 60$ GeV (blue line) is adopted to remove them. In addition, the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement of the $W+c$ selection criteria is relaxed from > 25 GeV to > 20 GeV, so as to not overly reduce the size of the data sample. As a check, a χ^2-test (footnote 5) is performed and the shape of the predicted and data distributions are found to be in good agreement (χ^2/dof=27.3/20). This indicates that the $W+c$ simulation is representative of the data for the modelling of $W \rightarrow \mu\nu$ decay. For this reason, in the unfold of the measured W+c yield to extract the cross section (see section 7.1) no uncertainty is added due to the harsher m_{T}^{W} selection requirement. Moreover, the $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$ requirements adopted in the $W+c$ analysis are found to not increase the cross section systematic uncertainties originating from the unfolding (section 7.2).

5.4 Kinematical distributions of the signal sample

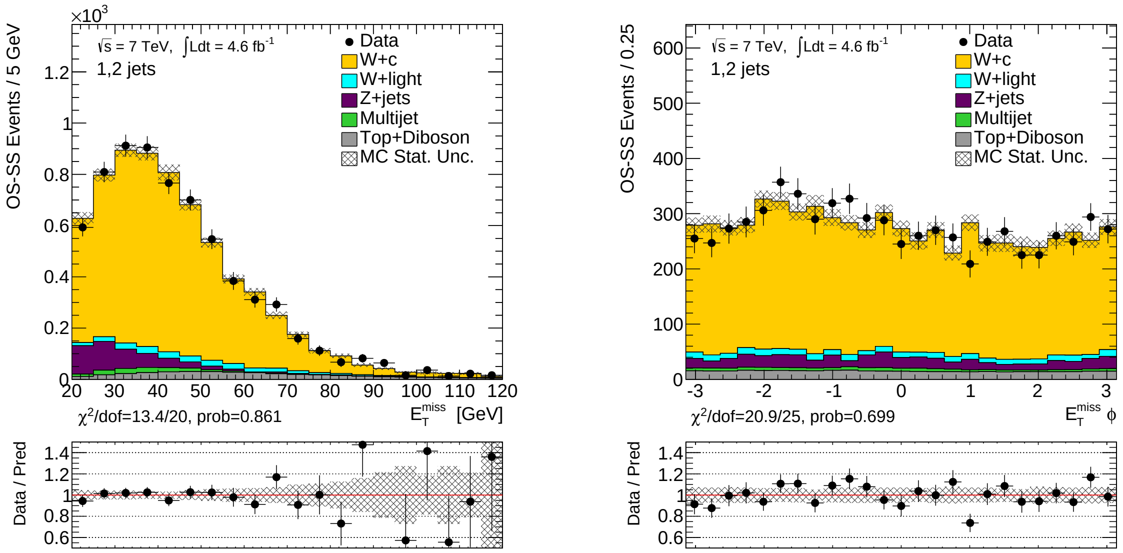
The selected data, signal modelling and background estimation are checked with the distributions of several kinematical variables in figures 5.9 to 5.12. Data distributions are compared with the predictions evaluated from the sum of signal and backgrounds (multijet, W+light, Z+jets, top quark and diboson production), for OS–SS events with one or two jets.

Distributions are determined from simulation for all processes except the multijet production, for which data events of the inverted isolation sample are used (section 5.1.1). The number of multijet, W+light and Z+jets events are normalised to the data-driven estimations in table 5.16. The $W+c$ distributions are normalised to the measured yield and therefore the predictions are normalised to data. The uncertainties assigned to the distributions of the data and simulated events are statistical. The ratio of the predicted distribution to data is shown in the lower panels.

A χ^2-test (footnote 5) is performed to compare the predicted and observed shapes and good agreement is found (expect for the single case below and the χ^2/dof results are shown in the lower panels). This indicates that the signal simulation is representative of the data and provides a validation of the background estimation. The data distribution of the soft muon's χ^2_{match}, figure 5.12 (d), is found not to be well described by simulation. This quantity is used in soft muon identification requirements. However, as described in section 4.2.3, the simulations are corrected to match the SMT tagging efficiency measured in data. Therefore, no uncertainty is added to the cross section measurement due to this mismodelling.

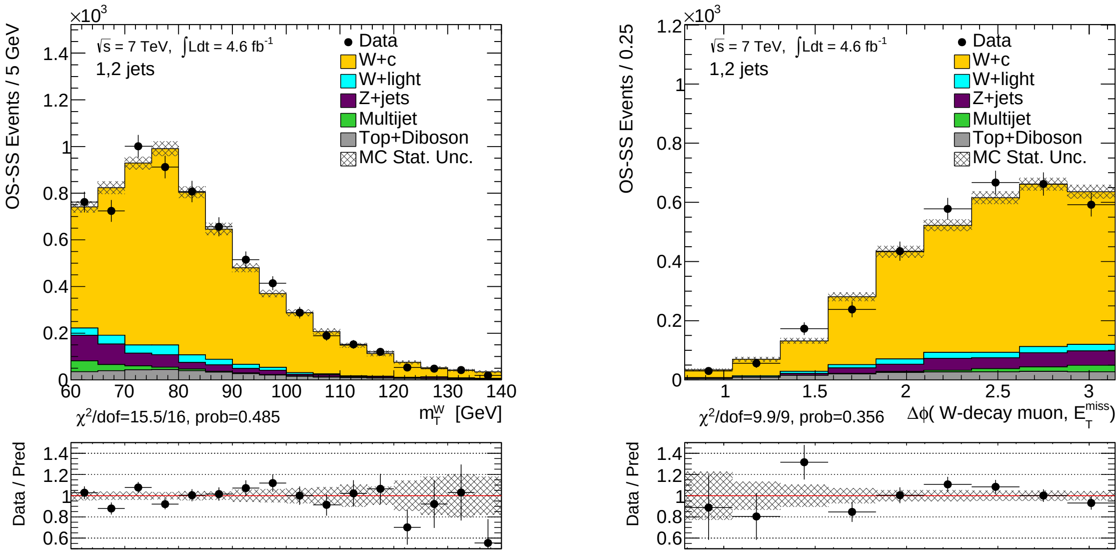


(a) W-decay muon p_T (b) W-decay muon ϕ

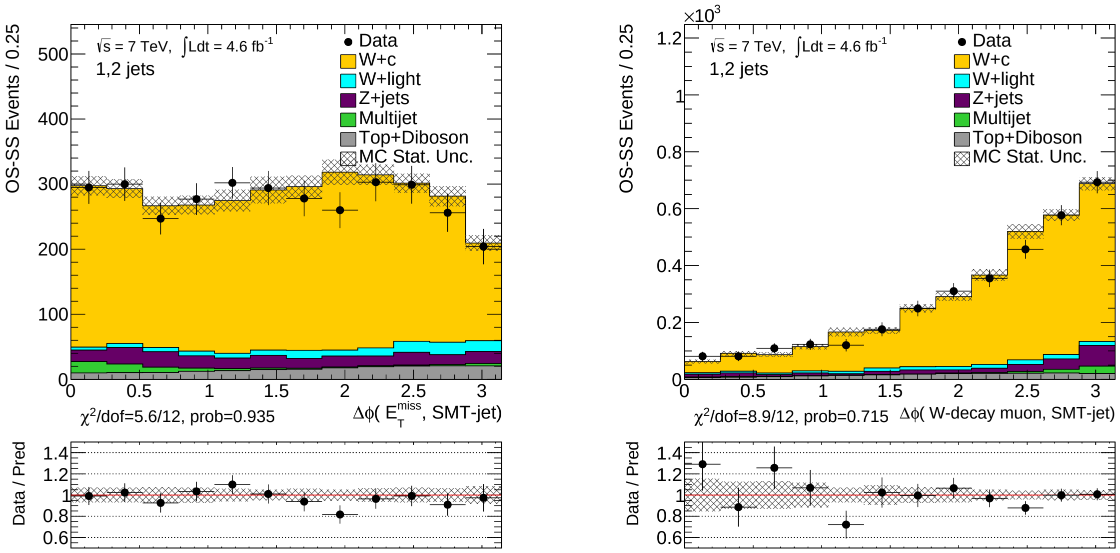


(c) missing energy $E_\mathrm{T}^\mathrm{miss}$ (d) missing energy ϕ

Figure 5.9: Distributions of p_T and ϕ for the W-decay muon and for the missing energy. OS–SS events selected from data with one or two jets are used. The predictions are obtained by summing signal and backgrounds and are normalised to data. The lower panels show the data-to-prediction ratio and the result of the χ^2-test.

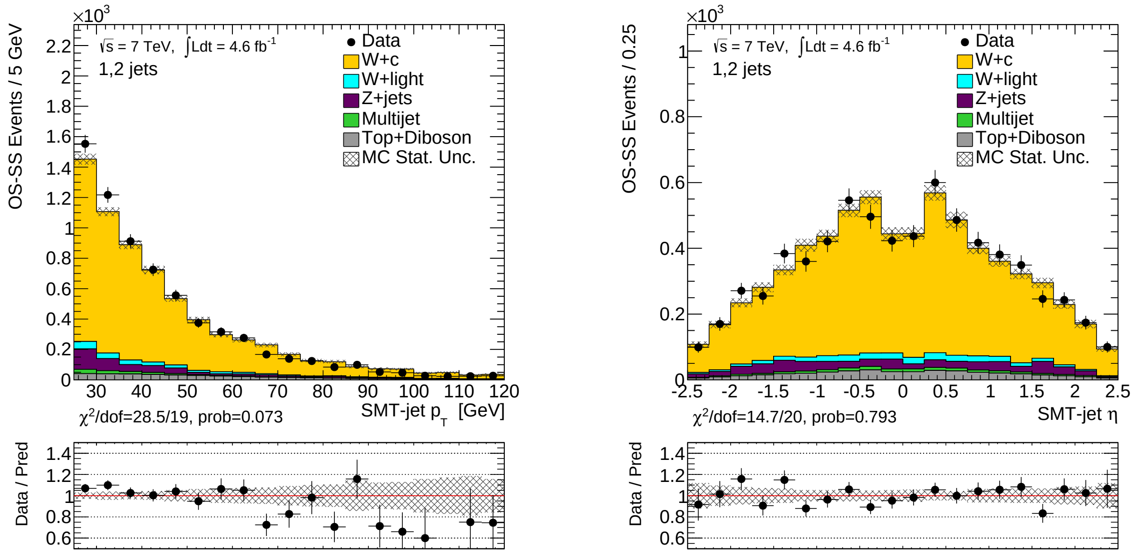


(a) m_{T}^{W} (b) $\Delta\phi$(W-decay muon, $E_{\mathrm{T}}^{\mathrm{miss}}$)

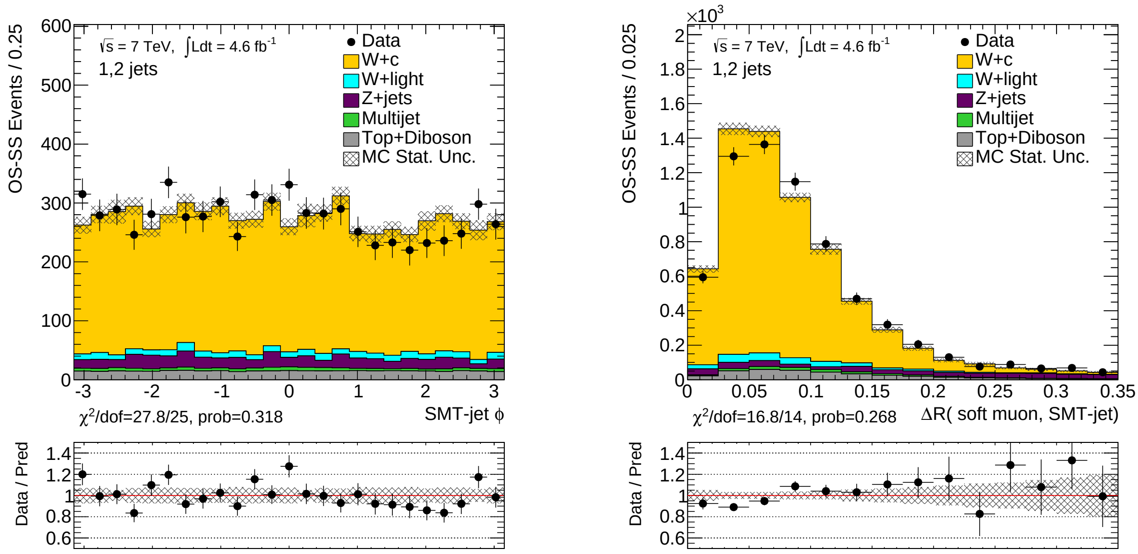


(c) $\Delta\phi$($E_{\mathrm{T}}^{\mathrm{miss}}$, SMT-jet) (d) $\Delta\phi$(W-decay muon, SMT-jet)

Figure 5.10: Distributions of m_{T}^{W} and of $\Delta\phi$ between the W-decay muon, missing energy and SMT-jet. OS–SS events selected from data with one or two jets are used. The prediction are obtained by summing signal and backgrounds an are is normalised to data. The lower panels show the data-to-prediction ratio and the result of the χ^2-test.

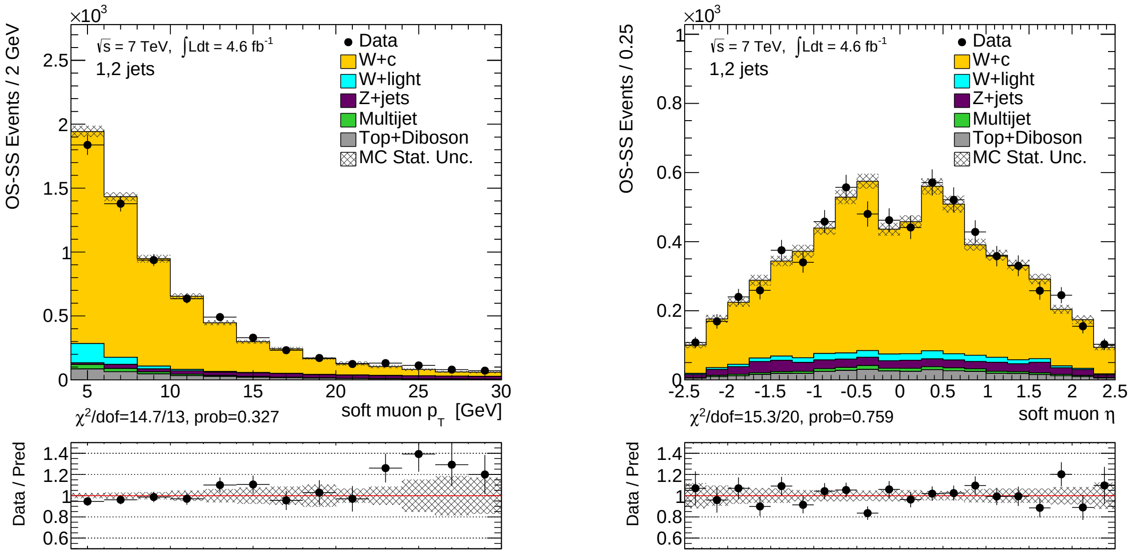


(a) SMT-jet p_{T} (b) SMT-jet η



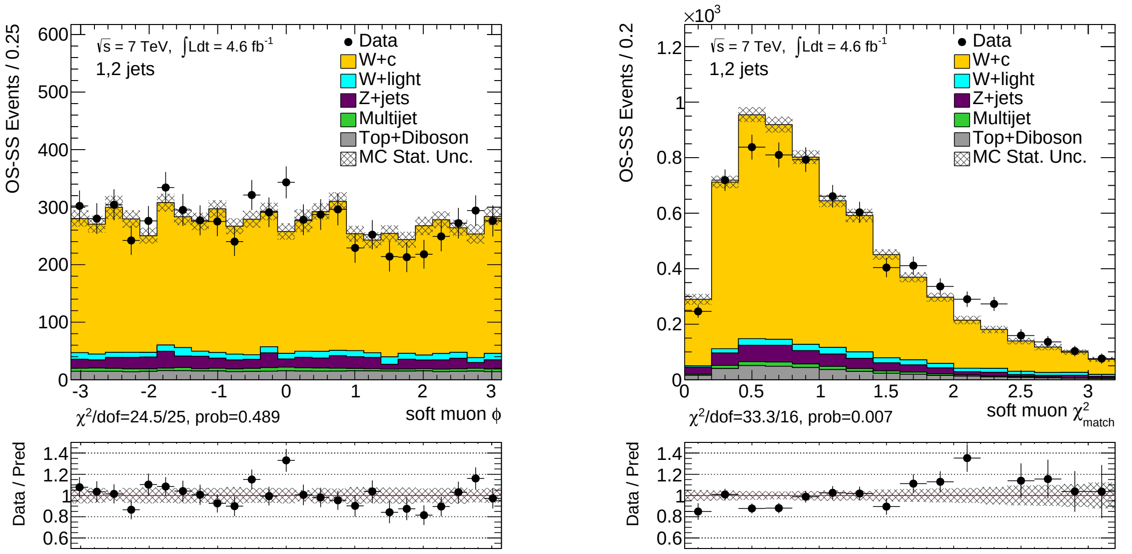
(c) SMT-jet ϕ (d) ΔR(soft muon, SMT-jet)

Figure 5.11: Distributions of p_{T}, η, and ϕ for the SMT-jet and of ΔR between the soft muon and SMT-jet. OS–SS events selected from data with one or two jets are used. The prediction are obtained by summing signal and backgrounds and are normalised to data. The lower panels show the data-to-prediction ratio and the result of the χ^2-test.



(a) soft muon p_{T}

(b) soft muon η



(c) soft muon ϕ

(d) soft muon χ^2_{match}

Figure 5.12: Distributions of p_{T}, η, ϕ and χ^2_{match} of the soft muon. OS–SS events selected from data with one or two jets are used. The prediction obtained by summing signal and backgrounds is normalised to data. The lower panels show the data-to-prediction ratio and the result of the χ^2-test.

Chapter 6

Determination of the charge-divided W+c yields as a function of pseudorapidity

This chapter presents the measurement of the $W+c$ event yields as a function of the W boson's charge and of the W-muon's pseudorapidity. The same data sample selected to extract the inclusive yield of $W+c$ events is used (chapter 5).

The yields of signal events are measured from the data after estimating the background contributions, which are evaluated with data-driven methods for the most important processes. The analysis described in this chapter is entirely the own work of the author of this thesis and the results are published in reference [1].

6.1 Determination of the $\mathbf{W^{+}+\bar{c}}$ and $\mathbf{W^{-}+c}$ production yields

The data events selected with the criteria described in section 4.2 are referred to as *inclusive sample*. The *charge-divided samples* are subsequently obtained from it by separating the events by the W-decay muon's charge. The yields of $W^{+}+\bar{c}$ and $W^{-}+c$ events are measured as $N_{\mathrm{data}}^{\mathrm{OS-SS},\pm} - N_{\mathrm{bkg}}^{\mathrm{OS-SS},\pm}$, where the positive and negative signs refer to the W boson charge. The positive-to-negative *W-charge* *ratio* of the $W+c$ event yield is measured in data as

$$R^{\pm,\,\mathrm{yield}}_{W+c} = \frac{N^{\mathrm{OS-SS},+}_{\mathrm{data}} - N^{\mathrm{OS-SS},+}_{\mathrm{bkg}}}{N^{\mathrm{OS-SS},-}_{\mathrm{data}} - N^{\mathrm{OS-SS},-}_{\mathrm{bkg}}} \,. \tag{6.1.1}$$

The background contribution to the charge-divided data samples, $N^{\mathrm{OS-SS},+}_{\mathrm{bkg}}$ and $N^{\mathrm{OS-SS},-}_{\mathrm{bkg}}$, is evaluated starting from the background estimation in the inclusive sample (chapter 5), as described below. This approach is adopted, instead of independently estimating $N^{OS-SS,+}_{\mathrm{bkg}}$ and $N^{OS-SS,-}_{\mathrm{bkg}}$ in the charged-divided samples, for two reasons: (i) the number of events of the charge-divided samples is too small to use the data-driven methods applied in the inclusive sample; (ii) this method offers a simpler way to evaluate the correlation between $N^{OS-SS,+}_{\mathrm{bkg}}$ and $N^{OS-SS,-}_{\mathrm{bkg}}$ is simpler.

The contribution of each background to the charge-divided samples can be expressed as

$$\begin{aligned} N^{\mathrm{OS-SS},+} &= \frac{R^{\pm}}{1+R^{\pm}} \cdot N^{\mathrm{OS-SS}} \\ N^{\mathrm{OS-SS},-} &= \frac{1}{1+R^{\pm}} \cdot N^{\mathrm{OS-SS}} \,, \end{aligned} \tag{6.1.2}$$

where $R^{\pm}$ and $N^{\mathrm{OS-SS}}$ are the positive-to-negative charge ratio and the yield in the inclusive sample, respectively. Hence, the total background contributions are evaluated from the relations

$$\begin{aligned} N^{\mathrm{OS-SS},+}_{\mathrm{bkg}} &= \sum_i \frac{R^{\pm}_i}{1+R^{\pm}_i} \cdot N^{\mathrm{OS-SS}}_i \,, \\ N^{\mathrm{OS-SS},-}_{\mathrm{bkg}} &= \sum_i \frac{1}{1+R^{\pm}_i} \cdot N^{\mathrm{OS-SS}}_i \,, \end{aligned} \tag{6.1.3}$$

where the sum index i includes the multijet, W+light, Z+jets, top quark and diboson productions. The yields $N^{\mathrm{OS-SS}}_i$ are estimated as described in chapter 5 and listed in table 5.16.

6.1.1 Estimation of charge ratio of the backgrounds

The charge ratio of the candidate W-decay muon in the background processes is evaluated using data-driven methods for the multijet and W+light productions and from simulation for the other backgrounds, as described below.

Multijet background

The data-driven method similar to that explained in section 5.1.1 for the OS/SS asymmetry is used to estimate the $R^{\pm}_{\text{multijet}}$. Two samples with high-purity of multijet events are selected from data, by "inverting" the selection requirements on the W-decay muon's isolation variables and on m_{T}^{W}, respectively. In both control samples, the events are divided according to the charge of the candidate W-decay muon and the small contribution of W+jets and Z+jets events evaluated from simulation is subtracted. The ratio of the corrected control data yields is taken as $R^{\pm}_{\text{multijet}}$.

The results obtained from the inverted isolation and inverted m_{T}^{W} control samples are shown in table 6.1.

	$R^{\pm}_{\text{multijet}}$		
Control sample	1-jet	2-jets	1,2-jets
Inverted isolation	1.12 ± 0.08	1.13 ± 0.09	1.13 ± 0.06
Inverted m_{T}^{W}	1.02 ± 0.08	0.98 ± 0.09	1.01 ± 0.06
Average	1.07 ± 0.08	1.06 ± 0.09	1.07 ± 0.06

Table 6.1: Reconstructed W-charge ratio in the multijet events of the inverted isolation and inverted m_{T}^{W} control data samples. The average value is taken as the best estimation of $R^{\pm}_{\text{multijet}}$. The uncertainties include the statistical and systematic components.

The uncertainties include the statistical and systematic contributions. The latter is evaluated by assigning uncorrelated uncertainties of 25% to the W+jets and Z+jets contributions. The average of the two control samples is taken as the best estimation of $R^{\pm}_{\text{multijet}}$ and is assigned the same uncertainty as that of the single measurement. To reduce the statistical uncertainties in the estimation of $R^{\pm}_{\text{multijet}}$, the results are obtained with OS+SS events. The same charge ratio is expected to be reconstructed in OS and SS events of the multijet production. This is checked by calculating $R^{\pm}_{\text{multijet}}$ in OS, SS and OS–SS events; the results are found to be compatible (in the 1,2-jets sample, $R^{\pm}_{\text{multijet}}$ is 1.04 ± 0.07, 1.10 ± 0.09, and 0.95 ± 0.18 for OS, SS and OS–SS events, respectively).

W+light background

A data-driven method similar to that explained in section 5.1.2 for the OS/SS tracks asymmetry is used to estimate $R^{\pm}_{W+\text{light}}$. The W-charge ratio is first evaluated in pretag events selected from data and then extrapolated to the signal sample according to the equation

$$R^{\pm}_{W+\text{light}} = k_{R^{\pm}} \cdot R^{\pm,\,\text{pretag}}_{W+\text{light,data}}, \tag{6.1.4}$$

in which the correction factor $k_{R^{\pm}}$ is obtained from simulation.

The W-charge ratio of pretag events in data can be expressed as $R^{\pm,\,\text{pretag}}_{\text{data}} = \sum_i F_i \cdot R^{\pm,\,\text{pretag}}_i$, where the summation includes the W+light, W+heavy, Z+jets, multijet, top quark and diboson events. For each of these processes, F_i and $R^{\pm,\,\text{pretag}}_i$ are the fraction of pretag events in data and the pretag W-charge ratio, respectively. Consequently, the W+light production is calculated as:

$$R^{\pm,\,\text{pretag}}_{W+\text{light,data}} = \frac{1}{F_{W+\text{light}}} \left(R^{\pm,\,\text{pretag}}_{\text{data}} - \sum_{i \neq W+\text{light}} F_i \cdot R^{\pm,\,\text{pretag}}_i \right). \tag{6.1.5}$$

The event fractions and pretag charge ratios used in equation 6.1.5 are shown in tables 5.10 and 6.2, respectively. For the data, the ratio is measured in the pretag charge-divided samples. The ratios $R^{\pm,\,\text{pretag}}_i$ are evaluated from simulation for all processes except the multijet production, for which the matrix method described in section 5.1.1 is used. The uncertainties in table 5.10 are statistical except for those on $R^{\pm,\,\text{pretag}}_{W+\text{heavy}}$, which include the systematic contribution, evaluated as done in section 5.1.2 for $A^{\text{tracks}}_{W+\text{heavy}}$.

The results of the W-charge ratio for the W+light production are shown in table 6.3. The pretag ratio measured in data is 1.50±0.15 for the 1,2-jets sample and the uncertainties includes the statistical and systematic components. The extrapolation factor is evaluated as

$$k_{R^{\pm}} = \frac{R^{\pm}_{W+\text{light,MC}}}{R^{\pm,\,\text{pretag}}_{W+\text{light,MC}}}, \tag{6.1.6}$$

where $R^{\pm}_{W+\text{light,MC}}$ ($R^{\pm,\,\text{pretag}}_{W+\text{light,MC}}$) is obtained from simulated events of the signal (pretag) region. The factor $k_{R^{\pm}}$ is evaluated in OS+SS events and the uncertainties are statistical. The OS+SS sample is used to reduce the large uncertainties

	$R^{\pm,\text{ pretag}}$		
	1-jet	2-jets	1,2-jets
Data	1.4129 ± 0.0021	1.423 ± 0.005	1.4147 ± 0.0019
W+heavy	1.09 ± 0.06	1.18 ± 0.11	1.11 ± 0.07
Multijet	1.146 ± 0.005	1.130 ± 0.005	1.143 ± 0.006
Z+jets	1.124 ± 0.007	1.098 ± 0.009	1.119 ± 0.006
Singletop	1.688 ± 0.012	1.509 ± 0.011	1.587 ± 0.008
$t\bar{t}$	0.992 ± 0.010	1.014 ± 0.005	1.009 ± 0.005
Diboson	1.074 ± 0.005	1.139 ± 0.006	1.100 ± 0.004

Table 6.2: Pretag W-charge ratios used in equation 6.1.5 to evaluate $R^{\pm,\text{ pretag}}_{W+\text{light}}$. The uncertainties are statistical except for those on the W+heavy production, which include the systematic component.

	1-jet	2-jets	1,2-jets
$R^{\pm,\text{ pretag}}_{W+\text{light,data}}$	1.49 ± 0.15	1.54 ± 0.16	1.50 ± 0.15
$R^{\pm}_{W+\text{light,MC}}$	1.54 ± 0.10	1.82 ± 0.12	1.63 ± 0.08
$R^{\pm,\text{ pretag}}_{W+\text{light,MC}}$	1.529 ± 0.004	1.607 ± 0.007	1.541 ± 0.003
$k_{R^{\pm}}$	1.01 ± 0.06	1.13 ± 0.07	1.06 ± 0.05
$R^{\pm}_{W+\text{light}}$	1.50 ± 0.18	1.74 ± 0.21	1.58 ± 0.18

Table 6.3: Results of the W-charge ratio for the W+light production evaluated from equation 6.1.5. The ratio $R^{\pm,\text{ pretag}}_{W+\text{light,data}}$ is measured in pretag data events and includes the statistical and systematic uncertainties. This is extrapolated the to signal sample by means of the multiplicative factor $k_{R^{\pm}}$. The latter is obtained dividing $R^{\pm}_{W+\text{light,MC}}$ for $R^{\pm,\text{ pretag}}_{W+\text{light,MC}}$ and its uncertainties are statistical.

which arise from the small number of simulated W+light events passing the requirement of one SMT-jet. The extrapolation factor calculated with OS, SS and OS–SS events is checked for compatibility with that in the table (in the 1,2-jets sample, $k_{R^{\pm}}$ is 1.57±0.09, 1.70±0.12 and 1.0±0.6 for OS, SS and OS–SS events, respectively). Finally, $R^{\pm}_{W+\text{light}}$ is estimated to be 1.58±0.18 in the 1,2-jets sample, whose uncertainty dominated by the estimation of the charge ratio in the pretag data events.

Owing to the larger density of u-quarks in pp collisions, the W-charge ratio in the W+light events is expected to be larger than 1. For example, the processes $u+g \to d+W^{+}$ and $u+\overline{d} \to g+W^{+}$ give a larger contribution to the W+light production than the processes $d+g \to u+W^{-}$ and $d+\overline{u} \to g+W^{-}$, respectively.

Z+jets background

The ratio $R^{\pm}_{Z+\text{jets}}$ is evaluated from simulation and checked in the Z resonance control data sample. The results obtained from simulation are shown in table 6.4 for the 1-jet, 2-jets and 1,2-jets samples and the uncertainties are statistical. To improve the precision, OS+SS events are used for $R^{\pm}_{Z+\text{jets}}$. The same W-charge ratio is expected to be reconstructed in the OS and SS events of the Z+jets production. This is checked by calculating the charge ratio also with OS, SS and OS–SS events and the results are found to be compatible with those in the table (in the 1,2-jets sample, $R^{\pm}_{Z+\text{jets}}$ is 1.03±0.06, 1.12±0.11 and 0.99±0.09 for OS, SS and OS–SS events, respectively).

As a check of the simulation results, $R^{\pm}_{Z+\text{jets}}$ is evaluated also in the Z-resonance control sample described in section 5.1.3. The contributions of non-Z+jets events ($W+c$, W+light and diboson production) are subtracted from the yield of data events reconstructed with positive and negative W-decay muons. The charge ratio of the control sample, whose uncertainty includes the statistical and systematic components, is shown in table 6.4 and found to be compatible with that of the signal region.

$R^{\pm}_{Z+\text{jets}}$		
1-jet	2-jets	1,2-jets
1.09 ± 0.08	0.99 ± 0.06	1.05 ± 0.05
Z-resonance control sample		
1.13 ± 0.12		

Table 6.4: Charge ratio of the reconstructed W-decay in Z+jets events. The values evaluated for the 1-jet, 2-jets and 1,2-jets samples are obtained from simulation. As a check, the ratio is also estimated in the Z-resonance control region.

Top quark and diboson backgrounds

The ratios $R^{\pm}_{t\bar{t}}$, $R^{\pm}_{\text{single-top}}$ and $R^{\pm}_{\text{diboson}}$ are evaluated from simulation and the results are shown in table 6.5 with the statistical uncertainties. The ratios for the top quark production are computed using OS+SS events to reduce the uncertainties. The compatibility of the ratios calculated in the OS, SS and OS–SS events is checked (in the 1,2-jets sample: $R^{\pm}_{t\bar{t}}$ is 0.998±0.017, 0.97±0.10 and 0.996±0.020 for OS, SS and OS–SS events, respectively, and similarly $R^{\pm}_{\text{single-top}}$ is 1.61±0.04, 1.70±0.05 and 1.42±0.14). In the case of the diboson production, the OS and the SS samples have different charge ratios due to different fractions of WW and WZ events present in the two samples. For this reason, $R^{\pm}_{\text{diboson}}$ is evaluated using OS–SS events ($R^{\pm}_{\text{diboson}}$ is found to be 1.15±0.04 and 1.35±0.08 in OS and SS events, respectively).

	1-jet	2-jets	1,2-jets
$t\bar{t}$	0.958 ± 0.033	1.010 ± 0.014	0.996 ± 0.020
Single-top	1.79 ± 0.05	1.585 ± 0.034	1.648 ± 0.028
Diboson	1.02 ± 0.08	1.11 ± 0.09	1.07 ± 0.06

Table 6.5: Charge ratio of the W-decay reconstructed in $t\bar{t}$, single-top and diboson events. The values are obtained from simulation and the uncertainties are statistical.

6.1.2 Results

The event yields measured for the W^{+}+$\bar{c}$ and W^{-}+c productions and their ratios are presented in table 6.6. The uncertainties associated with the selected data samples are shown in the table. The background yields are calculated from equation 6.1.3 and the uncertainties include the statistical and systematic uncertainties. The yields of W^{+}+$\bar{c}$ and W^{-}+c events are measured from data after subtracting the contributions from backgrounds and the W-charge ratio measured for W^{+}+$\bar{c}$ events is shown in the last row. The W-charge ratio measured for the $W+c$ production is smaller than 1 as expected (see also section 7.3). The data statistical uncertainties are larger than those arising from the background estimation.

	1-jet	2-jets	1,2-jets
$N_{\text{data}}^{\text{OS}-\text{SS},+}$	2488 (76)	886 (62)	3374 (97)
$N_{\text{bkg}}^{\text{OS}-\text{SS},+}$	350 ± 50	291 ± 34	650 ± 70
Measured $N_{W^{+}+\bar{c}}^{\text{OS}-\text{SS}}$	2130 ± 90	600 ± 70	2720 ± 120
$N_{\text{data}}^{\text{OS}-\text{SS},-}$	2473 (69)	1011 (55)	3484 (89)
$N_{\text{bkg}}^{\text{OS}-\text{SS},-}$	290 ± 40	255 ± 24	540 ± 50
Measured $N_{W^{-}+c}^{\text{OS}-\text{SS}}$	2190 ± 80	760 ± 60	2950 ± 100
$R_{W+c}^{\pm,\,\text{yield}}$	0.97 ± 0.05	0.79 ± 0.10	0.92 ± 0.04

Table 6.6: Event yields of the charge-divided samples and W-charge ratio of $W+c$ production yields. The number of events selected in data are shown with the associated statistical uncertainties in parentheses. The uncertainties on the backgrounds, the measured charge divided $W+c$ yields and their ratio include the statistical and systematic contributions.

The breakdown of the uncertainties associated with the background estimation is shown in table 6.7 for the 1-jet and the 2-jets samples and in table 6.8 for the 1,2-jets sample. The uncertainties on the yields of OS–SS events in the inclusive sample, which are estimated as detailed in chapter 5, are larger than the uncertainties on the charge ratios. The dominating source of uncertainty in the background estimation is the estimation of the W+light OS–SS yield.

1-jet Uncertainty		$\delta N^{\mathrm{OS-SS}}_{W^+{+}\bar{c}}$ [%]	$\delta N^{\mathrm{OS-SS}}_{W^-{+}c}$ [%]	$\delta R^{\pm,\,\mathrm{yield}}_{W{+}c}$ [%]
Statistical		3.6	3.2	4.8
Multijet	OS–SS yield	0.7	0.6	0.1
	charge ratio	0.1	0.1	0.1
W+light	OS–SS yield	2.2	1.4	0.8
	charge ratio	0.3	0.3	0.6
Z+jets	OS–SS yield	0.5	0.5	<0.1
	charge ratio	0.2	0.2	0.4
Top & diboson	OS–SS yield	0.8	0.6	0.2
	charge ratio	<0.1	<0.1	0.1
Backgrounds		2.5	1.6	1.0
Total		4.4	3.6	4.9

2-jets Uncertainty		$\delta N^{\mathrm{OS-SS}}_{W^+{+}\bar{c}}$ [%]	$\delta N^{\mathrm{OS-SS}}_{W^-{+}c}$ [%]	$\delta R^{\pm,\,\mathrm{yield}}_{W{+}c}$ [%]
Statistical		10.4	7.3	12.7
Multijet	OS–SS yield	1.7	1.3	0.4
	charge ratio	0.2	0.2	0.4
W+light	OS–SS yield	4.3	1.9	2.3
	charge ratio	0.2	0.2	0.4
Z+jets	OS–SS yield	1.4	1.0	0.3
	charge ratio	0.5	0.4	0.9
Top & diboson	OS–SS yield	2.9	1.8	1.2
	charge ratio	0.1	0.1	0.2
Backgrounds		5.8	3.0	2.8
Total		11.9	7.9	13.0

Table 6.7: Sources of uncertainty on $W^+{+}\bar{c}$ and $W^-{+}c$ event yields and their ratio for the 1-jet and 2-jets samples. The OS–SS yields and the background yields in the inclusive (i.e. not charge-divided) sample. The charge ratio is the ratio of events reconstructed with a W^+ and a W^- boson.

Uncertainty		1,2-jets $\delta N^{\text{OS}-\text{SS}}_{W^{+}+\bar{c}}$ [%]	$\delta N^{\text{OS}-\text{SS}}_{W^{-}+c}$ [%]	$\delta R^{\pm,\,\text{yield}}_{W+c}$ [%]
Statistical		3.6	3.0	4.7
Multijet	OS–SS yield	0.7	0.6	0.1
	charge ratio	0.1	0.1	0.2
W+light	OS–SS yield	2.0	1.2	0.8
	charge ratio	0.3	0.2	0.5
Z+jets	OS–SS yield	0.6	0.6	0.1
	charge ratio	0.2	0.2	0.4
Top & diboson	OS–SS yield	0.7	0.6	0.2
	Charge	0.1	0.1	0.1
Backgrounds		2.4	1.6	1.0
Total		4.3	3.4	4.8

Table 6.8: Sources of uncertainty on $W^{+}+\bar{c}$ and $W^{-}+c$ event yields and their ratio for the 1-jet and 2-jets samples. The OS–SS yields and the background yields in the inclusive (i.e. not charge-divided) sample. The charge ratio is the ratio of events reconstructed with a W^{+} and a W^{-} boson.

6.2 Determination of the W+c production yield as a function of $|\eta^{\mu}|$

The yields of the $W^{+}+\bar{c}$ and $W^{-}+c$ events are measured separately in 11 bins[1] of the absolute value of the W-decay muon pseudorapidity, $|\eta^{\mu}|$. The distributions for the background processes are evaluated from simulation in all cases except the multijet production, for which these are derived from data.

The $|\eta^{\mu}|$ distribution of the multijet production is taken as the average of those obtained from the inverted isolation and inverted m_{T}^{W} control data samples. The distributions of the control data samples are corrected for the small contributions of W+jets and Z+jets events estimated from simulation. Figure 6.1 (a) shows the shapes obtained from the control region and their average for the events containing one or two jets. The uncertainty on the average distribution, which is displayed by the error band, is taken as half the difference of the two measurements. The distributions central "jump" is an effect of the size of the bin $|\eta^{\mu}| \in [1.05, 1.37]$, which is larger than the other bins. To reduce the uncertainties, these distributions are derived in the OS+SS samples. For the multijet production, OS and SS events are expected to be reconstructed with same $|\eta^{\mu}|$ shape. The shapes in OS and SS events have been verified as being compatible.

The $|\eta^{\mu}|$ distributions for non-multijet backgrounds are evaluated from simulation. For the W+light, single-top and $t\bar{t}$ productions, OS+SS events are used to reduce the statistical uncertainties. The normalised distributions derived for three processess in OS and SS events are shown in figures 6.1 (b) (c) (d). A χ^{2}-test (section 5.1.3) is performed and the OS shape is found to be compatible with the SS shape (χ^{2}/dof is 5.2/11, 13.5/11 and 8.6/11 for W+light, single-top and $t\bar{t}$ production, respectively). The $|\eta^{\mu}|$ shape of diboson and Z+jets productions are evaluated using OS–SS events. In both cases, the OS (black circles) and SS (red squares) samples are expected to be composed of events with different η^{μ} shapes.

The $|\eta^{\mu}|$ distributions of the data are shown in figure 6.2 for OS–SS events

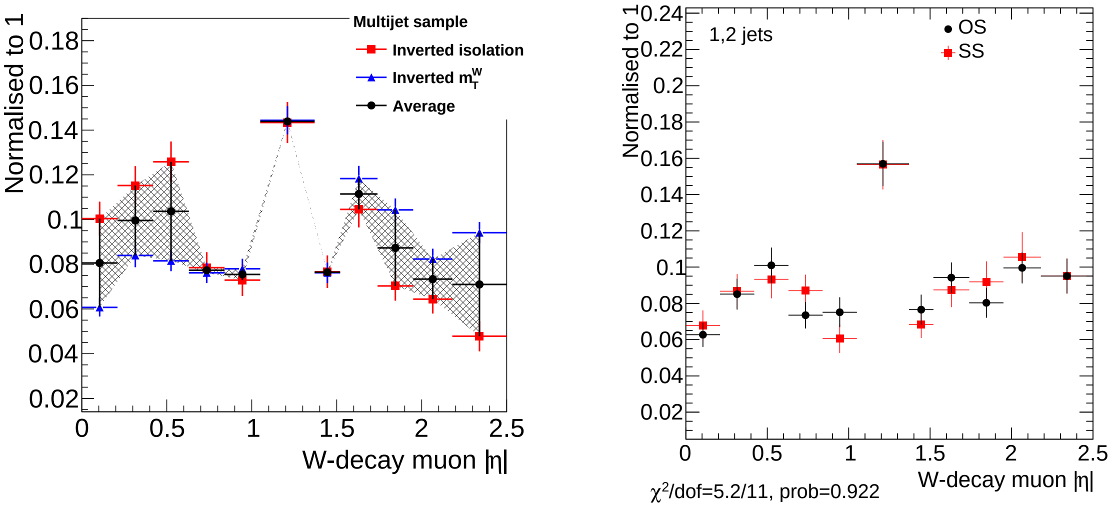
---

[1] The bins of $|\eta^{\mu}|$ are defined from the values {0, 0.21, 0.42, 0.63, 0.84, 1.05, 1.37, 1.52, 1.74, 1.95, 2.18, 2.5}. The bins have variable size and the bin $[1.05, 1.37]$ is the largest.

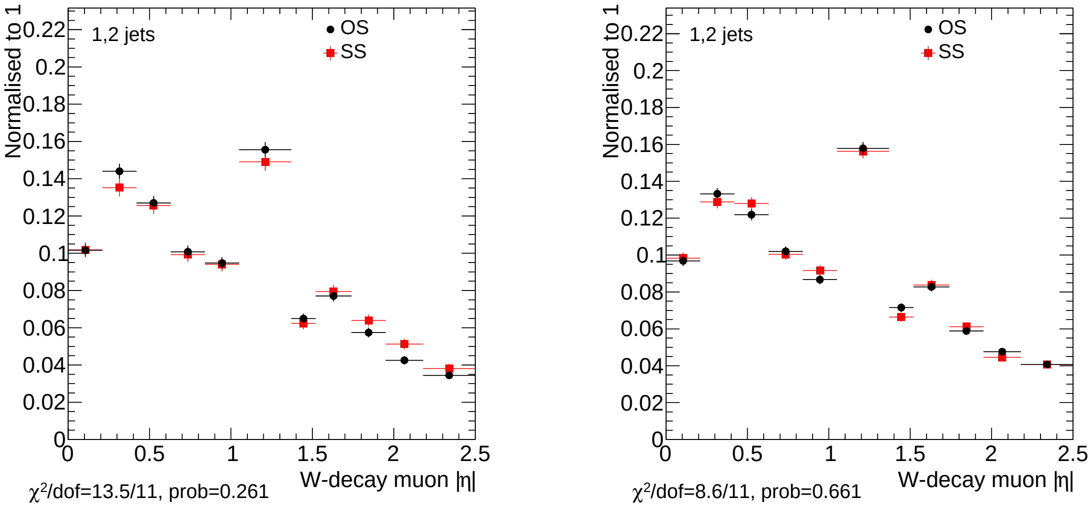
containing: (a) W^+ boson and one jet; (b) W^- boson and one jet; (c) W^+ boson and two jets; (d) W^- boson and two jets. Similarly, in figure 6.3 events of the 1,2-jets sample with (a) W^+ boson and (b) W^- boson are shown. The uncertainties on the data distributions are statistical. The W+c distributions are measured from the data by subtracting the estimated background in each $|\eta^\mu|$ bin.

The backgrounds are evaluated as described above and the distributions of multijet, W+light and Z+jets events are normalised to their data-driven estimations explained in section 6.1. The uncertainties on the backgrounds are the sum of the statistical and systematic components. The latter arise from the PDFs uncertainty and are evaluated as explained in section 7.2. Other uncertainties on the simulated shapes of the background processes have been found to be negligible. The total uncertainty on the background distributions are represented by the error bands.

The distributions of the data are compared with the predictions obtained as the sum of the backgrounds and signal distributions. The latter are taken from the $W+c$ simulation and normalised to the measured yields (table 6.6). A χ^2-test is performed between data and prediction and good agreement in found in all cases (the values of χ^2/dof are shown in the figures). This indicates that the $W+c$ simulation is representative of the data for the modelling of $|\eta^\mu|$ and provides a validation of the background estimation.

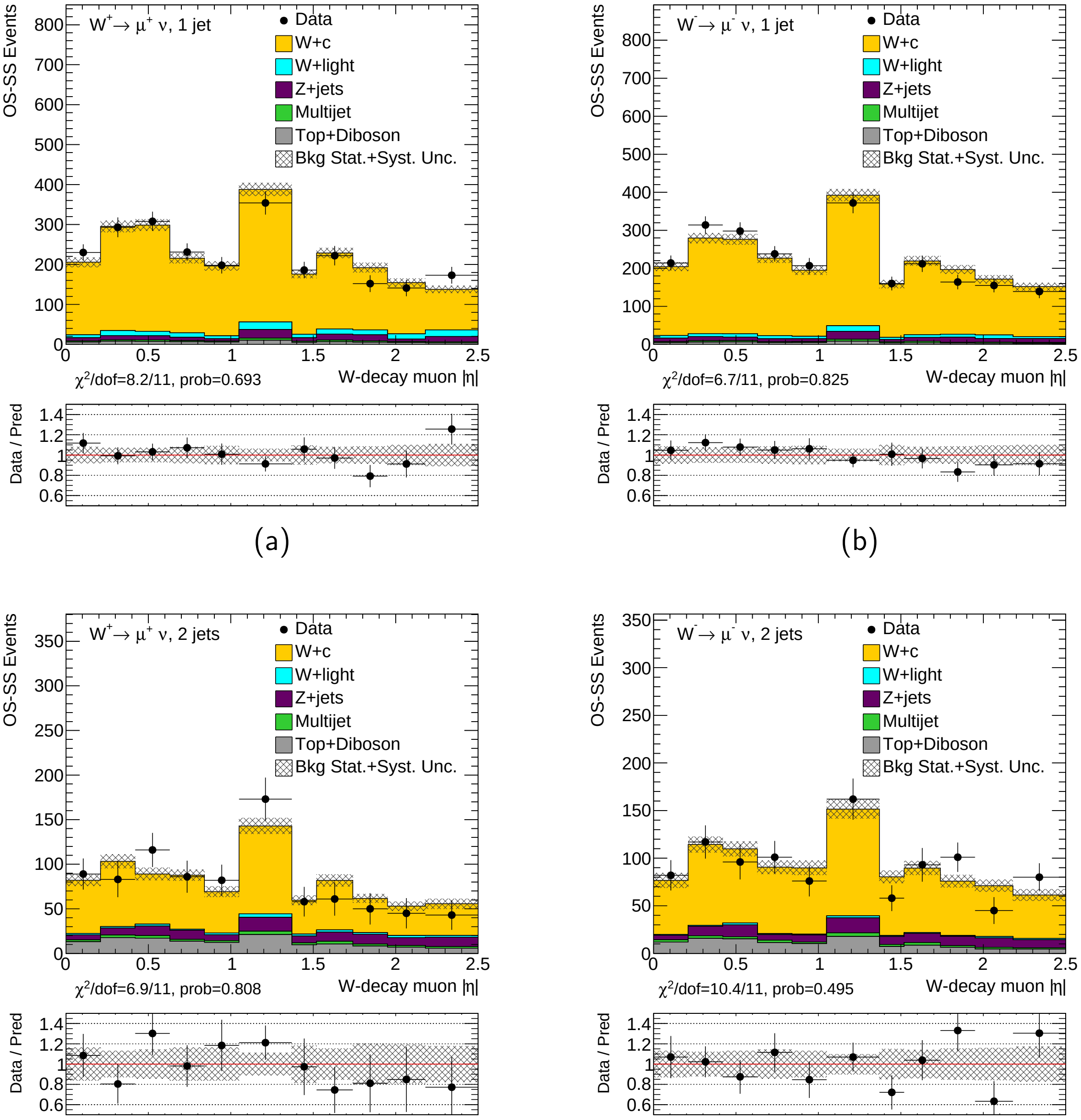


(a) Multijet (b) W+light



(c) Single top (d) $t\bar{t}$

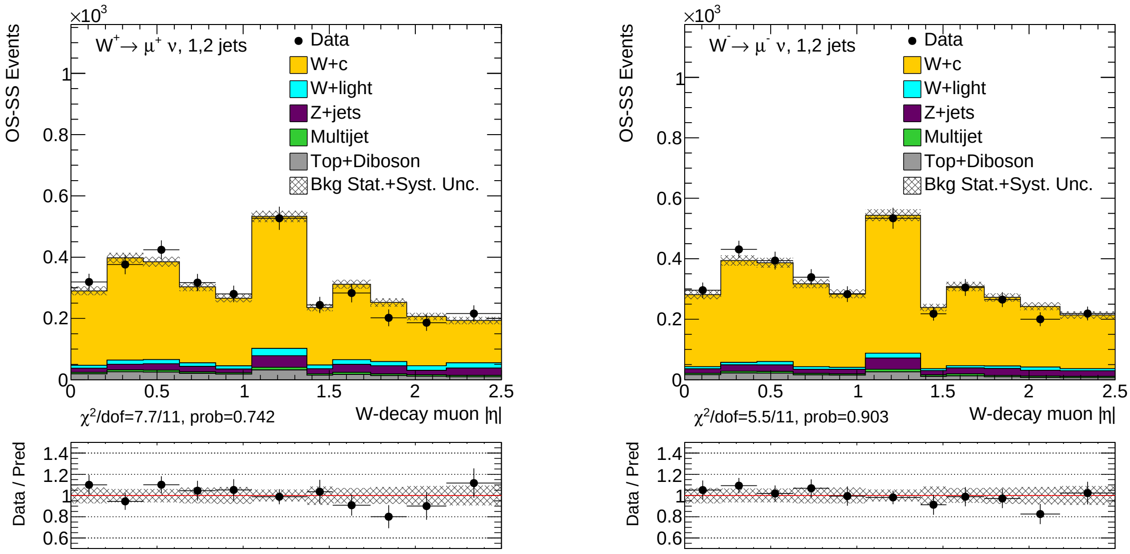
Figure 6.1: Normalised distributions of $|\eta^{\mu}|$ for the multijet, W+light and top quark productions in events with one or two jets. In (a), the shapes obtained for multijet events of the inverted isolation (red squares) and inverted m_{T}^{W} (blue triangles) control data samples are shown. Their average (black circles) and the uncertainty are represented by the black points and the error band, respectively. The distributions obtained from OS and SS events are compared for (b) W+light (c) single-top and (d) $t\bar{t}$ events with a χ^2-test and are found to be compatible.



(a) (b)

(c) (d)

Figure 6.2: Distributions of $|\eta^\mu|$ measured in the data for OS–SS events with: (a) W^+ boson and one jet; (b) W^- boson and one jet; (c) W^+ boson and two jets; (d) W^- boson and two jets. The distributions for W^++$\bar{c}$ and W^-+c productions are measured from the data by subtracting the estimated background. The uncertainties on data are statistical and those of the background include statistical and systematic contributions. The predicted distributions, i.e. the sum of background and signal, are also shown. The $W+c$ shapes are taken from simulation and normalised to the measured yields. The lower panels show the ratio of data to prediction and the result of their χ^2-test comparison.



(a) (b)

Figure 6.3: Distributions of $|\eta^{\mu}|$ measured in the data for OS–SS events of the 1,2-jets sample with (a) W^+ boson and (b) W^- boson. The distributions for $W^+ + \bar{c}$ and $W^- + c$ productions are measured from the data by subtracting the estimated background. The uncertainties on data are statistical and those of the background include statistical and systematic contributions. The predicted distributions, i.e. the sum of background and signal, are also shown. The $W+c$ shapes are taken from simulation and normalised to the measured yields. The lower panels show the ratio of data to prediction and the result of their χ^2-test comparison.

Chapter 7

Determination of the W+c production cross section

This chapter presents the results of the measurement of the W+c production cross section and the comparison with theoretical predictions.

Sections 7.1 to 7.3 report the extraction of the $W+c$ cross section from the event yields, which are presented in chapters 5 and 6 as a function of: (i) the jet multiplicity; (ii) the W boson charge; (iii) the pseudorapidity of the W-decay muon. The unfolding of the measured yields and the evaluation of the associated systematic uncertainties reported in references [1,2] and are summarised here. The results of the inclusive and differential cross section measurements are presented.

Additionally, section 7.4 presents the statistical combination of the W+c cross section measured in the *muon channel* and the *electron channel* analyses. The former refers to the analysis described in this thesis and the latter is the similar measurement using W boson decays to electrons, which is described in reference [1]. This combination of the cross sections is based on the profile likelihood method and is entirely the own work of the author of this thesis.

In section 7.5 the combined cross section is compared with the theoretical predictions calculated in reference [1]. Different PDF sets are used: the measurement is found to be sensitive to the strange quark content of the proton. Finally, section 7.6 reports additional results obtained from measurement of the $W+c$ production in this thesis. The W-charge ratio and the differential cross sections measured in the muon and electron channel analyses have been combined, and the results have been compared to theoretical predictions. The results of the muon and electron channels analyses have also been combined with those of the $WD^{(*)}$ analysis [1]. The latter is a measurement of the W+c production which employs a complementary method for the charm quark identification. The strange-to-down PDF ratio has been extracted from the combined $W+c$ production results.

7.1 Fiducial cross section definition

The W+c production cross section is measured in a *fiducial region*, or *fiducial phase space*. Fiducial cross sections are independent from the detector (particle level) and allow the comparison with theory predictions and among experimental results. The event yields of the signal are *unfolded* to the fiducial region, i.e. corrected for detector effects such as acceptance, efficiency and resolution. In order to reduce the extrapolation of the measured yields, the fiducial region closely matchs the selection criteria applied to the data (detector level).

The definition of the fiducial region for this analysis is summarised in table 7.1 and explained below.

The momentum of the W-decay muon is considered before final state radiation (section 3.1). Jets are constructed from stable particles, i.e. those with lifetime in excess of 10 ps, using the anti-k_T clustering algorithm with radius parameter R=0.4 (section 4.2.2). The muon and neutrino originating from the W-boson decay are excluded from the jet reconstruction.

The W+c production cross section is calculated separately for events containing one c-jet with no additional jets (*1-jet exclusive*), with exactly one additional non-c-jet (*2-jet exclusive*) or with any number of additional non-c-jets (*1-jet inclusive*). The c-quarks are identified through a c-hadron embedded in the reconstructed particle-level jets. The c-jets are those containing a c-hadron with $p_{\mathrm{T}} > 5$ GeV within a cone $\Delta R = 0.3$.

$W \to \mu\nu$ decay	
Muon:	$p_{\mathrm{T}} > 20$ GeV ; $\lvert\eta\rvert < 2.5$
Neutrino:	$p_{\mathrm{T}} > 25$ GeV
Transverse mass:	$m_{\mathrm{T}}^{W} > 40$ GeV
c-quark identification	
Jets :	Exactly 1, exactly 2, or ≥ 1 jets
	$p_{\mathrm{T}} > 25$ GeV ; $\lvert\eta\rvert < 2.5$
c-jet :	Exactly 1 jet containing a *c*-hadron
	$p_{\mathrm{T}}^{c-\mathrm{hadron}} > 5$ GeV
	ΔR(*c*-hadron, *c*-jet)< 0.3
OS–SS events using the *W* boson and the *c*-quark charge	

Table 7.1: Fiducial region of the $W{+}c$ production cross section. The selection criteria are explained in the text.

Integrated cross section

The integrated cross section in the fiducial region is calculated from the equation

$$\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}} = \frac{N_{\mathrm{data}}^{\mathrm{OS-SS}} - N_{\mathrm{bkg}}^{\mathrm{OS-SS}}}{U \cdot L}, \tag{7.1.1}$$

where $N_{\mathrm{data}}^{\mathrm{OS-SS}}$, $N_{\mathrm{bkg}}^{\mathrm{OS-SS}}$ and L are the yield, the estimated background, and the integrated luminosity of the data sample, respectively. The cross section $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ is the product of the total cross section and the branching ratio $BR(W \to \mu\nu)$.

The quantity U is the *unfolding factor*, which corrects the measured yield of signal events for the detector efficiency, acceptance and resolution. The unfolding factor is evaluated from the $W{+}c$ simulation (section 4.3) as

$$U = \frac{N_{\mathrm{W+c}}^{\mathrm{OS-SS}}(\text{detector level})}{N_{\mathrm{W+c}}^{\mathrm{OS-SS}}(\text{particle level})}, \tag{7.1.2}$$

where the numerator and denominator are the number of OS–SS events passing the detector selection criteria applied to the data (table 4.1), and the number evaluated in the defined fiducial region, respectively. The charge of the W boson and of the *c*-quark are used to classify OS and SS events in $N_{\mathrm{W+c}}^{\mathrm{OS-SS}}$(particle level).

The magnitude of U is about 2%, due to the small fraction of the c-quarks decaying to a muon.

Cross sections W-charge ratio

The W-charge ratio of the $W{+}c$ fiducial cross sections is defined and computed as

$$R^{\pm}_{W+c} \equiv \frac{\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}(W^{+}{+}\bar{c})}{\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}(W^{-}{+}c)} = \frac{R^{\pm,\,\mathrm{yield}}_{W+c}}{U^{+}/U^{-}}\,, \tag{7.1.3}$$

where $R^{\pm,\,\mathrm{yield}}_{W+c}$ is the W-charge ratio of the signal yields measured in data (equation 6.1.1). The unfolding factors U^{+} and U^{-} are calculated as per equation 7.1.2 for $W^{+}{+}\bar{c}$ and $W^{-}{+}c$ events, and their ratio is found to be 1.008 ± 0.005 in the 1-jet inclusive sample.

Differential cross section as a function of $|\eta^{\mu}|$

The differential cross section is calculated from equation 7.1.1 in intervals of the absolute pseudorapidity of the W-decay muon, $|\eta^{\mu}|$. The "bin-by-bin" unfolding is sufficiently accurate because the resolution of the $|\eta^{\mu}|$ measurement is high with respect to the bins width used.

7.2 Systematic uncertainties

The sources of systematic uncertainties on the fiducial cross section of equation 7.1.1 are: i) background estimation; ii) luminosity measurement; iii) detector response and its modelling in the simulation; iv) signal modelling in the simulation.

The sources and the magnitudes of systematic uncertainty evaluated for $\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}$ and $R^{\pm}_{W+c}$ in the 1-jet inclusive sample are found in table 7.2. The total systematic uncertainty on $\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}$ and $R^{\pm}_{W+c}$ are 5.3% and 2.0%, respectively. The two largest contributions to the cross section uncertainty arise from the modelling of the c-hadrons decay and from the jet energy scale. In the measured ratio, most of the systematic uncertainties either cancel out or are significantly reduced, being highly correlated in $W^{+}{+}\bar{c}$ and $W^{-}{+}c$ events.

Systematic uncertainty source	$\delta\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ [%]	$\delta R_{W+c}^{\pm}$ [%]
Backgrounds	1.9	1.0
Luminosity	1.8	-
W-decay muon trigger and reconstruction	0.8	-
W-decay muon p_{T} scale and resolution	0.6	0.6
$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction	0.3	0.3
Jet energy resolution	0.1	0.1
Jet energy scale	2.1	0.6
Soft muon tagging	1.4	-
c-quark fragmentation	1.6	-
c-hadrons decay	3.0	-
PS, UE and PDFs	0.2	1.4
MC signal sample size	1.4	0.5
Total	5.3	2.0

Table 7.2: List of the systematic uncertainties on $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ and $R_{W+c}^{\pm}$ in the 1-jet inclusive sample. The uncertainty on the backgrounds is shown in chapters 5 and 6 and that on the luminosity is determined in [34]. All other uncertainties are taken from [1] and their evaluation is outlined in the text.

The uncertainties related to the background estimation are shown in chapters 5 and 6 and are own work of the author of this thesis. The luminosity measurement is reported in [34]. The evaluation of the other systematic uncertainties is described in [1] and outlined below.

Uncertainties on the detector response and its modelling in the simulation are related to the event selection criteria applied to the data (section 4.2). These uncertainties have been assessed by recalculating the numerator of U in equation 7.1.2 and taking the difference of the measured $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ (or $R_{W+c}^{\pm}$).

The trigger and reconstruction efficiencies of the W-decay muon have been varied in the simulation within their respective uncertainties determined from data [58, 59]. A similar procedure has been used for the uncertainties due to the scale and resolution of the muon's p_{T}. The impact of the uncertainties due to the jet energy scale and resolution has been evaluated by varying each of these in the simulation within their respective uncertainties as determined from data [64]. The JES uncertainty ranges from less than 1% to about 7%, depending on jet p_{T} and η, with an additional 2% added in quadrature assigned to c-jets. The uncertainties of the momentum scale and resolution of the W-decay muon and the energy and resolution of jets have been propagated in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. In the $E_{\mathrm{T}}^{\mathrm{miss}}$ also the uncertainties stemming from the soft-jets and the calorimeter cells not associated with any physic object have been also taken into account as per [65]. The soft muon's tagging efficiency and mistag rate have been varied in the simulation within the range determined from the tagger calibration, which is summarised in section 4.2.3.

The evaluation of U depends on the modelling of the $W+c$ production and decay in the simulation. The modelling of c-quark fragmentation and c-hadrons decay are particularly relevant for the measurement with the SMT technique, as this affects the number of simulated c-hadrons, their fraction which decay to a muon, and the p_{T} spectrum of this muon. These uncertainties have been assessed by recalculating U and taking the difference in the measured $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ (or $R_{W+c}^{\pm}$).

The c-quark fragmentation systematic uncertainty is specific to the two corrections (section 4.3) applied to the signal simulation: (i) type and relative population of the c-hadrons resulting from c-quark fragmentation; (ii) fraction of the c-hadron's p_{T} carried by the reconstructed c-jet. The uncertainties on these two corrections are derived from the data in [80] and in [87], respectively. The c-hadron decay systematic uncertainty includes two sources: (i) the branching ratios of the c-hadrons' decays to a muon [82]; (ii) the p_{T} spectrum of the outgoing soft muons, whose uncertainty is evaluated as the difference of EvtGen and Herwig.

The systematic uncertainty due to the jet multiplicity of the simulated $W+c$ events has been estimated by varying the amount of initial and final state radiation in the parton shower (PS). The uncertainty related to the modelling of the underlying event (UE) is evaluated from simulation. The choice of the PDF set influences the signal simulation kinematics and the systematic uncertainty is assessed by using different PDF sets. Finally, the uncertainty on U due to the limited size of the simulated signal sample is 1.4%.

7.3 Cross section results

Table 7.3 presentes the fiducial cross section measured for the $W+c$, $W^{+}+\overline{c}$ and $W^{-}+c$ productions and the derived W-charge ratio. The definitions of $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ and $R^{\pm}_{W+c}$ are found in equations 7.1.1 and 7.1.3, respectively. The measurements are shown separately for the 1-jet, 2-jets and 1-jet inclusive samples. The two uncertainties quoted represent the statistical and systematic component, respectively.

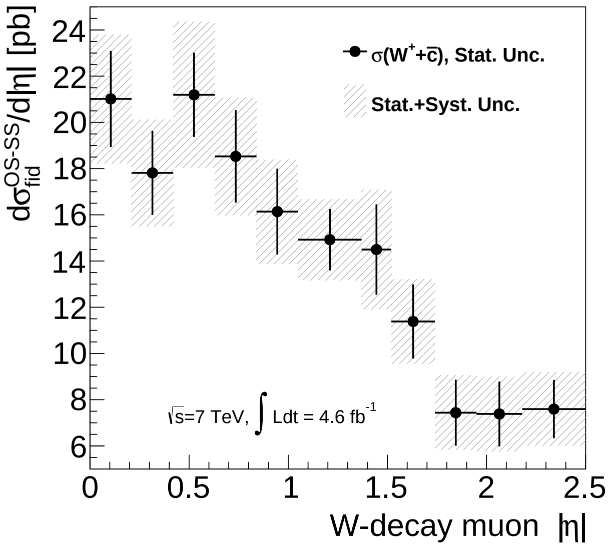
The $W+c$ production cross section measured in the fiducial region is found to be 72.6 pb, with a total uncertainty of 5.8%, for events with at least one jet. The systematic uncertainties on the measured cross sections are larger than the statistical uncertainties. In this fiducial region, $R^{\pm}_{W+c}$ is measured to be 0.92 with a total uncertainty of 4.6%, for events with at least one jet. As expected, the plus-to-minus W-charge ratio is slightly smaller than 1, owing to the larger fraction of $gd \to W^{-}c$ events than $g\overline{d} \to W^{+}\overline{c}$ events produced in proton-proton collision, as explained in section 7.6. The measurement of the W-charge ratio is limited by the statistical uncertainty.

The fiducial cross section as function of the $|\eta^{\mu}|$ is shown in figures 7.1 (a) and (b) for the $W^{+}+\overline{c}$ and $W^{-}+c$ events. The statistical and systematic uncertainties are represented by error bars and error bands, respectively. The uncertainties are dominated by the statistical component, especially in the forward $|\eta^{\mu}|$ region.

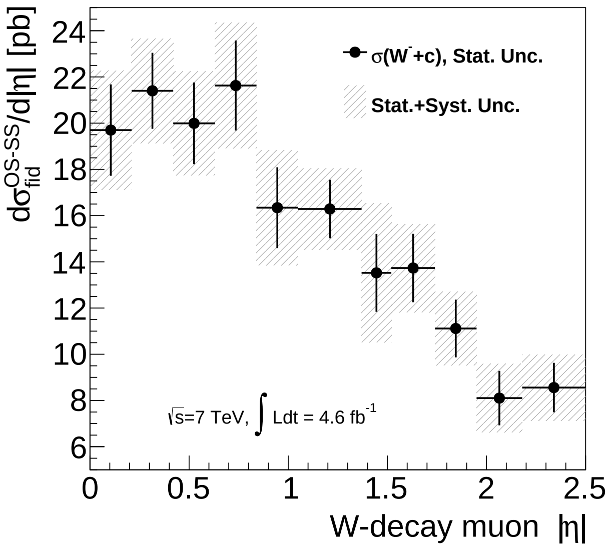
These measurements are compared with theoretical predictions in section 7.5, after the combination with the results of the electron channel analysis.

$\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W+c)$ [pb]	
1-jet	$54.1 \pm 1.3 \pm 3.1$
2-jets	$14.5 \pm 0.9 \pm 1.2$
1-jet inclusive	$72.6 \pm 1.7 \pm 3.9$
$\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^{+}+\overline{c})$ [pb]	
1-jet	$26.5 \pm 1.0 \pm 1.6$
2-jets	$6.4 \pm 0.7 \pm 0.6$
1-jet inclusive	$34.7 \pm 1.2 \pm 2.0$
$\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^{-}+c)$ [pb]	
1-jet	$27.7 \pm 0.9 \pm 1.6$
2-jets	$8.0 \pm 0.6 \pm 0.7$
1-jet inclusive	$37.9 \pm 1.1 \pm 2.0$
$R_{W+c}^{\pm} \equiv \sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^{+}+\overline{c}) \,/\, \sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^{-}+c)$	
1-jet	$0.96 \pm 0.05 \pm 0.02$
2-jets	$0.80 \pm 0.10 \pm 0.03$
1-jet inclusive	$0.92 \pm 0.04 \pm 0.02$

Table 7.3: Results for the measured fiducial cross sections and W-charge ratio, obtained in the 1-jet, 2-jets and 1,2-jets samples. The uncertainties shown correspond to the statistical (first) and the systematic (second) component.



(a)



(b)

Figure 7.1: Results of measured fiducial cross section as a function of $|\eta^{\mu}|$ for (a) $W^{+}+\overline{c}$and (b) $W^{-}+c$ events with at least one jet. The points represent the measured cross sections. The error bars and error bands indicate the statistical and total uncertainties, respectively.

7.4 Combination with the electron channel analysis

A measurement of the W+c production similar to that of described in this thesis, but using the electron channel decay of W bosons, is reported in reference [1]. The muon and electron channel analyses select independent samples of events from the ATLAS data (section 4.1) and measure the cross section in the identical fiducial region (section 7.1). Their results for $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}$ and $R_{W+c}^{\pm}$ are found to be compatible, and their statistical and systematic uncertainties are similar in magnitude [2]. The same sources of systematic uncertainty (section 7.2) have been considered and most of them are correlated between the two measurements.

This section describes the statistical combination of the cross section measured in the two channels, as done by the author of this thesis using the *profile likelihood ratio* method [88, 89]. The combination is performed to obtain a common result and to improve the precision of the W+c cross section measurement.

7.4.1 Profile likelihood ratio method

Likelihood functions are used to estimate parameters of statistical models from observed data. Models can contain both *parameters of interest* and *nuisance parameters*; the latter are not of interest but are introduced to make the model more representative of the observed data.

Given the likelihood function $\mathcal{L}(a, \underline{b})$, where a represents the parameter of interest[1] and $\underline{b}$ the nuisance parameters, the *profile likelihood ratio* is defined as

$$\lambda(a) \equiv \frac{\mathcal{L}(a, \hat{\hat{\underline{b}}})}{\mathcal{L}(\hat{a}, \hat{\underline{b}})} \,. \tag{7.4.1}$$

The numerator of equation 7.4.1 is the likelihood function maximised with respect to the parameters $\underline{b}$, for a given fixed value of a. In $\mathcal{L}(a, \hat{\hat{\underline{b}}})$ the likelihood function is "profiled" with respect to the nuisance parameters and $\hat{\hat{\underline{b}}}$ are known as the *conditional* maximum likelihood (ML) estimators of $\underline{b}$. The denominator $\mathcal{L}(\hat{a}, \hat{\underline{b}})$ represents the likelihood function maximised with respect to all parameters and ($\hat{a}$,

---

[1] The case of one parameter of interest is considered in this analysis, but the method can be extended to the case with more than one.

$\widehat{\underline{b}}$) are known as ML estimators of the parameters. The profile likelihood ratio is defined such that $\lambda(a) = 1$ for $a = \widehat{a}$.

Using the principle of maximum likelihood, the best estimation of the parameter a from the observed data is given by $\widehat{a}$. The uncertainty on the ML estimator is evaluated from the profile likelihood ratio. The estimators of the one standard deviation interval $\left[\widehat{a} - \delta^{-}_{\widehat{a}}, \widehat{a} + \delta^{+}_{\widehat{a}}\right]$ satisfy the relation

$$-\ln\lambda\left(\widehat{a} \pm \delta^{\pm}_{\widehat{a}}\right) = \frac{1}{2}\,. \tag{7.4.2}$$

The profile likelihood ratio method is used in the presence of a large number of nuisance parameters, which are added to the model to account for the systematic uncertainties in the measurement. The uncertainty on the parameter of interest extracted from the ML fit is larger when nuisance parameters are included, because the estimators $\widehat{a}$ and $\widehat{\underline{b}}$ are correlated and hence the contour defined in equation 7.4.2 becomes broader.

7.4.2 Likelihood function for the cross section combination

This section describes the likelihood function used to combine the $W+c$ production cross section measured in the muon and the electron channel analyses. The parameter of interest is the combined fiducial cross section in the 1-jet inclusive sample, which is hereafter referred to as σ.

The observed data of the model are the yields of OS–SS events selected from the data, $N^{c}_{\mathrm{obs}} = N^{\mathrm{OS-SS},c}_{\mathrm{data}}$, where $c = e, \mu$ refers to the electron and muon channel analysis. The observed yields are modeled with Gaussian distributions with expectation value N^{c}_{exp} and standard deviation s^{c}. Gaussian distributions are used owing to the large number of observed events and to account for the fact that the data yields are evaluated as OS–SS events. The parametrisation of N^{c}_{exp} is explained below and the standard deviation is fixed to the statistical uncertainty observed in the data, $s^{c} = \sqrt{N^{\mathrm{OS},c}_{\mathrm{data}} + N^{\mathrm{SS},c}_{\mathrm{data}}}$.

Equations 7.4.3 to 7.4.5 show the parametrisation of the expected number of events N^{c}_{exp} , which is expressed as the sum of the expected yield for the signal,

N^c_{sig}, and the background, N^c_{bkg}:

$$N^c_{\text{exp}} = N^c_{\text{sig}} + N^c_{\text{bkg}}\,, \tag{7.4.3}$$

$$N^c_{\text{sig}} = \sigma \cdot L \cdot U^c = \sigma \cdot L_0 \cdot U^c_0 \left(1 + \alpha_L \cdot \delta_L + \sum_{j\,\in\text{correl}} \alpha_j \cdot \delta^c_j + \sum_{k\,\in\text{uncorr}} \alpha^c_k \cdot \delta^c_k\right)\,, \tag{7.4.4}$$

$$N^c_{\text{bkg}} = N^c_{\text{bkg,0}}\left(1 + \alpha^c_{N\text{bkg}} \cdot \delta^c_{N\text{bkg}}\right)\,. \tag{7.4.5}$$

The signal yield N^c_{sig} is computed as the product of the combined cross section σ, integrated luminosity L and acceptance unfolding factor U^c (equation 7.1.1). To account for the systematic uncertainties on L, U^c and N^c_{bkg}, nuisance parameters are introduced into the model as explained below.

The luminosity nuisance parameter, α_L, is defined such that $\alpha_L = 0$ corresponds to the measured value for the luminosity, L_0, and $\alpha_L = \pm 1$ to a one standard deviation shift. The relative uncertainty on the luminosity measurement, δ_L, is fixed to the experimental uncertainty value ($\delta_L \equiv \delta_{L0}$).

Similarly, U^c_0 represents the estimated unfolding factor and nuisance parameters are added to account for the sources of systematic uncertainty. The uncertainties on U^c are considered to be either totally correlated or totally uncorrelated between the muon and electron channel analyses. Each correlated source introduces a single nuisance parameter in the model, which is indicated in equation 7.4.4 with α_j. Each uncorrelated source adds two nuisance parameters, one for the muon and one for the electron channel, which are indicated in equation 7.4.4 with α^c_k. Finally, the nuisance parameters $\alpha^c_{N\text{bkg}}$ are introduced in equation 7.4.5 to account for the uncertainty stemming from the background estimations.

The uncorrelated systematic uncertainties are modelled as independent Gaussian random variables. This corresponds to constraining the variation of each nuisance parameter α_i with an independent normal Gaussian distribution, $\text{Gaus}(0|\alpha_i, 1)$.

The likelihood function of the fit to extract the combined cross section σ is

$$\mathcal{L}(\sigma, \underline{\alpha}) = \prod_{c\,=\,e,\mu} \text{Gaus}\left(N^c_{\text{obs}}\,|\,N^c_{\text{exp}}(\sigma,\underline{\alpha})\,, s^c\right) \times \prod_{i\,\in\,\text{nuisances}} \text{Gaus}(0|\alpha_i, 1)\,, \tag{7.4.6}$$

where the index i includes the 16 nuisance parameters of the model.

Table 7.4 presents the observed yields, their statistical uncertainties, the observed luminosity, unfolding factors and background yields used in the likelihood function. The uncertainties on L_0, U_0^c and $N^c_{\mathrm{bkg},0}$ are listed in table 7.5. For the muon channel analysis, the observed and the background yield are evaluated in section 5.2 and the values of the unfolding factor and systematic uncertainties are reported in section 7.2. The analogous quantities in the electron channel analysis are taken from [2].

The sources of systematic uncertainty which are uncorrelated between the muon and electron channels are those associated with the identification of the lepton from the W-decay and with the background estimation. All other sources of systematic uncertainties are considered as correlated and are indicated with an asterisk (*) in table 7.5.

The ML fit is executed with the *RooFit/RooStats* software packages [90,91], in which the "negative log likelihood", $-\ln\mathcal{L}$, is numerically minimised by the *Minuit* software package [92].

	e-channel	μ-channel
N^c_{obs}	5918	6858
s^c	132	132
L_0 [pb^{-1}]	4580	
U_0^c	0.0167	0.0170
$N^c_{\mathrm{bkg},0}$	820	1189

Table 7.4: Observed yields (N^c_{obs}) with uncertainties (s^c) and the measured luminosity (L_0), unfolding factors (U_0^c) and backgrounds ($N^c_{\mathrm{bkg},0}$), for the muon and the electron channel analyses. These quantities are input to the maximum likelihood fit.

	Uncertainty [%]	
Source	e-channel	μ-channel
Luminosity *	1.8	
$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction *	0.8	0.3
Jet energy resolution *	0.1	0.1
Jet energy scale *	2.4	2.1
Soft muon tagging *	1.4	1.4
c-quark fragmentation *	2.0	1.6
c-hadron decays *	2.8	3.0
PS, UE and PDFs *	0.9	0.2
MC signal sample size	1.4	1.4
W-lepton trigger and reconstruction	0.7	0.8
W-lepton p_{T} scale and resolution	0.5	0.6
W-lepton charge misidentification	0.2	-
Background	24.8	9.3

Table 7.5: Uncertainties on the measured luminosity, unfolding factors and backgrounds, for the electron and muon channel analyses. The sources of systematic uncertainties which are correlated between the muon and electron channel analyses are indicated with (*). These quantities are input to the maximum likelihood fit.

7.4.3 Result of the maximum likelihood fit

The $W+c$ production cross section is measured to be $72.6 \pm 1.7 \pm 3.9$ pb in the muon channel (section 7.3) and $66.8 \pm 1.7 \pm 4.4$ pb in the electron channel [2], where the first uncertainty is statistical and the second is systematic. The combined cross section extracted from the ML fit is:

$$\widehat{\sigma} = 70.9\,^{+4.1}_{-3.8}\ \text{pb} = 70.9 \pm 1.2\,(\text{stat.})\,^{+3.9}_{-3.6}\,(\text{syst.})\ \text{pb}. \qquad (7.4.7)$$

The statistical and the systematic uncertainties of the combined result are smaller than those of the two individual measurements, as expected. The total uncertainty of the combined cross section corresponds to $+5.7\%/-5.3\%$, and the statistical and systematic components individually are 1.7% and $+5.5\%/-5.0\%$, respectively.

The statistical uncertainty on $\widehat{\sigma}$ is obtained from a ML fit without systematic uncertainties, i.e. by fixing $\underline{\alpha} \equiv 0$. In this case, the fit result is $\widehat{\sigma}_{\text{stat}} = 69.8 \pm 1.2\,(\text{stat.})$ pb, which, as expected, corresponds to the average of the two cross sections measurements weighted with their statistical uncertainties. The systematic uncertainty on $\widehat{\sigma}$ is computed by subtracting in quadrature the statistical uncertainty from the total uncertainty, where the latter is that obtained of the ML fit with both the statistical and systematic uncertainties.

The results of the ML fit can be seen in figure 7.2, in which the negative log profile likelihood ratio, $-\ln\lambda(\sigma)$, is shown as a function of the parameter of interest σ. The function minimised in the ML fit is shown by the solid blue curve. The ML estimator $\widehat{\sigma}$ corresponds to its minimum and the one standard deviation interval on $\widehat{\sigma}$ is indicated by the solid red lines. Similarly, the results of the ML fit including only statistical uncertainties is shown with dashed lines.

The presence of nuisance parameters in the fit broadens the shape of $-\ln\lambda(\sigma)$ and hence increases the uncertainty on $\widehat{\sigma}$ (equation 7.4.2). This reflects the "loss of information" in the data due to the systematic uncertainties in the cross section measurements. The profile likelihood of the fit with only statistical uncertainties is a quadratic function of σ and therefore $-\ln\lambda(\sigma)$ has a parabolic shape. If nuisance parameters are included in the fit, the coefficients of the quadratic function depend on the particular value of σ and, consequently, the $-\ln\lambda(\sigma)$ shape is only approximately a parabola. [2]

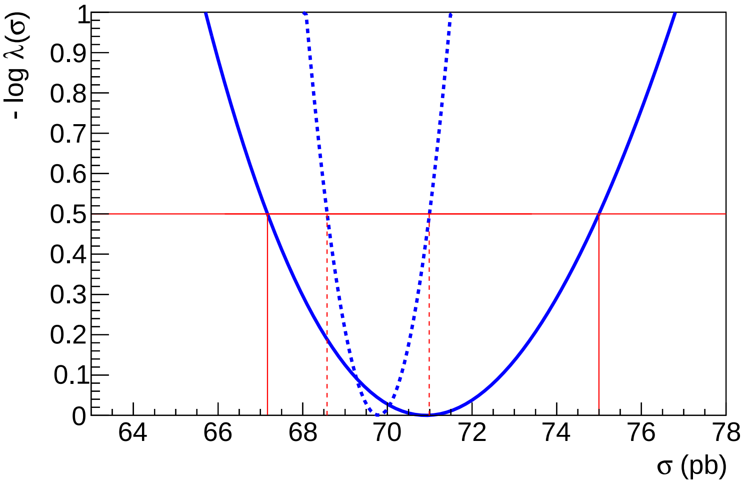


Figure 7.2: Negative log of $\lambda(\sigma)$ minimised in the maximum likelihood fit to extract the combined cross section (solid blue curve). Its minimum corresponds to the ML estimators of σ and the red solid lines represent the one standard deviation interval. Similarly, the dashed curve shows the result of the fit including only the statistical uncertainty.

Monte Carlo pseudo-experiments are used to verify that the estimator $\widehat{\sigma}$ has no bias and that its uncertainty is accurately extracted from the fit (see section 7.4.4). Moreover, it is also shown that the confidence level associated with the uncertainty interval can be taken as Gaussian, i.e. 68%.

Finally, the values and uncertainties of the 16 nuisance parameters extracted by the fit are listed in table 7.6. The fitted values correspond to the ML estimators $\widehat{\underline{\alpha}}$ and their uncertainties are derived from the Hessian matrix of $\mathcal{L}(\sigma, \underline{\alpha})$ at the global minimum. [3] The nuisance paramaters representing uncorrelated sources of systematic uncertainty in the two channels are varied by the fit. The largest shifts are found to be on the background yields, which result in a variation of 0.5 and -0.9

---

[2] This is because of the conditional ML estimators $\widehat{\widehat{\underline{b}}}$, which maximise the likelihood function for each given value of a.

[3] Given the likelihood function $\mathcal{L}(\underline{\theta})$, the inverse of the parameters' covariance matrix, Σ^{-1}, can be estimated as $(\Sigma^{-1})_{ij} = \partial^2 \mathcal{L}(\underline{\theta})/\partial\theta_i\partial\theta_j$, where the Hessian matrix of $\mathcal{L}(\underline{\theta})$ is calculated at $\theta = \widehat{\underline{\theta}}$.

standard deviations from the measured value in the muon and the electron channel, respectively. These two nuisance parameters present the largest variations because they are the less constrained, i.e. have the largest measurement uncertainties (see table 7.5). Furthermore, since the cross section measured in the muon channel is larger than that of the electron channel, nuisance parameters associated with the muon channel have positive shifts and those of the electron channel have negative shifts.

7.4.4 Fit validation with Monte Carlo pseudo-experiments

The validity of the results obtained from the ML fit is checked by generating an ensemble of 100,000 Monte Carlo pseudo-experiments. The pseudo-data are generated from the model's distribution $\mathrm{Gaus}(N^{c}_{\mathrm{obs}} \,|\, N^{c}_{\mathrm{exp}}(\sigma, \underline{\alpha}), s^{c})$, with nuisance parameters $\underline{\alpha} \equiv 0$ and with an arbitrary value of the cross section $\sigma_{\mathrm{true}} = 64.04$ pb. The latter is the baseline for the theoretical prediction of the $W+c$ cross section (see section 7.5). The generated pseudo-data N^{μ}_{obs} and N^{e}_{obs} are shown in figure 7.3 (a) and (b). To check the correct generation of the pseudo-data, the distributions are fitted to a Gaussian function and are found to be in agreement with the expectation.

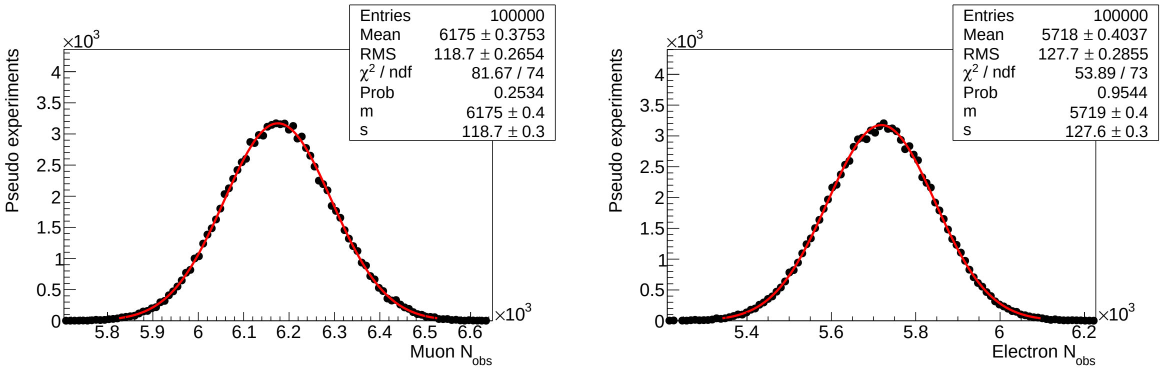
Similarly, the pseudo-measurements of U^{c}_{0}, $N^{c}_{\mathrm{bkg},0}$ and L_0 are generated from Gaussian distributions with centre and width equal to the measured value and uncertainty for each of these quantities (table 7.4). The ensembles of U^{μ}_{0} and U^{e}_{0} are generated taking into account the sources of systematic uncertainty which are correlated between the muon and electron channel (table 7.5). The generated distributions are shown in figures 7.3 (c) to (g). To check the correct generation of the pseudo-measurements, their distributions are fitted to Gaussian functions and found to be in agreement with the expectations.

The ML fit described in section 7.4.3 is executed on the ensemble of 100,000 pseudo-experiments and the *pull* is calculated from the result of each fit. The pull is defined as

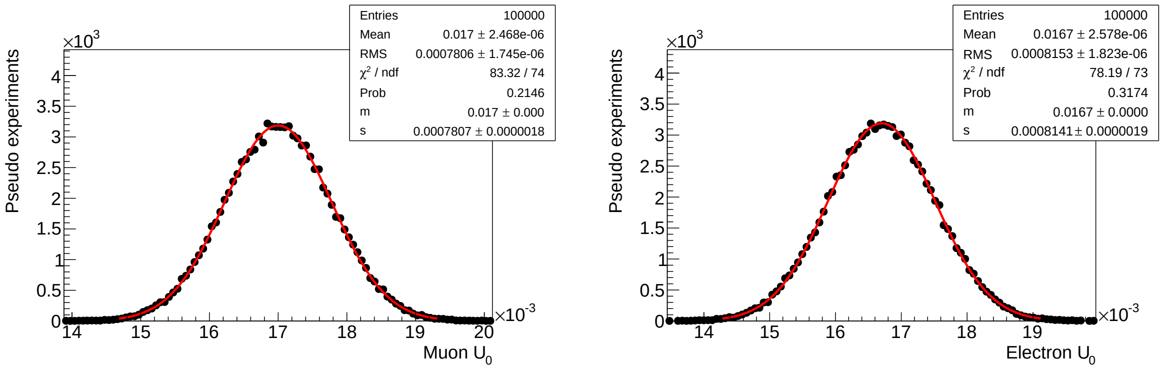
$$\begin{cases} (\sigma_{\mathrm{fit}} - \sigma_{\mathrm{true}})/\delta^{+}_{\mathrm{fit}} & \text{if } \sigma_{\mathrm{fit}} > \sigma_{\mathrm{true}} \; , \\ (\sigma_{\mathrm{fit}} - \sigma_{\mathrm{true}})/\delta^{-}_{\mathrm{fit}} & \text{otherwise} \, , \end{cases} \tag{7.4.8}$$

Nuisance parameter		Fit result Value	 Uncertainty
Luminosity		0.0	1.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction		-0.1	1.0
Jet energy resolution		0.0	1.0
Jet energy scale		-0.1	1.0
Soft muon tagging		0.0	1.0
c-quark fragmentation		-0.1	1.0
c-hadron decays		0.0	1.0
PS, UE and PDFs		-0.2	1.0
MC signal sample size	e-channel	-0.3	1.0
	μ-channel	0.3	1.0
W-decay lepton trigger and reconstruction	e-channel	-0.2	1.0
	μ-channel	0.2	1.0
W-decay lepton p_{T} scale and resolution	e-channel	-0.1	1.0
	μ-channel	0.2	1.0
W-decay lepton charge misidentification	e-channel	0.0	1.0
Background	e-channel	-0.9	0.8
	μ-channel	0.5	0.9

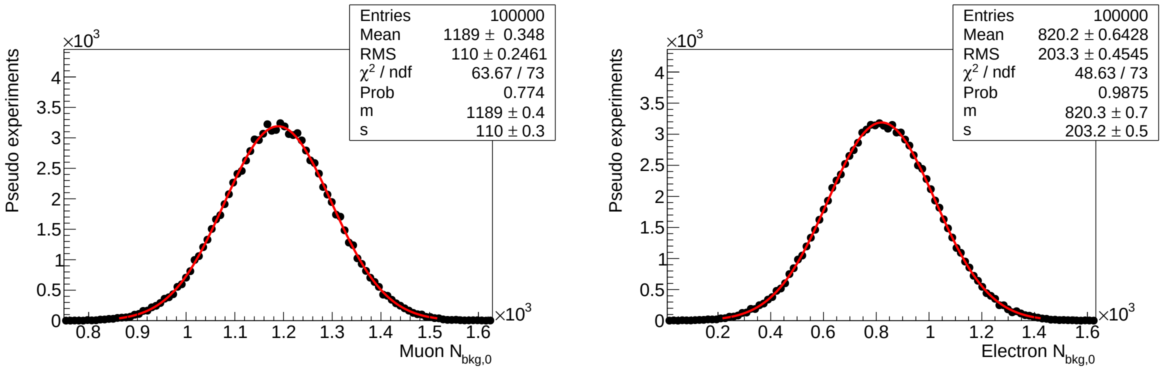
Table 7.6: Values and uncertainties of the nuisance parameters resulting from the maximum likelihood fit.



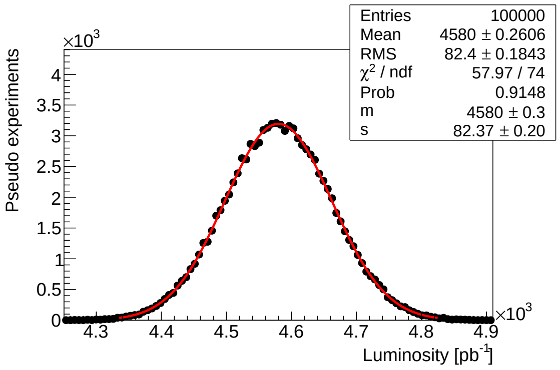
(a) N_{obs} μ-channel (b) N_{obs} e-channel



(c) U_0 μ-channel (d) U_0 e-channel



(e) $N_{\text{bkg},0}$ μ-channel (f) $N_{\text{bkg},0}$ e-channel



(g) L_0

Figure 7.3: Distributions of N^{μ}_{obs}, N^{e}_{obs}, L_0, U^{e}_{0}, U^{μ}_{0}, $N^{\mu}_{\text{bkg},0}$ and $N^{e}_{\text{bkg},0}$ of Monte Carlo pseudo-experiments. To check the correct generation of the ensembles, the distributions are fitted to Gaussian functions with mean m and standard deviation s.

where σ_{fit} is the cross section obtained from each fit and $\delta^{+}_{\text{fit}}/\delta^{-}_{\text{fit}}$ are the fit uncertainties. The resulting pull distribution is presented in figure 7.4. The root mean square (RMS) is found to be compatible with 1, showing that the fit uncertainties are an accurate estimation of the standard deviation of σ_{fit}. The mean value of the pull distribution indicates that the estimator σ_{fit} has a very small bias, which is neglected since it corresponds to only 0.0168 standards deviations or, equivalently, to 0.066 pb. [4] This small bias is introduced by the nuisance parameters, since the estimator was found to be unbiassed when the same validation procedure was carried out by fitting with statistical uncertainties only.

An unbinned maximum likelihood fit to a Gaussian with central value m and standard deviation s is performed on the pull distribution. The result shows that the pull is approximately Gaussian-distributed and hence the confidence level associated with the σ_{fit} uncertainty interval corresponds to approximately 68%. In addition, the parameters of the fitted gaussian (m and s) are found to be compatible with the distribution's mean and RMS, and the slightly asymmetric shape is found to have a skewness[5] of -0.151 ± 0.008.

---

[4] From the result of the combination of measured cross sections in equation 7.4.7, the bias is evaluated as: $\langle \text{pull} \rangle \cdot (\delta^{+} + \delta^{-})/2 = 0.0168 \cdot (4.1 + 3.8)/2$ pb.

[5] Given a random variable x, the skewness is $E\left[(\frac{x-\mu}{\sigma})^3\right]$, where $\mu = E[x]$ and $\sigma = Var[x]$.

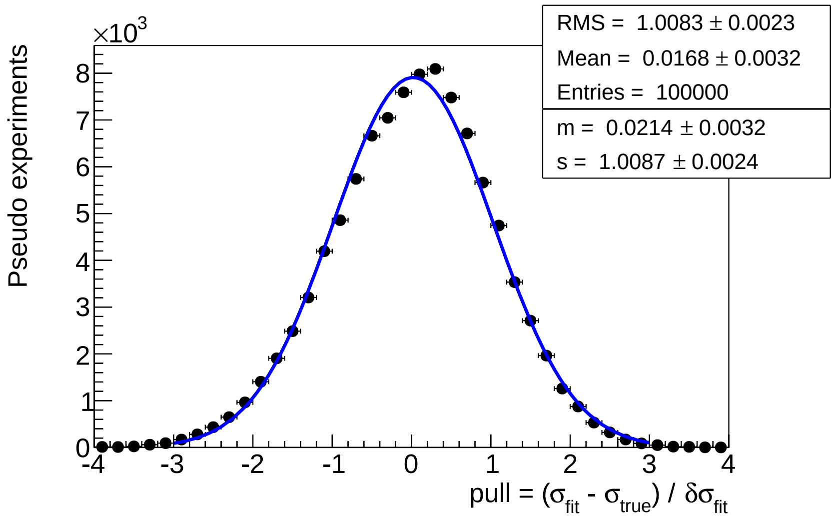


Figure 7.4: Distribution of the pull obtained from the fits to extract the combined cross section in the pseudo-experiments ensamble. The curve is the result of the unbinned maximum likelihood fit to a Gaussian with central value m and standard deviation s.

7.5 Comparison with theoretical predictions

This section compares the $W{+}c$ production cross section obtained from the combination of the muon and electron channel measurements with the QCD NLO theoretical predictions taken from reference [1].

The predicted $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}$ have been obtained with the aMC@NLO [93] generator, for the computation of the matrix element, interfaced with Herwig++ [81] for simulation of the parton shower, hadronisation and underlying event. The corrections to the simulated c-quark fragmentation explained in section 4.3 are applied to the predicted results.

The dependence of the NLO prediction on the choice of the renormalisation and factorisation scales has been evaluated by independently halving and doubling their nominal value, chosen as the sum of the transverse mass of all final-state particles. The prediction uncertainty due to the QCD scales has been found to be $+8\%/-4\%$. The fiducial cross section in the 1-jet inclusive sample ensures the smallest prediction uncertainties due to the PS simulation. This uncertainty is assessed by comparing the modelling of the Herwig++, Herwig and Pythia generators and is found to be 1%.

Six theoretical predictions have been calculated by using the six PDF sets introduced in section 3.3: CT10, MSTW2008, HERAPDF1.5, NNPDF2.3, NNPDF2.3coll and ATLAS-epWZ12. All parton distributions are computed at NLO, with the exception of ATLAS-epWZ12 which is an NNLO PDF set. For each prediction, the uncertainty due to the PDFs has been calculated according to the prescription of each PDF set. In the NNPDF2.3 and MSTW2008 analyses, the s-quark PDF is suppressed with respect to that of the d-quark sea. The CT10 and HERAPDF1.5 analyses show a similar, although smaller, suppression. The s-quark PDF is non-suppressed in the ATLAS-epWZ12 and NNPDF2.3coll analyses.

Figure 7.5 presents the comparison of the measured $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}$ for the $W+c$ production (equation 7.4.7) with the theoretical predictions. The vertical blue line is the measured cross section, whose inner yellow and outer green error bands correspond to the statistical and to the total uncertainties, respectively. The cross sections predicted using the six PDF sets are shown with markers and are found to differ by as much as 25%. The inner and outer error bars on the predictions represent the PDFs uncertainty and the total uncertainty, respectively. This total uncertainty on the predictions is the quadrature sum of those associated with the PDFs, QCD scales and parton shower.

The predictions for the ATLAS-epWZ12, NNPDF2.3coll, CT10 and HERAPDF1.5 PDF sets are found to be in agreement with the data within 1-sigma. In these PDF analyses, the s-quark contribution to the proton sea is non-suppressed or slightly suppressed. The predictions for the NNPDF2.3 and MSTW2008 PDF sets, in which the s-quark component is more heavily suppressed, are found to be less favoured by the data.

The measurement of the W+c production cross section in this thesis disfavours PDF sets with suppressed s-quark density.

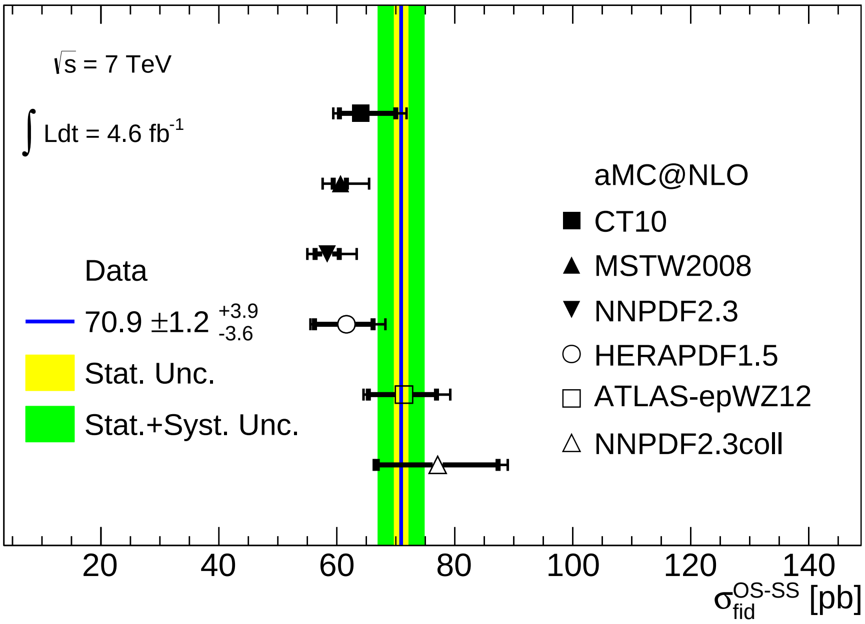


Figure 7.5: Measured fiducial cross section of the $W+c$ production compared to the theoretical predictions of aMC@NLO for six PDF sets. The blue vertical line shows the measurement, the yellow inner error band corresponds to its statistical uncertainty and the green outer error band to the quadratic sum of its statistical and systematic uncertainties. The six predictions are shown with markers, whose inner error bars are the uncertainties associated with the PDFs and outer error bars are the total prediction uncertainty (quadratic sum of PDFs, QCD scales and parton shower). The data show better agreement with the predictions where the *s*-quark density is non-suppressed or slightly suppressed.

7.6 Additional results on the W+c production

The $W+c$ analysis described in this thesis has been combined with those using additional W boson and c-quark decay channels. The combined results are found in reference [1] and are reported in this section.

The muon and the electron channel analyses have been combined to obtain a common measurement for the: (i) W-charge cross section ratio $R^{\pm}_{W+c}$ measurement; (ii) differential cross section as a function of $|\eta^{\mu}|$; (iii) differential cross section as a function of the jet multiplicity. The combined results have been compared to QCD NLO theoretical predictions.

The $W+c$ production has also been measured by tagging the c-quark from hadronic decays of the D and D^{*} mesons in the ATLAS data. This measurement, which is reported in the reference [1], is referred to as the $WD^{(*)}$ analysis. The cross sections measured using soft muon tagging, which are here referred to as Wc-jet muon and electron analyses, have been combined with those of the $WD^{(*)}$ analysis to determine a common value of the strange-to-down PDFs ratio.

Cross section W-charge ratio

The combined measurement of $R^{\pm}_{W+c}{=}\sigma^{\text{OS}-\text{SS}}_{\text{fid}}(W^{+}{+}\overline{c})/\sigma^{\text{OS}-\text{SS}}_{\text{fid}}(W^{-}{+}c)$ is compared to theoretical predictions in figure 7.6. The predictions and their uncertainties have been calculated as summarised in section 7.5 for $\sigma^{\text{OS}-\text{SS}}_{\text{fid}}$.

The measured ratio is found to be smaller than unity. This is expected in proton-proton collisions because, owing to the valence d-quarks, the $gd \to W^{-}c$ production is larger then the $g\overline{d} \to W^{+}\overline{c}$ production (section 3.2). Also a larger density of s-quarks relative to $\overline{s}$-quarks in the sensitive range of the measurement would lower $R^{\pm}_{W+c}$. Such an asymmetry between $s(x)$ and $\overline{s}(x)$ is suggested by neutrino data (section 3.3) and is implemented in the MSTW2008 and NNPDF2.3 PDF sets.

The measured $R^{\pm}_{W+c}$ is found to be compatible with the six theoretical predictions within 1-sigma. The statistical uncertainty is dominant and most of the systematic uncertainties cancel out or are reduced when evaluating the cross section ratio (section 7.3). Similarly, the uncertainties due to the QCD scales and to the parton shower cancel out in the predicted ratios.

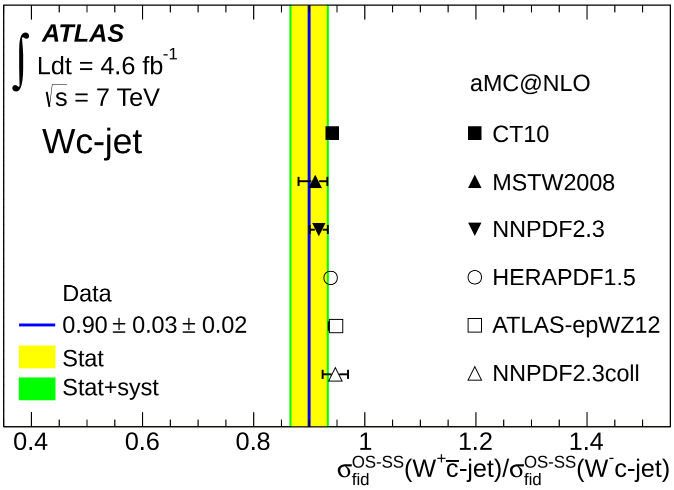


Figure 7.6: Measured cross section W-charge ratio compared to the theoretical predictions of aMC@NLO for six PDF sets. The blue vertical lines shows the measured ratio, the yellow inner error band corresponds to its statistical uncertainty and the green outer error band to the quadratic sum of its statistical and systematic uncertainties. The six predictions are shown with markers, whose error bars represent uncertainties associated with the PDFs. Taken from [1].

An estimation of the $s/\bar{s}$-quark PDF asymmetry, $A_{s\bar{s}}$, can be extracted from the measurement of $R^{\pm}_{W+c}$. The expected ratio can be approximated as

$$R^{\pm}_{W+c} \equiv \frac{\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^+ + \bar{c})}{\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^- + c)} \approx \frac{\bar{s} + |V_{cd}|^2 \cdot \bar{d}}{s + |V_{cd}|^2 \cdot d}\,, \tag{7.6.1}$$

where $|V_{cd}|^2 = 0.225$ and the PDFs are averaged over the phase space of the measurement (see also [40]). In PDF sets such as CT10, the strange densities are symmetric and hence $s \equiv \bar{s}$. Therefore the Cabibbo-suppressed processes $gd(\bar{d}) \to Wc(\bar{c})$ are the only mechanism which can lower $R^{\pm}_{W+c}$(CT10). Assuming that the ratios of d-quark to s-quark densities from CT10 are representative of the data, the strange PDF asymmetry can be estimated as

$$A_{s\bar{s}} \equiv \frac{s - \bar{s}}{s} \approx R^{\pm}_{W+c}(\text{CT10}) - R^{\pm}_{W+c}(\text{data}) = (2 \pm 3)\%\,, \tag{7.6.2}$$

where $R^{\pm}_{W+c}$(CT10) and $R^{\pm}_{W+c}$(data) are the cross section ratios obtained from the simulation and data, respectively. The uncertainty resulting on the estimation is largely due to the data statistical component.

The conclusions which can be drawn from this measurement of the W-charge ratio of the $W{+}c$ production are: (i) the data are consistent with the hypothesis of symmetric $s/\bar{s}$-quark PDFs; (ii) more data would be necessary to increase the sensitivity to the $s/\bar{s}$-quark PDF asymmetry further.

Differential cross sections as a function of $|\eta^{l}|$

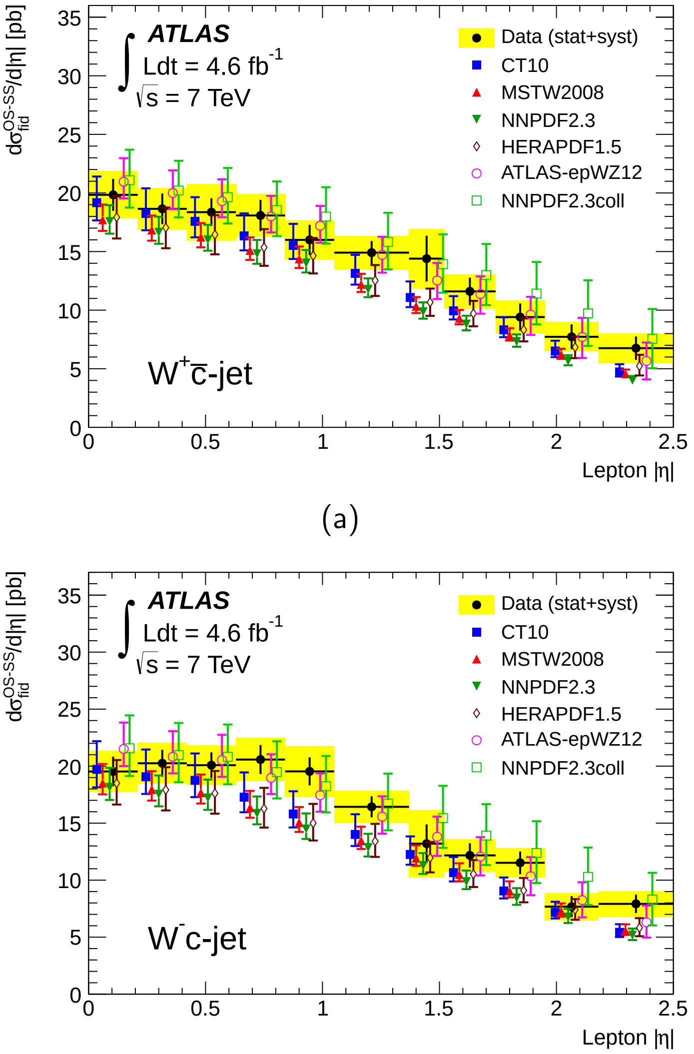
The dependence of $\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}(W^{+}{+}\bar{c})$ and $\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}(W^{-}{+}c)$ cross sections on $|\eta^{l}|$, where l is the lepton from the W decay, is shown in figure 7.7. Similar theoretical predictions of the $|\eta^{l}|$ shapes are obtained with the six PDF sets. The predicted distributions differ mainly in their normalisation and the predicted shapes are found to describe that of the data.

Differential cross section as a function of the jet multiplicity

In addition to the fiducial cross section with exactly one c-jet and any number of additional jets, the $\sigma^{\mathrm{OS-SS}}_{\mathrm{fid}}$ is measured by requiring either exactly one or exactly two jets only one of which is identified as a c-jet. The results of the muon channel analysis are presented in section 7.3 and have been combined with those of the electron channel.

The fiducial cross sections as a function of the jet multiplicity are compared to theoretical predictions in figure 7.8. The band on the data points show the sum of statistical and systematic uncertainties. The predictions have been calculated with both the aMC@NLO NLO generator (red solid) and the Alpgen LO generator (blue dashed). The Alpgen simulation is described in section 4.3 and is normalised to the inclusive W boson NNLO cross section in [69]. In the lower panel, the ratio of the simulated distribution to data is shown.

The one-to-two-jets ratio is found to be not well described by aMC@NLO. The prediction uncertainties due to QCD scales, parton shower and PDFs cancel out when considering the cross section ratio. The (NNLO-normalised) prediction of



(a)

(b)

Figure 7.7: Measured differential cross section as a function of W-lepton $|\eta|$ compared to predictions obtained using various PDF sets, for the (a) $W^{+}+\bar{c}$ (a) and (b) $W^{-}+c$ production. The measurements are shown by the black filled circles. The error bars give the statistical uncertainty, and the quadrature sum of the statistical and systematic uncertainties is shown by the yellow error band. The theory predictions using the six PDF sets are shown with different markers, whose error bars represent the total prediction uncertainties (sum in quadrature of PDFs, parton shower and QCD scales uncertainties). Taken from [1].

Alpgen are found to describe well the measured one-to-two-jets ratio although the data are underestimated for both the 1-jet and 2-jets samples. This result provides a test of the perturbative QCD calculations implemented in these simulations.

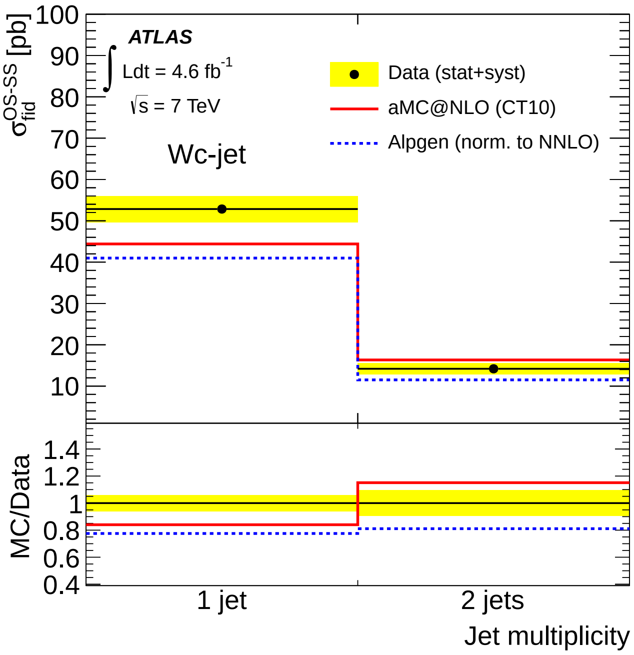


Figure 7.8: Measured $W{+}c$ production cross section as a function of the jet multiplicity, compared to the theoretical predictions from aMC@NLO (red solid) and Alpgen (blue dashed). The band on the data points shows the sum of statistical and systematic uncertainties. In the lower panel, the ratio of the simulated distribution to data is shown. Taken from [1].

Strange-to-down PDF ratio

The measurement of the $W+c$ production obtained from the $WD^{(*)}$ analysis is reported in reference [1]. In this analysis, the c-quark is tagged from the hadronic decays of D/D^* mesons to kaons and pions. The signal events selected in the $WD^{(*)}$ analysis are independent to those of the Wc-jet muon and electron channel analyses. The complementary results obtained from the Wc-jet and $WD^{(*)}$ analyses are found to be consistent.

The results of the two analyses have been combined to evaluate the *strange-to-down PDF ratio*, $r_s(x) \equiv [s(x)+\bar{s}(x)]\,/\,2\bar{d}(x)$, exclusively from the ATLAS

Wc-jet/$WD^{(*)}$ data. This calculation involves the minimisation of a χ^2 function containing: (i) the measurements of the W-lepton differential cross section of the $WD^{(*)}$ and Wc-jet analyses; (ii) the theoretical predictions obtained with HERAPDF1.5, in which the s-quark population in the proton sea is a floating parameter.

Figure 7.9 shows the value of r_s obtained as a function of x. In the HERAPDF1.5 PDF set (green), the s-quark sea density is lower than the d-quark sea density over the whole x range relevant to this measurement (10^{-3} to 10^{-1}). The ATLAS Wc-jet/$WD^{(*)}$ data (yellow) instead favour a non-suppressed s-quark PDF. Similarly, the ATLAS-epWZ12 analysis (red), which combines W and Z boson cross sections measured by ATLAS with ep collision data from the HERA collider, shows a symmetric composition of the light-quarks in the proton sea. The average of the strange-to-down PDF ratio over the x phase space is found to be

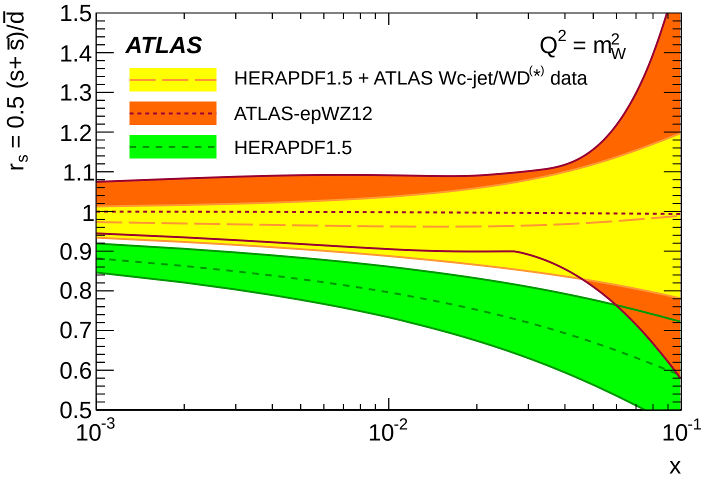


Figure 7.9: Ratio of the strange-to-down sea-quark PDFs as a function of x, obtained from: (yellow) the HERAPDF1.5 PDFs set using the ATLAS Wc-jet/WD$^{(*)}$ data to constrain the s-quark distribution; (green) the HERAPDF1.5 PDF set; (red) the ATLAS-epWZ12 analysis, which combines an ATLAS W/Z boson cross section measurement with HERA ep data. The error band include the measurement uncertainty (ATLAS Wc-jet/WD$^{(*)}$ and ATLAS-epWZ12) and/or that of the predictions (ATLAS Wc-jet/WD$^{(*)}$ and HERAPDF1.5). The ratio is shown for $Q^2 = m_W^2$. Taken from [1].

$$r_s \equiv \frac{s+\bar{s}}{2\bar{d}} = 0.96\,^{+0.26}_{-0.30}\,, \tag{7.6.3}$$

where the energy scale is $Q^2 = 1.9$ GeV^2 and the uncertainties stem from both the measurements and the theoretical prediction. These results for r_s support the hypothesis of a symmetric composition of light-quarks of the proton sea. This analysis results are found to be compatible with those of the ATLAS-epWZ12 analysis, which measures $r_s = 1.00^{+0.25}_{-0.28}$ from independent W and Z bosons data.

In conclusion, the measurement of the W+c production described in this thesis is combined with those using additional W boson and c-quark decay channels. The s-quark content of the proton is probed in the x range from 10^{-3} to 10^{-1} and with energy scale $Q^2 \sim m_W^2$. The analysis results disfavour PDF sets where the strange density is suppressed, supporting the hypothesis of a symmetric composition of light-quarks in the proton sea.

The strange distribution is only loosely determined using existing nucleon-neutrino DIS data. This measurement of the $W+c$ production provide further constraint for the determination of the strange quark density. Accurate determinations of the parton density functions further the knowledge of the composite structure of the proton and are important for all future analyses of LHC data. .

Conclusions

The standard model describes the nature of the electromagnetic, weak and strong interactions to an extraordinary degree of success. However, many standard model processes are very difficult to calculate from theory and require measurements using experimental data. Such processes include those which enable insight into the proton structure and those forming backgrounds to Higgs and discovery analyses. Therefore, precision measurements of the standard model are a fundamental part of the research programme at the LHC. One such measurement, which is the subject of this thesis, is the production cross section of a W boson with a single charm quark.

The $W+c$ production is of particular interest as a probe of the strange quark density in the proton. Flavour-symmetric strong interactions imply the same density of up, down and strange quarks in the proton sea. However, due to its larger mass, the density of strange quarks could be suppressed relative to that of the other light-quarks in the proton sea. The determination of the strange quark density is only loosely constrained by data. The $W+c$ events provide additional data to further constrain this parton density. Prior to the LHC, the $W+c$ production cross section could only be measured with a precision of 20–30%. The larger production rate available at the LHC provides the first opportunity to measure this process with sufficient precision to constrain the strange quark density in the range of $x = 10^{-3} - 10^{-1}$ at $Q^2 \sim m_W^2$.

In this thesis the $W+c$ production cross section has been measured using $4.6\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 7$ TeV collected with the ATLAS detector at the LHC. In events with a reconstructed $W \to \mu\nu$ decay, the charm quark was identified using soft muon tagging, that is by its semileptonic decay to a muon within a hadronic jet. The events analysed contained at most two jets, only one of which had to be tagged as a charm-jet. The analysis result were combined with those obtained from $W \to e\nu$ events. The combined results have been compared to various theoretical predictions for the: i) strange quark density; ii) s- and $\bar{s}$-quark PDF asymmetry; iii) shape of the strange quark PDF; iv) perturbative QCD calculation scheme.

The measured $W{+}c$ production cross section was found to be

$$\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W{+}c) \;=\; 70.9 \pm 1.2\,(\text{stat.})\;^{+3.9}_{-3.6}\,(\text{syst.})\text{ pb}\,.$$

This result has been compared with predictions of NLO QCD calculations obtained using various PDF sets. The data has been found to disfavour those predictions where the density of strange quarks is suppressed relative to that of the other light-quarks in the proton sea.

In order to investigate the s- and $\bar{s}$-quark PDF asymmetry, the ratio of the $W^{+}{+}\bar{c}$ and $W^{-}{+}c$ production cross sections has been measured. The ratio was found to be

$$R^{\pm}_{W+c} = 0.90 \pm 0.03\,(\text{stat.})\;\pm 0.02\,(\text{syst.})\,,$$

which is consistent with the hypothesis of symmetric s- and $\bar{s}$-quark PDFs. However, this result shows that a larger LHC data sample is necessary to achieve the sensitivity required by the small values of the predicted asymmetry.

Predictions for the shape of the strange quark PDF can be assessed from the dependence of the $W+c$ production cross section on the pseudorapidity of the lepton from the W boson decay. The measured $W+c$ production cross section as a function of the pseudorapidity was found to be well described by predictions using various PDF sets.

The dependence of the $W{+}c$ production cross section on the jet multiplicity has also been measured to test the perturbative QCD calculation scheme of the Alpgen and aMC@NLO generators. Alpgen implements a tree-level scheme while aMC@NLO a fixed-order scheme. The measured one-to-two-jets cross section ratio is well described by Alpgen, but not by aMC@NLO, thereby indicating that the tree-level scheme for the simulation of the $W{+}c$ events is more accurate.

The results of the analysis presented in this thesis have also been combined with those using complementary decay channels for the charm quark identification. A combined measurement of the strange quark density has been determined from the strange-to-down sea-quark PDF ratio, $r_s(x) \equiv [s(x) + \bar{s}(x)] \,/\, 2\bar{d}(x)$. The x-averaged ratio at $Q^2 = 1.9$ GeV^2 was found to be

$$r_s = 0.96\,^{+0.26}_{-0.30}.$$

This result supports the hypothesis of a symmetric composition of light-quarks in the proton sea. This result is also in agreement with the ratio determined in a recent QCD analysis by the ATLAS Collaboration using W and Z boson production data.

In conclusion, this thesis describes the analysis using, for the first time, data recorded by the ATLAS detector to measure the $W+c$ production cross section. The precision allows the results of this measurement to be sensitive to the strange quark density of the proton and, as such, used to further constrain it. This advances the knowledge of the fundamental structure of the proton, a component of every atom of ordinary matter in the Universe.

Appendix A

Estimation of the multijet background for a measurement of the top quark pair production

This appendix presents an estimation of the multijet background for the measurement of the top quark pair production cross section reported in reference [4]. The background is estimated extending the studies of the W+c analysis (section 5.1.1). The data-driven background estimation described here is the own work of the author and uses some inputs from the ATLAS top quark analysis group. The author has implemented the events selection and the soft muon tagging algorithm which are used in the analysis.

The $t\bar{t}$ cross section is measured with the pp collisions at $\sqrt{s} = 7$ TeV recorded by ATLAS at the LHC. Top quark pair events are selected in the single-lepton channel, and the b-quarks originating from the t-quarks decays are identified using the SMT. The event topology is similar to that of the $W + c$ production; the main difference is the higher jet multiplicity of the $t\bar{t}$ events. The cross section is measured with a total uncertainty of 10% and is to be found in agreement with theoretical predictions and the other $t\bar{t}$ production measurements at the LHC.

Introduction

Precision measurements of the $t\bar{t}$ production cross section, $\sigma_{t\bar{t}}$, are motivated by three main reasons: (i) test theoretical predictions of perturbative QCD; (ii) improve background estimations for measurements and searches performed at the LHC; (iii) potentially find hints for physics processes beyond the standard model. Moreover, the unprecedented number of $t\bar{t}$ events produced at the LHC (tens of thousands) allows to improve of the precision achieved by the measurements at the Tevatron.

A t-quark decays to a b-quark and a W boson. The analysis described in the reference. exploits the $t\bar{t}$ single-lepton channel, in which one of the W bosons decays to a muon or to an electron. Events are selected by requiring a single, isolated, high-p_{T} muon or electron produced in association with large $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets. The cross section is measured separately on the μ+jets and the e+jets samples.

The b-quarks are identified by requiring that at least one of the selected jets contains a soft muon. The total $b \rightarrow \mu X$ branching ratio, which includes the $b \rightarrow cX \rightarrow \mu X'$ sequential decay, is approximately 20%, and consequently 36% of the $t\bar{t}$ events contain at least one b-quark decaying to a muon. The measurement of $\sigma(t\bar{t})$ with the SMT algorithm is complementary to those using lifetime-based b-quark identification.

The data sample of $\sqrt{s} = 7$ TeV LHC collisions is described in section 4.1 and corresponds to an integrated luminosity of 4.66 fb^{-1}. The event selections of the e+jets analysis are discussed in reference [4]. The events of the μ+jets analysis are selected with the requirements and physics object definitions explained in section 4.2, and therefore are similar to those of the W+c analysis, except for:

- *Triangular cut:* the event selection requirement $E_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{T}}^{W} > 60$ GeV, which is known as the triangular cut, substitutes that made on the W boson's transverse mass, $m_{\mathrm{T}}^{W} > 60$ GeV;

- *Number of selected jets:* events with at least 3 jets (≥3-jets) or at least 4 jets (≥4-jets) are considered, and the analysis is performed separately in the two samples;

- *Number of SMT-jets:* events are required to contain at the least one SMT-jet. The samples selected with or without this requirement are referred to as *pretag* and *tagged*, respectively.

The cross section is calculated as

$$\sigma_{t\bar{t}} = \frac{N_{\text{data}} - N_{\text{bkg}}}{L \cdot U \cdot BR}, \tag{A.1}$$

where N_{data} is the event yield selected from the data, N_{bkg} is the estimated background contribution, L is the integrated luminosity of the data sample (4.66fb^{-1}) and U is the unfolding factor of the $t\bar{t}$ single-lepton channel. The branching ratio of the the $t\bar{t}$ single-lepton channel is $BR = 0.543$.

Multijet background estimation for μ+jets analysis

The multijet production processes contribute to the background yield of the cross section calculation. The yield of multijet events in the tagged sample, $N^{\text{tagged}}_{\text{multijet}}$, is estimated with a data-driven method by multiplying the yield in the pretag sample, $N^{\text{pretag}}_{\text{multijet}}$, for the SMT tagging rate, $R^{\text{SMT}}_{\text{multijet}}$:

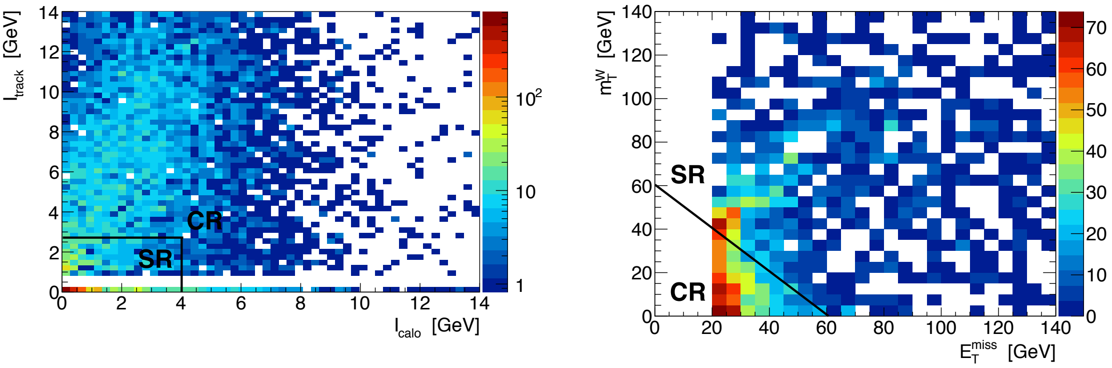
$$N^{\text{tagged}}_{\text{multijet}} = N^{\text{pretag}}_{\text{multijet}} \cdot R^{\text{SMT}}_{\text{multijet}}. \tag{A.2}$$

The matrix method described in section 5.1.1 is used to evaluate $N^{\text{pretag}}_{\text{multijet}}$. Two different sets of efficiencies, which have been measured by the ATLAS top quark analysis group, are used to obtain two estimations of the yield. The central value of $N^{\text{pretag}}_{\text{multijet}}$ is taken as the average of the two estimations, and results in 27000 ± 5400 and 6300 ± 1300 events for the $\geq$3-jets and $\geq$4-jets, respectively. An uncertainty of 20% covers half the difference of the two estimations and is assigned as the systematic uncertainty on $N^{\text{pretag}}_{\text{multijet}}$ in both samples.

The tagging rate $R^{\text{SMT}}_{\text{multijet}}$ is measured separately in two control data samples defined as

- *Inverted isolation:* $I_{\text{track}} > 2.5$ GeV and $I_{\text{calo}} > 4$ GeV;
- *Inverted triangular:* $E_{\text{T}}^{\text{miss}} + m_{\text{T}}^{W} < 60$ GeV.

All other selection criteria for the two control samples are identical to those of the data. Figure A.1 shows the expected number of multijet events, evaluated as explained below, as a function of the W-muon isolation variables and of $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{T}}^{W}$. The inverted isolation and inverted triangular control samples present a large number of QCD events.



(a) (b)

Figure A.1: Expected number of multijet events as a function of (a) the W-decay muon isolation variables I_{calo} and I_{track}; (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W}. Tagged events with at least three jets are shown. The values which define the signal (SR) and the two control (CR) regions are displayed.

The estimated fraction of non-multijet events in the two control samples are listed in table A.1. The inverted isolation sample has a 99% purity of multijet events and the inverted triangular sample has 60–68% of non-multijet events. The yield of $t\bar{t}$, W+jets and Z+jets events in the pretag and tagged ≥3-jets samples are estimated from simulation. [1] The number of non-multijet events from other sources, such as the single-top and the diboson productions, is negligible. The uncertainties include the statistical and systematic contributions. The latter are evaluated by assigning uncertainties of 25% (section 5.1.1) and 10% (section 5.1.4) to the W/Z+jets and $t\bar{t}$ predictions, respectively.

---

[1] The Monte Carlo simulation simulations used are described in section 4.3, and, in addition, they are corrected for the branching ratios of the b-quark decays to a muon as described in reference [4].

		Fraction of events [%]			
Control sample		$t\bar{t}$	W+jets	Z+jets	Total
Inverted isolation	Pretag	0.37 ± 0.06	0.71 ± 0.18	0.05 ± 0.01	1.13 ± 0.19
	Tagged	1.10 ± 0.16	0.29 ± 0.07	0.04 ± 0.01	1.42 ± 0.18
Inverted triangular	Pretag	6.5 ± 1.0	17.0 ± 4.3	8.1 ± 2.0	31.6 ± 4.8
	Tagged	24.9 ± 3.7	7.9 ± 2.0	6.8 ± 1.7	39.6 ± 4.6

Table A.1: Fractions of non-multijet events in the inverted isolation and the inverted triangular samples. Events with at least three jets are considered separately for the pretag and the tagged samples. The yields for the $t\bar{t}$, W+jets and Z+jets productions are evaluated from simulation. The uncertainties include the statistical and the systematic component.

The SMT tagging rate is calculated in each of the control samples as

$$R^{\text{SMT}}_{\text{multijet}} = \frac{(N_{\text{data}} - N_{t\bar{t}} - N_{W+\text{jets}} - N_{Z+\text{jets}})^{\text{tagged}}}{(N_{\text{data}} - N_{t\bar{t}} - N_{W+\text{jets}} - N_{Z+\text{jets}})^{\text{pretag}}}, \tag{A.3}$$

where N_{data}, $N_{t\bar{t}}$, N_{W+jets} and N_{Z+jets} are yields of data, $t\bar{t}$, W+jets and Z+jets events, respectively. The results are shown in table A.2, where the uncertainties on the estimations obtained from the inverted isolation and the inverted triangular samples include the statistical and the systematic contributions. The latter are evaluated by considering uncorrelated uncertainties on $N_{t\bar{t}}$, $N_{W+\text{jets}}$ and $N_{Z+\text{jets}}$ and between the pretag and tagged events.

As explained in section 5.1.1, the different tagging rate observed in the inverted isolation and inverted triangular samples is explained in terms of their different content of heavy-flavour jets. The tagging rate of the two control samples is expected to "bracket" that of the signal sample. Therefore, $R^{\text{SMT}}_{\text{multijet}}$ is taken as the unweighted average of the control sample values and its uncertainty as half the difference.

The final results for the multijet background in the muon channel analysis for the $t\bar{t}$ cross section measurement are presented in table A.3. The uncertainty on $N^{\text{tagged}}_{\text{multijet}}$ is obtained by propagating those associated with $N^{\text{pretag}}_{\text{multijet}}$ and $R^{\text{SMT}}_{\text{multijet}}$, and results in 27% in the ≥3-jets sample and 29% in the ≥4-jets sample.

	$R^{\mathrm{SMT}}_{\mathrm{multijet}}$ [%]	
Control sample	≥3 jets	≥4 jets
Inverted isolation	5.67 ± 0.11	7.66 ± 0.21
Inverted triangular	4.02 ± 0.50	5.41 ± 1.12
Average	4.85 ± 0.83	6.54 ± 1.13

Table A.2: Tagging rate of the multijet production evaluated from the events of either the inverted isolation or the inverted m_{T}^{W} samples. The results for the ≥3-jets and ≥4-jets samples are shown separately, whose uncertainties include the statistical and systematic contributions. The unweighted average of the two measurements and half the difference are taken best as the estimation of $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ in the signal sample and the uncertainty, respectively.

Multijet background estimation		
	≥3 jets	≥4 jets
Pretag yield $N^{\mathrm{pretag}}_{\mathrm{multijet}}$	27000 ± 5400	6300 ± 1300
SMT tagging rate $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ [%]	4.85 ± 0.83	6.54 ± 1.13
Tagged yield $N^{\mathrm{tagged}}_{\mathrm{multijet}}$	1310 ± 350	410 ± 120

Table A.3: Results of the multijet background estimation for the $t\bar{t}$ cross section measurement, in the ≥3-jets and ≥4-jets samples. The values of $N^{\mathrm{pretag}}_{\mathrm{multijet}}$ and $R^{\mathrm{SMT}}_{\mathrm{multijet}}$ are used in equation A.2 to obtain $N^{\mathrm{tagged}}_{\mathrm{multijet}}$. The uncertainties include the statistical and systematic components.

Top pair cross section results

The top pair production cross section measured in this analysis is found to be [4] :

$$\sigma_{t\bar{t}} = 165 \pm 2(\mathrm{stat}) \pm 17(\mathrm{syst})\ \mathrm{pb}.$$

This result is obtained by combining the cross section measured in the μ+jets and e+jets analyses, using events with ≥3-jets. The total uncertainty is found to be 10%, dominated by the systematic component. The most relevant contributions to the total uncertainty arise from the estimation of the W+jets and multijet backgrounds and from the jet energy scale.

The result obtained is in agreement with theoretical predictions based on the QCD NNLO predictions in references [74, 94] and with the other $t\bar{t}$ cross section measurements at the LHC. This is shown in figure A.2, where the theoretical prediction is compared to the results of the cross sections measurements from different ATLAS and CMS analyses [95]. The measurement here presented is that labelled as "ATLAS, l+jets, $b \to X\mu\nu$" in the lower part of the plot, which is found in agreement with the theoretical predictions and the other LHC measurements shown.

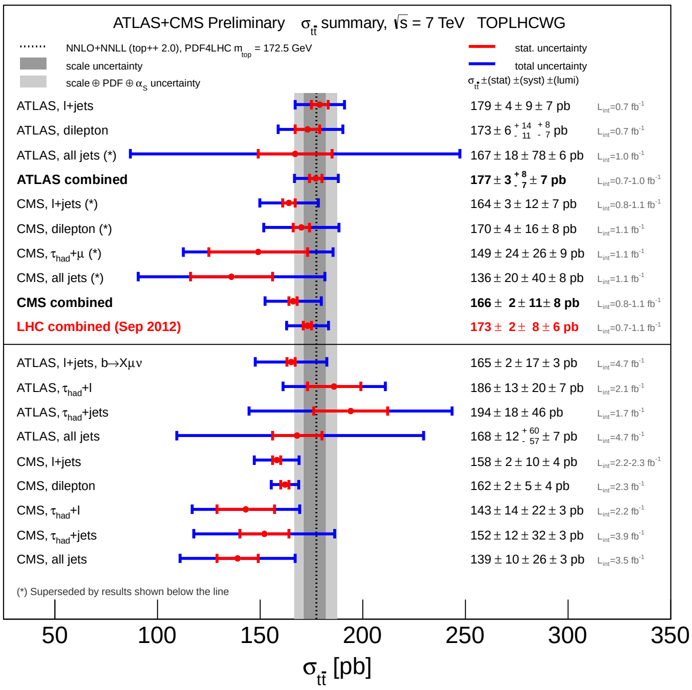


Figure A.2: Top quark pair production cross section measurements by ATLAS and CMS at the LHC. The blue (red) error bars represent the statistical (total) uncertainties. The error band shows the NNLO QCD prediction whose uncertainties are due to the renormalisation and factorisation scale, parton density functions and the strong coupling. The measurement here presented is that labelled as "ATLAS, l+jets, $b \to X\mu\nu$" in the lower part of the plot, which is found in agreement with the theoretical predictions and the other LHC measurements shown. Taken from [95].

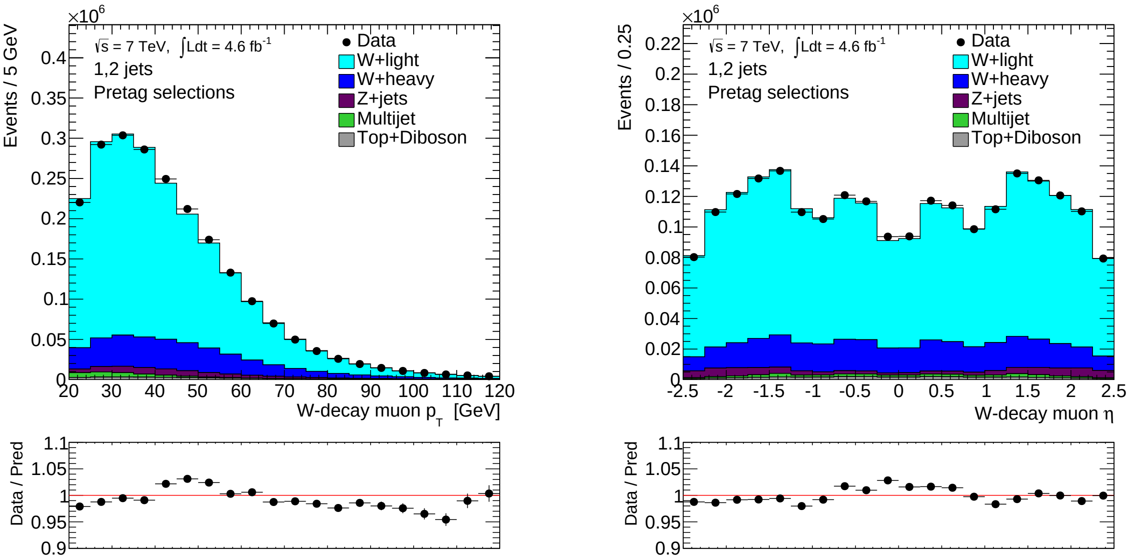
Appendix B

Kinematical distributions of the data pretag sample

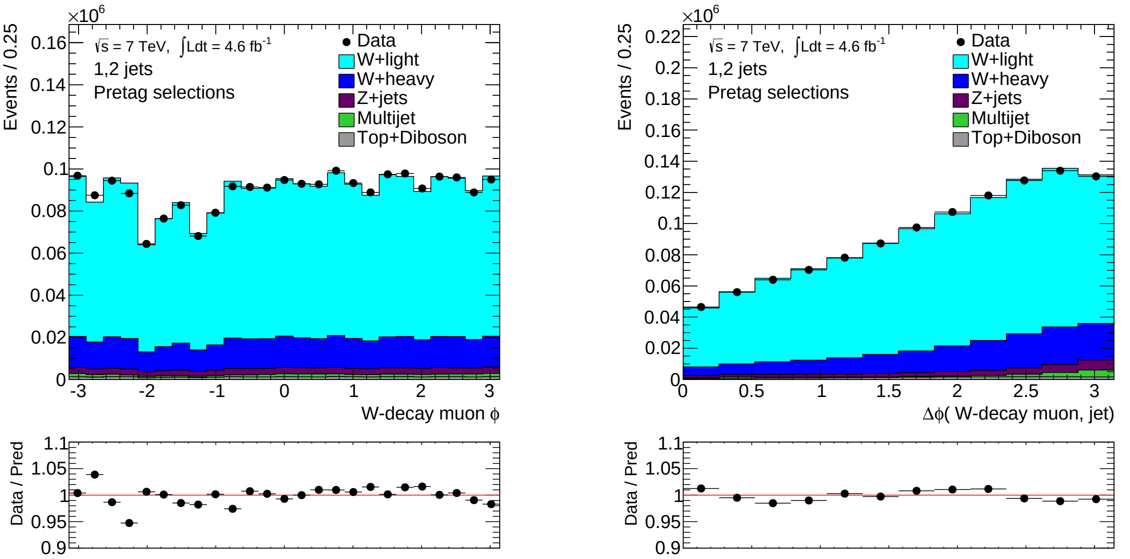
This appendix presents the distributions of several kinematical quantities for the pretag events selected from data. The latter are compared with the predictions obtained as the sum of W+light, W+heavy, multijet, Z+jets, top quark and diboson events.

The W+heavy events are those originating from the $W+b\bar{b}$, $W+c\bar{c}$ and $W+c$ productions. The distributions for the multijet events are evaluated from data with the matrix method described in section 5.1.1. All the other distributions are obtained from simulated events. The yields of W+jets events is normalised to match that of data and the fraction of W+light events is evaluated in a data-driven method, as described in section 5.1.2. The event yield of each process is listed in tables 5.7 and 5.8. The 1,2-jets sample is used and, for the events containing two jets, the leading jet is considered for the respective distributions.

The ratio of the data to predicted distributions is shown in the lower panels. The predictions are found to be representative of the data, with differences of the order of the percent.

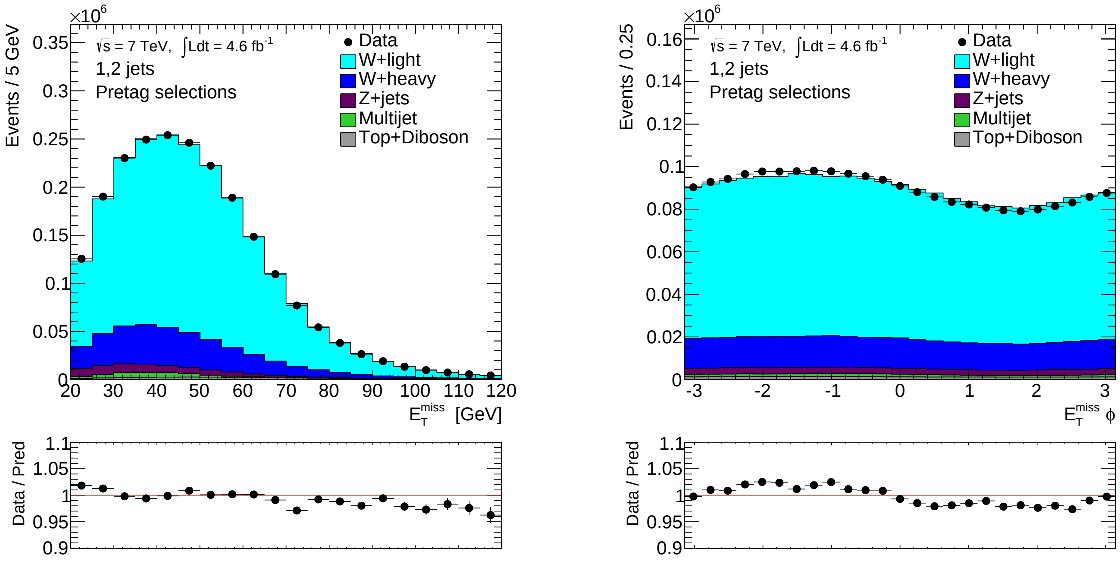


(a) W-decay muon p_{T} (b) W-decay muon η

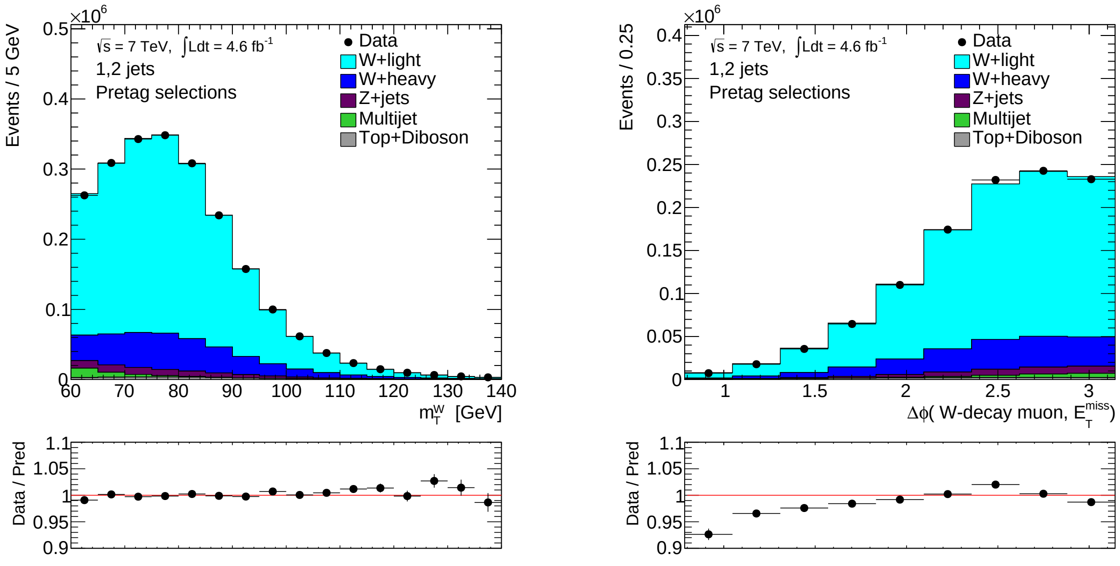


(c) W-decay muon ϕ (d) $\Delta\phi$(W-decay muon, leading jet)

Figure B.1: Distributions of p_{T}, η and ϕ of W-decay muon and of $\Delta\phi$(W-decay muon, leading jet) in events selected from data with the pretag selection criteria. Events of the 1,2-jets sample are used. Data distributions are compared with the predictions normalised to data event yield. The ratio of the data to predicted distribution is shown in the lower panel.

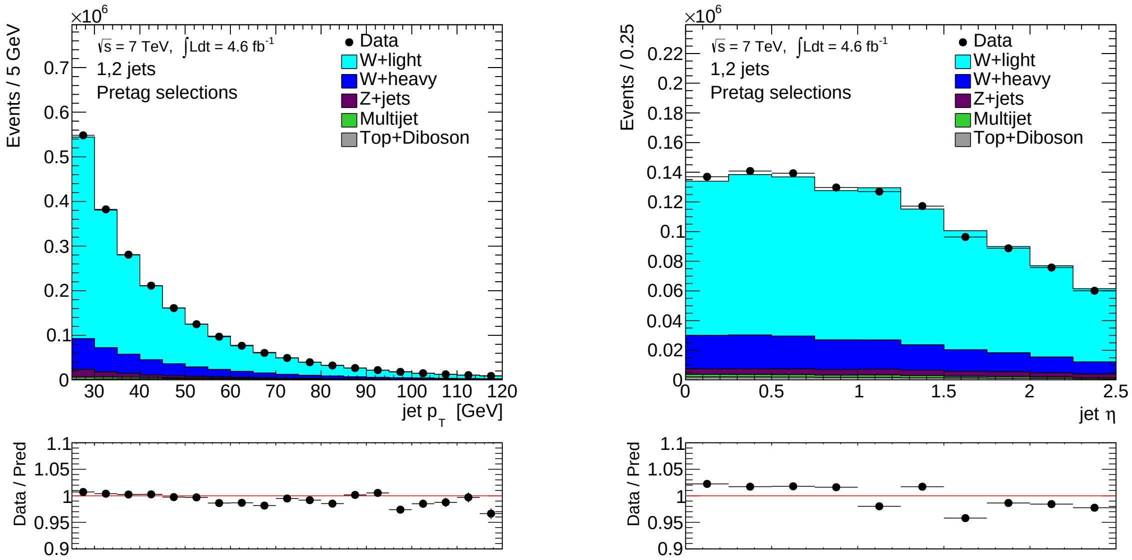


(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ ϕ



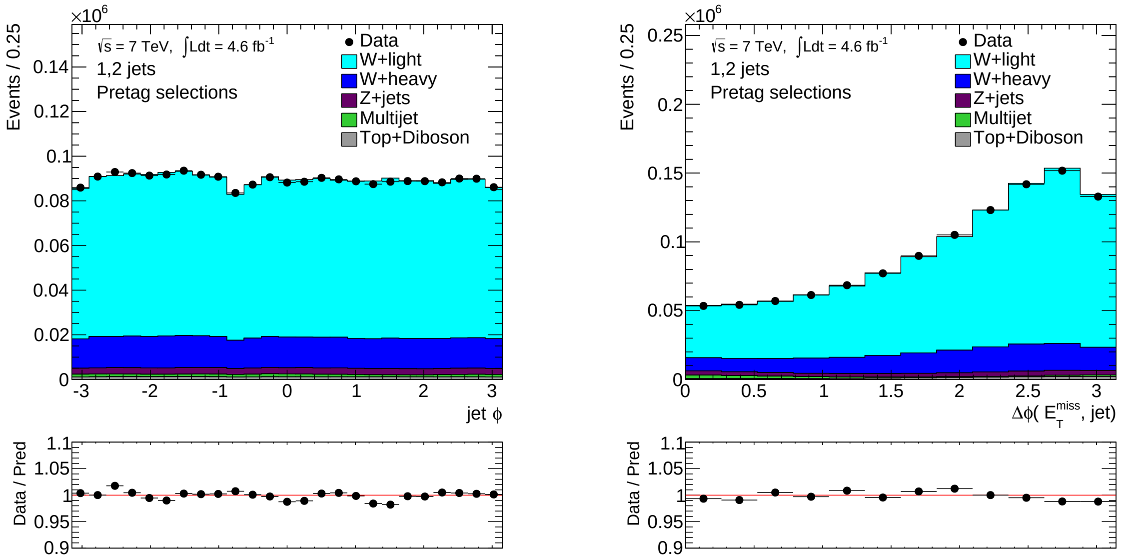
(c) m_{T}^{W} (d) $\Delta\phi$($E_{\mathrm{T}}^{\mathrm{miss}}$, W-decay muon)

Figure B.2: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}$ ϕ, m_{T}^{W} and $\Delta\phi$($E_{\mathrm{T}}^{\mathrm{miss}}$, W-decay muon) in events selected from data with the pretag selection criteria. Events of the 1,2-jets sample are used. Data distributions are compared with the predictions normalised to data event yield. The ratio of the data to predicted distribution is shown in the lower panel.



(a) leading jet E_{T}

(b) leading jet η



(c) leading jet ϕ

(d) $\Delta\phi$($E_{\mathrm{T}}^{\mathrm{miss}}$, leading jet)

Figure B.3: Distributions of p_{T}, η and ϕ of the leading jet and of $\Delta\phi$($E_{\mathrm{T}}^{\mathrm{miss}}$, leading jet) in events selected from data with the pretag selection criteria. Events of the 1,2-jets sample are used. Data distributions are compared with the predictions normalised to data event yield. The ratio of the data to predicted distribution is shown in the lower panel.

Acknowledgements

I thank the Queen Mary School of Physics and Astronomy for the Dr Edmund Gray PhD Studentship and the Queen Mary Doctoral College for the Postgraduate Research Fund related to my one year placement at CERN.

I owe my supervisor Lucio a big thank-you for instilling the enthusiasm and constantly guiding me throughout my doctoral research; without these two aspects this thesis would not have been possible. I also would like to thank Murrough Landon for patiently advising me on the L1Calo trigger analysis.

I am in debt with Georges Aad and Marco Vanadia for their invaluable help and precious collaboration in the hard work we shared for the $W+c$ analysis. I am grateful to Giuseppe Salamanna of the $t\bar{t}$ analysis team for introducing me to the ATLAS experiment during the first year of my PhD. Many thanks also to Kristin Lohwasser for her expert insights on the $W+c$ analysis PDF results.

Finally I must thank my parents Maria Rosa and Mario, and my grandfather Adolfo, for their never-ending support and encouragement. The last but probably more important person to thank is Sarah, for her positive attitude, ready smile, "bearing with me" and also for tirelessly proofreading this thesis.

Ad maiora!

Bibliography

[1] ATLAS Collaboration, *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **05** (2014) 68.

[2] G. Aad, H. Arnold, L. Cerrito, K. Lohwasser, G. Snidero, M. Vanadia, and W. Weiser, *Measurement of the production cross-section of a W boson in association with a charm jet using 4.6 fb^{-1}of data at $\sqrt{s} = 7$ TeV*, ATL-COM-PHYS-2012-1403, ATLAS internal analysis report.

[3] J. T. Childers and others (G. Snidero), *The ATLAS Level-1 Calorimeter Trigger: PreProcessor implementation and performance*, *JINST* **7** no. 12, (2012) P12008.

[4] ATLAS Collaboration, *Measurement of the top quark pair production cross-section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV in the single-lepton channel using semileptonic b-decays*, ATLAS-CONF-2012-131 (2012). `http://cds.cern.ch/record/1478370`.

[5] J. Blanco, V. Boisvert, L. Cerrito, N. Cooper-Smith, J. D. Morris, M. Rose, G. Salamanna, G. Snidero, and M. Vanadia, *Calibration of the χ^2_{match}-based Soft Muon Tagger algorithm*, ATL-COM-PHYS-2012-008, ATLAS internal analysis report.

[6] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579–588.

[7] A. Salam, *Weak and Electromagnetic Interactions*. N. Svartholm, p.367, Almqvist & Wiksell, 1968. Proceedings of the eighth Nobel symposium.

[8] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[9] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, *Phys. Rev.* **D2** (1970) 1285–1292.

[10] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, *Phys. Lett.* **B47** (1973) 365–368.

[11] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343–1346.

[12] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346–1349.

[13] `http://www.isgtw.org/sites/default/files/Standard_model_infographic.png`, May 2014.

[14] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, *Phys. Rev.* **145** (1966) 1156–1163.

[15] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323.

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

[17] CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626–2631.

[18] DØ Collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632–2637.

[19] Particle Data Group, *The mass and the width of the W boson – Review of Particle Physics*, *Phys. Rev.* **D86** (2012) 470–471.

[20] 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** (2012) 1–29.

[21] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61.

[22] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, *Phys. Lett.* **B726** (2013) 120–144.

[23] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, *Phys. Lett.* **B726** (2013) 88–119.

[24] Particle Data Group, *Higgs bosons: theory and searches – Review of Particle Physics*, *Phys. Rev.* **D86** (2012) 501–536.

[25] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390.

[26] Particle Data Group, *Hypothetical Particles and Concepts – Review of Particle Physics*, *Phys. Rev.* **D86** (2012). Sections listed in `http://pdg.lbl.gov/2013/reviews/contents_sports.html`.

[27] J. M. Butterworth, G. Dissertori, and G. P. Salam, *Hard Processes in Proton-Proton Collisions at the Large Hadron Collider*, *Ann. Rev. Nucl. Part. Sci.* **62** (2012) 387–405.

[28] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.

[29] `http://espace.cern.ch/acc-tec-sector/default.aspx`, May 2014.

[30] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[31] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[32] LHC*b* Collaboration, *The LHCb experiment at the CERN LHC*, *JINST* **3** (2008) S08005.

[33] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[34] ATLAS Collaboration, *Improved luminosity determination in pp collisions at* $\sqrt{s} = 7$ TeV *using the ATLAS detector at the LHC*, *Eur. Phys. J.* **C73** (2013) 2518.

Figure from `http://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[35] D. Boumediene, E. Dubreuil, and D. Pallin, *Calibration of the ATLAS Tile Calorimeter channels using the LASER system*, ATL-COM-TILECAL-2013-057, ATLAS internal analysis report.

[36] Particle Data Group, *Quantum Chromodynamics, Review of Particle Physics*, *Phys. Rev.* **D86** (2012) 120–135.

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

[38] V. Gribov and L. Lipatov, *Deep inelastic* e-p *scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[39] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298.

[40] W. J. Stirling and E. Vryonidou, *Charm production in association with an electroweak gauge boson at the LHC*, *Phys. Rev. Lett.* **109** (2012) 082002.

[41] W. T. Giele, S. Keller, and E. Laenen, *QCD corrections to* W *boson plus heavy quark production at the Tevatron*, *Phys. Lett.* **B372** (1996) 141–149.

[42] A. Kusina, T. Stavreva, S. Berge, F. I. Olness, I. Schienbein, et al., *Strange Quark PDFs and Implications for Drell-Yan Boson Production at the LHC*, *Phys. Rev.* **D85** (2012) 094028.

[43] NuTeV Collaboration, *Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data*, *Phys. Rev. Lett.* **99** (2007) 192001.

[44] NuTeV Collaboration, *Precise measurement of dimuon production cross-sections in muon neutrino Fe and muon anti-neutrino Fe deep inelastic scattering at the Tevatron*, *Phys. Rev.* **D64** (2001) 112006.

[45] CCFR Collaboration, *Determination of the strange quark content of the nucleon from a next-to-leading order QCD analysis of neutrino charm production*, *Z. Phys.* **C65** (1995) 189–198.

[46] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C63** (2009) 189–285.

[47] H1 and ZEUS Collaborations, F. D. Aaron et al., *Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA*, *JHEP* **01** (2010) 109.

[48] L. Hung-Liang et al., *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024.

[49] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B867** (2013) 244–289.

[50] ATLAS Collaboration, *Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections*, *Phys. Rev. Lett.* **109** (2012) 012001.

[51] S. Catani, D. de Florian, G. Rodrigo, and W. Vogelsang, *Perturbative Generation of a Strange-Quark Asymmetry in the Nucleon*, *Phys. Rev. Lett.* **93** (2004) 152003.

[52] U. Baur, F. Halzen, S. Keller, M. L. Mangano, and K. Riesselmann, *The Charm content of W + 1 jet events as a probe of the strange quark distribution function*, *Phys. Lett.* **B318** (1993) 544–548.

[53] CDF Collaboration, *First measurement of the production of a W boson in association with a single charm quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **100** (2008) 091803.

[54] CDF Collaboration, *Observation of the Production of a W Boson in Association with a Single Charm Quark*, Phys. Rev. Lett. **110** (2013) 071801.

[55] DØ Collaboration, *Measurement of the ratio of the $p\bar{p} \to W^+c^-$ jet cross section to the inclusive $p\bar{p} \to W+$ jets cross section*, Phys. Lett. **B666** (2008) 23–30.

[56] CMS Collaboration, *Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **1402** (2014) 013.

[57] Particle Data Group, *Summary tables of particle properties – Review of Particle Physics*, Phys. Rev. **D86** (2012) 43, 27.

[58] ATLAS Collaboration, *Performance of the ATLAS muon trigger in 2011*, ATLAS-CONF-2012-099 (2012). `http://cds.cern.ch/record/1462601`.

[59] ATLAS Collaboration, *Muon Performance in Minimum Bias pp Collision Data at $\sqrt{s} = 7$ TeV with ATLAS*, ATLAS-CONF-2010-036 (2010). `http://cds.cern.ch/record/1277675`.

[60] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, ATL-PHYS-PUB-2011-006 (2011). `http://cds.cern.ch/record/1345327`.

[61] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002 (2008). `http://cdsweb.cern.ch/record/1099735`.

[62] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. **C72** (2012) 1909.

[63] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **0804** (2008) 063.

[64] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C73** (2013) 2304.

[65] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, *Eur. Phys. J.* **C72** (2012) 1844.

[66] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020 (2012). `http://cds.cern.ch/record/1430034`.

[67] T. Sjostrand, *Monte Carlo Generators*, *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006* (2007), `arXiv:hep-ph/0611247 [hep-ph]`.

[68] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J.* **C70** 823–874.

[69] C. Anastasiou, L. Dixon, K. Melnikov, and F. Petriello, *High-precision QCD at hadron colliders: electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev.* **D69** (2004) 094008.

[70] 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** (2003) 001.

[71] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **0605** (2006) 026.

[72] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010.

[73] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029.

[74] M. Aliev et al., *HATHOR - HAdronic Top and Heavy quarks cross section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034–1046.

[75] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comput. Phys. Commun.* **184** no. 3, (2013) 919–985.

[76] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* **D83** (2010) 091503.

[77] N. Kidonakis, *Next-to-next-to-leading-logarithm resummation for s-channel single top quark production*, *Phys. Rev.* **D81** (2010) 054028.

[78] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, *Phys. Rev.* **D82** (2010) 054018.

[79] J. M. Campbell and R. Ellis, *MCFM for the Tevatron and the LHC*, *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10–15.

[80] E. Lohrmann, *A Summary of Charm Hadron Production Fractions*, `arXiv:1112.3757 [hep-ex]`.

[81] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J.* **C58** (2008) 639–707.

[82] Particle Data Group, *Charmed Mesons; Charmed Barions – Review of Particle Physics*, *Phys. Rev.* **D86** (2012) 885, 1383.

[83] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[84] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C71** (2011) 1577.

[85] ATLAS Collaboration, *Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. **C73** (2013) 2261.

[86] 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 (2012). `http://cds.cern.ch/record/1435197`.

[87] Belle Collaboration, *Charm hadrons from fragmentation and B decays in e^+e^- annihilation at $\sqrt{s} = 10.6$ GeV*, Phys. Rev. **D73** (2006) 032002.

[88] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** no. 2, (2011).

[89] Particle Data Group, *Statistics – Review of Particle Physics*, Phys. Rev. **D86** (2012) 390.

[90] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, (2003), `arXiv:physics/0306116`.

[91] Moneta, L. and Belasco, K. and Cranmer, K. and Kreiss, S. andin the p Lazzaro, A. and Piparo, D. and Schott, G. and Verkerke, W. and Wolf, M., *The RooStats Project*, *Proceedings of ACAT2010 Conference* (2010), `arXiv:1009.1003`.

[92] F. James, *MINUIT Reference Manual*, CERN Program Library Writeup D506. `http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html`.

[93] R. Frederix et al., *Scalar and pseudoscalar Higgs production in association with a top-antitop pair*, Phys. Lett. **B701** (2011) 427–433.

[94] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, `arXiv:1112.5675`.

[95] Top LHC Working Group. `http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/TOP`, May 2014.