RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Valerie Susanne Lang

Measurement of the W boson production in association with jets using $\sqrt{s} = 8\,\mathrm{TeV}$ proton-proton collisions at the ATLAS detector

Dissertation

HD-KIP 16-57

KIRCHHOFF-INSTITUT FÜR PHYSIK

Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Valerie Susanne Lang

born in Heppenheim

Oral examination on July 26, 2016

Measurement of the W boson production in association with jets using $\sqrt{s} = 8\,\text{TeV}$ proton-proton collisions at the ATLAS detector

Referees: Prof. Dr. Hans-Christian Schultz-Coulon
Prof. Dr. Johanna Stachel

Abstract

The production of W bosons in association with jets serves as a precision test of perturbative quantum chromodynamics (pQCD). Split into W^+ and W^- production, it provides access to the valence quark composition in the proton. The measurement of W + jets production presented here uses $20.3\,\mathrm{fb}^{-1}$ of proton-proton collision data at a center-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ recorded with the ATLAS detector. The cross section measurement is performed in the decay channel $W \to e\nu$ for the charge-independent W production as well as W^+ and W^- production separately, as a function of the jet multiplicity and differentially as a function of six observables in association with at least one or two jets. The results are compared to leading-order multi-leg and next-to-leading order pQCD calculations. The cross section ratio $(W^+ + \mathrm{jet})/(W^- + \mathrm{jet})$ is shown to be sensitive to the ratio of valence quark distribution functions in the proton, in a range of the momentum fraction $x \sim 0.1 - 0.5$. This range is accessible so far mainly by fixed-target deep inelastic scattering experiments and the W asymmetry measurements at the Tevatron. The ratio of W^+ to W^- production is measured as a function of the jet multiplicity and six observables in the presence of at least one or at least two jets, providing essential input to global fits of parton distribution functions.

Zusammenfassung

Die Produktion von W Bosonen in Verbindung mit Jets ermöglicht einen präzisen Test störungstheoretischer Berechnungen der Quantenchromodynamik (pQCD). Aufgeteilt in W^+ und W^- Produktion, erlaubt sie Rückschlüsse auf die Valenzquark-Struktur des Protons. Die hier beschriebene Messung der W + jets Produktion basiert auf $20{,}3\,\mathrm{fb}^{-1}$ an Daten von Proton-Proton Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 8\,\mathrm{TeV}$, aufgezeichnet mit dem ATLAS Detektor. Die Messung der Wirkungsquerschnitte wird im Zerfallskanal $W \to e\nu$ für die ladungsunabhängige W Produktion, sowie W^+ und W^- Produktion, als Funktion der Anzahl an Jets und differentiell als Funktion von sechs Observablen in Verbindung mit einem oder zwei Jets durchgeführt. Die Ergebnisse werden mit pQCD Berechnungen führender und nächst-führender Ordnung verglichen. Der Quotient der Wirkungsquerschnitte $(W^+ + \mathrm{jet})/(W^- + \mathrm{jet})$ wird als sensitiv befunden hinsichtlich des Verhältnisses der Valenzquark-Verteilungsfunktionen im Proton in einem Bereich des Impulsanteils von $x \sim 0{,}1 - 0{,}5$. Dieser Bereich ist bisher hauptsächlich mittels Fix-Target-Experimenten zu tiefinelastischer Streuung sowie W Asymmetriemessungen am Tevatron zugänglich. Der Quotient aus W^+ und W^- Produktion wird als Funktion der Anzahl an Jets sowie sechs Observablen bei mindestens einem oder zwei vorhandenen Jets gemessen und liefert wichtigen Input für globale Fits von Parton-Verteilungsfunktionen.

Contents

CHAPTER 1

Introduction

Particle collisions in modern accelerators test the structure of matter on very short distances and at high energies. They provide a unique view into the structure and the interactions of particles and allow to obtain a deeper understanding of the laws of nature.

The *Large Hadron Collider* (LHC) collides protons at unprecedented energies and at enormous rates, providing the energy to create new particles and producing the large data samples needed to measure known processes with previously unachievable precision. In the past years, its center-of-mass energy $\sqrt{s}$ increased from 7 TeV to 8 TeV and now to 13 TeV with the largest data sample so far obtained at $\sqrt{s} = 8\,\text{TeV}$ in 2012.

The production of W bosons is one of the most abundant processes at the LHC. Its total production cross section as a function of $\sqrt{s}$ is shown in Figure 1.1. Only the cross sections for jet production and inelastic collisions are larger. At $\sqrt{s} = 8\,\text{TeV}$, the cross section of W production with subsequent leptonic decay ($W \to \ell\nu$) is approximately 12 nb [2]. Given the integrated luminosity of $22.8\,\text{fb}^{-1}$ delivered by the LHC in 2012, this makes around 270 Mio W bosons. Assuming an average instantaneous luminosity of roughly $4 \cdot 10^{33}\ \text{cm}^{-2}\ \text{s}^{-1}$, this corresponds to approximately 50 W bosons per second.

W bosons are charged particles and the production of W^+ and W^- bosons depends on the content of the colliding protons. The combined content of two colliding protons is however not charge symmetric, leading to differences in the W^+ and W^- production cross sections, in contrast to the symmetric production cross section at the proton-anti-proton ($p\bar{p}$) collider Tevatron. The measurement of W^+ and W^- cross sections at the LHC therefore allows insights into the composition of the proton and places experimental constraints on parton densities in the proton.

In approximately 10 % of the W boson events, the W bosons are produced in association with hadronic jets (W + jets). Since jet emissions are described by perturbatively computable *quantum chromodynamics* (pQCD), W + jets cross section measurements are a powerful way to test these predictions and drive the development of higher-order calculations. Previous measurements of W + jets production have been performed by the CDF and DØ collaborations at the Tevatron [3–6] as well as the ATLAS and CMS collaborations at the LHC [7–10]. These measurements have tested pQCD predictions in a variety of observables, jet multiplicities and center-of-mass energies.

This analysis presents the measurement of n_{jets}-dependent W + jets production cross sections up to 7 jets at $\sqrt{s} = 8\,\text{TeV}$. In association with at least 1 or 2 jets, the W cross section measurement is performed furthermore differentially as a function of the following six observables:

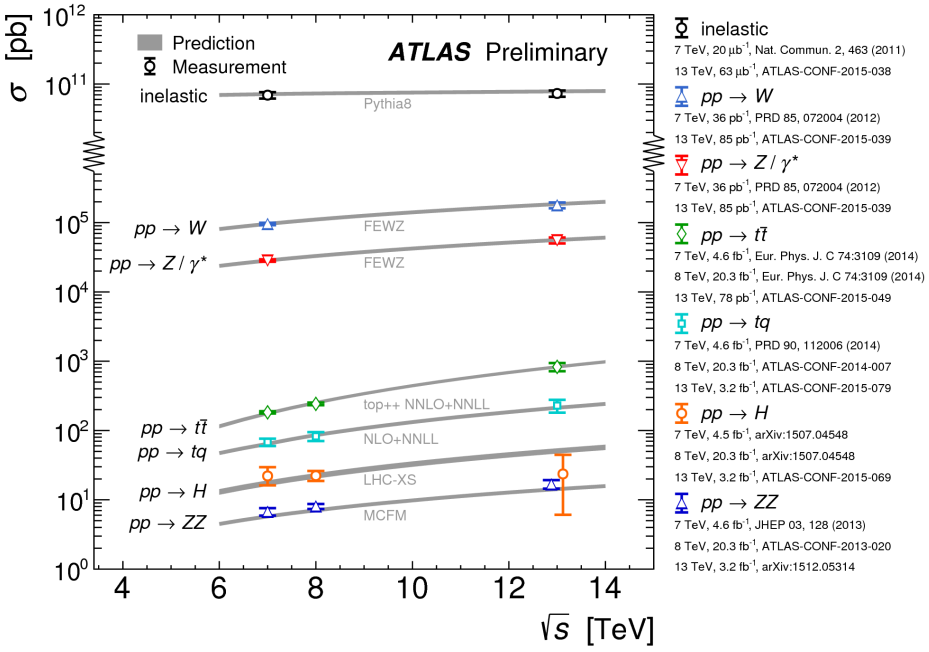


Figure 1.1: Summary of cross section measurements by ATLAS presented as a function of center-of-mass energy $\sqrt{s}$ from 7 TeV to 13 TeV for a few selected processes [1]. Jet production which has a higher cross section than W production is not shown here.

- H_{T},
- W boson p_{T},
- Leading jet p_{T} and rapidity,
- Second leading jet p_{T} and rapidity.

The measurement of the W boson p_{T} presented here extends to previously unmeasured energy scales. The results are compared to leading-order multi-leg predictions as well as to a next-to-leading order prediction for $W + 1$ jet. The measurement is performed in the decay channel $W \to e\nu$ only.

In view of the sensitivity of $W^{\pm}$ production to the content of the proton, the analysis is performed in addition for $W^+ +$ jets and $W^- +$ jets production separately. The relative $W^+ +$ jets vs. $W^- +$ jets cross sections are presented as $\frac{W^+ + \text{jets}}{W^- + \text{jets}}$ ratio. The selection of measured observables has been studied to be sensitive to the proton composition. In this way, this analysis aims at providing valuable input to constraints on the proton composition. The presence of jets here increases the energy scale needed for the production of the measured final state and thus probes the proton at higher values of the Bjorken x than accessible in inclusive $W^{\pm}$ production.

Overview

This thesis is the first publication of $W^{(\pm)} +$ jets cross sections measured at $\sqrt{s} = 8$ TeV with the ATLAS detector. The document is structured as follows.

The Standard Model, the computation of $W+$jets cross sections, the expected impact of the W^+/W^- ratio measurement on the knowledge of the proton composition and the used Monte

Carlo simulation are introduced in Chapter 2. Chapter 3 describes the LHC and the ATLAS detector. Particle reconstruction and selection in the $W+$jets analysis as well as the analysis strategy are presented in Chapter 4. The estimation of backgrounds from other processes which are selected in data and the comparison of data to the combined signal and background prediction is shown in Chapter 5. In Chapter 6, the unfolding procedure converting measured distributions in data to $W+$jets cross sections as well as a variety of tests of the performance of this procedure are described. The considered systematic uncertainties as well as their propagation through the unfolding procedure and the uncertainty handling in the W^+/W^- ratio are explained in Chapter 7. Most importantly, the measurement of the $W^{(\pm)}+$jets cross sections and the comparison to theory predictions as well as theory studies on the impact of the W^+/W^- measurement and the measurement of the W^+/W^- ratio are reported in Chapter 8 and Chapter 9, respectively. Finally, the thesis is concluded in Chapter 10.

Author's contribution

The work presented in this thesis has been realized within the ATLAS collaboration. It relies in several places, for example, the reconstruction, identification and calibration of the selected particles, the simulation of signal and background processes, etc. on ATLAS performance studies and measurements as well as the related ATLAS software. The collaboration effort is, of course, even more prominent in terms of the actual recording of the analyzed data from 2012 and the operation of the detector. Inputs, in particular from performance groups in ATLAS, providing the calibration and uncertainties of the selected particles, have been indicated in this thesis where possible.

The $W+$jets analysis in 2012 ATLAS data has been performed by the author as presented in this thesis. This includes the experimental measurement as well as the next-to-leading order predictions and the theory studies of the sensitivity to the proton composition discussed in the final result chapters. Further, smaller studies have been conducted by the author, but are not discussed in this thesis in the interest of focus. These studies include the optimization of the signal region definition regarding the isolation of the selected electron, the definition of the control region used in the multi-jet background estimation as well as the modelling of missing transverse momentum in the absence of jets.

Prior to the $W+$jets analysis, the author has contributed to a search for the direct production of charginos, neutralinos and staus in final states involving at least two hadronically decaying τ-leptons in 2012 ATLAS data [11]. She estimated the background from diboson production including experimental and theoretical systematic uncertainties.

Concerning the calibration of the ATLAS detector, the author has performed the precision synchronization of the Level-1 Calorimeter trigger in collision data from 2010, 2011 and 2012 [12]. The method for this calibration is based on fits of calorimeter trigger tower signals [13] and was delevoped to operation status by the author, starting from existing initial ideas.

CHAPTER 2

Theory

Understanding what the world is made of and how it can interact has been a major enterprise of humanity, dating back at least around 2400 years to the first atomic model in Ancient Greece. The modern version of this attempt lead to the development of the Standard Model (SM) of particle physics which summarizes our current knowledge of the fundamental particles and their interactions in a consistent theoretical framework. The Standard Model in its current form was formulated in the early/mid-1970s and has been subject to scientific scrutiny in various experiments since. So far, however, no unresolvable flaws were discovered. Instead, its predictions are confirmed by the discoveries of the W and Z^0 bosons in 1983 [14–17], the top-quark in 1995 [18, 19], the tau neutrino in 2000 [20] and, most recently, the Higgs boson in 2012 [21, 22]. Its continuing success has been honored with several Nobel prizes, the last one in 2013 to François Englert and Peter Higgs – *for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider* [23].

Astronomical observations of *dark matter* and the predominant presence of matter instead of anti-matter in the universe are examples where the Standard Model lacks an explanation. They are thus strong arguments to believe that the Standard Model is incomplete. In addition to direct searches for *something new*, precise measurements of SM processes are meant to reveal the point where its description breaks down.

This thesis presents the measurement of the W + jets production. This process is well described by the Standard Model and allows to deepen the knowlegde on SM processes and predictions, contributing indirectly to the attempt to find the 'point of failure'.

In the following, the Standard Model will be briefly summarized in Section 2.1 and the cross section determination of W + jets production will be exemplary described in Section 2.2. Section 2.3 presents an introduction how the measurement of W + jets can help to gain insights into the structure of protons. Finally, Section 2.4 gives a short overview of the simulation of W + jets production in Monte Carlo generators, including an explanation of the Monte Carlo generators used in this analysis.

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is a relativistic quantum field theory which describes the elementary particles of matter and their interactions by the strong, weak and electromagnetic forces via the symmetry groups $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$. Gravity – the fourth known

force – is not included in the framework, but also does not play a role on the short distances of particle interactions.

The Standard Model groups particles into fermions, gauge bosons and the Higgs boson according to their intrinsic angular momentum – the spin – with values $1/2$, 1 and 0, respectively. An overview of the particle content in the Standard Model is shown in Figure 2.1.

Fermions are the building blocks of matter and are further distinguished into quarks (q) and leptons (ℓ), both arranged in three *generations* of two particles each. The quark doublets consist of an *up-type* and a *down-type* quark, carrying an electric charge of $+2/3$ and $-1/3$ in units of the electron charge, respectively. In the order of increasing mass, the quarks are: up (u), down (d), strange (s), charm (c), bottom (b) and top (t), where the masses range from a few MeV up to 173 GeV. In addition to the electric charge, quarks carry weak and strong charges, where the latter is also called the colour charge and comes in the values *red*, *blue* and *green*. Quarks form composite states, called *hadrons*, of either three quarks (*baryon*), or of a quark and its *anti-quark* ($\bar{q}$) (*meson*), since only colour-neutral objects can exist freely. *Anti-particles* exist for every fermion and have the same mass, but opposite-sign electric and, in the case of quarks, also inverse colour charge. The quark-combinations *uud* and *udd* constitute the proton and the neutron, respectively.

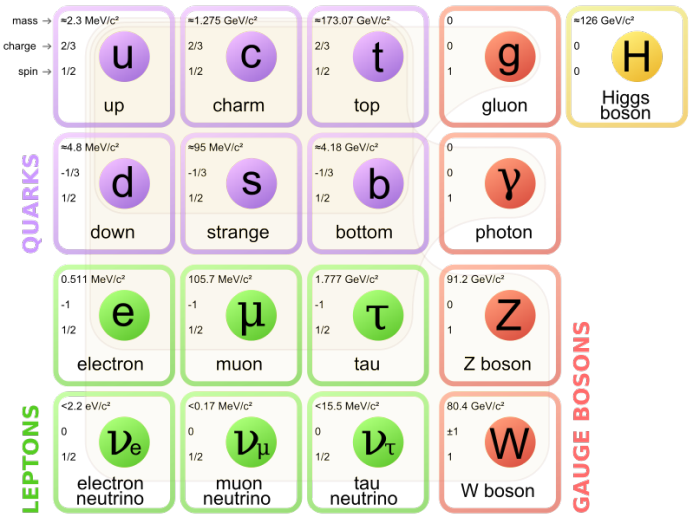


Figure 2.1: Particle content of the Standard Model. The light brown loops mark which bosons (red) couple to which fermions (purple and green) [24].

Leptons are arranged in left-handed $\mathrm{SU(2)_L}$-doublets of one negatively-charged lepton and one electrically neutral neutrino, or right-handed weak isospin singlets. The latter consist only of the charged lepton. While quarks interact via the strong interaction, leptons do not and therefore exist freely in nature. The charged leptons are: electron (e^-), muon (μ^-) and tau lepton (τ^-), with anti-particles e^+, μ^+ and τ^+, and masses of 511 keV, 106 MeV and 1.78 GeV, respectively. Their electric charge is either -1 or $+1$. The corresponding neutrinos are denoted as electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). Though originally conceived as massless, they have been proven by neutrino-oscillation experiments to possess a non-zero mass. The exact values are yet to be determined and only upper bounds exist, but they already demonstrate that the neutrino mass-scale is far below the masses of the other SM particles. As a result of their zero electric charge, neutrinos cannot interact electromagnetically and are sensitive only to the weak force.

The strong, weak and electromagnetic interactions are described in the Standard Model as the exchange of mediator particles – the gauge bosons. The gauge bosons result from the requirement of the local gauge invariance of the Lagrangian density $\mathcal{L}$, where local gauge changes of fermion *spinors* ($\Psi(x)$) are counterbalanced by an appropriate co-transformation of the boson gauge. The best-known gauge boson is the photon which is the mediator of the electromagnetic interaction as formulated in *quantum electrodynamics* (QED). The photon is

massless – giving the electromagnetic interaction an infinite reach – and couples to the electric charge of particles. The weak interaction is mediated by three massive gauge bosons, $W^{\pm}$ and Z^0 which couple to the weak charge of the particles. Although the coupling constant $\alpha_W \sim 1/30$ is in fact larger than the electromagnetic coupling $\alpha = 1/137$, the interaction appears weaker due to the heavy masses of its propagating particles of $m_W = 80.4\,\text{GeV}$ and $m_Z = 91.2\,\text{GeV}$ [25]. The weak interaction has been unified with the electromagnetic interaction to the *electroweak* interaction by Glashow, Salam and Weinberg in the 1960s.

The theory of strong interactions is called *quantum chromodynamics* (QCD) and its propagators are eight gluons. Gluons are massless and carry colour and anti-colour, meaning they can self-interact. While the coupling constant of the strong interaction is the largest of the coupling constants $\alpha_S \sim 1$ (at energy scales close to the proton mass), the self-coupling of gluons limits the interaction to short distances. Two characteristics of QCD are furthermore related to the gluon self-coupling and the strong coupling α_S: *confinement* and *asymptotic freedom*. Confinement describes the phenomenon that upon growing separation between two coloured particles the energy stored in the field between them increases linearly with the distance. This continues until it becomes energetically favourable to create a new $q\bar{q}$ pair out of the field energy. These then form again color-neutral states with the originally to-be-separated pair, thus ruining the attempt to obtain single quarks. Asymptotic freedom is a consequence of the fact that α_S is energy dependent in a particular way. While α_S is large (~ 1) at around 1 GeV, it reduces as the energy scale of the interaction increases. At the electroweak scale of 100 GeV which is easily reached by modern particle colliders, α_S reduces to about 0.1. This allows the quarks and gluons inside a high-energetic proton to undergo interactions as quasi-free – a major prerequisite for cross section calculations in a perturbative approach.

The Brout-Englert-Higgs mechanism describes that particles acquire mass through the interaction with the universally permeating Higgs field. The mechanism explains in particular the mass of the $W^{\pm}$ and Z^0 bosons, while the photon which is combined with the three heavy gauge bosons in the electroweak theory can remain massless. This is achieved by *spontaneous breaking* of the local $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ gauge symmetry which generates mass terms for $W^{\pm}$ and Z^0 bosons from a non-zero vacuum expectation value of the Higgs field. Additionally, mathematical terms for a massive scalar particle emerge – the Higgs boson – whose mass scale is set by the same vacuum expectation value. It is therefore no surprise that the mass of the recently discovered Higgs boson was measured as approximately 125 GeV, i.e. close to the masses of the weak gauge bosons.

The Standard Model comprises many interesting aspects, but only two are explained in the following in more detail, as they are of particular relevance for this thesis. These are the charged weak interaction, i.e. interactions involving the $W^{\pm}$ boson, as well as strong interactions, in particular the emission of QCD radiations. These are essential in understanding the structure of the proton and the generation of jets.

2.1.1 Charged weak interactions

Charged weak interactions are described in the Standard Model as the exchange and the interaction of particles with a W boson. Without going into the details of electroweak unification, many features of the interaction, including the understanding of W production cross sections, can be obtained in analogy to QED, while at the same time demonstrating the important contrasts to QED. Introductions to charged weak interactions can be found in many textbooks, while the overview in the following mainly draws on explanations in Refs. [26, 27].

The weak current describes the interaction of a W^- with an electron and an electron anti-neutrino, or of a W^+ with a positron (e^+) and an electron neutrino, as in the case of the W decay – exemplary shown in Figure 2.2. The W^- current, for example, can be expressed in terms of the Lorentz four-vector J^μ as:

$$J^\mu \propto \bar{\psi}_e \Gamma^\mu \psi_\nu$$

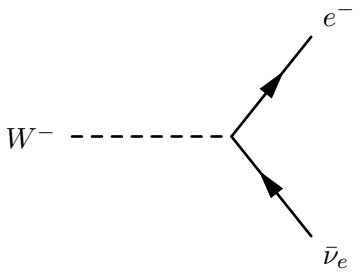


Figure 2.2: W^- boson coupling to electron and anti-electron neutrino as in the leptonic W boson decay.

where ψ_e and ψ_ν correspond to the four-component Dirac spinors, representing the propagating electron and neutrino. Γ^μ specifies the properties of the interaction vertex. In QED, it corresponds to the Dirac γ-matrices, γ^μ, and results in a parity conserving interaction.

From experiments like the measurement of nuclear β-decay of polarised Co^{60} by Wu et al. [28] or the measurement of the charged pion leptonic decay rates [29], it is known that the weak interaction is (maximally) parity violating by means of a vector-minus-axial-vector ($V-A$) structure. Requiring Lorentz invariance of the interaction current in addition to the $V-A$ type restricts the options of defining Γ^μ considerably. Γ^μ is therefore constructed from the Dirac γ-matrices as $\frac{1}{2}\gamma^\mu(1-\gamma^5)$ with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The charged weak current can then be written as:

$$J^\mu = \frac{g_W}{\sqrt{2}}\bar{\psi}_e\frac{1}{2}\gamma^\mu(1-\gamma^5)\psi_\nu = \frac{g_W}{\sqrt{2}}\cdot\frac{1}{2}\left[\bar{\psi}_e\gamma^\mu\psi_\nu - \bar{\psi}_e\gamma^\mu\gamma^5\psi_\nu\right] ,$$

where $-i\frac{g_W}{\sqrt{2}}\frac{1}{2}\gamma^\mu(1-\gamma^5)$ is identified with the weak interaction vertex and g_W denotes the weak coupling constant.

Under parity transformations, the two terms in the second formulation transform as:

$$\bar{\psi}\gamma^\mu\psi = \begin{pmatrix}\bar{\psi}\gamma^0\psi\\ \bar{\psi}\gamma^i\psi\end{pmatrix} \xrightarrow{P} \begin{pmatrix}+\bar{\psi}\gamma^0\psi\\ -\bar{\psi}\gamma^i\psi\end{pmatrix} \quad\text{and}\quad \bar{\psi}\gamma^\mu\gamma^5\psi = \begin{pmatrix}\bar{\psi}\gamma^0\gamma^5\psi\\ \bar{\psi}\gamma^i\gamma^5\psi\end{pmatrix} \xrightarrow{P} \begin{pmatrix}+\bar{\psi}\gamma^0\gamma^5\psi\\ +\bar{\psi}\gamma^i\gamma^5\psi\end{pmatrix} ,$$

i.e. like a four-vector and an axial vector, respectively – thus fulfilling the $V-A$ requirement for the charged weak current.

The term $\frac{1}{2}(1-\gamma^5)$ and its counter-part $\frac{1}{2}(1+\gamma^5)$ also have the properties of projection operators, and are denoted as *chiral* operators P_L and P_R. Projections have to fulfil $P_{L/R}^2 = P_{L/R}$, $P_L + P_R = 1$ and $P_L P_R = 0$ which is true for the terms above. The charged weak current can therefore be understood as acting only on the left-handed spinor components $\psi_L = P_L\psi = \frac{1}{2}\gamma^\mu(1-\gamma^5)\psi$ of particles and the right-handed spinor components of anti-particles.

Using the common notation with particle spinors $\psi(\mathbf{x},t) = u(E,\mathbf{p})\cdot e^{-i(Et-\mathbf{x}\cdot\mathbf{p})}$ and anti-particle spinors $\psi(\mathbf{x},t) = v(E,\mathbf{p})\cdot e^{+i(Et-\mathbf{x}\cdot\mathbf{p})}$, the charged current for the W^- decay can be written as:

$$J^{\mu,-} \propto \bar{u}_{e,L}\gamma^\mu v_{\nu,R} = \bar{u}_{e,L}\gamma^\mu u_{\nu,L} ,$$

where the outgoing anti-neutrino has been replaced by an incoming neutrino.

The W boson is a spin-1 particle and is described in terms of a plane wave with polarization four-vector: $W^\mu = \epsilon^\mu_\lambda \cdot e^{-i(Et-\mathbf{x}\cdot\mathbf{p})}$. In contrast to the photon, the W boson is massive and thus possesses three orthogonal polarization states λ. If the W boson is travelling in z-direction, they can be written as: $\epsilon^\mu_\pm = \mp\frac{1}{\sqrt{2}}(0,1,\pm i,0)$ for the transverse polarization states and $\epsilon^\mu_L = \frac{1}{m_W}(p_z,0,0,E)$ for the longitudinal polarization, with $\epsilon_\mu p^\mu = 0$ as required for

polarization states. Assuming that the polarization of the W boson is unknown which is the typical situation if the W boson is produced in a hard interaction at a collider, the polarization states need to be summed over. It can be shown that in this case the completeness relation holds:

$$\sum_{\lambda} \epsilon_{\lambda}^{*\mu} \epsilon_{\lambda}^{\nu} = -g^{\mu\nu} + \frac{q^{\mu} q^{\nu}}{m_W^2} \,,$$

where q^{μ} is the four-momentum vector of the propagating W boson and $g^{\mu\nu}$ is the Minkowski space-time metric. The weak interaction as it is mediated by a virtual W boson is therefore described by the propagation term:

$$\frac{-i\left(g^{\mu\nu} + \frac{q^{\mu}q^{\nu}}{m_W^2}\right)}{q^2 - m_W^2} \,,$$

or, if the W boson is considered to decay within a finite lifetime:

$$\frac{-i\left(g^{\mu\nu} + \frac{q^{\mu}q^{\nu}}{m_W^2}\right)}{q^2 - m_W^2 + i m_W \Gamma_W} \,.$$

In the electroweak unification, the Lagrangian density $\mathcal{L}$ from which the equations of motion are derived by the Euler-Lagrange equations is required to be invariant under local SU(2) phase transformations. The generators of the SU(2) group can be written in terms of the Pauli spin matrices $\mathbf{T} = \frac{1}{2}\boldsymbol{\sigma}$ and the electron and neutrino or up- and down-type quarks can be arranged in weak isospin doublets:

$$\varphi_e(x) = \begin{pmatrix} \psi_{\nu}(x) \\ \psi_e(x) \end{pmatrix}_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \; \varphi_q(x) = \begin{pmatrix} \psi_u(x) \\ \psi_{d'}(x) \end{pmatrix}_L = \begin{pmatrix} u_L \\ d'_L \end{pmatrix} \,,$$

where the down-type quark includes the effect of the CKM matrix in order to relate weak eigenstates to the mass eigenstates of the down-type quarks.

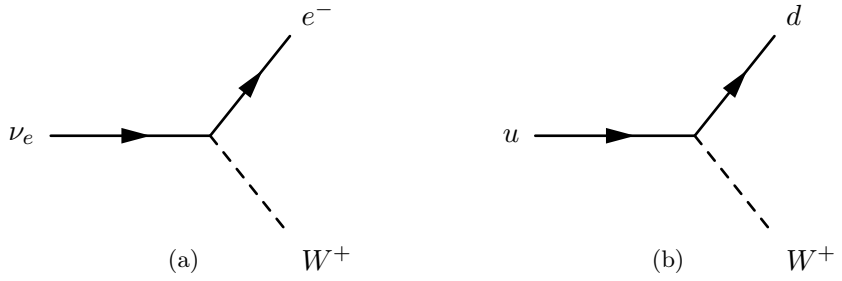


Figure 2.3: W^+ emission from a neutrino which transforms in the interaction into an electron (a) or from an u-quark which transforms into a d-quark according to the V_{ud} entry of the CKM matrix (b).

The emission of a W^+ boson as shown in Figure 2.3 from a neutrino is described by the following *charge-lowering* weak current:

$$J^{\mu} = \frac{g_W}{\sqrt{2}} \bar{e} \frac{1}{2} \gamma^{\mu}(1 - \gamma^5)\nu = \frac{g_W}{\sqrt{2}} \bar{e}_L \gamma^{\mu} \nu_L = \frac{g_W}{\sqrt{2}} \begin{pmatrix} \bar{\nu}_L & \bar{e}_L \end{pmatrix} \gamma^{\mu} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} = \frac{g_W}{\sqrt{2}} \bar{\varphi}_L \gamma^{\mu} \sigma_- \varphi_L$$

where $\sigma_{\pm}$ are weak isospin raising/lowering operators and are constructed from the group generators as $\sigma_{\pm} = \frac{1}{2}(\sigma_1 \pm i\sigma_2)$. The description is analogous for the emission of W^+ from a u-quark, but introduces the CKM matrix in addition. The current above is equivalent to

the current for the W^- decay, in which case neutrino and W switch sides and convert into their respective anti-particles. The physical W bosons are then given as linear combinations $W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} \mp iW_\mu^{(2)}\right)$ of the first two SU(2) gauge fields $W_\mu^{(k)}$.

Without going into further detail on the inclusion of the neutral currents mediated by the Z^0 boson and the photon which are represented by a combination of $W_\mu^{(3)}$ and an additional $\mathrm{U}(1)_Y$ weak hypercharge field B_μ, the full electro-weak Lagrangian can be written as:

$$\mathcal{L} = \bar{\varphi}_L\gamma^\mu\left[i\partial_\mu - \frac{g}{2}\overrightarrow{\sigma}\cdot\overrightarrow{W_\mu} - g'\frac{Y_L}{2}B_\mu\right]\varphi_L + \bar{\psi}_R\gamma^\mu\left[i\partial_\mu - g'\frac{Y_R}{2}B_\mu\right]\psi_R - \frac{1}{4}\overrightarrow{W_{\mu\nu}}\overrightarrow{W^{\mu\nu}} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,.$$

g and $g'/2$ denote the electroweak coupling strengths of the $W_\mu^{(k)}$ and B_μ fields, and $Y_{L/R}$ indicates the weak hypercharge for left/right-handed particles. The kinetic energy terms for the fields contain the definitions $\overrightarrow{W_{\mu\nu}} = \partial_\mu\overrightarrow{W_\nu} - \partial_\nu\overrightarrow{W_\mu} - g\overrightarrow{W_\mu}\times\overrightarrow{W_\nu}$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. The interaction of the right-handed spinor components ψ_R remains in the form of singlets and occurs only with the hypercharge field B_μ. Furthermore, the Lagrangian does not contain mass terms neither for the gauge bosons nor for the fermions which in the electroweak unification can only be introduced in a gauge invariant way via the Brout-Englert-Higgs mechanism.

2.1.2 Strong interactions

The theory of strong interactions (QCD) is based on the requirement of invariance of the Lagrangian density under local SU(3) gauge transformations. This necessitates the introduction of gauge fields which compensate the respective spinor transformations. For QCD, eight fields $G_\mu^a(x)$ with $a = 1, ..., 8$ are introduced – one for each of the eight generators $T^a = \frac{1}{2}\lambda^a$ of the SU(3) symmetry group. The λ^a are the 3×3 dimensional Gell-Mann matrices. The quark spinors are then arranged in three component vectors for the three color states *red*, *green*, *blue* to match the λ^a-dimensionality: $\psi^T = (\psi_r, \psi_g, \psi_b)$, or alternatively in terms of the color basis: $c_1 = r = (1,0,0)$, $c_2 = g = (0,1,0)$ and $c_3 = b = (0,0,1)$ as $\psi \propto c_i u(p)$.

The quark current for a QCD interaction vertex shown in Figure 2.4 is written as:

$$J_{ij}^{\mu,a} = \bar{u}(p_3)c_j^\dagger\left[\frac{g_S}{2}\lambda^a\gamma^\mu\right]c_i u(p_1) = \frac{g_S}{2}\lambda_{ji}^a\left[\bar{u}(p_3)\gamma^\mu u(p_1)\right]\,.$$

The term $-i\frac{g_S}{2}\lambda^a\gamma^\mu$ or $-i\frac{g_S}{2}\lambda_{ji}^a\gamma^\mu$ specifies the QCD vertex factor with the strong coupling constant g_S and $\lambda_{ji}^a = c_j^\dagger\lambda^a c_i$ as the ji-th element of λ^a. The second formulation is possible because the colour part of the current factorises from the remainder of the current. This is a general property of the colour part of the interaction, allowing it to be separated from the spinor-and-γ-matrices-based terms in the calculation of process cross sections. It is denoted as *colour factor* of the respective diagram or interaction.

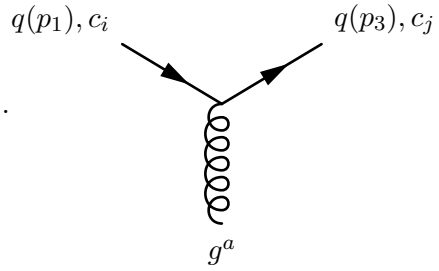


Figure 2.4: QCD quark-gluon vertex with incoming quark of momentum p_1 and color i, outgoing quark with momentum p_3 and color j as well as the emitted gluon of type a.

The dimensionless variable $\alpha_S = \frac{g_S^2}{4\pi}$ is commonly used to define the QCD coupling strength. When calculating the value of the coupling strength for comparison with measurements, loop corrections to the gluon propagation need to be taken into account as shown in Figure 2.5. These need to be summed up to all orders, i.e. considering from 1 to an infinite number of loops. Applying the method of *renormalization*, this can be absorbed in the definition of α_S instead of modifying the gluon propagator term itself, which effectively modifies the coupling

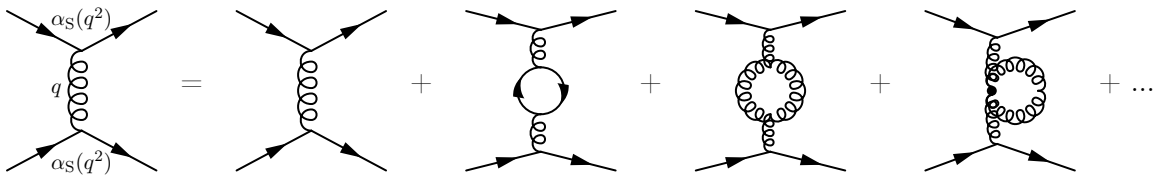


Figure 2.5: Renormalization of loop terms in the gluon propagator into the strong coupling constant, leading to the running of α_S, analog to [27].

strength. The loops in the propagator introduce terms dependent on q^2 – the momentum of the propagating gluon – which means that the coupling becomes energy scale dependent. Although this procedure gives an improved description of the coupling and thus the strong current, it does not remove the need to consider loop-contributions again when computing terms on the order of α_S^2 or higher.

Given a reference energy μ^2, the strong coupling constant after renormalization becomes:

$$\alpha_S(q^2) = \frac{\alpha_S(\mu^2)}{1 + \frac{11n_c - 2n_f}{12\pi}\,\alpha_S(\mu^2)\ln\left(\frac{q^2}{\mu^2}\right)},$$

where $n_c = 3$ is the number of colours and $n_f \leq 6$ the number of quark flavours. The $b_0 = \frac{11n_c - 2n_f}{12\pi}$ is then positive, leading to a reduction of α_S with increasing q^2. At energies equal to the mass of the Z^0 boson, i.e. $q^2 = m_Z^2$, the strong coupling has decreased to $\alpha_S(m_Z^2) = 0.1185 \pm 0.0006$ [25] which is sufficiently small for perturbative methods to be applicable. The energy dependence of α_S is also called *running of* α_S and explains the *asymptotic freedom* of quarks and gluons in the proton – mentioned above as one of the key features of the strong interaction.

Energy scales below about 1 GeV where α_S is large are often referred to as *non-perturbative regime*, while energy scales above are termed *perturbative regime* and allow the use of perturbative computations of QCD (pQCD). This corresponds to the distinction between *soft*, *long-distance* and *hard*, *short-distance* physics. Defining Λ_{QCD} as the scale where $\alpha_S(Q^2)$ formally diverges – with $Q^2 = -q^2$, it specifies the boundary line between perturbative and non-perturbative descriptions and allows α_S to be rewritten as:

$$\alpha_S(Q^2) = \frac{1}{b_0 \ln\left(Q^2/\Lambda_{\mathrm{QCD}}^2\right)}.$$

Assuming that the scale of the hard interaction Q is much larger than $\Lambda_{\mathrm{QCD}} \sim 250\,\mathrm{MeV}$, the description of the interaction factorizes into terms for the hard and the soft interaction components. This is also called *factorization theorem* and is the basis of any process predictions for proton colliders as the LHC (described in Section 3.1).

QCD radiation and the content of the proton

In addition to high-energetic interactions involving quarks and gluons, a sound understanding of QCD radiation is of relevance for the prediction of QCD behaviour at the LHC.

In the presence of QCD radiation, a quark with momentum x, given as a fraction of the proton momentum, could have originated from a quark with a higher momentum fraction $y > x$ which has radiated a gluon. This radiation reduced the momentum fraction of the

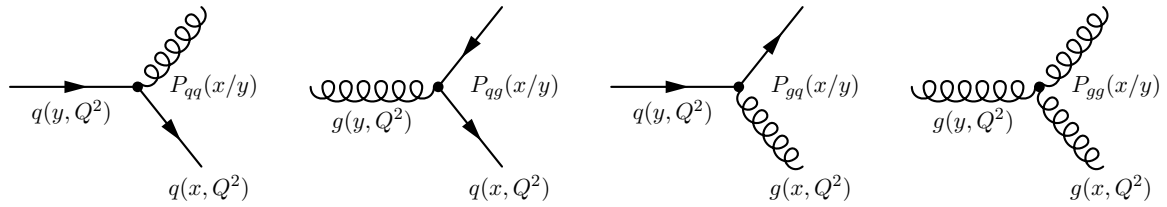


Figure 2.6: Possible QCD radiations from quarks or gluons with original momentum fraction y, post-radiation momentum fraction x and splitting kernels $P_{pp}(x/y)$ with $p = q$ or g.

quark from y to x where the quark is encountered. The splitting, i.e. the radiation from the original quark is described by the so-called splitting kernel $P_{q\leftarrow q}$, or often denoted as P_{qq}. Assuming that there is a certain probability to encounter a quark with momentum fraction x, namely $q(x)$, this probability can be described as a function of the probability to find the former quark $q(y)$ and the splitting kernel P_{qq}. In QCD, this radiation process is however not the only option to obtain a quark with momentum fraction x. It could also have come from a gluon with momentum fraction $y > x$ which split into a $q\bar{q}$ pair according to the splitting kernel $P_{q\leftarrow g} = P_{qg}$. Denoting the probability for encountering such a gluon as $g(y)$, the same arguments as above can be applied to the gluon. Given a gluon with momentum fraction x, $g(x)$ is described by the probability to find a quark with higher momentum fraction $q(y)$ which radiated the gluon according to P_{gq} and, due to gluon self-coupling, the probability to encounter a higher momentum gluon $g(y)$ which split into two gluons according to P_{gg}. The different types of radiations with the related splitting kernels are shown in Figure 2.6. Without further derivation, the splitting kernels can be written as:

$$
\begin{aligned}
P_{qq}(z) &= C_F \left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z) \right] \\
P_{qg}(z) &= T_R \left[z^2 + (1-z)^2 \right] \\
P_{gq}(z) &= C_F \frac{1+(1-z)^2}{z} \\
P_{gg}(z) &= 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \frac{11C_A - 4n_f T_R}{6}\delta(1-z)
\end{aligned}
$$

with $z = x/y$ being the momentum ratio of the quark or gluon after vs. before the radiation. C_F, T_R and C_A here are the QCD colour factors with values $4/3$, $1/2$ and 3, respectively, and the '+' subscript indicates a specific way to regularize the divergence $z \to 1$. The δ-terms correspond to contributions from virtual gluon diagrams around the emission vertex. Since C_A is the largest, gluon-splittings into gluons occur most frequently and lead to higher number of particles in a splitting cascade if it started from a gluon instead of a quark.

If changes in the probabilities $q(x)$ or $g(x)$ depend on the splittings, the question is what determines the sensitivity to such changes, i.e. as a function of which parameter are the changes observed. The answer here is the energy scale of the hard interaction Q^2. In a high-energetic interaction, the content of the proton can be resolved in more detail than in a lower-energetic interaction. More structure, i.e. possible QCD radiations are visible with higher Q^2 by means that they are not viewed as part of the interacting quark or gluon. The quark and gluon probabilities are thus Q^2 dependent: $q(x) \to q(x,Q^2)$, $g(x) \to g(x,Q^2)$. The explained relations are combined in the so-called DGLAP[1] equations [30–32]:

[1] Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

$$\frac{dq_i(x,Q^2)}{d\log Q^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dy}{y}\left[q_i(y,Q^2)P_{qq}(x/y) + g(y,Q^2)P_{qg}(x/y)\right]$$
$$\frac{dg(x,Q^2)}{d\log Q^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dy}{y}\left[\sum_i q_i(y,Q^2)P_{gq}(x/y) + g(y,Q^2)P_{gg}(x/y)\right]$$

where i indicates the different quark flavours. The momentum fraction x is defined within 0 and 1 and y is within x and 1.

Recurring radiations inside the proton mean that the naive picture of the proton as consisting of two u-quarks and one d-quark is over-simplified. Instead, gluons carry a fraction of around 50 % of the proton momentum as measured in deep inelastic scattering experiments. Via $q\bar{q}$-splittings, the gluons lead furthermore to the presence of $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$ and $b\bar{b}$ quarks in the proton where the latter two are heavily suppressed by their larger mass, but have to be taken into account for some processes at the energies of the LHC. These $q\bar{q}$ pairs are called *sea quarks*.

If $u(x)$ and $d(x)$ are defined as the probability density functions to find any up- or down-quark independent of its origin in the proton, then $u_v = u - \bar{u}$ and $d_v = d - \bar{d}$ are called the *valence quark* distributions and represent the original, naive uud-composition of the proton. Their DGLAP equation consists only of the first term in the upper equation, since they cannot originate from a gluon splitting. They are thus decoupled from the gluon and sea quark evolution. Both quarks and gluons – being proton constituents – are referred to as partons. The distributions $q_i(x)$ and $g(x)$ are therefore named *parton distribution functions* (PDFs).

The DGLAP equations describe the evolution of the PDFs from a starting scale Q_0^2 to a higher energy scale Q^2. They however do not state the dependence of the PDFs on the momentum fraction x. This cannot be deduced from QCD calculations, but is only accessible from experimental data. Measurements of fixed-target deep inelastic scattering, including neutrino-nucleon scattering, as well as ep collisions at HERA, $p\bar{p}$ collisions at the Tevatron and now pp collision data from the LHC are needed to constrain the PDF x-dependence. Often a parameterization in x is assumed whose free parameters are determined in a global PDF fit to data of a large selection of measurements. The following parameterization was for example used for the valence quark PDFs in the global fit resulting in the CT10 PDF set, at the initial scale $Q_0 = 1.3\,\text{GeV}$:

$$q_v(x,Q_0^2) = a_0 x^{a_1}(1-x)^{a_2}\exp\left(a_3 x + a_4 x^2 + a_5\sqrt{x}\right)$$

where $a_0, ..., a_5$ are the fitted parameters [33].

Results for the proton PDFs in such global fits can however differ between collaborations depending on the included data sets, the parametetrization and/or the fitting procedure as well as the order in α_S to which the splitting kernels are computed. The given splitting kernels correspond to the leading order (LO) DGLAP evolution, i.e. they contain only the lowest possible number of QCD vertices.

Although the splitting kernels are discussed here in the context of PDFs, they are in fact universal and are used to describe QCD radiation before and after the hard interaction which is not considered as part of the hard interaction, but also not as part of the proton. This will come up in the cross section computation for W + jets production at the LHC.

2.2 W + jets cross section in the Standard Model

W bosons have been discovered in 1983 by the UA1 and UA2 collaborations [14, 15]. At the LHC, W bosons are produced in approximately 10 % of the cases together with at least one emitted quark or gluon. These partons radiate due to their colour charge and eventually form colour-neutral hadrons – a process denoted as *fragmentation* and *hadronization* and explained in more detail later. The arising collimated sprays of hadrons are called *jets*. W boson production in association with jets provides an ideal testing ground for pQCD predictions which describe the frequencies of additional quark or gluon production and radiation.

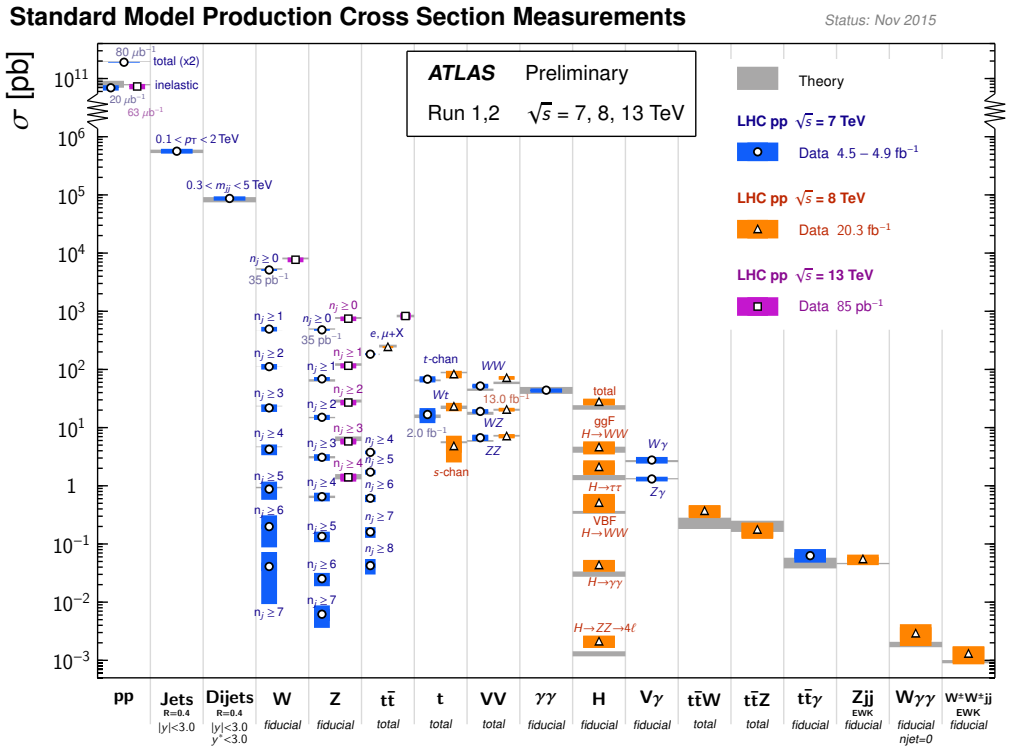


Figure 2.7: Summary of several Standard Model total and fiducial production cross section measurements using ATLAS 2010, 2011, 2012 and 2015 data-sets, corrected for leptonic branching fractions and compared to the corresponding theoretical expectations [1]. The integrated luminosity of the measurements are indicated if they differ from the total data set of the corresponding year.

At a center-of-mass energy of the colliding protons of $\sqrt{s} = 8\,\text{TeV}$ at the LHC, the production cross section of W bosons with subsequent leptonic decay is about 12 nb [2]. It is the process with the second largest production cross section at the LHC after the di-jet production. This can be seen in Figure 2.7 which shows a variety of SM cross sections measured by the ATLAS experiment so far. The production of jets as the signature of emitted quarks or gluons is suppressed by the strong coupling α_S. This reduces the production cross section for each jet by roughly 0.5 to 1 order of magnitude and results in W $+n$ jet production reaching from one of the highest cross sections down to cross sections of the order of Higgs production in one of the channels where it was discovered, i.e. as $H \to \gamma\gamma$.

The computation of W + jets production, involving the theories of electroweak and strong interactions, is highly non-trivial. Only recently, the first W +1 jet computation at next-to-next-to-leading order (NNLO) in α_S has become available [34]. Next-to-leading order

(NLO) calculations exist for W with up to 5 partons, however, so far cannot be matched to simulations of additional QCD radiation needed to describe the formation of jets [35, 36]. The latter is so far only possible at leading order (LO) in α_S with up to 5 partons in the final state. Above these parton multiplicities, predictions are extremely computing intensive and time consuming.

To get an idea of the complexity of W + jets cross section predictions, the example of W production in association with 1 jet at LO, generated from an incoming u-quark and a gluon is discussed in the following. This subprocess represents the largest contribution to the production of $W^{\pm}$ +1 jet whose measurement is the focus of this thesis.

The cross section σ of a particle interaction can be computed as the ratio of the transition rate Γ_{fi} between initial state i and final state f and the incoming particle flux Φ_{in}, i.e. $\sigma = \Gamma_{fi}/\Phi_{\text{in}}$. Following Fermi's golden rule, the transition rate is given by the transition matrix element $\mathcal{M}_{fi}$ squared and the integral over the final state phase space. For a $2 \to 2$ process with a and b being the incoming and 1 and 2 the outgoing particles, this is given in Lorentz-invariant form as:

$$\Gamma_{fi} = \frac{(2\pi)^4}{4E_a E_b} \int |\mathcal{M}_{fi}|^2 \delta^{(4)}(p_a + p_b - p_1 - p_2) \cdot \frac{d^3\mathbf{p}_1}{2E_1(2\pi)^3} \frac{d^3\mathbf{p}_2}{2E_2(2\pi)^3}$$

where the $p_{a,b,1,2}$ are the four-momentum vectors of the incoming and outgoing particles and $\mathbf{p}_1$ and $\mathbf{p}_2$ are the three-vector momenta of the outgoing particles (e.g. $p_a = (E_a, \mathbf{p}_a)$). The four-dimensional δ-function ensures both energy and momentum conservation. Integrating over the two-particle phase space in the center-of-mass frame of the collision, with p_i and p_f being the absolute values of incoming and outgoing momenta $p_i = |\mathbf{p}_a| = |\mathbf{p}_b|$ and $p_f = |\mathbf{p}_1| = |\mathbf{p}_2|$, the total and differential cross sections are:

$$\sigma = \frac{1}{64\pi^2 s} \frac{p_f}{p_i} \int |\mathcal{M}_{fi}|^2 \, d\Omega \qquad \text{and} \qquad \frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 s} \frac{p_f}{p_i} |\mathcal{M}_{fi}|^2 \,.$$

So far, the cross section definition is universal, mainly depending on the number of final state particles in the phase space integral, whose complexity, of course, increases with the number of final state particles. The process specific information is contained in the matrix element $\mathcal{M}_{fi}$.

Consider now the specific process of a u-quark and a gluon producing a W^+ boson and an outgoing d-quark as shown in Figure 2.8. The matrix element is composed of terms for all in- and outgoing particles, as well as the vertices and the propagating quark. Following the discussion above on the theory of weak and strong interactions, it is clear that the vertex terms are $-i\frac{g_W}{\sqrt{2}} V_{ud}^* \gamma^\mu \frac{1}{2}(1-\gamma^5)$ and $-i\frac{g_S}{2}\lambda^a \gamma^\mu$, respectively. The first term includes the respective CKM matrix element describing the probability of the u-quark being converted into a d-quark. Also the terms of the incoming and outgoing quarks are relatively simple and are given by their spinors together with the colour charge, i.e. $c_i u(p_a)$ and $c_j^\dagger \bar{u}(p_2)$.

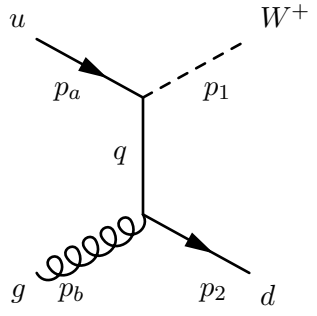


Figure 2.8: W^+ production in association with an outgoing d-quark from an incoming u-quark and a gluon. Incoming momenta are p_a and p_b, while the outgoing momenta are p_1 and p_2. q is the momentum of the propagating quark.

The W boson and similarly the gluon are described by their polarization vectors. The incoming gluon is denoted by ϵ_λ^μ with $\lambda = \pm 1$ for the two transverse polarization states of a massless spin-1 boson and the outgoing W is $\epsilon_\kappa^{\mu\,*}$ with polarization states $\kappa = \pm 1$ or 0. It

remains only the propagator of the quark, i.e. a spin-$\frac{1}{2}$ particle. According to the Feynman rules, this is given by $\frac{i(\gamma_\mu p^\mu + m)}{p^2 - m^2}$. It can be determined by solving the Dirac equation for its Green's function which describes the effect of a propagation of a particle probability wave produced at some point in space to another point.

As mentioned before, the colour sum for the participating quark and gluon flavours can be separated from the rest of the matrix element and runs over possible colour configurations and is averaged over the incoming colour possibilities. This gives a colour factor $f_c = \langle C^* C \rangle = 1/6$ in this example, with $C = \sum_{i,j} \sum_a \frac{1}{2} c_j^\dagger \lambda^a c_i$.

Combining the terms as $-i\mathcal{M}$=(gluon to d-quark current)×(propagator)×(u-quark to W boson current), the color averaged squared matrix element for the process in Figure 2.8 is:

$$\langle |\mathcal{M}|^2 \rangle = f_c \cdot \left| \bar{u}_s(p_2) \left(-i \frac{g_S}{2} \gamma^\nu \right) \epsilon_\nu^\lambda(p_b) \cdot \frac{i(\gamma_\rho q^\rho + m)}{q^2 - m^2} \cdot \epsilon_\mu^{\kappa\,*}(p_1) \left(-i \frac{g_W}{\sqrt{2}} V_{ud}^* \gamma^\mu \frac{1}{2}(1 - \gamma^5) \right) u_r(p_a) \right|^2 .$$

The discussed diagram corresponds to the t-channel version of this process, while also a s-channel diagram exists. The t and s in this context are the Mandelstam variables $t = (p_a - p_1)^2 = (p_b - p_2)^2$ and $s = (p_a + p_b)^2 = (p_1 + p_2)^2$. The s-channel diagram needs to be added to $\mathcal{M}$ before squaring the amplitude, in order to take interference effects into account. Typically it can furthermore not be distinguished whether the jet originating from the outgoing quark was started by a d-quark as in this configuration or e.g. by an s-quark. Therefore the same diagrams but with V_{us}^* have to be added. Since there is no interference between outgoing d- and s-quark diagrams, this sum can be done at the level of $|\mathcal{M}|^2$. Also, the description above has introduced the indices s and r for the spin states of the incoming and outgoing quarks whose configurations have to be summed and averaged over the incoming quark spins, using $|\mathcal{M}|^2$. The same is true for the polarizations of the incoming gluon and the outgoing W boson.

The description above does not yet consider the fact that the W boson decays. This would modify the W boson term to (W propagator)×(electron to neutrino current) assuming the decay $W^+ \to e^+ \nu_e$:

$$\epsilon_\mu^{\kappa\,*}(p_1) \to \frac{-i\left(g_{\mu\sigma} + q_\mu q_\sigma / m_W^2\right)}{q^2 - m_W^2 + i m_W \Gamma_W} \cdot \bar{u}_{s'}(p_{1,1}) \left(-i \frac{g_W}{\sqrt{2}} \gamma^\sigma \frac{1}{2}(1 - \gamma^5) \right) v_{s''}(p_{1,2}) \,,$$

where the spins s' and s'' of neutrino and positron are determined by the W polarization and the momenta $p_{1,1}$ and $p_{1,2}$ split the W momentum p_1, obeying momentum conservation.

While this prescription is already lengthy, it does not yet include other diagrams as e.g. $u\bar{d} \to W^+ g$ which can produce effectively the same final state, i.e. after conversion of the outgoing partons into jets. If the momentum of the emitted gluon transverse to the quark, p_T, is sufficiently small, the interaction can actually be factorized into the hard interaction $u\bar{d} \to W^+$ and an additional emission $q \to gq$. This factorization happens at the level of the cross sections which means that the two incidents: hard interaction and gluon emission, behave as independent occurrences. The description of the splitting is already known. It is the same splitting kernel P_{qq} as discussed above for the PDFs. Defining σ_0 as the cross section without emission, the cross section with the additional emission σ_1 is given as:

$$\sigma_1 = \sigma_0 \cdot \int_{Q_0^2}^{Q^2} \frac{dp_\mathrm{T}^2}{p_\mathrm{T}^2} \int_{z_\mathrm{min}}^{z_\mathrm{max}} dz \frac{\alpha_\mathrm{S}}{2\pi} P_{qq}(z) \propto \frac{\alpha_\mathrm{S}}{2\pi} \log^2 \left(\frac{Q^2}{Q_0^2} \right) .$$

The $1/p_\mathrm{T}^2$ dependency comes from the gluon propagator which is proportional to $1/q^2$ similar to the W propagator, but with mass zero, and the cut-off Q_0^2 limits the emissions to resolvable

p_{T}-values. Since higher order loops in the propagation of the emitted gluon can again be renormalized into the running of $\alpha_{\mathrm{S}} \to \alpha_{\mathrm{S}}(p_{\mathrm{T}}^2)$, the cut-off scale has to be well away from the non-perturbative divergence of α_{S}, so $Q_0^2 \gg \Lambda_{\mathrm{QCD}}^2$. Typically it is around 1.4 GeV. The second formulation, in particular the exponent of 2 of the logarithm results from $P_{qq} \sim 1/z$ and from the z-borders being related to Q^2.

The procedure of gluon emission generating an additional jet is applicable to all parton multiplicities, i.e. $\sigma_n \to \sigma_{n+1}$. It can also be applied repetitively with the subsequent emission with lower ordering parameter, here p_{T}, than the previous emission. This means, the subsequent emission boundary $Q^2 \to Q^2_{\mathrm{max}}$ in the p_{T}^2-integral is the integration parameter of the previous emission and thus $p_{\mathrm{T}}^{(1)} > p_{\mathrm{T}}^{(2)} > ... > p_{\mathrm{T}}^{(n)}$. This allows for a cascade of splittings until the resolution cut-off is reached and is important in the implementation of radiation processes in Monte Carlo simulation. Depending on the size of the radiation p_{T}, or the angle θ of the splitting, the radiation either induces the production of an additional jet or describes the evolution of an existing jet.

In the presence of large terms of $\alpha_{\mathrm{S}}/2\pi \log^2(Q^2/Q_0^2)$, α_{S} is not a good parameter for the perturbative series anymore. Instead, the entire term is used in the perturbative expansion, denoted as leading-log (LL) approximation. The leading-log here implies that in the z-integration of the splitting kernel only the dominant (logarithmic) terms are used. Instead of explicitly assuming radiations, one can compute the probability of no emission P_0, and determine the probability for an emission as $1 - P_0$. The probability for no emission is given by Poisson statistics $P(n = 0; \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}|_{n=0} = e^{-\lambda}$ and is called the *Sudakov form factor* $\Delta(Q^2, Q_0^2)$. With λ here specifying the additional emission probability, it is given by:

$$\Delta(Q^2, Q_0^2) = \exp\left[-\int_{Q_0^2}^{Q^2} \frac{dp_{\mathrm{T}}^2}{p_{\mathrm{T}}^2} \int dz \frac{\alpha_{\mathrm{S}}}{2\pi} P_{qq}(z) \right].$$

If the p_{T} of the emission is large, the assumptions contained in the *collinear* approximation however break down and the radiation has to be considered as part of the matrix element. In addition to increasing the complexity of the matrix element, this also increases the complexity of the phase space integration which can only be done numerically for n-body final states.

For further improvements in precision, higher orders in α_{S} need to be considered in the matrix element. For NLO predictions, this includes virtual, i.e. loop corrections, as well as real emissions. Real emissions simply correspond to the emission of another parton, i.e. it equals the LO prediction with $n+1$ partons in the final state. The NLO cross section with n partons in the final state can thus be split into a *LO Born-part* with the n-body phase space integral plus the virtual correction within the same phase space plus the real emission with a $n+1$ phase space integral. For NLO precision, the virtual corrections are the most difficult part to calculate. The difficulty is the treatment of singularities in the virtual term which are usually compensated by singularities in the real term, as long as both are computed together. A simultaneous treatment is not the case if existing LO $n+1$ predictions are used for the real term. That implies that the divergencies in the real term have to be captured in counter terms which can then be added back to the virtual term.

The same problem again arises for NNLO calculations. These have only become possible recently with the development of a new scheme to capture the structure of QCD singularities to cancel the divergencies in the NNLO virtual amplitudes, while allowing at the same time a partitioning of the phase space and relatively high computational efficiency [34]. The latter point implies that the subtraction scheme allows the use of existing NLO $n+1$ predictions and is interfaced to existing NLO tools.

For the final cross section prediction of the W + jets production at the LHC one important step has been neglected so far, namely the fact that the incoming partons in the interaction do not exist freely, but are confined in the accelerated protons. According to the factorization theorem the descriptions of the proton and the hard interaction decouple. Relativistic time dilation makes the time scale of interactions of the partons in the proton much larger than the time scale of the hard interaction with an external particle, i.e. a parton from the opposing proton.

The cross section for W + jets production can therefore be written as:

$$\sigma_{pp\to W+\text{jets}} = \sum_{a,b} \int_0^1 dx_a dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) \cdot \hat{\sigma}_{ab\to W+\text{jets}}(\mu_F, \mu_R) \,.$$

It factorizes into the description of the partonic cross section $\hat{\sigma}$ of the two incoming partons a and b and the probabilities to encounter these partons in the proton, given by the PDFs $f_a(x_a, \mu_F)$ and $f_b(x_b, \mu_F)$. The cross section is summed over all active parton flavours contributing to the production process and integrated over the momentum fractions x_a and x_b.

Both PDFs and the partonic cross section are dependent on the energy of the interaction Q^2 by means of the value of $\alpha_S(Q^2)$, the upper limit of radiations, etc. These energy dependencies are captured into two parameters, the factorization scale, μ_F, and the renormalization scale, μ_R. Summed to all orders in α_S, the physical cross sections are independent of μ_F and μ_R. This means, the dependencies of the cross section on μ_F and μ_R mark the fact that the calculations in the perturbative series of α_S are truncated at some order. The best choice for these parameters is non-trivial and typically presents one of the important theoretical uncertainties in the cross section. Large renormalization and factorization scale uncertainties usually indicate the need for higher order calculations of a process. Since no first-principles-derivation of good values for μ_F and μ_R exists, larger factorization and renormalization scale uncertainties can also point to the use of an improper choice for these parameters.

For the recently published NNLO cross section calculation of $W^+ + 1$ jet, the cross section prediction was found to shift by $+40\,\%$ from LO to NLO, but only by $-1\,\%$ from NLO to NNLO. The uncertainties from the scales, μ_F and μ_R, however reduce from $\pm 7\,\%$ at NLO to percent level at NNLO [34].

2.3 W^+/W^- for PDF constraints

$W^\pm$ bosons in association with one jet are mainly produced at the LHC via the partonic subprocesses gu and gd. For W bosons with some momentum transverse to the direction of the incoming partons from the recoil against a jet ($\gtrsim 60\,\text{GeV}$), this subprocess constitutes more than $50\,\%$ of the total production. According to the factorization theorem, the production cross sections for W^+ and W^- bosons are directly related to the values of the u, d and g PDFs. In contrast to u and d, the gluon part is the same for W^+ and W^- production and can therefore not cause differences in production cross section between W^+ and W^-. In order to produce the W boson together with a hard parton emission, the incoming partons furthermore need to have a significant momentum fraction x which increases with the considered transverse W momentum or the corresponding jet observable. The probability of encountering 'higher' x values, i.e. $x \sim 0.1 - 0.5$, is largest for valence quarks which implies that the W^+ and W^- cross sections are directly related to the values of the $u_v(x)$ and $d_v(x)$ valence quark PDFs. This can also be deduced from Figure 2.9 which shows one of the most recent PDF sets, the

CT14 set [37] computed at NNLO at the energy scale $Q = 100\,\text{GeV}$, close to the mass of the W boson. At x-ranges above 0.1, $u_\text{v}(x)$ and $d_\text{v}(x)$, visible as the differenes between $u(x)$ and $\bar{u}(x)$ as well as $d(x)$ and $\bar{d}(x)$ in Figure 2.9, clearly dominate and provide the momentum needed to produce the W in association with a jet.

The relative amounts of u_v and d_v therefore influence the size of the W^+ and W^- cross sections and, in particular, the ratio of W^+ and W^- cross sections. Theoretical studies [38] demonstrate that different PDF sets yield different predictions of the W^+/W^- ratio, or especially the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio, as displayed in Figure 2.10. The obtained difference is a couple of percent and is largest compared to other vector boson ratios like $W^\pm/Z^0$ which is almost insensitive to variations in the PDFs.

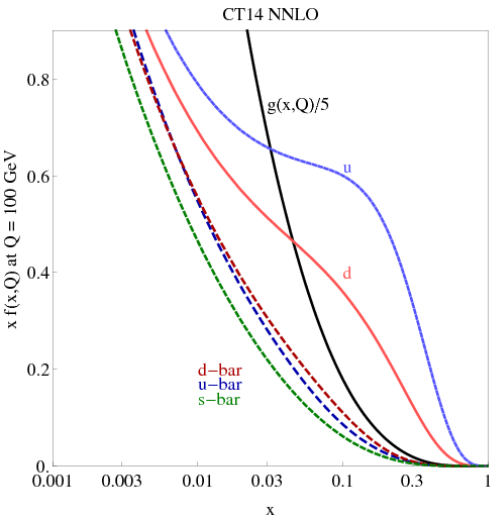


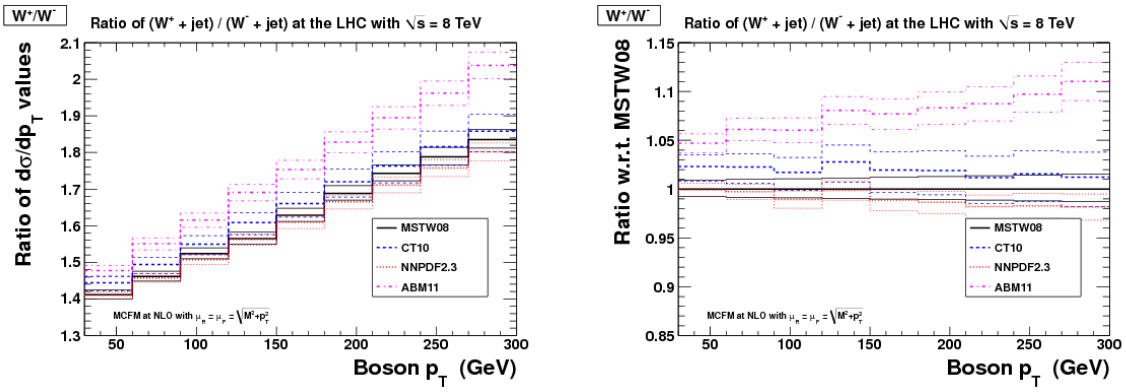
Figure 2.9: CT14 PDF set for u-, d- and s-quarks, gluons (g) as well as anti-quarks $\bar{u}$, $\bar{d}$ and $\bar{s} = s$ at $Q = 100\,\text{GeV}$ [37]. Valence quark PDFs *uval* and d_v are obtained as $u - \bar{u}$ and $d - \bar{d}$.

The ratio of W^+/W^- production has the further advantage that it is perturbatively stable, shown for example in Refs. [2, 39]. This is comprehensible even from a naive point of view, considering that both electroweak and strong interactions are identical for W^+ and W^-. Any higher order correction relevant for W production – apart from PDF information – should be indepedent of the W charge and thus cancel in the ratio. This is different for ratios with respect to the Z^0 boson, where mass differences and differences in the coupling, etc. can induce disparities which do not cancel in a ratio.

The measurement of ratios has the experimental advantage that correlated systematic uncertainties cancel, reducing the overall uncertainty in the measurement. For a measurement of $W^+ + \text{jet}$ and $W^- + \text{jet}$, the present jet is typically the source of large, in most cases dominating systematic uncertainties. The jet uncertainties however cancel to a large extend in the ratio, similar to the cancellation of higher order QCD effects. This represents an important experimental advantage compared to the measurement of absolute cross sections.

Inverting the chain of arguments, PDFs do not only influence the cross section predictions, but instead the measurement of cross sections can be used to constrain PDFs, if the precision of the measurement is high enough. In the case of the W^+/W^- ratio, sensitivity to the valence quark PDFs, or more precisely their ratio $u_\text{v}(x)/d_\text{v}(x)$ has the potential to constrain these. The actual impact of a measurement however depends only on the one hand on its precision, but on the other hand on the number and precision of already existing measurements which constrain PDFs in the same range of the momentum fraction x. For the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio, the same theoretical studies as mentioned above [38] as well as studies done for this analysis showed that x-ranges around 0.1-0.5 can be probed. The probed x-value increases with the energy scale of the $W^\pm$ production, i.e. kinematic variables related to this slightly shift the x-range mostly around and within the given window.

Constraints on u and d quark PDFs at $x > 0.1$ in modern PDF sets, for example the CT14 set, are obtained from fixed-target deep inelastic scattering (DIS) experiments, supplemented by $W^\pm$ asymmetry data from the CDF and DØ collaborations. Tensions in the latter and of the latter to the global data included in the CT14 PDF fit are only alleviated now compared



(a) Cross section ratio $(W^+ + \text{jet})/(W^- + \text{jet})$ (b) Normalized to the MSTW08 PDF prediction

Figure 2.10: Ratio of cross sections for W^+ +jet vs. W^- +jet production for several PDF sets (a) and normalized to the MSTW08 PDF set (b). Given are the central values and the PDF uncertainties (thinner lines) for the predictions as a function of the W momentum transverse to the initial direction of incoming partons (p_{T}) [38].

to previous fits with an update of electron charge asymmetry data from DØ, superseeding the previous measurement. The fixed-target DIS data was gathered mostly in the 1990s. It has to be preselected to remain mostly free of non-perturbative effects such as higher twists or nuclear corrections which can only be computed from phenomenological models and are associated with hard-to-control uncertainties. In collider-data-only PDF fits which have been performed e.g. by the NNPDF collaboration (NNPDF 2.3 set) [40] this data cannot be included, leaving only the CDF and DØ lepton asymmetry measurements to constrain u and d PDFs in that x-range. Although the agreement of the DØ electron asymmetry data set with the global data set, including CDF electron asymmetry and DØ muon asymmetry data, is not still particularly great, it has an important impact on the d-quark PDF and a moderate influence on the u-quark distribution in the CT14 PDF [37]. This demonstrates that already a single data set covering this x-range impacts central u and d PDF predictions, whereas both types of data sets in this x-range are associated with difficulties. Additional measurements covering this x-range would therefore certainly help in constraining the PDFs and resolving the present issues.

Asymmetry measurements by the LHC experiments, ATLAS, CMS and LHCb (for a discussion of the experiments see the next chapter), are included in most recent PDF fits as well. As a result of the higher collision energy of the LHC compared to the Tevatron accelerator (CDF and DØ experiments), the probed x-range is however much lower for ATLAS and CMS. A recent publication by CMS using data from $\sqrt{s} = 8\,\text{TeV}$ collisions [41] demonstrates the ability to constrain PDFs in the range of $x \sim 10^{-3} - 10^{-1}$. W asymmetry measurements by LHCb in principle can extend to higher x values than the respective ATLAS or CMS measurements, since LHCb covers a more forward region which is related to an increased x. LHCb however cannot provide the same amount of data as ATLAS or CMS, as indicated in Section 3.1.

The measurement of the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio as part of the presented W + jets analysis can thus provide valuable information for constraints of u_{v} and d_{v} PDFs, or more precisely the $u_{\mathrm{v}}/d_{\mathrm{v}}$ ratio, complementary to ATLAS and CMS lepton asymmetry measurements. Equivalent to the W^+/W^- ratio is the separate measurement of production cross sections for W^++jet and W^-+jet, if information on the correlation of systematic uncertainties is provided

in addition. This has the advantage that also the absolute cross section values and not only the ratio can be used as input. The ratio then mainly serves as visualization, including a visualization of the uncertainty cancellation.

Consequently, the measurement of $W^{\pm}$ boson production accompanied with at least one jet is of special importance for this analysis. Distributions shown are therefore mainly for $W^{(\pm)}+ \geq 1$ jet, and the measured $W^{(\pm)}+ \geq 1$ jet cross sections are the most important result of this analysis. Its potential inclusion in future PDF fits is facilitated by the recent achievement of NNLO W +jet calculations mentioned before – a prerequisite for NNLO PDF fits.

In the presence of two jets, another production channel for W bosons, namely via the gg initial state, opens up with a small contribution. This production channel does not differentiate between W^+ and W^- production and is sensitive to the gluon PDF, following similar arguments as above. While this is not the focus of this analysis, results for $W^{(\pm)}+ \geq 2$ jets are provided as well in the analysis presented in this thesis.

2.4 Simulation of W + jets in Monte Carlo

The calculation of production cross sections including effects of the matrix element, parton showers, PDFs, etc. is implemented in Monte Carlo-based software programs. These Monte Carlo (MC) generators can either provide total cross section calculations, predictions of kinematic distributions and/or full simulated collision events with four-momentum particle kinematics and polarization effects for a variety of SM processes. Depending on the generator, it either focusses on specific parts in the simulation of the process, e.g. a dedicated simulation of parton radiation or a particle decay, or provides a full description of the entire production chain up to the point where it can be interfaced to a detector simulation framework. The aim of all generators is to provide predictions of a scattering process which can be compared to measurements from the large particle detectors recording collisions from the LHC, the Tevatron, HERA, etc.

The simulation of, for example, W+jets production is split into different components which correspond to energy regimes where certain approximations hold. When the approximations of one regime are known to break down, the simulation of the event is typically passed on to another dedicated software, which models this regime using a different set of approximations. Often the splitting of the simulation into different sub-simulations is motivated by applications of the factorization theorem, but the matching of one simulation regime to another is non-trivial. Very roughly the simulation of proton-proton collisions at the LHC is split into the following components:

- Hard parton-parton interaction→ This is computed by the matrix element and the corresponding phase space integration and – depending on the generator – includes full polarization treatment of initial, intermediate and final state particles (bosons, leptons, partons). Moreover, depending on the generator and the settings, it includes n final state particles and is computed at LO, NLO or NNLO. The LO generators incorporating the real emission amplitudes from higher order calculations by means of simulating a larger number ($\sim 4-5$) of final state partons are called *multi-leg* generators.

- Additional QCD radiations before and after the hard collision → These radiations are called *inital state* (ISR) and *final state radiation* (FSR) and are simulated by parton shower programs. They are typically computed at LL precision.

- The structure of the colliding proton → This information is given by PDFs and is interfaced to MC generators to extract probabilities for certain intial state configurations. The precision of the PDF set, i.e. LO, NLO or NNLO, should match the precision of the matrix element calculation.
- Hadronization of the coloured final state quarks and gluons into colour-neutral baryons and mesons → This non-perturbative process is simulated by dedicated hadronization models which are tuned to data.
- Particle decays → The generated baryons and mesons often undergo decays. In general, decays are simulated until only stable particles with lifetimes larger 30 ps remain, i.e. for example electrons, muons, photons, charged pions as well as protons and neutrons.
- Electromagnetic radiations of charged particles, in particular electrons and muons → This is similar to QCD ISR and FSR with the difference that now photons are radiated. For electrons, the FSR radiation as well as bremsstrahlung effects in the detector are important.
- Underlying event → The partons of both protons which did not participate in the hard interaction can – with a small, but non-negligible probability – undergo another, softer parton-parton interaction, giving rise to *multiple parton interactions* (MPI). In addition, the protons broken up by the hard interaction are still colour-connected to the rest of the event. The description of the *proton remnant* thus includes additional radiations and hadronization.
- *Pile-up* → Simultaneous softer proton-proton (pp) collisions occur as a result of the large number of protons which are grouped into *bunches* and collided at the same time to increase the interaction probability. In 2012, the number of simultaneous interactions was on average 20, but reached a peak value of 72.

In nature, the event of a pp collision starts with the protons. The partons in the proton often radiate and interact with the opposing proton producing a certain final state according to the interaction cross section. Final state partons again radiate and eventually hadronize.

In simulation, however, the same sequence – proton, radiation, hard interaction, radiation, hadronization – hinders the generation of reasonable MC statistics of low cross section processes such as Higgs boson production, while being overwhelmed with the computing of high cross section processes. Instead, the MC generation is therefore centered around the matrix element computation as the core of the simulation of a specific process. 'Around' that calculation, the event is dressed with radiations via the parton shower. While the FSR occurs and evolves in time, the ISR emissions need to evolve backwards in time. This means, starting from the hard interaction, the partons evolve radiating up to higher values of the momentum fraction x, until further emission is suppressed by the rapidly falling PDFs with increasing x. Both in backward time direction (ISR) and in forward time direction (FSR), the parton shower emissions are cut off at a scale of approximately 1 GeV ($\gg \Lambda_{\mathrm{QCD}}$) before non-perturbative effects of the proton remnant or the hadronization, respectively, set in.

For the parton showers implemented in the simulation the ordering parameter used to determine the hierarchy of the collinear emissions is important. Theoretically, collinear emissions ordered in the virtuality of the original parton, the momentum of the emitted parton transverse to the original one (p_{T}) or the corresponding angle (θ) are identical. But in the implementation in MC parton showers, they can give different results, in particular towards the edges of the collinear phase space.

In addition to the collinear divergence in the probability of additional parton emission, the splitting kernel $P_{qq}(z)$ also diverges for $z \to 1$. This corresponds to the emission of soft gluons. The soft gluon is emitted as if coming from the scattering process as a whole. In a collinear parton shower, this is correctly considered, e.g. if angular ordering is used. The soft gluons are emitted first with wide angles before the hardest collinear emission takes place.

A non-trivial step in the simulation is the matching of the matrix element calculation with a certain number of outgoing partons to the parton shower (*ME-PS matching*), in particular for NLO matrix element calculations. For a LO matrix element, the two most known matching and merging methods, which ensure no double counting of final state parton configurations, are the CKKW [42, 43] and the MLM [44] matching schemes.

After parton showering, the colour-connected quarks and gluons need to be converted into colour-neutral hadrons. The process – called *hadronization* – includes the *fragmentation* of the colour strings and is typically described by two schemes: the *Lund string* [45] or the *cluster hadronization* [46, 47] model.

The state when only stable final state particles exist in the simulated event is denoted as *particle level*. It is different from the *parton level* which describes the direct outcome of the hard interaction and does not include e.g. radiations. The particle level corresponds to the state of particles directly before hitting the detector.

The simulation of pile-up is a problem of experimental origin. One distinguishes between *in-time* and *out-of-time* pile-up. In-time pile-up are additional soft pp interactions contributing to detector signals from the hard interaction. Their average contributions need to be considered, e.g. for a proper estimate of jet energies. These collisions are usually simulated separately, and simulation of the hard interaction is then overlaid with pile-up events according to the measured number of average interactions before simulating the detector response. Out-of-time pile-up is the result of the signal duration time in the detector which is longer than the time between collisions. It characterizes the effect of the overlay of residual signals from previous, but also from the following collisions of proton bunches during the triggered one. Both in- and out-of-time pile-up need to be considered for a good match of data and simulation and are relevant for the precise measurement of detector signals.

A detailed description of the different components of MC simulation and their interplay is given in Ref. [48] for general-purpose event generators providing a full simulation of the collision event. A dedicated description of measurements of single vector boson production at $\sqrt{s} = 7\,\text{TeV}$, including an explanation of their simulation, is given in Ref. [49].

MC simulation in the W + jets analysis

In the W + jets analysis, MC simulation is used to provide predictions for the W + jets signal as well as to model several background processes, based on the following MC generators:

- **Alpgen** [50] is a multi-leg LO generator which simulates W + jets and Z+jets events with up to 5 partons in the final state. It is interfaced to Pythia [51] for parton showering. Pythia is a LO multi-purpose generator for $2 \to 2$ processes and follows the p_{T}-ordered approach for the parton shower. Differential predictions are scaled to the total NNLO cross section prediction obtained with DYNNLO [52] interfaced to the PDF set MSTW2008 [53].

- **Sherpa** [54, 55] is a general purpose multi-leg LO generator which simulates up to 4 partons in the final state. It has its own showering approach based on Catani-Seymour dipoles [56] merged to the matrix element using the CKKW approach [57] and its own

hadronization, realized with a cluster model [58]. In the used samples, c- and b-quarks are simulated as massive. The used SHERPA version is 1.4.1. The SHERPA samples are normalized to the same NNLO cross section as the ALPGEN samples.

- **MCFM** [59] is a NLO generator which can simulate up to two additional partons, but for W production only either 1 or 2 partons at NLO, not both at the same time. The predictions are at parton level only.
- **POWHEG** [60] is a NLO generator which realizes a proper matching of the NLO matrix element to a showering generator via the POWHEG method [61, 62]. For the parton shower, PYTHIA is used. For $t\bar{t}$ production, the POWHEG distributions are normalized to the NNLO+NNLL cross section predictions from TOP++ 2.0 [63].
- **HERWIG** [64] is a LO multi-purpose generator with angular ordered parton showers and the cluster hadronization model. The distributions are normalized to NLO cross sections, calculated in [2].

Events simulated with these generators, except MCFM, are passed to a full simulation of the ATLAS detector [65], based on the GEANT4 toolkit [66].

ALPGEN, SHERPA and MCFM are used to obtain predictions for W + jets production, where ALPGEN is the main generator in this analysis and SHERPA is used in the determination of systematic uncertainties. MCFM is compared to the measured W + jets cross sections, however without correcting for the difference between parton level (MCFM prediction) and particle level (data). More details on the MCFM predictions are given in Chapter 8 in the context of the cross section measurement.

The other MC generators are used for the background estimates. Z and W related background processes are estimated with ALPGEN like the W + jets signal. t-quark processes, like $t\bar{t}$ and single top production are estimated with POWHEG, and HERWIG is used for the simulation of diboson production. For more details on the backgrounds in this analysis and their estimation see Chapter 5.

CHAPTER 3

The ATLAS experiment at the LHC

Since the successful investigation of atomic structure by the Rutherford experiment [67], Particle Physics' approach to new insights into the structure of matter and its interactions has been the collision of particles at ever increasing energies and the study of their behaviour. The current frontier in this attempt is formed by the *Large Hadron Collider* (LHC), designed to collide protons at the unprecedented energy of 14 TeV and the detectors ATLAS[1], CMS[2], LHCb[3] and ALICE[4] built to measure them. ATLAS and CMS are multipurpose detectors covering almost 4π in solid angle and precisely measure the SM, culminating so far in the discovery of the Higgs boson in 2012 [21, 22], as well as search for any indication of physics phenomena not contained in the SM. LHCb is a single-arm forward detector specialized to measure rare decays of charm and bottom mesons and study CP violations. ALICE is a general purpose detector with a central region and one-sided forward geometry, constructed to resolve the extreme particle multiplicities which are generated from a quark-gluon plasma in heavy-ion collisions.

The data set, this analysis is based on, consists of proton-proton collisions at a center-of-mass energy of 8 TeV, produced by the LHC in 2012 and recorded by the ATLAS detector. In the following, the LHC accelerator is described in Section 3.1 and the ATLAS detector in Section 3.2.

3.1 The Large Hadron Collider

First ideas of a large hadron collider at CERN[5] date back to 1977, when the planning of the Large Electron-Positron Collider, *LEP*, was ongoing. The possibility of a multi-TeV hadron collider as a successor of LEP was one of the arguments for the large circumference of 27 km of the LEP tunnel. The LHC was seen as discovery machine in the tradition of the pp and $p\bar{p}$ colliders like *ISR*[6], *SpS*/*Sp$\bar{p}$S*[7] and *Tevatron*. These had lead, among other things, to the measurement of single particles with high p_{T} in 1973, the discovery of the $W^{\pm}$ and Z^0 bosons in 1983 and the discovery of the top-quark in 1995.

[1] ATLAS: A Toroidal LHC ApparatuS
[2] CMS: Compact Muon Solenoid
[3] LHCb: Large Hadron Collider beauty
[4] ALICE: A Large Ion Collider Experiment
[5] CERN: Conseil Européen pour la Recherche Nucléaire, Geneva, Switzerland
[6] ISR: Intersecting Storage Rings
[7] SpS: Super proton Synchrotron, Sp$\bar{p}$S: Super proton-antiproton Synchrotron

The LHC was approved in 1997 with the design energy of $\sqrt{s} = 14\,\text{TeV}$ and a nominal luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ and the aim of discovering the Higgs boson and investigating the high energy region where BSM physics was expected. In 2008, operation of the LHC started for the first time.

In 2012, the LHC was running with a center-of-mass energy of 8 TeV, reaching peak luminosities of $7.7 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$. As intended, the LHC is located in the former LEP tunnel between 70 to 140 m below the surface in the area of Geneva and the Jura mountains. Due to the relatively small tunnel diameter, separate storage rings for the clock- and counter-clock-wise rotating proton beams were impossible and the LHC magnets are designed as twin-aperture magnets where two separate beam pipes share a common return yoke and cryostat. A picture of the magnetic field structure required to bend the counter-rotating beams is given in Figure 3.1. 1232 dipole magnets using Niobium-Titanium (NbTi) coils generate a magnetic field of 8.3 T from currents of $\sim$ 11 kA to keep the protons on track in the eight LHC arcs with an effective bending radius of 2804 m. 120 t of liquid helium are in turn needed to ensure an operating temperature of 1.9 K for the superconducting NbTi wires. At this temperature, helium is superfluid and can thus permeate the NbTi coils, acting as thermally highly conductive heat dissipator without the need for large-scale fluid circulation. In addition to the 14.3 m long dipole magnets, 382 quadrupole magnets for beam optics and further 3700 special magnets, including the kicker magnets for beam insertion and extraction are needed for beam operation.

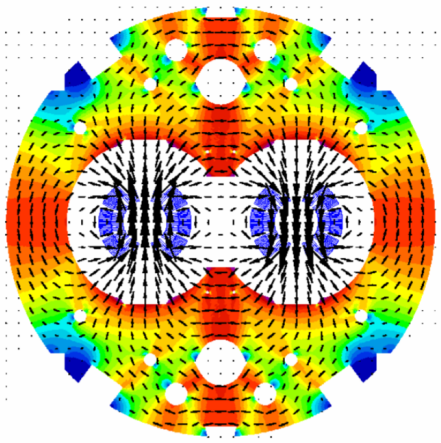


Figure 3.1: Lateral view of a LHC dipole magnet. The blue semi-arcs in the center indicate the superconducting coils which are surrounded by non-magnetic (white) steel collars and the (colored) iron return yoke. The direction of the magnetic flux is indicated by black arrows and the color displays the residual field strengths in the return yoke. The entire displayed object (apart from the beam areas between the coils) corresponds to the dipole cold mass and is cooled by superfluid helium [68].

Between the eight LHC arcs, eight straight sections of approximately 530 m length serve as insertion region (IR) of machine equipment for the LHC itself or as interaction points (IP) where the beams are brought into collision and the four main detectors are placed. Figure 3.2 shows a schematic view of the LHC geometry. In the straight sections around the IPs, the two beams share the same beam pipe for about 130 m. As visible in Figure 3.2, ATLAS and CMS are located at the opposite high-luminosity interaction points IP1 and IP5, while ALICE and LHCb are placed at IP2 and IP8 where also the beam insertion takes place. IR3 and IR7 house collimation systems for beam quality control, IR6 the beam dump system and IR4 the accelerating radio-frequency (RF) systems. In order to maintain approximately the same circumference for both beams, the beams cross over at the four detector IPs.

The LHC collision energy is reached building on a set of preaccelerators. Protons extracted from hydrogen gas are first accelerated - grouped together in so-called *bunches* of $\sim 11^{11}$ protons per bunch - by the linear LINAC2 accelerator to 50 MeV. Subsequently, they are transferred to the PS BOOSTER which accelerates them in four stacked rings to 1.4 GeV and then to the PS which increases the energy to 25 GeV. The PS in addition prepares the bunch structure by splitting 6 bunches from the PS Booster into 72 bunches with 25 ns spacing required by the LHC at nominal running conditions. Up to four PS fills can then be passed on to the SPS which accelerates the protons to the LHC insertion energy of 450 GeV. About

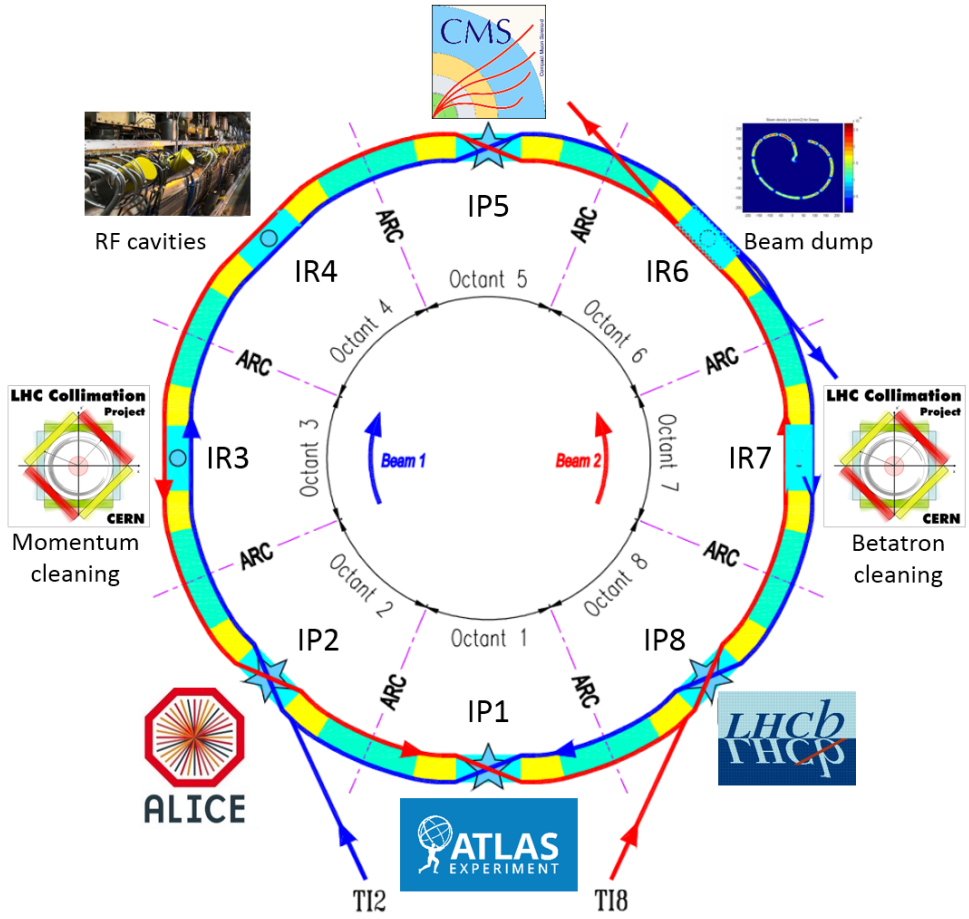


Figure 3.2: Structure of the LHC ring. Adapted from [68] with inputs from *CERN*.

12 SPS fills are transferred to the LHC where eight superconducting RF cavities per beam operating at 400.8 MHz with accelerating field strenghts of 5.5 MV/m capture and accelerate the protons to the final collision energy. The cavities are cooled to 4.5 K and provide an energy gain per turn of $\sim 0.5\,\mathrm{MeV}$[8]. In total, the LHC can be filled with 2808 bunches, separated by 25 ns or 7.5 m, and grouped together in trains of 288 proton bunches.

The luminosity, $\mathcal{L}$, of the LHC - a measure for the *beam brightness* or the proportionality factor between the interaction cross section and the actual produced number of events - is given as:

$$\mathcal{L} = \frac{N_1 N_2 f_{\mathrm{rev}} n_b}{2\pi\sqrt{\sigma_{1x}^2 + \sigma_{2x}^2}\sqrt{\sigma_{1y}^2 + \sigma_{2y}^2}} \cdot F \cdot W\,,$$

where N_1 and N_2 are the number of protons in the n_b colliding bunches. f_{rev} is the revolution frequency of the beams in the ring, and σ is the beam size in horizontal (x) and vertical (y) planes for beam 1 and 2 at the IP. F and W are two luminosity reduction factors where F considers the beam crossing angle and W the transverse offset at the collision point[9]. A finite beam crossing angle ϕ is needed to avoid unwanted additional collisions up- and

[8] For comparison, the energy loss per turn due to synchrotron radiation is 7 keV.

[9] W is relevant for LHCb and ALICE to control and keep the luminosity at tolerable levels for these two detectors. For ATLAS and CMS, $W = 1$.

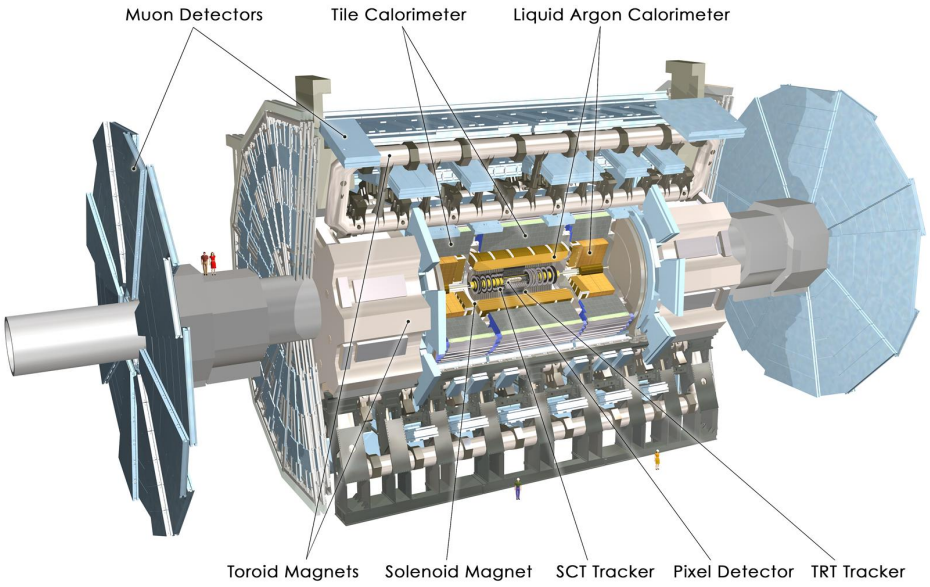


Figure 3.3: The ATLAS detector [69].

downstream of the actual IP. At design operation parameters[10] of the LHC, this leads to a factor of $F = (1 + 2\frac{\sigma_s^2}{\sigma_{1x}^2+\sigma_{2x}^2}\tan^2\frac{\phi}{2})^{-1/2} = 0.836$ reduction in the luminosity compared to head-on collisions. The integrated luminosity is the time integral of $\mathcal{L}$ and amounted in 2012 to 22.8 fb^{-1} delivered by the LHC and 21.3 fb^{-1} recorded by ATLAS (see e.g. Figure 4.4 in Chapter 4).

3.2 The ATLAS detector

The ATLAS detector [70] is build in cylindrical shape around the interaction point IP1 of the LHC ring. It is designed in an onion-like structure with tracking detectors on the inside, followed by the calorimeter and the muon system on the outside. An overview is shown in Figure 3.3. In total, the ATLAS detector is 44 m long and 25 m in diameter and weights about 7000 t.

The tracking system provides information about the flight path and the momenta of the produced charged particles, as well as particle identification and primary and secondary vertex reconstruction. The calorimeters - split into an electromagnetic (EM) calorimeter on the inside and a hadronic (HAD) calorimeter on the outside cover the full solid angle until very close to the beam pipe and provide the energy measurement of the generated charged and neutral particles. Shower shapes as well as track association allow for further particle identification, and global energy sums determine the signatures of weakly interacting particles as missing energy balance in the calorimeter. The muon spectrometer determines the flight paths and momenta of muons which are the only particle species penetrating enough to reach the detector up to this point and leave a signal. If possible, the muon-system tracks are combined with measurements from the inner tracking detectors for improved precision due to the large lever arm.

Both the inner tracking detectors as well as the muon system are immersed in magnetic

[10] Bunch length $\sigma_s =\sim 0.075$ m, Beam size at IP $\sigma_x =\sim 17\,\mu$m, Crossing angle $\phi = 284\,\mu$rad

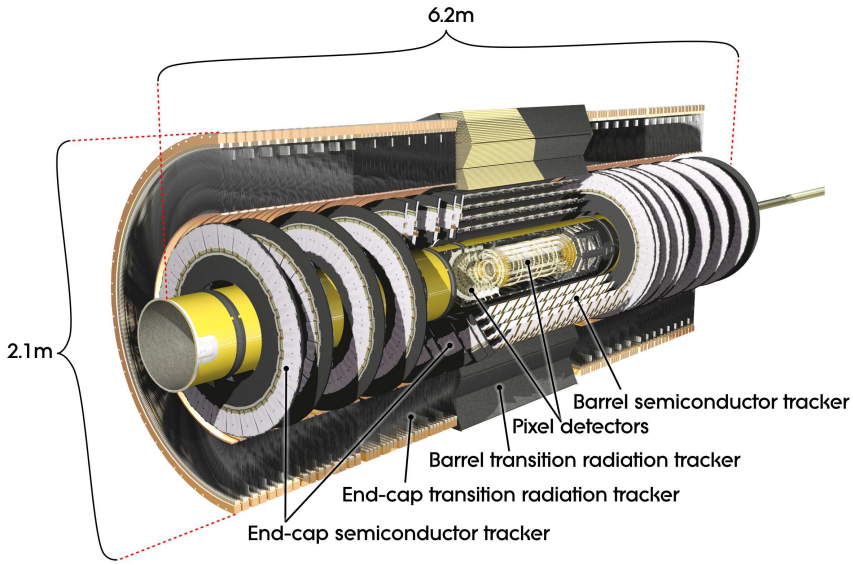


Figure 3.4: View of the ATLAS Inner Detector [70].

fields to measure the charged particle momenta from the track curvature in the magnetic field. The fields are generated by a solenoid magnet for the inner detector and a system of three toroid magnets for the muon spectrometer, reaching field values of 2 T and $0.5 - 1$ T, respectively.

The ATLAS detector uses a right-handed, orthogonal coordinate system with the origin at the nominal interaction point in the center of the detector. The x-axis points towards the center of the LHC ring, the y-axis points upwards towards the surface of the Earth and the z-axis points into the direction of beam 2. Following the cylindrical shape of ATLAS, often cylindrical coordinates (r, ϕ, z) are used, where the azimuthal angle ϕ is measured from the x-axis around the beam pipe. The polar angle θ is defined as the angle with respect to the positive x-axis and is in the following mostly given in terms of the pseudo-rapidity η with $\eta = -\ln\tan(\theta/2)$. The pseudo-rapidity corresponds to the massless or high-energy approximation of the rapidity $y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$ of a particle with energy E and momentum $\mathbf{p}$. Rapidity differences Δy are invariant under Lorentz boosts. This is advantageous for the description of kinematics at the LHC where the z-boost of the colliding partons in the proton is unknown. p_T and E_T are projections onto the transverse detector plane ($x - y$ plane) of the momentum and the energy, respectively, i.e. $E_\mathrm{T} = E \cdot \sin\theta$. Angular distances between particles are given as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

3.2.1 Inner Detector

The ATLAS Inner Detector (ID) was designed to provide tracking with a momentum resolution of $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\,\% \cdot p_\mathrm{T} \oplus 1\,\%$ for charged particles with $p_\mathrm{T} > 0.5$ GeV and a transverse impact parameter resolution of 10 μm for high momentum particles in the central detector region. To achieve this purpose, the ID consists of three subsystems: The pixel detector (PD), the semi-conductor tracker (SCT) and the transition radiation tracker (TRT), covering radial distances from the beam axis of approximately $5 - 15$ cm, $30 - 56$ cm and $56 - 107$ cm, respectively. Along the beam axis, the sensitive areas of the ID subsystems extend 65 cm in positive and negative z-direction for the PD and 2.7 m for both the SCT and TRT where the TRT surrounds the SCT radially over the entire length. This corresponds to a pseudo-rapidity

coverage of $|\eta| < 2.5$ for the PD and SCT and $|\eta| < 2.0$ for the TRT. An overview of the ID system is given in Figure 3.4. All three subsystems consist of barrel components, where the sensors are arranged in cylindrical shape around the beam axis, and end-cap regions where they are placed on disks stacked perpendicular to the beam axis.

The PD provides the highest granularity with a minimum pixel size in $(R-\phi) \times z$ of $50 \times 400 \mu m^2$, resulting in intrinsic accuracies, i.e. typical position resolutions of $10\,\mu$m in $(R-\phi)$ both in the barrel and end-cap regions and $115\,\mu$m in $(z)/(R)$ for the barrel/end-caps. The total area covered with active silicon sensors is approximately $1.7\,\mathrm{m}^2$ and the total number of readout channels is 80.4 million. Typically, each track crosses 3 pixel layers.

The SCT uses silicon stereo strips with a mean pitch width of about $80\,\mu$m which are placed back-to-back with an angle of 40 mrad, thus allowing to determine both coordinates of the track position. Intrinsic accuracies of $17\,\mu$m are obtained in $(R-\phi)$ both in the barrel and end-cap regions and $580\,\mu$m in $(z)/(R)$ for the barrel/end-caps. In total, the SCT covers a silicon surface area of $63\,\mathrm{m}^2$ and provides approximately 6.3 million readout channels. On average, each track crosses the SCT modules in 4 places (corresponding to 8 strip layers).

The TRT consists of 4 mm diameter straw tubes with a $31\,\mu$m thin gold-plated tungsten wire as central anode and filled with a Xenon-based gas mixture. The 144 cm (37 cm) long straw tubes are placed in 73 (160) layers parallel (radially outwards) to the beam direction in the barrel (end-cap) regions. At operation voltages of about $-1.5\,$kV, the maximum electron charge collection time from ionization in the gas is approximately 48 ns and the obtained intrinsic accuracy in $(R-\phi)$ is $130\,\mu$m per straw. Transition radiation photons emitted from relativistic electrons yield much larger signals than other, minimally ionizing particles and thus allow for electron identification for electron p_{T}'s in the range of $0.5 - 150\,$GeV. The TRT covers a gas volume of $12\,\mathrm{m}^3$ and provides about 351 000 readout channels. Typically, traversing charged particles produce 36 hits in the TRT.

Very high tracking efficiency is an important ingredient for many measurements, if only for the reconstruction of the primary vertex. In Run1 of the LHC[11], a track reconstruction efficiency in the central region above 98 % was obtained [71] and a primary vertex position resolution of less than $20\,\mu$m and approximately $30\,\mu$m in the transverse and longitudinal directions, respectively [72]. The overall hit efficiency measured e.g. in the SCT in 2012 was $(99.74 \pm 0.04)\%$, where the uncertainty is systematic and the statistical uncertainty is negligible [73].

3.2.2 The Calorimeter system

The calorimeter measures the energy of both charged and neutral particles produced in the interaction. It is located outside the ID and the surrounding solenoid magnet, covering radial distances of about $1.4 - 4.3\,$m from the beam axis and a large pseudo-rapidity range $|\eta| < 4.9$. The calorimeter is supposed to stop incoming particles (apart from muons and weakly interacting particles like neutrinos) and is thus constructed to contain the full length of the particle showers induced by the particles generated in the collision. An overview is displayed in Figure 3.5. The calorimeter is the detector component which is mainly relied upon for the reconstruction of electron and jets as well as the computation of the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ which are the final state quantities important for this analysis.

As customary, the calorimeter is split into an electromagnetic and a hadronic part, respecting the different interactions of the original particles in the detector material, i.e. electromagnetic showers in case of electrons and photons and hadronic showers including nuclear reactions

[11] *Run1* refers to the LHC operation years 2009-2013.

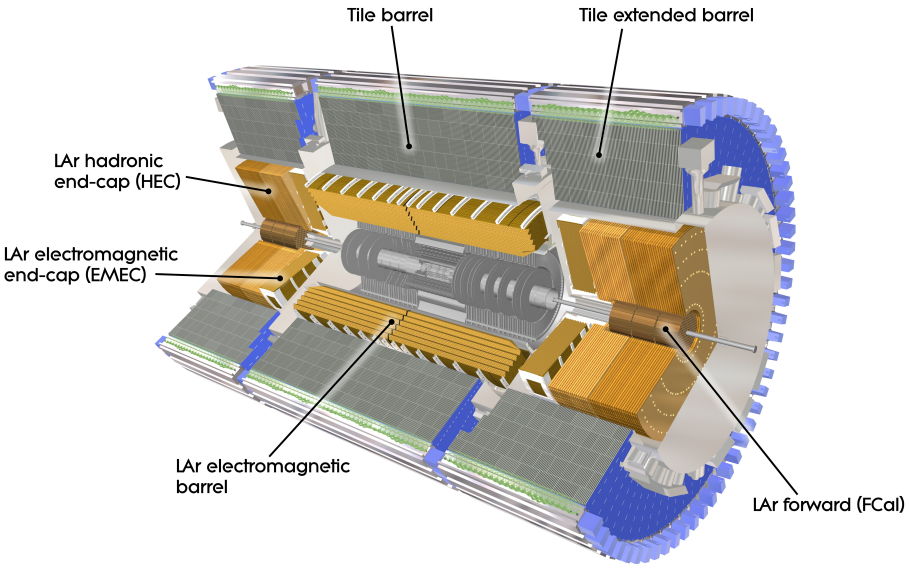


Figure 3.5: View of the ATLAS Calorimeter [70].

in case of e.g. charged pions, protons and neutrons. The depth of the ATLAS calorimeter was chosen to be above 22 radiation lengths X_0 [12] for the electromagnetic calorimeter and about 10 interaction lengths λ [13] for the combined electromagnetic and hadronic calorimeter.

Typically, the energy resolution of the calorimeter can be parameterized as

$$\frac{\sigma(E)}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C\,,$$

where N denotes the noise term, S the stochastic and C the constant term. The noise term includes detector and electronic noise, as well as contributions from additional simultaneous collisions (pile-up) and is mostly relevant at low energies. The stochastic term reflects statistical fluctations in the shower evolution and its sampling in the detector. The constant term results from non-linearities in the signal response, signal losses due to not-captured shower tails or non-instrumented regions in the detector and becomes the limiting factor for high energies. The design specifications for the ATLAS calorimeter demand a 10 % stochastic and 0.7 % constant term for the electromagnetic calorimeter and 50 % and 3 %, respectively, for the central hadronic calorimeter. The values for the different calorimeter regions measured in test beam setups mostly fulfil these requirements and are listed in Table 3.1 (for details on the calorimeter regions see below). The calorimeter is finely segmented to achieve the required energy resolution and measurement precision, yielding in total slightly more than 192 000 readout channels.

Electromagnetic calorimeter

The electromagnetic calorimeter consists of a barrel part, covering $|\eta| < 1.475$ (EMB) and two end-cap components which cover $1.375 < |\eta| < 3.2$ (EMEC). To minimize the material in

[12] A radiation length X_0 is defined as the distance after which the energy of a high energetic electron is reduced via bremsstrahlung to $1/e \cdot E_{\text{original}}$.

[13] The interaction length λ is the average distance which a relativistic hadron travels inside a medium before undergoing a nuclear interaction.

	Calorimeter component	Material	S [%]	C [%]
EM	Barrel	LAr + Pb	10	0.7
	End-Cap	LAr + Pb	10	0.7
	Forward	LAr + Cu	28.5	3.5
HAD	Tile	Scint.+Steel	56	5.5
	HEC	LAr + Cu	84	-
	Forward	LAr + W	94	7.5

Table 3.1: Stochastic (S) and constant (C) terms of the energy resolution for the components of the ATLAS calorimeter as determined from test beam measurements [70]. Additionally, the active and absorber materials of the calorimeter components are given.

front of the calorimeter, the EMB shares a common vacuum vessel with the central solenoid magnet. EMB and EMEC are based on LAr-Pb technology where liquid argon circulates as active material between accordion-shaped lead absorbers. Also see Table 3.1 for an overview of the active and passive materials of the calorimeter components. The accordion wave-structures of the absorbers run radially outwards in the case of the EMB and parallel to the beam direction for the EMEC, thus allowing a coverage without azimuthal (longitudinal) cracks for the EMB (EMEC).

The EMB is segmented into three layers in depth, corresponding to radiation lengths of $4.3X_0$, $16X_0$ and $2X_0$. The largest fraction of the energy in electromagnetic showers is deposited in the middle layer, while the first layer provides precision information on the origin position of the shower and the third layer collects the shower tail. Following these requirements, the layers provide a granularity of about 0.003×0.1, 0.025×0.025 and 0.05×0.025 in $\Delta\eta \times \Delta\phi$, respectively, in the most central region. Also see Figure 3.6a for an overview of the segmentation and granularity of the EMB calorimeter. In the region $|\eta| < 1.8$, the first layer is supplemented by a pre-sampler layer in front, in order to measure and correct for the energy of electrons and photons lost upstream of the calorimeter.

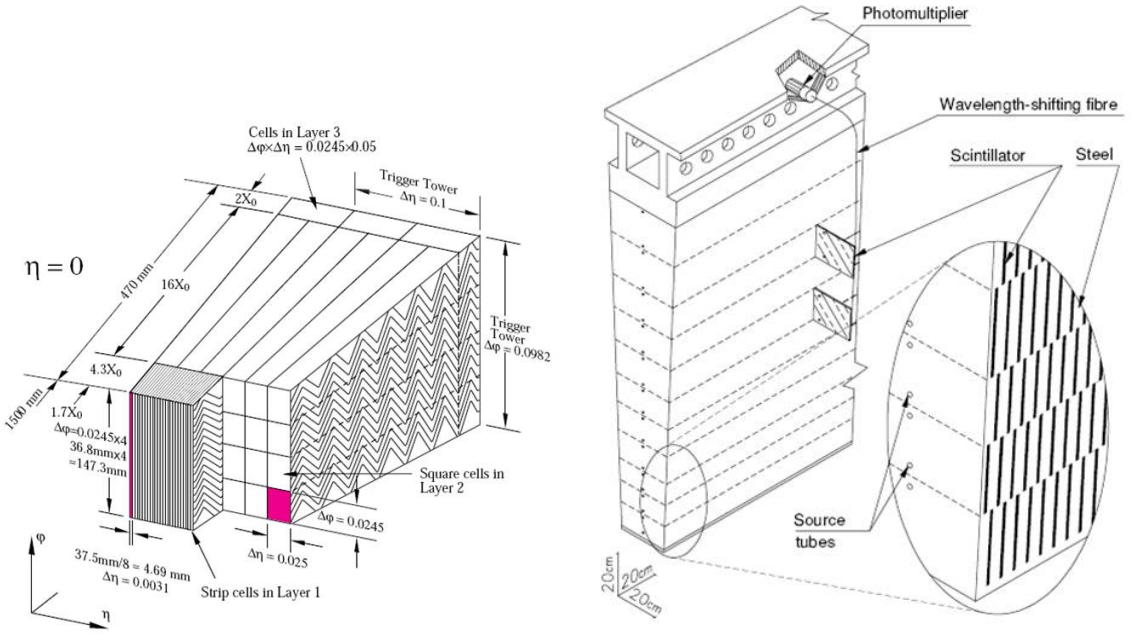
The EMEC is split into two separate wheels where the outer wheel covers $1.375 < |\eta| < 2.5$ and is segmented into three longitudinal layers with granularities in its central measurement region similar to the EMB. The inner wheel spans the pseudo-rapidity range $2.5 < |\eta| < 3.2$ and is segmented into two layers in depth only, with a coarser granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$.

The forward calorimeter (FCal) covers the pseudo-rapidity range $3.1 < |\eta| < 4.9$ and is integrated into the cryostat of the calorimeter end-caps, thus also using liquid argon as active medium. The electromagnetic part of the FCal utilizes copper as absorber material owing to its resolution and heat conductivity properties in the high rate forward region. The depth of the electromagnetic FCal corresponds to approximately 28 radiation lengths X_0.

Hadronic calorimeter

The hadronic calorimeter consists of the Tile calorimeter – placed outside the electromagnetic calorimeter –, the hadronic end-cap calorimeter (HEC) – installed directly behind the EMEC – and the hadronic forward calorimeter – sharing the EM end-cap cryostat.

The Tile calorimeter is build as sampling calorimeter with fluor-doped polystyrene scintillating tiles and steel as absorber material, and is split into a barrel part with $|\eta| < 1.0$ and two extended barrels with $0.8 < |\eta| < 1.7$. The scintillation light produced by the hadronic showers from traversing hadronic particles is read out with photomultiplier tubes via wave-



(a) Segmentation in the EMB

(b) Structure of the Tile

Figure 3.6: Segmentation and structure of the ATLAS LAr EMB calorimeter (a) and the steel-scintillator-based hadronic Tile calorimeter (b) [70].

length shifting fibers at two sides of the scintillating tiles. An overview of the tile structure and the light-readout is shown in Figure 3.6b.

The Tile calorimeter is segmented into three layers in depth, corresponding to interaction lengths of 1.5λ, 4.1λ and 1.8λ (1.5λ, 2.6λ and 3.3λ), respectively, for the barrel (extended barrel) region. The granularity is 0.1×0.1 in $\Delta\eta \times \Delta\phi$ for the first two and 0.2×0.1 for the last layer. The volume ratio of steel absorber and scintillating tiles is approximately 4.7:1 and the steel support structure also provides the flux return of the solenoid field.

The HEC employs liquid argon as active material and copper as absorber, using a flat-plate design and covering $1.5 < |\eta| < 3.2$. It thus has small overlaps with the Tile calorimeter at low $|\eta|$ and the FCal at high $|\eta|$. The HEC is segmented into four layers in depth and has a granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ up to $|\eta| < 2.5$ and twice that for the higher pseudo-rapidities.

The hadronic FCal is located behind the electromagnetic FCal and is split into two modules per detector side. Together with the electromagnetic FCal, it has a total depth of approximately 10λ, covering the same $|\eta|$-range as the electromagnetic part. While the active material is identical, tungsten is used as absorber instead of copper, in order to limit the lateral spread of the hadronic showers and ensure shower containment within the calorimeter.

Jets are used in this analysis up to $|\eta| = 4.4$, thus also relying on the FCal energy measurement.

3.2.3 Muon System

The muon system is the outer-most part of the ATLAS detector, specialized to measure muon tracks. The momentum resolution for muons depends on the strength of the magnetic field B bending the muon tracks and the lever arm L, i.e. the distance from interaction point to

the muon measurement.

$$\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} \propto \frac{1}{B \cdot L^2}$$

To achieve the goal of a momentum resolution of about 10 % for 1 TeV tracks, the barrel region of the muon system is located at distances of approximately $5 - 10\,\mathrm{m}$ from the interaction point, immmersed into a toroidal magnetic field with a bending power $\int Bdl$ of $1.5 - 5.5\,\mathrm{Tm}$. Additionally, two end-caps are formed by large wheels, located at distances of approximately $7.4 - 21.5\,\mathrm{m}$ in z-direction from the interaction point, again interspersed in the magnetic field generated by two toroidal end-cap magnets with bending power of $1 - 7.5\,\mathrm{Tm}$.

For the muon detection, four types of spectrometers are used, two for precision tracking, covering a pseudo-rapidity range of $|\eta| < 2.7$: Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC), and two for fast response needed for triggering within $|\eta| < 2.4$: Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC)[14]. Since muons are not used in this analysis, please refer to Ref. [70] for more information on the muon system.

3.2.4 Trigger System

At the design LHC luminosity, about 40 TB of data are produced per second by ATLAS, given crossings of pp-bunches every 25 ns (in 2012: 50 ns) and a storage size of 1 event with information from all ATLAS subdetectors of approximately 1 MB. Translated into the height of a stack of CD disks, the total generated pp ATLAS data in 2012 [15] corresponds roughly to half the distance between the earth and the moon.

However, while the total inelastic cross section is high, the interesting processes to be measured as part of the LHC physics program have cross sections at least about 10^6 orders of magnitude lower, as can be seen from Figure 1.1.

A three level trigger system is therefore employed in ATLAS to reduce the finally stored event rate *online*, i.e. in parallel to LHC running to a managable amount of a few hundred Hz or MB/s. The levels contain various algorithms which run successively with increased degree of refinement and detector data input. They search for signatures of high energetic electrons, photons, muons, hadronically decaying τ-leptons, jets as well as large global energy sums or missing transverse momentum in the just-occurred collision events. An overview of the trigger and data acquisition architecture is presented in Figure 3.7.

The first trigger level (L1) is realized in custom-built hardware and uses reduced granularity data from the calorimeter and the trigger-assigned muon chambers to obtain a decision on the event rejection within $2.5\,\mu\mathrm{s}$ [16]. During this time, the event information is stored in on- or near-detector pipeline buffers whose length provides the hard cutoff for the L1 latency. The L1-accept (L1A) decision is made by the Central Trigger Processor (CTP), based on the counting of object multiplicities above different thresholds or above-threshold flags for global quantities. The L1 trigger also identifies the bunch-crossing (BCID) which the detector signals date back to. This is a challenging task due to the width of the calorimeter signals which typically extend over four bunch-crossings and time-of-flight of the muons from generation in the collision to their measurement longer than the time between bunch-crossings. The L1 trigger reduces the original rate of 20 MHz in 2012 to approximately 75 kHz[17].

[14] The RPCs and TGCs also provide the measurement of the second track coordinate within $|\eta| < 2.7$.

[15] Considering only the fraction of 36.5 % of the 200 days of pp-running in 2012 with stable beam operation [74].

[16] Note that already $1\,\mu\mathrm{s}$ of the allowed L1 latency is consumed by the signal propagation in cables from the detector to the trigger system which is located in a cavern next to the detector.

[17] The given trigger rates for the different levels are average numbers for the detector operation in 2012.

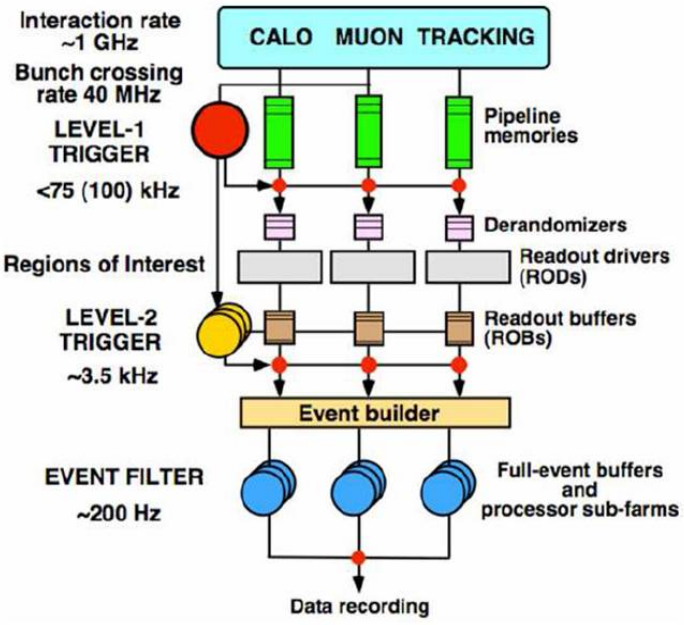


Figure 3.7: Schematic view of the three-level ATLAS trigger and data aquisition system [75]. The indicated numbers for the trigger rates are design values, while in the text, average values for 2012 data-taking are given.

The second (L2) and third (Event Filter, EF) trigger levels – together also denoted as High-Level Trigger (HLT) – implement software algorithms running on a network of largely commercially available PCs. After the L1A, the L2 trigger requests full granularity data in so-called Regions of Interest (RoI), defined by L1, from the data aquisition system (TDAQ). This selective data transmission limits the amount of transferred data to about $1 - 2\,\%$, thus enabeling refined trigger decisions with e.g. track reconstruction from additional inner detector data. In addition to this RoI-based approach, a full scan of the coarse granularity L1-calorimeter-trigger data at L2 was developed in 2011 and activated in 2012. This allows for running unseeded anti-k_t jet reconstruction at L2 (L2FS), recovering inefficiencies in particular in multi-jet topologies and for close-by jets [76]. The L2 trigger reduces the event rate to about 6 kHz within a processing time of $\lesssim 40$ ms.

After full event reconstruction in the Event Builder, the EF runs offline analysis procedures using the complete, high granularity detector information. The EF provides a trigger decision within approximately 4 s, after which the event is moved to the CERN computing center for permanent storage and offline analysis. The final trigger and storage rate is about 400 Hz.

Trigger prescales

To reduce the rate of individual triggers, so-called *trigger prescales* can be applied at each trigger level. A prescale of e.g. 10 for a specific trigger item has the effect that only every 10th event passing this trigger item is transferred to the next trigger level. This allows to downscale few-object low-threshold triggers without disabling them, while keeping trigger bandwidth for more refined and rare trigger selections. Data acquired with a prescaled trigger does not correspond to the full integrated luminosity in 2012. Instead the integrated luminosities within short time intervals, called *lumi-blocks*, where running conditions are deemed constant,

need to be summed up, weighted with the applied trigger prescales P.

$$L_{\text{int}}^{\text{tot,prescaled}} = \sum_{i}^{\text{lumi-blocks}} \frac{L_{\text{int}}^{i}}{P_i}$$

The integrated luminosity of a prescaled trigger item is thus unique for that particular trigger.

CHAPTER 4

Reconstruction, strategy and selection

Particles produced in pp-collisions at the LHC are registered by the ATLAS detector as signals in the various subdetector components. These individual signals need to be combined within the subdetectors and across the detector systems in order to *reconstruct* the traversing particle. Reconstruction procedures combine detector information in a way that yield the best possible reconstruction efficiency for a particular particle type. This means that different reconstruction procedures are employed for the different particle types.

Since the reconstruction uses all available detector signals as input, the reconstructed signatures need to be further categorized as resulting from the considered particle type. They are thus *identified* as electrons, muons, etc. Particle identification is based on differences in the detector response between particle species, for example, differences in the shapes of energy depositions. Depending on the considered quantity, these can however be subject to more or less strong fluctuations per particle. Selections on these quantities therefore entail inefficiencies in the particle identification which grow with increased tightness of the applied criteria. At the same time however the chance of a misidentification of particles of a different type is reduced. The identification criteria e.g. for electrons are thus a trade-off between selection efficiency and purity.

The measured properties of the reconstructed and identified particles, in particular the energy, should correspond to the properties of the initial particles. Losses in less/non-instrumented detector regions, different response to electromagnetic or hadronic particle showers as a result of the non-compensating nature of the ATLAS calorimeter, etc. however spoil a simple 1:1 correspondence. The particle properties and, in particular, the particle energies therefore need to be *calibrated*. In general, the calibration comprises and combines simulation-based procedures and *in-situ* measurements comparing data and simulation as well as calibrations in different detector regions and/or particle types. For jets, the calibration relates the jet energy to the incident parton energy which is an important, but non-trivial undertaking and is accompanied with uncertainties which are often the dominant uncertainty contribution in measurements involving jets.

The reconstruction, identification and calibration procedures are developed within specific ATLAS performance groups and are managed centrally in the ATLAS software framework [77]. This ensures a consistent particle definition for all analyses within the collaboration. For the particle identification, different levels of tightness are made available which are applied according to analysis needs.

The reconstruction, identification and calibration of particles relevant to this analysis, i.e. electrons, $E_{\mathrm{T}}^{\mathrm{miss}}$– the detector signature of neutrinos – as well as jets are discussed in

Section 4.1. Section 4.2 gives an overview of the strategy for the W + jets analysis and Section 4.3 explains the selections based on electron, $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets in the W + jets analysis.

4.1 Particle reconstruction and calibration

Reconstruction, identification and calibration for electrons, $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets are summarized in the following based on Refs. [78, 79] for the electron, Ref. [80] for $E_{\mathrm{T}}^{\mathrm{miss}}$ and Refs. [81–84] for jets.

4.1.1 Electrons

Electrons are charged leptons with a mass of 511 keV and interact electromagnetically with matter. In the ATLAS detector, they therefore leave a curved track in the inner detector and a narrow energy deposition within the EM calorimeter. Making use of these two features, electrons are reconstructed within the fine granularity region of the EM calorimeter and within acceptance of the tracking system, i.e. within $|\eta| < 2.47$. Shower shapes in the calorimeter as well as track properties including transition-radiation signals in the TRT are used to identify electrons and distinguish them in particular from jets and converted photons. A clean sample of electrons can, for example, be obtained from the production of Z bosons with subsequent decay into electrons ($Z \to ee$[1]) which is used among other methods for the energy calibration of electrons. The procedures for reconstruction, identification and calibration are described in the following in more detail.

Reconstruction

The reconstruction of electrons combines the energy deposits in the electromagnetic calorimeter with a track in the inner detector, following a three-step procedure. First, electron seeds are collected from clusters in the EM calorimeter. Then, electron tracks are built and associated with the seed cluster, and finally, the electron kinematics is recalculated with more refined algorithms from the cluster energy and the associated track. In more detail, the three steps involve the following points.

- **Seed-cluster reconstruction:** A *sliding window* algorithm [85] with a window size of 3×5 towers in $\eta \times \phi$ searches for clusters with $E_{\mathrm{T}} > 2.5\,\mathrm{GeV}$. For this purpose, the EM calorimeter is partitioned into a grid of towers of size 0.025×0.025 in $\Delta\eta \times \Delta\phi$ which corresponds to the granularity of the EM calorimeter middle layer. For simulated electrons in the detector barrel region, the clusters are reconstructed with an efficiency of $\varepsilon > 99\,\%$.

- **Track association:** Tracks are found using pattern recognition and fitting, applying the general pion hypothesis for energy loss at material surfaces or the electron hypothesis. The latter allows for up to 30 % energy loss due to bremsstrahlung at each material surface. The good-quality track best matched to the seed-cluster is chosen as primary electron track.

- **Rebuilding of the electron:** Starting from the middle EM calorimeter layer, electron candidates are rebuilt using 3×7 (5×5) cells in $\eta \times \phi$ in the barrel (end-cap) region of the EM calorimeter. The lateral cluster sizes are optimized to contain the energy

[1] $Z \to ee$ indicates the decay $Z \to e^+e^-$, but the electron charges are omitted in the following for brevity.

deposition including e.g. the final state radiation in the barrel region, and the layer-wise combination allows to take the distribution of the deposited energy into account, while determining the exact cluster position.

The final electron energy after the calibration is given by the cluster energy, while the η and ϕ coordinates are determined by the primary electron track [2]. The electron transverse momentum is obtained from the cluster energy as $p_{\mathrm{T}} = E_{\mathrm{T}} = E/\cosh(\eta)$.

Identification

The electron identification needs to separate prompt electrons e.g. of W boson decays from narrow hadronic jets, photon conversion electrons and electrons of semi-leptonic heavy flavor hadron decays, for example. These are reconstructed by the electron reconstruction procedure as well, but in most cases do not correspond to the targeted signal. The reconstruction procedure has been optimized for prompt electrons, but by itself does otherwise not apply any specific identification requirement on the 'electrons'.

To differentiate between background (*fake*) electrons and signal electrons from a $W \to e\nu$ decay in this analysis, the most straight-forward approach is a cut-based electron identification [78]. This consists of a sequential application of requirements on a number of selected variables and has been in use in ATLAS since the beginning of data-taking [86–88]. The dependence of shower shapes and track characteristics on the energy and the amount of passive detector material upstream of the calorimeter is taken into account by adjustments on the identification requirements in 10 bins in $|\eta|$ and 11 in E_{T}.

Due to fluctuations in energy depositions and similarities between signal and fake electrons, every set of identification criteria however comes with a loss in efficiency for signal electrons if criteria are tightened, or an increase in fake rates if criteria are loosened. Therefore, a set of three selections – *loose*, *medium* and *tight* – exist which build on each other, and whose usage is up to the analysis. In the W + jets analysis presented in this thesis, the *tight* selection is used and the variables employed for this selection are listed in Table 4.1. An overview of the identification efficiency for the cut-based selections, measured in data of pp-collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ is given in Figure 4.1. The identification efficiency for the *tight* selection is around 70 %, with slightly higher values in the central detector region and lower towards the end-caps where the amount of material in front of the calorimeter increases. A rather strong dependence is observed with respect to the electron E_{T}, where the identification efficiency rises towards higher energies, starting from about 72 % for $E_{\mathrm{T}} > 25\,\mathrm{GeV}$. The combined reconstruction and identification efficiency for the *tight* electron identification was measured as 68 % in data from 2012 [78].

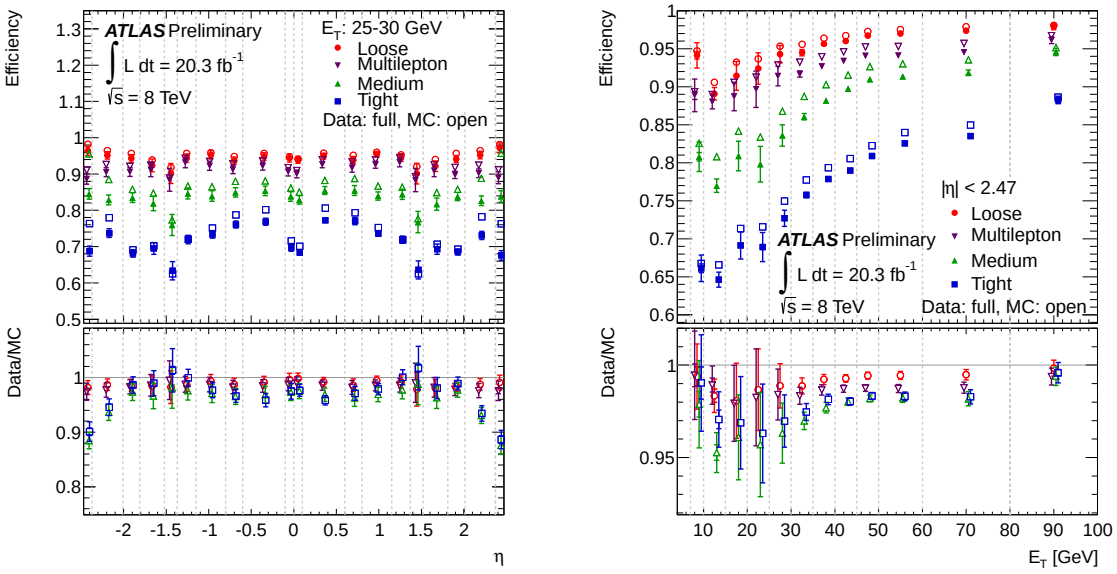
Calibration

The energy of the reconstructed and identified electron needs to be calibrated in order to ensure a correspondence of the energy the electron was produced with and the measured signal. This calibration is conducted using a step-wise MC-based and data-supplemented procedure, explained in detail in Ref. [79]. The key components are an improvement of the passive material modelling in front of the calorimeter, the intercalibration of the EM calorimeter layers, the electron cluster calibration, and the final energy determination from $Z \to ee$ events. In the following, the main points of these calibration steps are summarized.

[2] The exception are tracks without hits in the pixel or SCT detectors. In this case, the cluster position is used.

Category	Variable name	Description
Strip layer of EM calorimeter	w_{stot}	Shower width within $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$ (≈20 strips in η): $\sqrt{\frac{\sum E_i (i-i_{\max})^2}{\sum E_i}}$ (sums run over strips, $i_{\max}$ = strip index with highest energy)
	E_{ratio}	$\frac{E^{\text{largest}} - E^{\text{2nd largest}}}{E^{\text{largest}} + E^{\text{2nd largest}}}$, referring to the energy deposits in the cluster
Middle layer of EM calorimeter	$W_{\eta 2}$	Lateral shower width within 3×5 cells: $\sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$ (sums run over cells)
	R_η	$E^{3\times 7 \text{ cells}} / E^{7\times 7 \text{ cells}}$, centered at the electron cluster position
Back layer of EM calorimeter	f_3	$E^{\text{back-layer}_{em}} / E^{\text{total}_{em}}$
Hadronic leakage	$R_{\text{had(1)}}$	$E_{\text{T}}^{\text{had(1)}} / E_{\text{T}}^{\text{em-cluster}}$ with $E_{\text{T}}^{\text{had(1)}}$ denoting the energy in the (first layer of the) hadronic calorimeter, used for $0.8 < \lvert\eta\rvert < 1.37$ ($\lvert\eta\rvert < 0.8$ or $\lvert\eta\rvert > 1.37$)
Tracking and TRT identification	$n_{\text{b-layer}}$	Number of hits in the first layer of the pixel detector (=b-layer)
	n_{pixel}	Number of hits in the pixel detector
	n_{SCT}	Number of hits in the SCT
	d_0	Transverse impact parameter
	n_{TRT}	Number of hits in the TRT
	F_{HT}	$n_{\text{TRT}}^{\text{high-threshold}} / n_{\text{TRT}}$
Cluster-track matching	$\Delta\eta_1$	$\Delta\eta$ between cluster position in strip layer and extrapolated track
	$\Delta\phi_2$	$\Delta\phi$ between cluster position in middle layer and extrapolated track
	E/p	Cluster energy vs. track momentum
Conversions	isConv	Veto of electrons matched to reconstructed photon conversions

Table 4.1: List and description of variables used in the cut-based *tight* electron identification in 2012 data-taking.



(a) vs. η (b) vs. E_{T}

Figure 4.1: Electron identification efficiency measured in data from 2012 vs. η for an electron E_{T} within $25 - 30\,\mathrm{GeV}$ (a) and vs. electron E_{T} for $|\eta| < 2.47$ (b). The lower panel shows the data-to-MC efficiency ratios, which are applied as electron scale factors to the simulation in this analysis [78].

- **Passive material determination:** The fraction of energy deposited by an electron in the first or second calorimeter layer depends on the starting point of the shower with respect to the interaction vertex. More material upstream of the calorimeter leads to an earlier start of the shower and thus to an increase of the energy deposited in the first layer (E_1) compared to the second calorimeter layer (E_2). The ratio of the energy deposits in both layers is denoted as $E_{1/2}$. The difference of $E_{1/2}$ between data and simulation allows to estimate the simulated amount of passive material in front of the calorimeter. Using different particle types, even some information on the location of the material with respect to the interaction vertex can be obtained. Applying this method, an improved passive material simulation was obtained in 2012 with increases in the material amounts in the end-caps and in the region of service structures between the inner detector and the calorimeter cryostat.

- **Layer intercalibration:** The longitudinal segmentation of the EM calorimeter necessitates an equalization of energy scales, in order to be able to correctly extrapolate the detector response over the full electron p_{T}-range. This needs to be done prior to the overall energy scale determination and synchronized between data and simulation. The energy response among calorimeter layers in data is thus adjusted to the response in simulation. The correction factors are denoted as $\alpha_{1/2}$ and α_{PS}. $\alpha_{1/2}$ is derived as the ratio of the most probable energy deposits of muons in the calorimeter layers 1 vs. 2 for data compared to MC, i.e. $\alpha_{1/2} = \langle E_{1/2} \rangle^{\mathrm{data}} / \langle E_{1/2} \rangle^{\mathrm{MC}}$, and is measured from $Z \to \mu\mu$ decays. The correction is applied to the energies measured in the middle calorimeter layer, i.e. $E_2^{\mathrm{corr}} = E_2 \times \alpha_{1/2}$. The presampler (PS) energy scale α_{PS} is estimated from electrons from W and Z decays as the ratio of the PS energies (E_0) between data and MC simulation. The method makes use of the expected correlation between E_0 and $E_{1/2}$ under variations of the passive material upstream of the PS. Using muons and the correlated quantities, both methods are relatively insensitive to the amount of passive

material upstream of the calorimeter.

- **Electron cluster calibration:** Reconstructed electron clusters are calibrated by optimizing the estimate of the true particle energy from detector quantities, using a multivariate algorithm (MVA). The cluster energies are corrected for energy losses upstream of the calorimeter, in neighbouring cells, closeness of impact to cell boundaries, etc. - heavily relying on the assumed amount of passive material in front of the calorimeter. The calibration is applied both to data and simulation and is for electrons on the order of a few percent and only in the lower rapidity end-cap region reaches $10 - 30\,\%$ depending on E_{T}.

- **Final energy determination from $\mathbf{Z \to ee}$:** The final electron energy is calibrated in data as $E_{\mathrm{el}}^{\mathrm{data}} = E_{\mathrm{el}}^{\mathrm{MC}}(1 + \alpha_i)$ in bins (i) of pseudorapidity. The energy scale corrections α_i - along with energy resolution corrections c_i - are determined from a χ^2 minimization of differences in the electron pair invariant mass distribution from $Z \to ee$ events between data and simulation. The energy resolution term c is here defined as additional Gaussian constant term to the electron resolution: $\left(\frac{\sigma_E}{E}\right)^{\mathrm{data}} = \left(\frac{\sigma_E}{E}\right)^{\mathrm{MC}} \oplus c$. The simulation is varied by different values of α and c until data and MC simulation are found to match. The $Z \to ee$ energy calibration leads only to small adjustments of the electron energy scale and resolution with respect to previous calibrations. For the barrel region, very small uncertainties of $0.3 \cdot 10^{-3}$ and $0.3\,\%$ for energy scale and resolution, respectively, are obtained. The energy scale adjustments are then applied to data, while the resolution smearing is applied in the simulation.

4.1.2 $E_{\mathrm{T}}^{\mathrm{miss}}$

Weakly interacting particles like neutrinos cannot be detected within the ATLAS detector volume due to their low interaction cross section with matter. They however create a momentum imbalance in the detector, which is visible when summing the other measured signatures in the event. Since in a hadron collider the longitudinal momentum of the colliding partons within the proton is inherently unknown, vectorial momentum sums are only useful in the transverse detector plane. Here they should sum to zero - the transverse momentum of the partons prior to the collision, unless particles escaping detection like neutrinos have been present in the event.

The missing transverse momentum, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, is thus defined as the negative vectorial sum of the transverse momenta of all particles created in the collision. This is equivalent to $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ being the sum of vectorial transverse momentum imbalances, $\vec{E}_{\mathrm{T}}^{\mathrm{miss},i}$, considering each particle type i separately.

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \vec{E}_{\mathrm{T}}^{\mathrm{miss},e} + \vec{E}_{\mathrm{T}}^{\mathrm{miss},\gamma} + \vec{E}_{\mathrm{T}}^{\mathrm{miss},\tau} + \vec{E}_{\mathrm{T}}^{\mathrm{miss,jets}} + \vec{E}_{\mathrm{T}}^{\mathrm{miss},\mu} + \vec{E}_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$$

The first five terms are often grouped together as *hard term*, in contrast to the *soft term*, $\vec{E}_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, which collects detector signals that cannot be assigned to any of the reconstructed particle types.

The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, denotes the magnitude of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and ϕ^{miss} its azimuthal angle.

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{\left(E_{\mathrm{x}}^{\mathrm{miss}}\right)^2 + \left(E_{\mathrm{y}}^{\mathrm{miss}}\right)^2}$$

$$\phi^{\mathrm{miss}} = \arctan\left(E_{\mathrm{y}}^{\mathrm{miss}} / E_{\mathrm{x}}^{\mathrm{miss}}\right)$$

The electrons (e), photons (γ), the visible part of hadronically decaying taus (τ), jets and muons (μ) considered in the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation are reconstructed using the nominal particle reconstruction and calibration for this particle type. Depending on the particle type furthermore a few kinematic and identification criteria are applied. Relevant for this analysis is only the consideration of electrons and jets in the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation. For electrons in the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation, $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and the *medium* cut-based electron identification (see Section 4.1.1) is required to be passed. Jets are considered if they have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. If a jet is central and low p_{T} ($|\eta| < 2.4$ and $p_{\mathrm{T}} < 50\,\mathrm{GeV}$), it is furthermore required to have a jet vertex fraction > 0.25 - for an explanation of the jet vertex fraction, see Section 4.1.3. The final calibrations of electrons and jets passing the above criteria are propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation in the W + jets analysis.

The soft term, $\vec{E}_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, contains the remaining energy depositions not associated with any of the reconstructed particles. It comprises contributions from the hard parton interaction in the event as well as from the underlying event and pileup interactions. Several procedures how to calculate the soft term have been developed in ATLAS and have been in use for analyses based on 2012 data. For more details see Ref. [80].

The type of $\vec{E}_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ used in this analysis is based on tracks which are not associated to any of the other listed particles (TST - Track-based Soft Term). This has the advantage of good pileup stability, but misses contributions from soft neutral particles. The soft term is very difficult to measure, since it contributes only significantly in the absence of other high-energetic particles. It is therefore not calibrated which leads to an underestimation of the soft term contribution, relevant however only in the absence of jets or other well calibrated high-p_{T} particles. Nevertheless, this type of $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$-computation exhibits a pileup-robust performance, while maintaining reasonable linearity[3] and scale. The track-based soft term furthermore provides a better suppression of fake $E_{\mathrm{T}}^{\mathrm{miss}}$-contributions from mis-measurements of reconstructed particles than other calorimeter-based alternatives. This leads to an improved rejection of backgrounds without neutrinos in the final state, an important property for the background suppression in this analysis.

4.1.3 Jets

Jets are the fragmented and hadronized signatures of quarks or gluons which are generated in the collision. Since the LHC collides protons where quarks and gluons are inherently present, jets are the most abundant particle signature in ATLAS and their proper reconstruction and calibration is essential for any analysis either containing jets or rejecting jets.

In the ATLAS calorimeter, jets are visible as *groups of topologically-related energy deposits* [81]. These energy deposits are clustered first into local groups and then into jets – thus reconstructing the original jet. The reconstructed jet is then calibrated to the energy scale of truth jets[4] in the MC simulation. Since in the W + jets analysis, jets are an important ingredient, an overview of the jet reconstruction and calibration as used in this analysis is given in the following. Alternative jet definitions concerning jet size, calibration choices etc. are possible, but will not be discussed in the following. For more details, in particular on the jet calibrations, please refer to Refs. [81–84].

[3] The linearity is defined as the mean value of $\left(E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,\,Truth}}\right) / E_{\mathrm{T}}^{\mathrm{miss,\,Truth}}$ and is investigated as a function of $E_{\mathrm{T}}^{\mathrm{miss,\,Truth}}$.

[4] Truth jets are jets created from stable interacting particles in the simulation, i.e. particles with lifetimes larger than 30 ps, using the same jet reconstruction as applied on the clustered calorimeter energy deposits.

Reconstruction

The topological clustering groups energy deposits in the calorimeter together into topological clusters of calorimeter cells – so-called *topo-clusters* – using the (4/2/0)-scheme [85, 89, 90]. This means, cluster formation starts from a *seed cell* with a signal-to-noise ratio of $S/N > 4$. Neighbouring cells to the seed/forming cluster are added iteratively if they have $S/N \geq 2$ and finally, all calorimeter cells adjacent to the formed topo-cluster are enclosed as well. Since 2011 operations, the noise considered here includes both the electronic and pile-up noise.

$$\sigma_{\text{noise}} = \sqrt{\left(\sigma_{\text{noise}}^{\text{electronic}}\right)^2 + \left(\sigma_{\text{noise}}^{\text{pile-up}}\right)^2}$$

The topological clustering also includes a splitting step in order to improve the handling of close-by particles. For this purpose, the cells in the formed topo-clusters are searched for local maxima with energies above 500 MeV. These cells are then used as seeds for a new iteration of the topological clustering algorithm, leading to an improved set of topo-clusters.

The topo-clusters are calibrated using the local cluster weighting (LCW) method [89]. This corrects for the non-compensating nature of the ATLAS calorimeter as well as energy losses from out-of-cluster deposits or deposits in inactive detector areas. The key feature of the first correction is the classification or probability determination of topo-clusters as resulting from electromagnetic or hadronic showers. The calorimeter cells contained in the cluster are then reweighted accordingly, with the LCW corrections having been obtained from simulation of charged and neutral pions.

The procedure yields calibrated topo-clusters with energies equal to the sum of the contained calorimeter cell energies, zero mass and coordinates assuming an origin of the topo-cluster at the geometrical center of ATLAS.

The jet clustering defines jets in the final state using LCW topo-clusters as input. For a stable definition of jets, the jet algorithm is required to be collinear and infrared safe, i.e. it has to be independent of very low angle radiation emissions as well as soft gluon emissions between two jets. The anti-k_t algorithm [91] belongs to the group of *sequential recombination algorithms* and fulfils these requirements. It has become the standard jet algorithm used in ATLAS and CMS.

The anti-k_t algorithm groups input objects i, j repeatedly according to their distance d_{ij} until a stopping criterion is met. The distance measure is defined as:

$$d_{ij} = \min\left(\frac{1}{p_{\text{T},i}^2}, \frac{1}{p_{\text{T},j}^2}\right) \cdot \frac{\Delta R_{ij}^2}{R^2}\,.$$

$p_{\text{T},i}$ here denotes the transverse momentum of the input object i, ΔR_{ij} specifies the spatial distance between the objects with $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and R is the selectable distance parameter of the algorithm. If the distance to another object d_{ij} is smaller than the *beam distance* $d_{iB} = \frac{1}{p_{\text{T},i}^2}$, i.e. $d_{ij} < d_{iB}$, then the two objects are combined. This means, their four vectors are added and the new combined object is reinserted into the list of input objects. If however $d_{ij} > d_{iB}$ for all j, then the object is declared a jet and removed from the list of input objects. The algorithm starts with the two input objects with smallest d_{ij} and stops when all input objects have been combined and declared as jets.

The anti-k_t algorithm therefore starts with the most energetic and close-by objects and grows outwards collecting softer contributions. This results in almost circular jets in $y - \phi$

space (see e.g. [92]) compared to alternative algorithms of the sequential recombination type like the k_t [93–95] or the Cambridge/Aachen (C/A) [96, 97] algorithm.

Jets used in this analysis are reconstructed with the anti-k_t algorithm with a distance parameter of $R = 0.4$. The jet algorithm is implemented in the FastJet software [98, 99] and is run within the ATLAS software framework. Truth jets which are utilized in the general jet calibration as well as in the unfolding prodedure and the fiducial cross section measurement of this analysis (see Chapter 6) are determined using the same anti-k_t jet clustering, but with stable truth particles instead of calibrated topo-clusters as input.

Calibration

The response of reconstructed jets needs to be calibrated and equalized between data and MC simulation, relating the reconstructed jet energy to the truth jet energy. This is achieved in ATLAS using a combination of MC-based and data-driven methods. The jet calibration for data recorded in 2012 is documented in Refs. [81–84] and a short overview following mainly Ref. [81] is given in the following. A distinction is made between the jet energy scale (JES), i.e. the mean of the jet response distribution, and the jet energy resolution (JER), i.e. its standard deviation. Both need to be measured and, if required, calibrated individually. The JES calibration is however more involved and its uncertainty often dominates measurement uncertainties of analyses involving jets.

The JES calibration proceeds in a step-wise manner, starting with the calibration of the topo-clusters using the LCW method explained above. After jet reconstruction, the following calibration steps are applied:

- **Origin Correction:** Jets after reconstruction point by default to the center of the detector, as a result of the standard assigned direction for the topo-clusters. A better assumption for the origin of the jet is however the primary vertex[5]. The jet direction is corrected without affecting the jet energy. Since the luminous size of the beamspot in 2012 was around $4 - 5.5\,\text{cm}$ in the z-direction, while the longitudinal vertex position resolution is about $30\,\mu\text{m}$ [72], this leads to a large improvement of the jet resolution in η.

- **Pile-up Correction:** The jet energy is corrected for contributions from pile-up using the median energy density ρ in the event multiplied with the area A of the jets, following the idea of jet area corrections [100]. ρ is determined from jets reconstructed with the k_t algorithm in the central region ($|\eta| < 2$). The k_t algorithm starts clustering soft energy deposits first, and therefore in particular captures the pile-up in the event. The area A of the reconstructed (anti-k_t) jet is determined using a ghost-clustering procedure. This correction catches most of the pile-up dependence, and residual dependencies on in-time and out-of-time pile-up are parametrized depending on the number of primary vertices, N_{PV}, and the average number of interactions per bunch-crossing, $\langle\mu\rangle$. The pileup-corrected jet p_T is thus given as follows:

$$p_\text{T}^\text{corr} = p_\text{T} - \rho \cdot A - \alpha \cdot (N_{PV} - 1) - \beta \cdot \langle\mu\rangle\,,$$

  where α and β are constants determined from MC simulation.

[5]The primary vertex is defined as the vertex with the highest $\sum p_\text{T}^2$ of tracks with track $p_\text{T} > 400\,\text{MeV}$.

- **MC-based Energy Scale Correction:** Jets are calibrated to the LCW+JES scale using a jet response calibration function $f_{\text{calib}}(E^{\text{jet}}_{\text{LCW}})|_\eta$:

$$E^{\text{jet}}_{\text{LCW+JES}} = \frac{E^{\text{jet}}_{\text{LCW}}}{f_{\text{calib}}(E^{\text{jet}}_{\text{LCW}})|_\eta}$$

  $f_{\text{calib}}(E^{\text{jet}}_{\text{LCW}})|_\eta$ is determined per η bin[6] from isolated jets in a MC simulation of inclusive jet production. It relates the jet energy $E^{\text{jet}}_{\text{LCW}}$ to the average jet response $\left\langle \mathcal{R}^{\text{jet}} \right\rangle$ which is the mean of the response distribution $E^{\text{jet}}_{\text{LCW}}/E^{\text{jet}}_{\text{truth}}$. For more details see Ref. [89]. In addition a small correction to the already origin-corrected jet η is applied to compensate a small bias with respect to η_{truth} in poorly instrumented calorimeter regions.

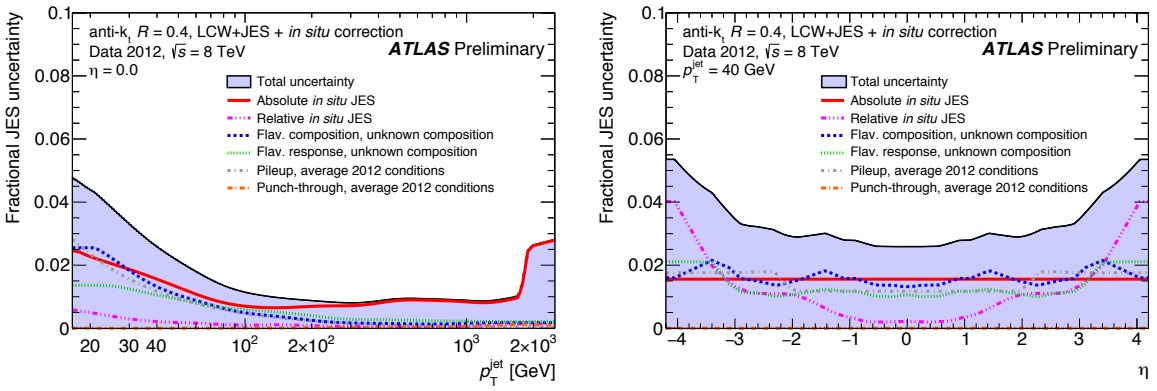
- **Global Sequential Correction:** The Global Sequential Correction (GSC) is a series of multiplicative corrections to the calibrated jet. They are based on jet observables and aim to improve performance without modifying the mean jet energy response. The two main targets are the flavour dependent jet response and high-p_{T} jets not fully contained in the calorimeter. The first arises from differences in the response to quark and gluon initiated jets and contributes a large uncertainty in the JES calibration. The difference is caused by a tendency of gluon initiated jets towards higher number of constituent particles with lower constituent energies and a wider transverse shower profile. The differences can be accessed and corrected, for example, using information on the tracks associated to the jet. The incompletely contained high-p_{T} jets are detected with *punch-through* to the muon system and are corrected based on the amount of activity in the muon system behind the jet. For more information, see Ref. [84].

- **In-situ JES Calibration:** Reference particles which balance the jet p_{T} in the transverse plane are used to measure the jet energy scale in both data and MC *in-situ*. These reference particles can either be a second jet, employed for the η-intercalibration from the central to the forward pseudorapidities, a Z boson, a photon or a system of low-p_{T} jets recoiling against the tested jet. In all cases, the double-ratio

$$\frac{\mathcal{R}_{\text{data}}}{\mathcal{R}_{\text{MC}}} = \frac{\left\langle p^{\text{jet}}_{\text{T}}/p^{\text{ref}}_{\text{T}} \right\rangle_{\text{data}}}{\left\langle p^{\text{jet}}_{\text{T}}/p^{\text{ref}}_{\text{T}} \right\rangle_{\text{MC}}}$$

  is measured to define the residual correction which is applied to data. The in-situ measurements cover different, yet partially overlapping jet p_{T} ranges: $20\,\text{GeV} \leq p_{\text{T}} \leq 200\,\text{GeV}$ for Z+jet, $30\,\text{GeV} \leq p_{\text{T}} \leq 800\,\text{GeV}$ for γ+jet and $300\,\text{GeV} \leq p_{\text{T}} \leq 1700\,\text{GeV}$ for multi-jets and are detailed in Refs. [82, 83]. The Z+jet, γ+jet and multi-jet balance in-situ measurements are combined using a weighted average for the in-situ JES calibration. At very high p_{T} ($> 1500\,\text{GeV}$), the statistics for the in-situ methods becomes limited. The calibration is thus extended by the single hadron response method which constrains the uncertainty in the calibration at high jet p_{T}. Here, the jet response is determined from the response of the single hadrons which constitute the jet. The method is described in detail in Ref. [101].

[6] The coordinate system with origin in the geometric detector center is used here as it relates directly to the region in the calorimeter.



(a) (b)

Figure 4.2: Total JES uncertainty as a function of p_T at $\eta = 0$ (a) and as a function of η for $p_T = 40\,\text{GeV}$ (b). The label *Absolute in situ JES* refers to the uncertainty from the Z+jet, γ+jet and multi-jet balance measurements, while *Relative in situ JES* denotes the uncertainty from the dijet η-intercalibration. The label *unknown composition* for the flavor uncertainties indicates that the composition of quark and gluon initiated jets is unknown and thus a composition of 50:50 quark:gluon jets is assumed with an uncertainty of 100 % [81].

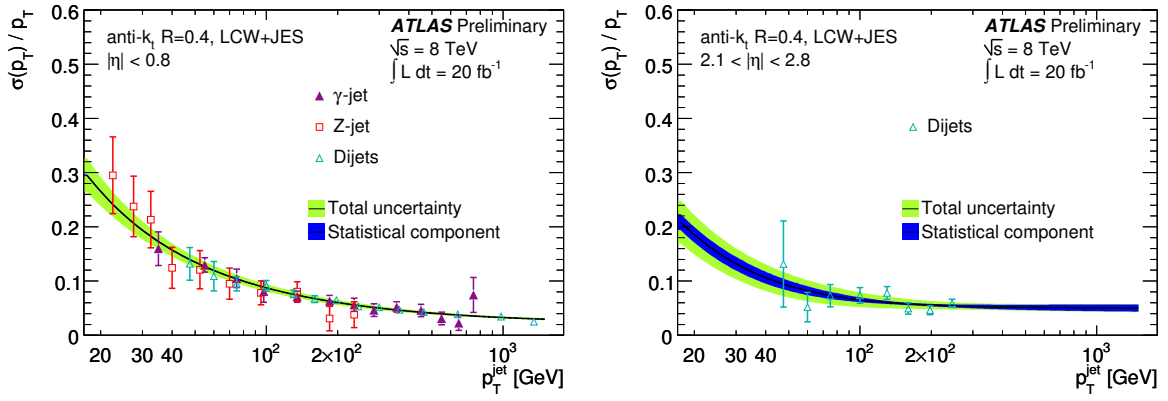
The uncertainty in the JES calibration results from a total number of 65 sources of uncertainty, among which 56 alone are statistical or systematic components of the in-situ measurements. The η-intercalibration measurement contributes two uncertainty sources. Four sources of uncertainty are associated to the pile-up corrections and one source of uncertainty each are the jet flavour composition and response (light vs. gluon initiated jets) as well as the single-hadron response. The uncertainty in the calibration of the final jet energy scale is shown in Figure 4.2. The total uncertainty is found to be about 3.5 % for a 30 GeV-jet at central pseudorapidity. It decreases to approximately 1 % for jets with $p_T > 100\,\text{GeV}$ and rises again to about 2.5 % at jet p_T around 1.5–2 TeV where the statistics from the in-situ JES calibration runs out. In the central pseudorapidity the JES uncertainty is roughly constant at 2.5–3 % for a 40 GeV-jet and rises towards the forward region to approximately 5.5 %.

Since the in-situ uncertainties depend only on a single parameter, p_T, the number of components in the JES uncertainty can be reduced by a diagonalization of the covariance matrix. This yields a new reduced set of uncertainties from the eigenvectors and eigenvalues of the covariance matrix, decreasing the 56 in-situ components to a new set of 6 uncertainty components. The first five are the uncertainty components with largest eigenvalues, while the last combines the remaining sources of uncertainty into a residual term. This diagonalization diminishes the total number of JES uncertainties to 15 with very little loss of correlation information, but much simpler treatment in the analysis.

The resolution of the jet energy (JER), i.e. the precision of the jet energy measurement can be parametrized as follows.

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C$$

N denotes the effect of electronic and pile-up noise, S the stochastic and C the constant term. The noise term is a significant component of the JER at low p_T. It has been measured in 2012 data with two novel methods in order to better constrain the impact of increased levels of pile-up and the resulting noise in 2012 pp-running. A combined fit of the JER with



(a) $|\eta| < 0.8$ (b) $2.1 < |\eta| < 2.8$

Figure 4.3: Jet resolution as a function of p_{T} for central jets (a) and more forward jets (b). The combined fit to data from three (one) in-situ measurements is shown with its associated statistical and total uncertainties for the central (more forward) jets [81].

fixed noise term is then realized in data from the in-situ measurements of the dijet and the Z/γ+jet balances. The measured and fitted jet resolution is shown in Figure 4.3 for a central and a more forward pseudorapidity bin. For the most central jets, a small JER uncertainty of approximately 2 % for a jet p_{T} of 30 GeV and less than 1 % uncertainty for jets above 100 GeV is obtained. For more forward jets, the JER itself reduces, however the uncertainty in the JER measurement grows because also only the dijet in-situ measurement remains with reasonable statistics in these η regions to constrain the fit.

Pile-up Jets

In addition to jets originating from the hard parton-parton interaction, simultaneous softer proton-proton collisions produce jets which are reconstructed in the same way. These pile-up jets do not contain information on the interesting process and thus have to be suppressed. At central rapidity, tracking information can be used for this purpose. The *jet vertex fraction* (JVF) measures the fraction of jet-associated track p_{T} where the tracks can be extrapolated back to the primary vertex (PV) compared to the p_{T}-sum of all associated tracks, i.e.

$$\mathrm{JVF} = \frac{\sum_i^{\text{PV jet-tracks}} p_{\mathrm{T}}^i}{\sum_i^{\text{all jet-tracks}} p_{\mathrm{T}}^i}.$$

The tracks are associated to the jet using the *ghost association procedure* [100, 102] and have $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ and $|\eta| < 2.5$, the latter in order to remain within the tracking acceptance. For more details see Ref. [103].

Since pile-up jets usually have low p_{T}, JVF requirements are recommended for jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ which are within tracking acceptance, i.e. $|\eta| < 2.4$.

4.2 Analysis strategy

The aim of this analysis is the measurement of the differential cross section for $W^{\pm}$ boson production in association with jets as well as the measurement of the W^+/W^- ratio.

The W + jets analysis is executed for the charge independent W, the W^+ and the W^- selection in parallel and as a function of the jet multiplicity. Events where a W boson is

produced are identified in this analysis via the decay $W \to e\nu$, i.e. by the selection of a single isolated electron and $E_{\mathrm{T}}^{\mathrm{miss}}$. The charge of the W boson is determined from the charge of the selected electron, and events passing the W boson selection are split exclusively in W^+ and W^-. Jets which are produced in association with the selected W are counted and the event is classified according to this as $W + (\geq)n$ jets, if exactly (at least) n jets are present – also denoted as exclusive (inclusive) jet multiplicities. Of special importance for PDF constraints are the jet multiplicities with at least one or at least two jets. In this thesis, distributions are therefore mainly shown for these jet multiplicities, but numbers are also given up to higher jet multiplicities.

The applied selection of W bosons has been optimized for high signal purity and low background contribution and is explained in detail in Section 4.3. In the presence of at least 1 jet, more than 80 % of the selected events in data are from W boson production.

The largest background sources are multi-jet production and also $t\bar{t}$ production for higher jet multiplicities. The contribution from multi-jet production after the W selection is estimated with a data-driven approach. A specialized event selection in data is needed as control sample for the data-driven multi-jet estimate and is outlined in Section 4.3.2. The background from $t\bar{t}$ production is estimated from MC simulation and its predictions are validated against data in a $t\bar{t}$-enriched selection which is listed in Section 4.3.3. Further minor backgrounds are estimated from MC simulation. The sum of the W+jets signal prediction and the backgrounds is denoted in the following as *total SM prediction*. Its general agreement with data is tested by comparison to data for a large number of kinematic distributions and jet multiplicities, in order to detect potential problems in the data description. Background estimation and comparison to data are discussed in Chapter 5. Technically, distributions like the leading jet p_{T} are measured in 2 dimensions with one axis being the variable – leading jet p_{T} – and the second axis corresponding to the jet multiplicity. Distributions for individual jet multiplicities as mentioned above are determined as projections of the 2-dimensional distributions.

For a selection of distributions which have been identified as potentially PDF-sensitive, data with subtracted backgrounds are *unfolded* from the detector level to particle level. An iterative Bayesian method is employed for the unfolding and tested in-depth to ensure a proper performance of the unfolding procedure. Systematic uncertainties calculated at the detector level are propagated through the unfolding. Uncertainties in the unfolding procedure are considered in addition. The unfolding method and tests are explained in Chapter 6, the handling and size of systematic uncertainties in Chapter 7. In the background estimation and the unfolding prodedure exclusive jet multiplicities are mostly used and distributions are therefore often displayed as a function of exclusive n_{jets}.

Differential cross sections for W, W^+ and W^- production in association with jets are measured at the particle level in a fiducial phase space similar to the selection at detector level. For the cross section measurement, inclusive jet multiplicities are most significant, in particular due to improved statistics in the high-p_{T} tails. Therefore, the cross section measurements are presented in association with ≥ 1 and ≥ 2 jets. The ratio of W^+/W^- in association with jets is determined from the measured differential cross sections and provides the best access among the obtained results to valence quark PDFs in the proton. The results of the cross section and ratio measurements are presented in Chapter 8 and Chapter 9, respectively.

4.3 Event selection

The selection of W + jets events aims at achieving a high signal purity, while suppressing the backgrounds from other physics processes. It mainly follows the selection in previous ATLAS measurements of W + jets production [7, 8], but adjusts criteria needed for improved background suppression. As a result of the roughly an order of magnitude larger multi-jet cross section, contamination from this process is a particular issue – also in view of the increased level of pile-up in 2012 compared to previous measurements. Stringent selections on the $W \to e\nu$ decay products, i.e. a well-isolated electron and the pile-up stable $E_{\mathrm{T}}^{\mathrm{miss}}$ type used in this analysis, provide improved rejection for this background.

The selection can be differentiated into event-, trigger- and particle-based selections. In general, the event-based selection guarantees good data-quality and thus proper particle reconstruction, while the trigger-based selection ensures that the event is recorded and the particle-based selection specifies the W + jets final state.

Event-based selection

The data used in the presented analysis has been delivered by the LHC and recorded by the ATLAS experiment in 2012. The cumulative luminosity vs. time in 2012 is shown in Figure 4.4. Inefficiencies in the data-aquisition system and the time between a protective stand-by state of the ATLAS detector until full detector operation with operating high-voltages produces a small reduction of the recorded luminosity compared to the delivered luminosity. The data-taking efficiency is with about 93 % however rather good.

The recorded data is examined for malfunctioning detector components or other issues during data-taking based on lumi-blocks[7]. This information is stored in so-called *Good Run Lists* [105] and can differ depending on which detector components are required to be fully functional for a given analysis. The Good Run List used in this analysis [106] requires the entire detector to be fully operational and leaves a total integrated luminosity of $20.3\,\mathrm{fb}^{-1}$ for analysis. The amount of data thus declared as *good for physics* is indicated in Figure 4.4 as well.

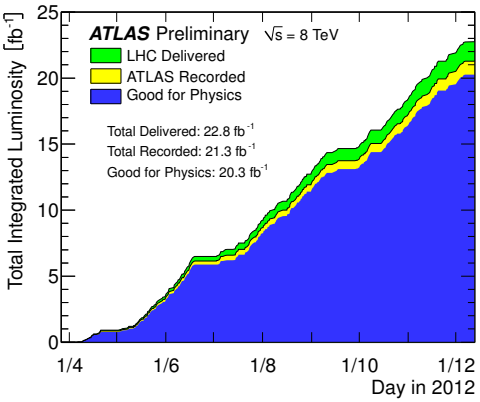


Figure 4.4: Total integrated luminosity of pp-collisions in 2012. Shown are the luminosities delivered by the LHC (green), recorded by ATLAS (yellow) and considered good for physics analysis (blue) as a function of time [104].

Further data quality selection is done on event-by-event basis. Corrupt events with errors e.g. from noise bursts in the liquid argon calorimeter, errors or trips in the tile calorimeter as well as incomplete event-readouts after partial detector restarts are rejected. To ensure a well measured primary vertex [107, 108], the presence of at least 3 tracks with $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ associated to the vertex is required. Furthermore, events with badly measured jets are discarded as they are likely to distort the $E_{\mathrm{T}}^{\mathrm{miss}}$. Badly measured jets here are jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ failing to satisfy the *looser* quality criteria defined in [109] or

[7] Lumi-blocks denoting short time-intervals with stable detector and beam conditions had a duration of about 60 s in 2012.

jets close to a specific noisy Tile calorimeter cell which was not masked in the reconstruction at the beginning of 2012 data-taking.

Trigger-based selection

Events in this analysis were triggered and subsequently stored for data analysis by selecting single electrons in the final state. In order to keep the trigger rate at a reasonable level, electron identification criteria together with additional electron selections like isolation or increased electron p_{T} have to be applied. The used trigger items are defined at Event Filter (EF) and are called `EF_e24vhi_medium1` and `EF_e60_medium1`. They require a single isolated electron with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ or a single (un-isolated) electron with $p_{\mathrm{T}} > 60\,\mathrm{GeV}$. In both cases, an electron identification similar to the offline *medium*, denoted as `medium1`, is employed in the trigger. The labels `v` and `h` refer to additional selections applied at Level-1 of the trigger. *Varied thresholds* (`v`) dependent on the calorimeter region implement an approximate dead material correction in the trigger, while the *hadronic veto* (`h`) suppresses an extension of the electromagnetic shower to the hadronic calorimeter. The latter causes inefficiencies towards larger electron p_{T}, hence the supplementation with the 60 GeV-trigger.

The letter `i` indicates the application of an isolation criterion around the electron at EF. Tracks which surround the electron within a radius $\Delta R < 0.2$ together must have less than 10 % of the electron E_{T}. This sum of transverse track momenta is called $p_{\mathrm{T}}^{\mathrm{cone20}}$ and closely follows a similar definition offline, except for small differences in the trigger-level track selection including a minimum p_{T} of 1 GeV for the tracks and slightly different hit requirements.

The electron selected offline (see below) is required to be matched to the electron identified at the trigger level within $\Delta R < 0.1$.

Particle-based selection

Events are selected offline, i.e. after data-taking and particle reconstruction, mainly by requiring the presence of exactly one well-isolated high-quality electron and sufficient $E_{\mathrm{T}}^{\mathrm{miss}}$. The full list of selection requirements for the W + jets signal region (SR) is summarized in Table 4.2.

The electron is demanded to fulfil $E_{\mathrm{T}} > 25\,\mathrm{GeV}$, $|\eta| < 2.47$ (excluding $1.37 < |\eta| < 1.52$) and to be of good quality. The $|\eta|$-region between 1.37 and 1.52 is excluded because of poor instrumentation in this overlap-region between the electromagnetic barrel and the end-cap calorimeter leading to worse electron purity. The electron quality criterion ensures a proper functioning of the detecting hardware. The electron must furthermore pass the *tight* electron identification and be isolated, i.e. the sum of calorimeter deposits and of track transverse momenta around the electron must be low. A good isolation of the selected electron contributes to the suppression of multi-jet events where a narrow jet can imitate the electron signature but additional particles apart from the main energy deposit or track in the jet increase the surrounding activity. The definition of the electron isolation used in this analysis is based both on tracking and calorimeter isolation.

The track-based isolation $p_{\mathrm{T}}^{\mathrm{cone30}}$ sums the transverse momenta of tracks within a cone of radius $\Delta R < 0.3$, excluding the electron track. The tracks considered in the sum have $p_{\mathrm{T}} > 400\,\mathrm{MeV}$, originate from the primary vertex and possess a suffient number of hits in the inner detector, in order to be well measured. The $p_{\mathrm{T}}^{\mathrm{cone30}}$ is required for this analysis to be smaller than 7 % of the electron E_{T}[8].

[8] For consistency with the trigger, the offline-computed $p_{\mathrm{T}}^{\mathrm{cone20}}$ has to be less than 10 % of the electron E_{T},

Pre-selection		
Data quality	Good Run List No data-corruption in the LAr and Tile calorimeters and complete event readout	
Vertex	Primary vertex with $n_{\text{tracks}} \geq 3$	
Valid $E_{\text{T}}^{\text{miss}}$	No jet with $p_{\text{T}} > 20\,\text{GeV}$ failing the *looser* quality criteria and no jet close to a noisy Tile calorimeter cell	
Triggers	`EF_e24vhi_medium1`, `EF_e60_medium1`	
$W \to e\nu$ Signal Region		
Electron	Exactly 1 electron with:	
	E_{T}	$> 25\,\text{GeV}$
	Pseudorapidity	$\lvert\eta\rvert < 2.47$ (excluding $1.37 < \lvert\eta\rvert < 1.52$)
	ID	*tight*
	Trigger matching	Matched to triggered electron within $\Delta R < 0.1$
	Isolation	$p_{\text{T}}^{\text{cone30}}/E_{\text{T}} < 0.07$ and $E_{\text{T}}^{\text{cone30}}/E_{\text{T}} < 0.14$
	Impact parameters	$\lvert d_0/\sigma_{d_0}\rvert < 5$ and $\lvert z_0\rvert \cdot \sin(\theta) < 0.5$
Z veto	No second electron with:	
	E_{T}	$> 20\,\text{GeV}$
	Pseudorapidity	As above
	ID	*medium*
Neutrino	$E_{\text{T}}^{\text{miss}}$	$> 25\,\text{GeV}$
W boson	m_{T}	$> 40\,\text{GeV}$
$t\bar{t}$ suppression	b-Jet veto	No b-jet identified with the MV1 b-tagger, 60 % efficiency working point

Table 4.2: Event selection criteria for the W + jets measurement in the $W \to e\nu$ decay channel.

The calorimetric isolation $E_{\text{T}}^{\text{cone30}}$ is defined as the sum of transverse energies from topo-clusters at the electromagnetic scale within $\Delta R < 0.3$ around the electron cluster position, again excluding the core electron contribution. The calorimeter isolation is corrected for leakage of the electron energy into the isolation cone as well as pile-up. The latter is achieved using a similar technique as described for the pile-up correction in the jet calibration (see Section 4.1.3). For more information on the electron isolation variables see Refs. [78, 110]. For the W + jets analysis, $E_{\text{T}}^{\text{cone30}}$ (with corrections applied) is required to be less than 14 % of the electron E_{T}.

Other sources of electrons which do not originate from the W boson decay and thus allow multi-jet events to enter the W + jets selection, are electrons from heavy flavour decays. Heavy flavour mesons or baryons produced in multi-jet events can decay with an electron in the decay chain. These electrons mostly are surrounded by the other decay products and are thus reduced by the isolation requirements above. Since heavy flavour mesons and baryons usually travel away from the the primary interaction vertex before decaying and thus create a secondary vertex, the impact parameters of the electron track to the primary vertex can be used to suppress this type of fakes even further. Requirements both on the transverse and the longitudinal impact parameter of the selected electron are therefore applied. The

as well.

significance of the transverse impact parameter $|d_0/\sigma_{d_0}|$ is required to satisfy $|d_0/\sigma_{d_0}| < 5$, while the longitudinal impact parameter z_0 has to fulfil $|z_0| \cdot \sin(\theta) < 0.5$. The combination of impact parameter and isolation requirements has been optimized in this analysis for best multi-jet background suppression.

The background from $Z \to ee$ events is minimized by vetoing events with a second good quality electron with p_{T} larger than 20 GeV, passing the same pseudorapidity requirements and the *medium* electron identification. Isolation criteria or impact parameters are not considered for this veto. $Z \to ee$ events passing the selection thus mainly correspond to cases where the second electron was not reconstructed and/or identified either due to inefficiencies or due to being out of the detector acceptance.

Leptonically decaying W bosons (here $W \to e\nu$) are selected further via requirements on the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, and the transverse mass, m_{T}. The transverse mass of the W boson is defined as follows.

$$m_{\mathrm{T}} = \sqrt{2 \cdot E_{\mathrm{T}}^{\mathrm{el}} \cdot E_{\mathrm{T}}^{\mathrm{miss}} \cdot (1 - \cos \Delta\phi_{\mathrm{el},E_{\mathrm{T}}^{\mathrm{miss}}})}$$

where $E_{\mathrm{T}}^{\mathrm{el}}$ is the transverse energy of the selected electron and $\Delta\phi_{\mathrm{el},E_{\mathrm{T}}^{\mathrm{miss}}}$ is the azimuthal angle between the electron and the direction of the missing transverse momentum ($\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$). Requirements of $E_{\mathrm{T}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$ and $m_{\mathrm{T}} > 40\,\mathrm{GeV}$ are applied, and significantly increase the contribution from W production compared to multi-jet events in the selection.

For the higher jet multiplicities, the suppression of contributions from $t\bar{t}$ production is of vital importance, as this was a limiting factor in the previous W + jets measurement [8]. Produced top-quarks decay almost exclusively via $t \to bW$ and can therefore be identified in the detector by the properties of b-quark originated jets like a slightly displaced secondary vertex.

For this analysis, b-jets are defined as reconstructed jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.5$, passing the *looser* jet quality criteria and identified (tagged) using the `MV1` algorithm. The `MV1` algorithm employs an artificial neutral network based on three simpler algorithms which utilize impact parameters, secondary vertex information and fitting algorithms taking advantage of b- and c-hadron decay topologies. For more information see Refs. [111, 112]. The b-tagging working point which achieves an efficiency of 60 % for real b-jets is used in this analysis. For the W + jets selection, events are rejected if any b-jet is found in the event.

Weighting of the MC simulation

While the MC simulation in general describes data correctly, it needs to be adjusted to consider the actual data-taking conditions in 2012 and to accommodate small data-MC differences in e.g. particle reconstruction and identification efficiencies. Two types of weights are therefore applied in the MC simulation to improve the data-MC agreement.

The first type of weights are based on global event quantities. These are the generator weights calculated by the respective MC generator to adjust the theoretical prediction, as well as pile-up and vertex position weights. The latter two correct the $\langle\mu\rangle$ distribution and the distribution of the vertex position in z-direction in MC to that present in data in 2012. In the case of the pile-up weights, the average activity per event is found to agree in data and the Pythia8 MC simulation which models the overlaid pile-up interactions if events with a given $\langle\mu\rangle$ in MC are compared to data with $1.09 \cdot \langle\mu\rangle$. The pile-up weights therefore match the $\langle\mu\rangle$ distribution in MC simulation to the one measured in data, but adjusted by this factor. The vertex position adjustments in z-direction have been determined in $Z \to ee$ events and have

only very small effect in the analysis. The global weights are considered in the scaling of the simulation to the integrated luminosity of $20.3\,\mathrm{fb}^{-1}$.

In addition to the global weights, simulated events are weighted to correct electron and jet b-tagging efficiencies to those measured in data. The weights depend on the kinematics of the selected electron or jet and are called *scale factors* (SF).

For the electron, scale factors for electron reconstruction, identification, trigger and isolation efficiencies are considered. The scale factors for electron reconstruction efficiencies have been measured in $Z \to ee$ events and are mostly within a few per mill of unity, but at most $1-2\,\%$ away from one. Scale factors for electron identification efficiency are determined and combined from $Z \to ee$, $Z \to ee\gamma$ and $J/\Psi \to ee$ events. They are in general larger than reconstruction efficiency scale factors, i.e. on the order of a few percent, but for the *tight* identification reach up to $10\,\%$ at low p_T and large η, or in extreme cases $20\,\%$ for the largest η values. More information on these measurements can be found in Ref. [78]. Trigger efficiencies and scale factors as well as isolation scale factors were measured as well in $Z \to ee$ events. While the trigger scale factors are found to be roughly of the same size as the identification scale factors, the impact of isolation scale factors is small.

The tagging efficiencies for b-jets or c-jets (= jets originating from a c-quark) as well as the mistagging rates for light jets have also been measured in dileptonic $t\bar{t}$ events, an inclusive jets sample with associated D^* mesons or simply an inclusive jet sample, respectively. Derived scale factors correcting the MC simulation to measured data efficiencies are provided for use in analyses. In general, the scale factors for b-tagging of b-jets, c-jets or mistagging of light jets are consistent with unity within about $2\,\%$ or $8-15\,\%$ uncertainty in the first two cases, and slightly larger than unity with a precision of $15-43\,\%$ in the last case. Since mistag rates are however small, very little effect is expected from this. For more details on these measurements see Refs. [111, 112]. The b-, c- and light-jet tagging SFs are applied to the MC simulation in the $W+\,$jets analysis for any selection requiring the presence or the absence of b-jets.

4.3.1 $W+$ jets signal region

Events in the $W+\,$jets SR are categorized according to their jet multiplicity and the charge of the W boson.

For the jet multiplicity classification, jets are counted which pass the following requirements. They must have $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 4.4$. To suppress pile-up jets, a jet vertex faction larger than $50\,\%$ is required for jets with $p_\mathrm{T} < 50\,\mathrm{GeV}$ and $|\eta| < 2.4$. Jets which overlap with the selected electron within $\Delta R < 0.5$ are not considered. The rather large ΔR-requirement is important to ensure that jets do not share calorimeter cells with the selected electron which would distort the energy measurement of both. Reconstructed jets passing these criteria are called *signal jets* and classify the event as $W+0$,1,2,... jets.

The distinction into W^+ and W^- events is realized based on the charge of the selected electron.

The number of events selected in the $W+\,$jets SR by applying the explained requirements are listed in Table 4.3 after the different selection requirements and in Figure 4.5 dependent on the data period.

About 67 Mio W events remain after the full $W+\,$jets selection in the SR among which about 39 Mio are W^+ and about 28 Mio are W^-. $84.4\,\%$ of the selected data events have been produced without any signal jet associated, $11.6\,\%$ are generated in association with exactly 1 signal jet, while $4.1\,\%$ have 2 or more signal jets created in the collision.

Selection	Events remaining	Fraction rejected [%]
Recorded (*Egamma* stream)	732 243 734	–
Good Run List (*)	–	4.2
Data quality (*)	–	0.2
Vertex (*)	–	0.3
Single electron selection (no 2nd electron, incl. trigger) (**)	102 495 690	–
Valid $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement	102 381 454	0.1
$E_{\mathrm{T}}^{\mathrm{miss}}$ selection	70 572 946	31.1
m_{T} selection	67 505 993	4.3
b-Jet rejection	66 806 152	1.0
W^+	38 907 623	58.2
W^-	27 898 529	41.8
W without jets	56 355 106	84.4
$W + 1$ jet	7 743 988	11.6
$W + \geq 2$ jets	2 707 058	4.1

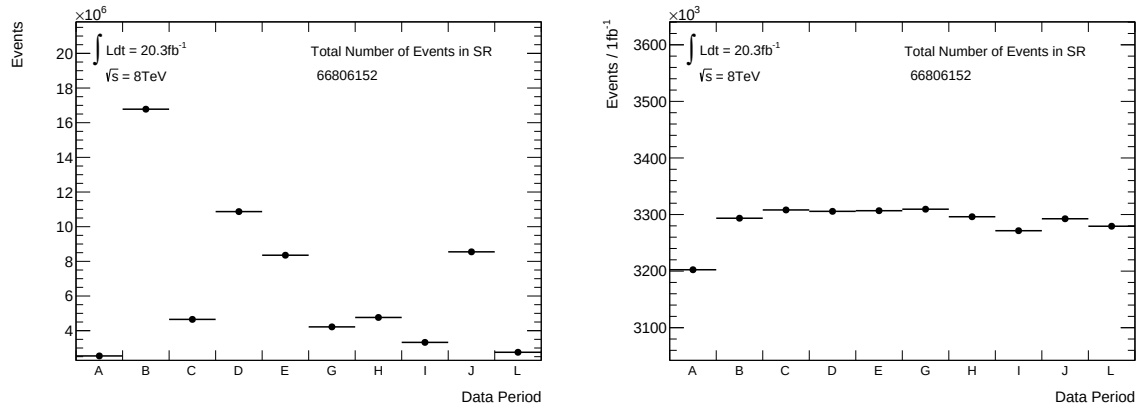
Table 4.3: Number of events counted in data after the different selection criteria for the W + jets SR as well as the fraction of events rejected at each requirement. For W^+ and W^- as well as the $W + n$ jets items, the given fractions are the respective fraction of events compared to the total number of W events in the SR. The *Egamma* data stream is a data-recording path with a trigger-preselection and stores events triggered by all active electron or photon triggers.
(*) The number of events after this requirement has only been determined on a smaller preselected data-sample and is therefore not listed. The corresponding fraction should however be approximately correct.
(**) As a result of the data-preselection, the number of events without trigger and electron pre-selection is not available anymore.

The highest integrated luminosity data-taking periods in 2012 were periods B, D and J. The number of events collected per data-taking period normalized to the respective integrated luminosity of that period is approximately constant, as visible in Figure 4.5. The largest difference in the event yield per fb^{-1} is visible for data period A and amounts to approximately 3 %. The electron identification applied in the trigger was re-optimized after period A, in order to limit event losses observed in the first 2012 data. The small mismatch between the period-A electron ID in the trigger and the final offline electron ID leads to the reduced normalized event yield in period A.

4.3.2 Multi-jet control region

For the data-driven background estimate of multi-jet production which is described in detail in Section 5.1, a separate exclusive selection is required to obtain a multi-jet enriched control sample in data.

The selection has been optimized using MC simulation to most closely mimick the multi-jet distribution shapes in the W + jets SR. It is based on both an inversion of the electron identification and the electron isolation compared to the SR. The criteria applied for the multi-jet control region (MJ-CR) are summarized in Table 4.4. In order to select events with un-isolated electrons, the low-threshold single electron trigger used in the W + jets SR cannot be employed, as it applies an isolation requirement at the trigger level. Therefore, the corresponding, yet prescaled trigger `EF_e24vh_medium1` without isolation requirement is used



(a) (b)

Figure 4.5: Contribution of event yields from the different data-taking periods in 2012 to the W + jets SR, considering the total number of events (a) and the number of events per $1\,\mathrm{fb}^{-1}$ (b). For the computation of events per fb^{-1}, the integrated luminosity per data period was computed using the official ATLAS tool for luminosity calculation with the same settings as for the official total integrated luminosity of $20.3\,\mathrm{fb}^{-1}$.

for the MJ-CR. The integrated luminosity collected with this trigger is $2.74\,\mathrm{fb}^{-1}$ and has been determined with the ATLAS tool for luminosity calculation [113].

Multi-jet control region	
Trigger	`EF_e24vh_medium1`
Electron ID	*medium*, but not *tight*
Anti-isolation	$p_{\mathrm{T}}^{\mathrm{cone30}}/E_{\mathrm{T}} > 0.07$
Impact parameters	None

Table 4.4: Requirements for the multi-jet control region which differ with respect to the SR selection as listed in Table 4.2.

4.3.3 $t\bar{t}$ validation region

The production of $t\bar{t}$-pairs generates a similar final state as $W + 4$ jets, if one W boson from the (anti)top-quark decay, $t \to bW$, decays as $W \to e\nu$, while the other decays hadronically. Although the contribution from $t\bar{t}$ production is minimized by rejecting events with b-jets, at larger jet multiplicities (≥ 4 jets) it becomes the largest background with rising contribution in the SR as n_{jets} increases. Since these multiplicities are not the main focus of this thesis, MC simulation is used to estimate the $t\bar{t}$ contribution. Due to the growing share of $t\bar{t}$ events with n_{jets}, the MC prediction is however specifically validated against data.

A separate selection enriched in $t\bar{t}$ events is defined for this purpose and denoted as $t\bar{t}$ validation region ($t\bar{t}$-VR). In contrast to the W + jets SR, the presence of at least one b-jet is required in the event. Other requirements on electron or $E_{\mathrm{T}}^{\mathrm{miss}}$ selections remain identical. For a reduction of the contamination from W production, for example in association with b-jets, at least 4 signal jets must furthermore be present in the event.

The `MV1` algorithm employing the 60 % efficiency working point is used to select the b-jets, identical to the b-jet definition utilized for the W + jets SR.

The criteria for the $t\bar{t}$-VR are summarized in Table 4.5.

	$t\bar{t}$ validation region
b-Jet	≥ 1 b-jet identified with MV1 b-tagger, 60 % efficiency working point
Signal jets	≥ 4

Table 4.5: Requirements for the $t\bar{t}$ validation region which differ with respect to the SR selection as listed in Table 4.2.

CHAPTER 5

Background estimation

Events selected in data result mainly from W production in association with jets, with subsequent W boson decay as $W \to e\nu$, also denoted as $W \to e\nu$ +jets in the following. In addition, however, other processes can mimick the W + jets signature or comply with the signature e.g. as a result of missed particles. The latter, for example, happens if one electron from $Z \to ee$ production is emitted outside of the detector acceptance and is thus not recorded. Moreover, $W \to \tau\nu$ events with subsequent τ-lepton decay as $\tau \to \nu_\tau e \bar{\nu_e}$ can end up in the W + jets signal region (SR), but are not considered as part of the signal.

For the measurement of the W + jets production cross section, these background events need to be subtracted from data before the measured distributions in data are unfolded to particle level. The precise estimate of the background events in the W + jets SR is therefore an essential step in the presented analysis, and will be described in the following.

The processes contributing to the background are sorted into three groups.

- **Multi-jet background:** Multi-jet events can pass the SR requirements if, for example, one jet mimicks the electron signature, i.e. it *fakes* the electron, and this mismeasurement furthermore gives rise to fake $E_{\mathrm{T}}^{\mathrm{miss}}$. Although fake probabilities are small, the large cross section of multi-jet production yields a multi-jet background contribution which is dominant at low jet multiplicities. The multi-jet background contribution is estimated using a data-driven method, described in Section 5.1.

- **$t\bar{t}$ background:** $t\bar{t}$ events match the required signature if one W boson (from $t \to bW$) decays via the electron decay channel, while the remaining decay products are reconstructed as four jets, two of which originate from the b-quarks. The $t\bar{t}$ contribution in the SR is therefore reduced with a veto on the presence of b-jets. Inefficiencies in the veto, but mainly the similar production cross sections of $W+ \geq 4$ jets and $t\bar{t}$ however lead to dominant $t\bar{t}$ contributions at higher jet multiplicities.

  The main focus of this measurement is however the production of $W + (\geq)1$jet, where the $t\bar{t}$ contribution is negligible. It is thus estimated using MC simulation and validated in data, as shown in Section 5.2, to ensure proper predictions at high jet multiplicities.

- **Other backgrounds:** Further backgrounds are $Z \to ee$, $W \to \tau\nu$, single top, diboson (WW, WZ, ZZ) as well as $Z \to \tau\tau$ events, approximately in the order of importance. Their contributions in the SR are suppressed by branching fractions, the probability of particle losses outside of the acceptance and/or smaller production cross sections, and are thus sub-dominant. These backgrounds are estimated from the MC simulation, as presented in Section 5.3.

The final predictions, including $W \to e\nu$ +jets as well as the estimated backgrounds, are compared to data in Section 5.4 for a variety of kinematic distributions for different jet multiplicities.

5.1 Multi-jet background

Multi-jet events by itself contain neither prompt isolated electrons nor prompt neutrinos which would give rise to genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The multi-jet events in the signal region therefore involve faked signatures of both the selected electron and the $E_{\mathrm{T}}^{\mathrm{miss}}$, where the latter can be a direct consequence of the first. While MC simulation has been developed to describe the kinematics and the detector response for real particles, its description of faked signatures is not very precise. A data-driven approach for the estimate of the multi-jet contribution is thus employed, similarly to previous versions of the W + jets analysis [7, 8].

The method consists in a binned extended maximum likelihood fit to the distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ in data. The fit is conducted separately for different jet multiplicities, in order to account for variations in the multi-jet contribution as a function of the jet multiplicity.

In the following, an introduction to the general method of the used maximum likelihood fit is given in Section 5.1.1, while the actual fit to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is shown in Section 5.1.2. Estimates for the statistical and the systematic fit uncertainties are explained there as well, together with the results for the multi-jet contribution in the SR.

5.1.1 Binned extended maximum likelihood fit

The maximum likelihood method[1] is a powerful tool to determine the values of unknown parameters from a set of data points. The basic idea is to contruct the joint probability for the obtained measurement in dependence of the parameters of interest $\boldsymbol{\theta} = (\theta_1, ..., \theta_P)$, called likelihood $L(\boldsymbol{\theta})$. The best estimate for the parameter values $\hat{\boldsymbol{\theta}}$ are those which maximize the likelihood. For this analysis, the parameter of interest is the number or fraction of multi-jet events in the W + jets signal region.

The normal maximum likelihood method is unbinned and considers each measured event individually in the likelihood. With about 67 Mio events in the SR, the likelihood would however become impossible to handle, thus a **binned** version of the maximum likelihood method is used.

Instead of considering the measured values per event of a random variable x with a probability density function (p.d.f.) $f(x; \boldsymbol{\theta})$, the measurement values are filled into a histogram with m bins. The number of entries in the histogram are $\mathbf{n} = (n_1, ..., n_m)$ and the total number of events, i.e. the sum of all bins is N. The probability to have n_1 events in the first bin, n_2 events in the second bin, etc. is described by a multinomial distribution where the probability that an event ends up in bin i is given as p_i:

$$f_{\mathrm{hist}} = \frac{N!}{n_1!...n_m!} \cdot p_1^{n_1} ... \, p_m^{n_m}.$$

For a multinomial distribution, the expectation value ν_i for bin i is given as $\nu_i = N \cdot p_i$, so p_i can be rewritten as $p_i = \nu_i / N$ with $\sum_{i=1}^{m} \nu_i = N$. The expectation value can however also be phrased in terms of the p.d.f. in that bin: $\nu_i = N \cdot f(x_i; \boldsymbol{\theta})$, and is thus dependent on the parameters $\boldsymbol{\theta}$. The bin-wise p.d.f. $f(x_i; \boldsymbol{\theta})$ can either be understood as the integral of a

[1] The explanation in this section is based on descriptions in Ref. [114] and Ref. [115] adjusted to the use-case in this analysis and partially follows the notation in the latter.

continuous p.d.f. within the bin boundaries $x_i^{\min}$ and $x_i^{\max}$: $f(x_i;\boldsymbol{\theta}) = \int_{x_i^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta})dx$ or as directly determined bin-wise. In both cases, the normalization is $\sum_{i=1}^{m} f(x_i;\boldsymbol{\theta}) = 1$, defining it as p.d.f.

The likelihood is then given as follows:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{m} \left(\frac{\nu_i(\boldsymbol{\theta})}{N} \right)^{n_i} ,$$

where the combinatorial terms independent of $\boldsymbol{\theta}$ are dropped.

In this analysis, the parameters of interest for the background estimation, $\boldsymbol{\theta}$, are the contribution of multi-jet events (MJ) and the contributions from all other processes (OP) in the W + jets signal region, where the latter includes the W + jets signal as well as the other background processes. Therefore, the p.d.f. can be written as the sum over the P parameters:

$$f(x_i;\boldsymbol{\theta}) = \sum_{j=1}^{P} \theta_j \cdot f_j(x_i) = \theta_{\mathrm{MJ}} \cdot f_{\mathrm{MJ}}(x_i) + \theta_{\mathrm{OP}} \cdot f_{\mathrm{OP}}(x_i)\,.$$

The parameters θ_j thus specify the relative contribution of the different processes or equivalently the relative contribution from the process p.d.f.s $f_j(x_i)$ per bin i. For a given number of events N in data, the parameters θ_j are related as $\sum_{j=1}^{P} \theta_j = 1$, i.e. in the case here: $\theta_{\mathrm{MJ}} = 1 - \theta_{\mathrm{OP}}$.

A fixed total number of events N is however not a good assumption in this analysis setup, since a different total number of events N' is likely to be obtained upon a potential repetition of the experiment. Instead, the total number of events N is therefore considered as a random variable from a Poisson distribution with mean ν and $\sum_{i=1}^{m} \nu_i = \nu$. This removes the constraint on the sum of the θ_j and modifies the likelihood – now called **extended likelihood** – to consider the additional random variable N as follows (again leaving out the factorials):

$$L_{\mathrm{ext}}(\boldsymbol{\theta}) = \prod_{i=1}^{m} \left(\frac{\nu_i(\boldsymbol{\theta})}{\nu} \right)^{n_i} \cdot e^{-\nu}\nu^N = \prod_{i=1}^{m} \left(\nu_i(\boldsymbol{\theta})\right)^{n_i} \cdot e^{-\nu_i}$$

The second formulation above can be obtained from the first by simple re-ordering and cancellation of terms and demonstrates that the extension of the multinomial likelihood with a global Possion is equivalent to independent Poisson distributions in each bin i.

Computationally, it is easier to handle sums instead of products, therefore instead of $L(\boldsymbol{\theta})$, $\log L(\boldsymbol{\theta})$ is usually maximized, or $-\log L(\boldsymbol{\theta})$ minimized. The log-likelihood then reads:

$$\log L_{\mathrm{ext}}(\boldsymbol{\theta}) = -\sum_{i=1}^{m} \nu_i(\boldsymbol{\theta}) + \sum_{i=1}^{m} n_i \log\left(\nu_i(\boldsymbol{\theta})\right)$$

Inserting the formulation of $\nu_i(\boldsymbol{\theta})$ in terms of the p.d.f.s, the first term in the $\log L_{\mathrm{ext}}(\boldsymbol{\theta})$ can be rephrased as $\sum_{i=1}^{m} \nu_i(\boldsymbol{\theta}) = \sum_{i=1}^{m} \nu \sum_{j=1}^{P} \theta_j f_j(x_i) = \sum_{j=1}^{P} \nu\theta_j \sum_{i=1}^{m} f_j(x_i) = \sum_{j=1}^{P} \nu\theta_j$, since also the process p.d.f.s are normalized to unity: $\sum_{i=1}^{m} f_j(x_i) = 1$. Defining $N_j = \theta_j\nu$ as the total number of events per process, the extended likelihood can be rewritten as:

$$\log L_{\mathrm{ext}}(\boldsymbol{\theta}) = -\sum_{j=1}^{P} N_j + \sum_{i=1}^{m} n_i \log\left(\sum_{j=1}^{P} N_j f_j(x_i)\right)$$

In the case of the multi-jet background estimation this becomes:

$$\log L_{\text{ext}}(\boldsymbol{\theta}) = -(N_{\text{MJ}} + N_{\text{OP}}) + \sum_{i=1}^{m} n_i \log \left(N_{\text{MJ}} f_{\text{MJ}}(x_i) + N_{\text{OP}} f_{\text{OP}}(x_i)\right)$$

The parameters which are fitted by minimizing the negative log-likelihood are N_{MJ} and N_{OP}, while the p.d.f.s $f_{\text{MJ}}(x_i)$ and $f_{\text{OP}}(x_i)$ are determined externally and are referred to as *templates* in the fit procedure.

In addition to estimates of the parameters, also estimates of the uncertainties $\hat{\sigma}_{\hat{\theta}_j}$ in these parameters are of importance. These are given as the square root of the corresponding parameter variances, $V[\hat{\theta}_j]$, and can be determined within the maximum likelihood method by several approaches. Two of those are used in this analysis.

The first approach uses the inverse of the second derivative matrix of the log-likelihood as the variance estimate, i.e. $(\widehat{V^{-1}})_{jk} = -\frac{\partial^2 \log L}{\partial\theta_j \partial\theta_k}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. It relies on the assumption of an approximately Gaussian shape of the likelihood around the maximum which is usually true for a sufficiently large data sample.

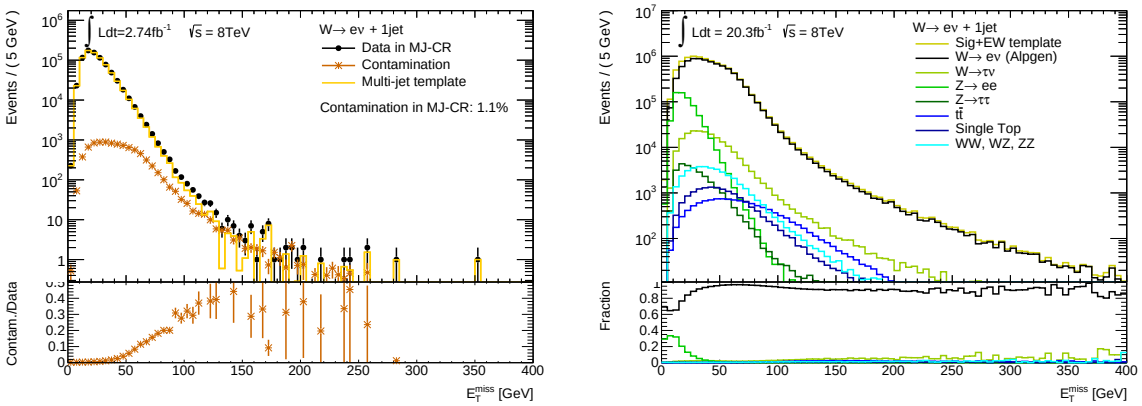
The alternative method obtains the range of likely parameter values by evaluating $\log L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \log L_{\max} - \frac{1}{2}$. It can be motivated by a Taylor expansion of the log-likelihood around the best parameter estimates $\hat{\boldsymbol{\theta}}$. This approach does not place assumptions on the likelihood shape and returned uncertainties can be asymmetric, but it converges to the same result as the first method if the likelihood is Gaussian. A comparison of both procedures gives an indication on how Gaussian the likelihood is around the maximum. For more details on the estimates of the parameter variances see e.g. Ref. [115] or Ref. [114].

The implementation of the binned extended maximum likelihood method within the RooFit framework [116] is used for the multi-jet background estimate in this analysis. This uses the Minuit [117] package contained within the ROOT Data Analysis Framework [118] for the negative log-likelihood minimization. Both mentioned approaches for uncertainty estimates are contained within the Minuit package and are denoted as Hesse and Minos uncertainties, respectively.

5.1.2 Fitting $E_{\text{T}}^{\text{miss}}$

The multi-jet background contribution in this analysis is fitted using the binned extended maximum likelihood approach as explained above, with the $E_{\text{T}}^{\text{miss}}$-distribution as random variable x. The $E_{\text{T}}^{\text{miss}}$ distribution provides reasonable discrimination between the multi-jet events and the other processes in the W + jets signal region. Multi-jet events typically do not contain or contain only very little $E_{\text{T}}^{\text{miss}}$ from mismeasurements (e.g. caused by faked signatures) or from instrumental defects. The $W \to e\nu$ + jets signal process, however, as well as other backgrounds like $t\bar{t}$, $W \to \tau\nu$, etc. mostly involve neutrinos and therefore genuine $E_{\text{T}}^{\text{miss}}$. The $E_{\text{T}}^{\text{miss}}$ distribution thus has a different shape and peaks at diffent values, namely at $\sim 20\,\text{GeV}$ for multi-jet events and at $\sim 40\,\text{GeV}$ for the combined signal and other background processes. In order not to lose the discrimiative power at the low $E_{\text{T}}^{\text{miss}}$ values, the SR requirement of $E_{\text{T}}^{\text{miss}} > 25\,\text{GeV}$ is removed for the fit. The other SR requirements, including the one on m_{T}, are however kept, in order not to introduce a bias. This will be denoted in the following as $E_{\text{T}}^{\text{miss}}$-relaxed SR. The $E_{\text{T}}^{\text{miss}}$-requirement is re-applied on the multi-jet $E_{\text{T}}^{\text{miss}}$ distribution after the fit in order to extrapolate to the W + jets SR and obtain the fraction of multi-jet events there.

The fit of the $E_{\text{T}}^{\text{miss}}$-distribution is conducted in bins of exclusive jet multiplicity from 0 to 5 jets and inclusive for ≥ 6 jets, the latter for statistics reasons. The chosen fit range covers the $E_{\text{T}}^{\text{miss}}$ values of $15 - 75\,\text{GeV}$.



(a) Multi-jet template

(b) Sig+EW template

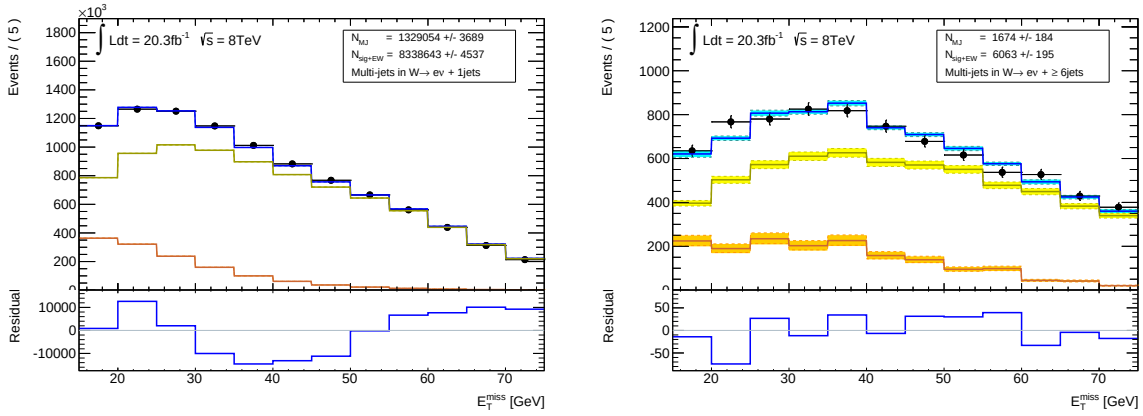
Figure 5.1: Templates for the $E_{\mathrm{T}}^{\mathrm{miss}}$ fit for the multi-jet background (a) and the signal and other background processes (b) in the case of 1 jet. The multi-jet template is obtained from data and the sig+EW template from MC simulation as described in the text. The lower panels show the fractional contribution of the contamination with respect to data in the MJ-CR (a) or the fractional contributions of the different processes to the combined template (b).

n_{jets}	$N_{\text{MJ-templ}}^{\text{evts}}$	$c_{\text{MJ-CR}}[\%]$
0	1 261 000 ± 1200	4.28
1	778 900 ± 890	1.11
2	218 700 ± 470	1.03
3	51 490 ± 230	1.06
4	11 760 ± 110	1.21
5	2429 ± 50	1.48
≥6	606.1 ± 25	2.18

Table 5.1: Number of events in the multi-jet template, $N_{\text{MJ-templ}}^{\text{evts}}$, as well as the subtracted contamination, $c_{\text{MJ-CR}}$, in % of the data counts in the MJ-CR.

The *templates* which are fitted to data in the $E_{\mathrm{T}}^{\mathrm{miss}}$-relaxed SR are obtained differently for the multi-jet background and the other processes (OP in the last section, but in the following denoted as *sig+EW* which indicates their origin better). The multi-jet template is obtained from data in a control region – the multi-jet control region (MJ-CR) listed in Section 4.3.2. The MJ-CR was optimized using multi-jet simulation to ensure that the kinematic distribution shapes and in particular the $E_{\mathrm{T}}^{\mathrm{miss}}$ shape in the MJ-CR reproduces the shapes expected in the SR. Contamination from the signal or other background processes in the MJ-CR is in general small (mostly $< 1.5\,\%$) and is subtracted using the simulation predictions for the signal and the other background processes. The sig+EW template is obtained from MC simulation in the $E_{\mathrm{T}}^{\mathrm{miss}}$-relaxed SR by adding the MC predictions taking into account their respective cross sections. Examples for the used templates in the likelihood fit are shown in Figure 5.1. A list of the number of events in the MJ-template as well as the contaminations in the MJ-CR is given in Table 5.1.

The results of the fit within the fit range (FR) are shown in Figure 5.2 for two jet multiplicities $n_{\mathrm{jets}} = 1$ and $n_{\mathrm{jets}} \geq 6$. For illustration purposes, two types of uncertainty estimates are shown for the total fit result as well as the individual templates. The uncertainty bands are provided by RooFit and are similar to the distinction between Hesse and Minos uncertain-



(a) 1 jet (b) ≥ 6 jets

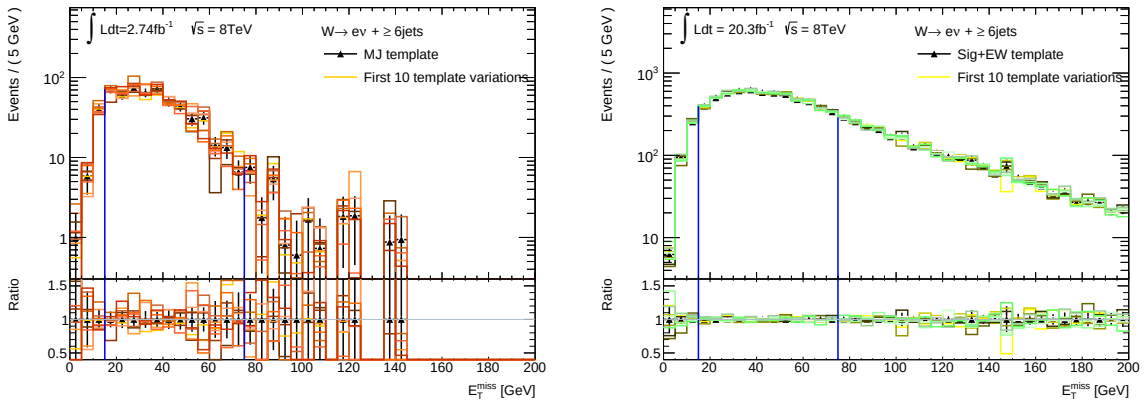
Figure 5.2: Result of the maximum likelihood fit to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for 1 jet (a) and ≥ 6 jets (b) in the fit range. The obtained fit result (blue), the fitted sig+EW template (yellow) and the fitted multi-jet template (orange) are displayed together with uncertainties. The residual in the lower panel is defined as the fit result minus data and is given as absolute event number difference.

ties, but differentially for the distribution and not only for the fitted parameters. The colored band visualizes a 1σ-uncertainty band which is determined from linear error propagation, taking into account the correlation between the fitted parameters, but relying on Gaussian shapes in the error propagation. The dashed lines, however, define the 68 %-central interval obtained by sampling curves from a multi-variate Gaussian probability density function around the fit result as constructed from the covariance matrix of the fit. Both uncertainty estimates are found to agree very well, hence, no strong asymmetries or non-Gaussian behaviour of the likelihood is observed.

n_{jets}	$N_{\mathrm{MJ,FR}}^{\mathrm{fitted}}$	$N_{\mathrm{sig+EW,FR}}^{\mathrm{fitted}}$	ρ	$f_{\mathrm{MJ,FR}}$[%]			$\mathrm{SF}_{\mathrm{MJ}}$	$\mathrm{SF}_{\mathrm{sig+EW}}$
				(HESSE)	(MINOS)			
0	3 860 000	± 5600	60 040 000	± 9300	−0.52	6.04	±0.01	$^{+0.009}_{-0.009}$	3.29 ± 0.01	1.04
1	1 329 000	± 3700	8 339 000	± 4500	−0.73	13.75	±0.04	$^{+0.038}_{-0.038}$	2.08 ± 0.01	1.02
2	499 100	± 2400	1 904 000	± 2700	−0.82	20.77	±0.10	$^{+0.102}_{-0.098}$	2.62 ± 0.02	1.02
3	117 500	± 1200	406 400	± 1400	−0.85	22.43	±0.24	$^{+0.237}_{-0.237}$	2.61 ± 0.04	1.04
4	27 650	± 640	91 710	± 680	−0.87	23.17	±0.53	$^{+0.533}_{-0.532}$	2.66 ± 0.09	1.04
5	6459	± 330	20 540	± 350	−0.89	23.92	±1.22	$^{+1.221}_{-1.224}$	3.00 ± 0.23	1.01
≥6	1674	± 180	6063	± 200	−0.89	21.64	±2.38	$^{+2.378}_{-2.378}$	3.13 ± 0.49	1.00

Table 5.2: Results of the likelihood fit in the fit range of $15 - 75\,\mathrm{GeV}$ for each fitted jet multiplicity. ρ is the correlation between the fitted parameters $N_{\mathrm{MJ,FR}}^{\mathrm{fitted}}$ and $N_{\mathrm{sig+EW,FR}}^{\mathrm{fitted}}$. For the fraction of multi-jet events in the fit range, $f_{\mathrm{MJ,FR}}$, both Hesse and Minos uncertainties are listed.

Table 5.2 lists the results of the $E_{\mathrm{T}}^{\mathrm{miss}}$ fit in the fit range (FR). The fitted number of multi-jet and sig+EW background events, $N_{\mathrm{MJ,FR}}^{\mathrm{fitted}}$ and $N_{\mathrm{sig+EW,FR}}^{\mathrm{fitted}}$, are converted into the fraction of multi-jet events in the FR, $f_{\mathrm{MJ,FR}}$, and the scale factors for the multi-jet and the sig+EW templates $\mathrm{SF}_{\mathrm{MJ}}$ and $\mathrm{SF}_{\mathrm{sig+EW}}$. The latter indicate by which amount the templates have to be scaled to comply with the fit result, starting from their original normalization. This normalization is arbitrary in the case of the multi-jet template, or to be more precise, it is the result of the integrated luminosity of the prescaled trigger used in the MJ-CR which is



(a) Multi-jet template

(b) Sig+EW template

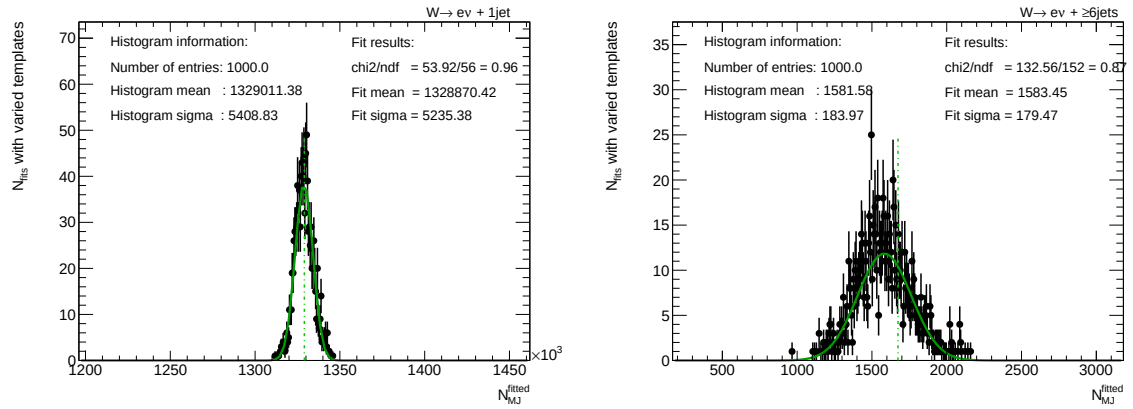
Figure 5.3: First 10 template variations for the multi-jet template (a) and the sig+EW template (b) for $n_{\text{jets}} \geq 6$, where statistics is smallest and the variations best visible. The lower panel shows the difference of the variations to the original template (black markers). The vertical blue lines indicate the fit range.

unrelated to the normalization in the SR. For the sig+EW template, the original normalization is however according to the process cross sections. For a valid fit, $\text{SF}_{\text{sig+EW}}$ is therefore expected to be close to unity, while SF_{MJ} can be anything, but should not be too large in order to limit the impact of statistical fluctations in the MJ-template on the fit. This is well fulfilled and SF_{MJ} is found to be $\approx 2-3$. For $f_{\text{MJ,FR}}$, both HESSE and MINOS uncertainties are listed and are found to be in good agreement. This demonstrates again a proper Gaussian and symmetric behaviour of the likelihood around its maximum. In the following, therefore, only the Gaussian-based, i.e. the HESSE uncertainties will be used further.

Statistical fit uncertainties

The extended maximum likelihood considers the possibility of fluctuations in the total number of events, or equivalently fluctuations of the measured number of events per bin via Poisson terms in the likelihood. The estimators of the parameter uncertainties obtained from the likelihood thus include the effect of statistical uncertainties in the fitted data on the fit result. The p.d.f.s used in the likelihood fit are however taken at nominal value. In this analysis, the p.d.f.s are obtained from data or MC simulation, as explained before, and thus have statistical uncertainties.

To estimate the effect of these statistical uncertainties on the final multi-jet estimate, the multi-jet fits are repeated 1000 times with modified templates. The templates are modified bin-wise by fluctuating the template contents using a Gaussian distribution centered at the original content of the template histogram with a σ corresponding to the statistical uncertainty per bin. The first 10 varied templates are shown in Figure 5.3 for both the multi-jet and sig+EW template for the case of ≥ 6 jets where statistical uncertainties are largest and the fluctuations best visible. The spread (σ) in the results obtained from the fits with varied templates is considered as additional statistical uncertainty, $\Delta N_{\text{MJ}}^{\text{stat,templ}}$, on the determined multi-jet contribution. Examples of the distribution of the number of multi-jet events are given in Figure 5.4 for $n_{\text{jets}} = 1$ and $n_{\text{jets}} \geq 6$. These distributions are fitted with a Gaussian and the mean and σ obtained from fit are in good agreement with histogram values. The original fit result furthermore agrees approximately with the mean of the distribution, demonstrating the existance of no or only very small bias in the fit result towards higher jet multiplicities.



(a) 1 jet (b) ≥ 6 jets

Figure 5.4: Fit results for the number of multi-jet events obtained from 1000 fits with varied templates for $n_{\text{jets}} = 1$ (a) and $n_{\text{jets}} \geq 6$ (b). The vertical dashed line indicates the original fit result. The distribution is fitted using a simple χ^2-fit with a Gaussian function (green) for illustration purposes.

Systematic uncertainties in the fit procedure

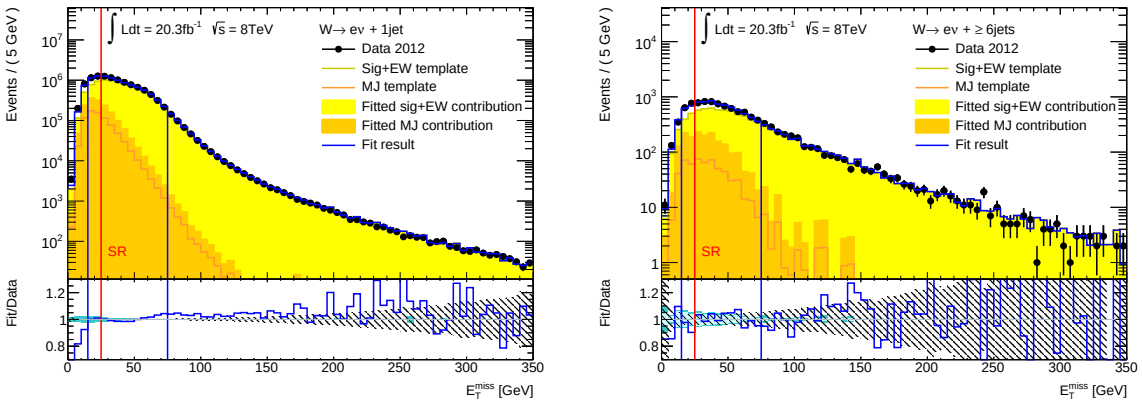
The results of the likelihood fit depend on choices made in the execution of the fit, namely the fit range and the definition of the MJ-CR. To estimate the impact of these choices, the fit is repeated with varied settings.

Starting from the default fit range of $15 - 75\,\text{GeV}$, both the upper and lower fit range boundaries are varied by $\pm 5\,\text{GeV}$ up and down and the obtained results are compared to the original fit results. The MJ-CR is varied by modifying the MJ-CR definition to use a different anti-isolation requirement on the selected electron: $p_{\text{T}}^{\text{cone20}}/E_{\text{T}} > 0.1$. This anti-isolation has been found to give a reasonable variation in the multi-jet distribution shapes without an overly large reduction in statistics in the MJ-CR. It corresponds both to a variation of the cone-size used to calculate the existing track p_{T} sum around the electron as well as a variation of the required threshold.

The variations of the upper and lower fit range boundaries are combined by using the maximum deviation of the lower or upper boundary up and down variations and then summed in quadrature together with the difference from the changed MJ-CR anti-isolation. The determined systematic uncertainties are discussed below, together with the final results on the multi-jet contribution in the W + jets SR.

Results for the multi-jet contribution

The determined multi-jet contributions are extrapolated to the W + jets SR by scaling the multi-jet template with the determined scale factor, SF_{MJ} (see Table 5.2), and re-applying the SR requirement of $E_{\text{T}}^{\text{miss}} > 25\,\text{GeV}$. Subsequently, the fraction of multi-jet events with respect to data in the SR, $f_{\text{MJ,SR}}$, is obtained and listed per fitted jet multiplicity in Table 5.3. $f_{\text{MJ,SR}}$ depends strongly on the jet multiplicity in particular for the lowest jet multiplicities. It ranges from about 3 % in the absence of jets, to approximately 8 % in the presence of 1 jet to around 15 % for 2 or more jets. Table 5.3 furthermore lists the statistical and systematic uncertainties in the fit results. It is found that both types of statistical uncertainties are roughly of the same order of magnitude, with $\Delta f_{\text{MJ,SR}}^{\text{stat,templ}}$ – the uncertainty reflecting the influence of template statistics – being slightly larger. Among the systematic uncertainties,



(a) 1 jet (b) ≥ 6 jets

Figure 5.5: Result of the multi-jet fit for $n_{\text{jets}} = 1$ (a) and $n_{\text{jets}} \geq 6$ (b). Shown is the data (black points), the sig+EW and MJ templates before the fit (yellow and orange lines), the sig+EW and MJ contributions after the fit (yellow and orange areas) as well as the combined fit result (blue line). The lower panel shows the ratio of combined fit vs. data in blue. The hashed area indicates the fractional size of the statistical uncertainties in the data, the blue band shows the total (stat⊕syst) fit uncertainty from the multi-jet contribution and the light blue band indicates the statistical fraction of it. Horizontal blue and red lines mark the fit range and the SR $E_{\text{T}}^{\text{miss}}$-requirement, respectively.

the variation of the lower fit range boundary, in particular to smaller values, provides the largest uncertainty in the fit result. This can be understood, since the $E_{\text{T}}^{\text{miss}}$ distribution is especially difficult to model at these low values.

n_{jets}	$f_{\text{MJ,SR}}$	$\Delta f_{\text{MJ,SR}}^{\text{stat,HESSE}}$	$\Delta f_{\text{MJ,SR}}^{\text{stat,templ}}$	$\Delta f_{\text{MJ,SR}}^{\text{FR 10-75GeV}}$	$\Delta f_{\text{MJ,SR}}^{\text{FR 20-75GeV}}$	$\Delta f_{\text{MJ,SR}}^{\text{FR 15-70GeV}}$	$\Delta f_{\text{MJ,SR}}^{\text{FR 15-80GeV}}$	$\Delta f_{\text{MJ,SR}}^{\text{anti-iso}}$
0	3.32	0.01	0.01	−0.02	+0.01	−0.00	+0.00	−0.05
1	8.36	0.02	0.03	+0.66	+0.02	−0.08	+0.05	−0.09
2	14.62	0.07	0.10	+0.50	−0.35	−0.07	+0.05	−0.08
3	15.73	0.17	0.21	+0.54	−0.37	−0.06	+0.14	−0.11
4	15.93	0.37	0.40	+0.65	−0.32	+0.09	+0.14	+0.04
5	16.07	0.82	0.89	+0.23	−0.33	−0.04	−0.94	+0.19
≥6	14.61	1.60	1.61	−0.83	−0.73	+0.78	−0.43	+0.56

Table 5.3: Fraction of multi-jet events, $f_{\text{MJ,SR}}$, in % extrapolated to the SR, per fitted jet multiplicity. Furthermore, listed are the statistical uncertainties in the fraction, $\Delta f_{\text{MJ,SR}}^{\text{stat,HESSE}}$ and $\Delta f_{\text{MJ,SR}}^{\text{stat,templ}}$, as well as the systematic uncertainties from the fit range variations, $\Delta f_{\text{MJ,SR}}^{\text{FR xx-yyGeV}}$, and from the MJ-CR with modified electron anti-isolation criterion, $\Delta f_{\text{MJ,SR}}^{\text{anti-iso}}$.

Figure 5.5 shows the fit result in a larger range of the $E_{\text{T}}^{\text{miss}}$ distribution for $n_{\text{jets}} = 1$ and $n_{\text{jets}} \geq 6$ – the first of particular importance for this analysis, while the latter demonstrates the fit behaviour under decreasing statistics. The fit result is found in reasonable agreement with the data distribution. The combined statistical and systematic uncertainties resulting from the multi-jet estimate are propagated to the final fit result and demonstrate the small size of the uncertainties from the fit used further in the analysis. Table 5.4 lists the obtained multi-jet contributions together with the combined statistical and systematic uncertainties and the fractional size of the uncertainties for a simpler overview. The systematic uncertainties are the dominant uncertainty contribution and lie in the range of about $1 - 8\,\%$, while the statistical uncertainties become relevant only for $\gtrsim 5$ jets. Corresponding tables for the results of the multi-jet fit for W^+ and W^- selections separately can be found in Appendix A.1.

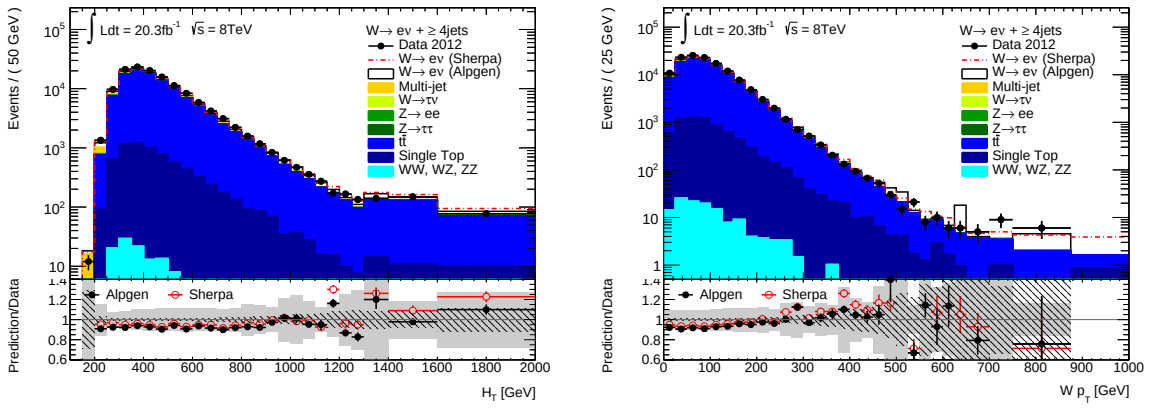
n_{jets}	$f_{\text{MJ,SR}}$	$\Delta f^{\text{stat}}_{\text{MJ,SR}}$	$\Delta f^{\text{syst}}_{\text{MJ,SR}}$	$\frac{\Delta f^{\text{stat}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$ [%]	$\frac{\Delta f^{\text{syst}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$ [%]
0	3.32	0.01	0.06	0.39	1.66
1	8.36	0.04	0.67	0.49	8.01
2	14.62	0.12	0.51	0.81	3.48
3	15.73	0.27	0.57	1.70	3.61
4	15.93	0.55	0.66	3.42	4.15
5	16.07	1.21	1.02	7.55	6.32
$\geq$6	14.61	2.27	1.26	15.53	8.64

Table 5.4: Fraction of multi-jet events, $f_{\text{MJ,SR}}$, in % extrapolated to the SR, per fitted jet multiplicity, as in Table 5.3. Listed in addition are the combined statistical and systematic uncertainty, $\Delta f^{\text{stat}}_{\text{MJ,SR}}$ and $\Delta f^{\text{syst}}_{\text{MJ,SR}}$, as well as their fractional size with respect to the determined multi-jet fraction.

5.2 $t\bar{t}$ background

The $t\bar{t}$ contributions are suppressed in the SR with a veto on the presence of b-jets. The used working point of the b-tagger has an efficiency of 60 % for real b-jets, and the tagger only operates within tracking acceptance. Hence, a non-negligible fraction of the produced b-originated jets are not b-tagged which allows the event to pass into the SR. Furthermore, while the inclusive W production cross section is almost three orders of magnitude larger than $t\bar{t}$ production (compare e.g. Figure 1.1), the production cross sections with ≥ 4 jets become roughly equal. This results from the fact that for $W+ \geq 4$ jets each jet emission is suppressed by a factor α_S, while for $t\bar{t}$ the presence of 4 jets corresponds to the leading-order signature in the case where one W from $t \to bW$ decays leptonically.

At low jet multiplicities, the $t\bar{t}$ contribution in the SR is small. Therefore, MC simulation is used to estimate the $t\bar{t}$ contribution. The used simulation has been tuned by the ATLAS working group for top-physics to describe ATLAS measurements of $t\bar{t}$ production [119]. It is furthermore validated in data in a $t\bar{t}$-enriched region which was defined in Section 4.3.3.



(a) H_{T} (b) W p_{T}

Figure 5.6: H_{T} (a) and W p_{T} (b) distributions for data and the total SM prediction in the $t\bar{t}$ validation region. More than 85 % of the events are from $t\bar{t}$ production. The lower panel shows the ratio of the total prediction over data. The total prediction is estimated using two different W + jets generators which however does not significantly change the ratio as a result of the overwhelming top contribution in the $t\bar{t}$-VR.

The process composition in the $t\bar{t}$ validation region (VR) is given in Table 5.5. For ≥ 4 jets, more than 90 % of the total prediction results from top-processes with more than 85 % from $t\bar{t}$ production alone. The multi-jet contribution in the $t\bar{t}$-VR is estimated by applying the same data-driven method as described for the signal region in Section 5.1. Figure 5.6 shows the comparison of the total prediction in the $t\bar{t}$-VR to data for two example kinematic distributions. The used $t\bar{t}$ simulation had been tuned to describe the p_{T} distribution of the $t\bar{t}$-system, but the p_{T} of the top-quark itself was found to be difficult to model satisfactory. This reflects in the good description of the H_{T} distribution, but a small slope in the prediction-to-data agreement for the W p_{T} distribution. In both cases, however, the MC description was found to agree with data within the uncertainties and was therefore used to describe the $t\bar{t}$ contribution in the W + jets signal region.

	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
WW, WZ, ZZ	0.6	0.6	0.5	0.2	0.1	0.1	<0.1	<0.1
$Z \to \tau\tau$	0.2	0.1	0.1	0.1	<0.1	<0.1	<0.1	<0.1
$Z \to ee$	1.0	1.1	1.1	0.9	0.7	0.6	0.5	0.5
$W \to \tau\nu$	0.9	0.8	0.5	0.3	0.2	0.1	0.1	0.1
Multi-jet	10.5	9.8	7.7	4.0	1.9	0.5	1.0	1.1
$W \to e\nu$	37.5	31.2	18.9	10.3	6.2	4.5	3.6	2.6
Single t	8.6	9.7	10.7	8.0	5.8	4.7	4.1	3.4
$t\bar{t}$	40.8	46.7	60.5	76.2	85.1	89.5	90.6	92.3

Table 5.5: Composition of top and other processes expected from MC simulation and the data-driven multi-jet estimate, if the presence of at least 1 b-jet is required. The numbers are in % with respect to the total SM prediction and are listed for different inclusive jet multiplicities. The ≥ 4 jet case corresponds to the $t\bar{t}$ validation region.

5.3 Other backgrounds

Other processes contribute $\lesssim 10\,\%$ to the total prediction in the signal region with even smaller contributions for ≤ 2 jets. Roughly in decreasing order of importance, the additional backgrounds are:

- **$\mathbf{Z \to ee}$:** This process passes the selection of the signal, if one of the electrons is emitted out of the detector acceptance or not detected due to detector inefficiencies. In both cases, this can lead to artificial $E_{\mathrm{T}}^{\mathrm{miss}}$, generating the $W \to e\nu$-signature. The $Z \to ee$ contribution reaches around 6 % maximum, but lies below 3 % for 1 jet.

- **$\boldsymbol{W \to \tau\nu}$:** The decay of the W boson into τ-lepton and neutrino passes the signal selection if the τ further decays as $\tau \to \nu_\tau e \nu_e$. While the W production is identical to the measured decay into electron and neutrino, the contribution is suppressed by the τ-branching fraction of 17.8 % [25] and the change in kinematics. Slightly smaller electron p_{T}, worse isolation and impact parameters (due to the secondary decay) as well as different $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} distributions lead to the $W \to \tau\nu$ suppression in the W + jets SR. For most jet multiplicities, the contribution is around $1.5 - 2\,\%$.

- **Single t:** Single top production via the t-channel, s-channel and Wt production contribute by means of the W boson which is produced either in the top-decay or directly.

The b-jet veto in the signal region suppresses this background similarly as the $t\bar{t}$ contribution. For the higher jet multiplicities, it reaches around $1.5 - 3\,\%$, but for ≤ 2 jets its contribution is in the sub-percent range.

- **Dibosons:** Diboson events contributing in the signal region come mainly from WW production where one W decays as the signal $W \to e\nu$. WZ and ZZ production are suppressed by the lower cross sections and the missing W in the latter case. The combined diboson contribution in the signal region is $\lesssim 1\,\%$ for all jet multiplicities.
- **$Z \to \tau\tau$:** The contribution of $Z \to \tau\tau$ events is heavily suppressed by branching-fractions and kinematics, leading to a sub-percent contribution for all jet multiplicities.

These backgrounds are estimated from MC simulation, due to the smallness of their contributions in the signal region and the fact that mostly real W boson decays and/or real electrons are involved which are well described by the simulation. A detailed list of contributions from each process, including the W + jets signal, is given in the next section.

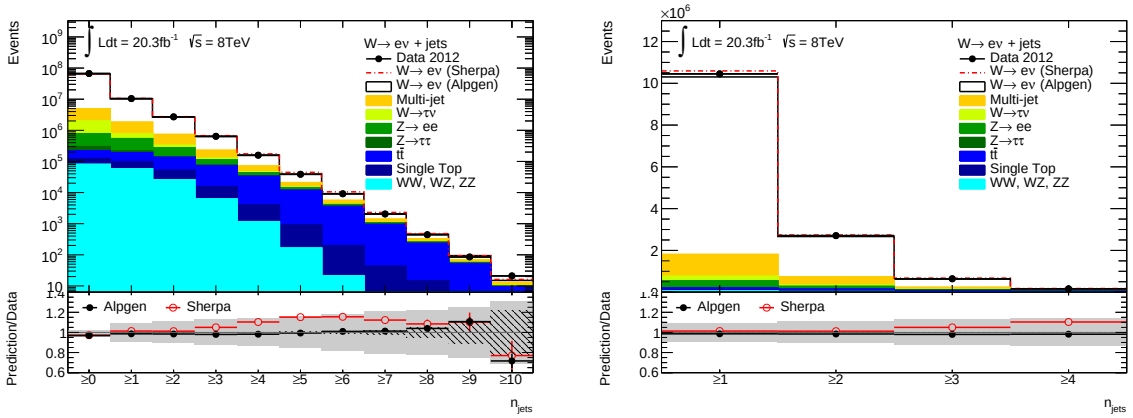
5.4 Detector level distributions

The background estimates plus the signal prediction are compared to data in the W + jets signal region in order to test the validity of the background estimates and investigate the quality of the signal prediction from two different MC generators. Figure 5.7 shows the jet multiplicity distribution in data compared to the total signal+background prediction. Table 5.6 lists the composition of signal and background contributions in the W + jets signal region for the different inclusive jet multiplicities. It is clear that for low jet multiplicities, the signal contribution by far dominates, while for larger jet multiplicities the $t\bar{t}$ production becomes competitive and eventually provides the largest contribution in the signal region. For ≥ 1 jet which is most interesting for this analysis, the signal contribution is about $83\,\%$, and for ≥ 2 jets it is still around $73\,\%$.

	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
WW, WZ, ZZ	0.1	0.6	1.0	1.0	0.7	0.4	0.2	0.1
Single t	0.1	0.4	1.0	1.5	1.8	1.9	1.9	1.9
$t\bar{t}$	0.1	0.9	3.1	9.1	18.6	28.1	36.1	43.0
$Z \to \tau\tau$	0.1	0.2	0.3	0.3	0.3	0.3	0.3	0.3
$Z \to ee$	0.7	3.1	4.8	5.8	6.0	5.8	5.6	5.5
$W \to \tau\nu$	1.8	2.1	2.1	2.0	1.9	1.6	1.4	1.2
Multi-jet	4.5	10.2	15.1	16.1	16.2	15.8	14.4	14.4
$W \to e\nu$	92.5	82.5	72.7	64.2	54.6	46.1	40.1	33.6

Table 5.6: Composition of signal and background processes in the W + jets signal region, as expected from MC simulation and the data-driven multi-jet estimate. The numbers are in % with respect to the total SM prediction and are listed for different inclusive jet multiplicities. For the signal numbers, the ALPGEN prediction has been used.

For further validation of the background estimates, several kinematic distributions are investigated which test the W boson and jet kinematics for ≥ 1 or ≥ 2 jets. Figure 5.8 displays the electron and W p_{T} distributions, the leading jet p_{T} and rapidity as well as the H_{T} distribution for ≥ 1 jet. Figure 5.9 presents two additional distributions demonstrating the description of the leading two jets, i.e. the invariant mass of the two jets $m_{1,2}$ and the angular separation $\Delta R_{1,2}$ for ≥ 2 jets.



(a) Display in log-scale

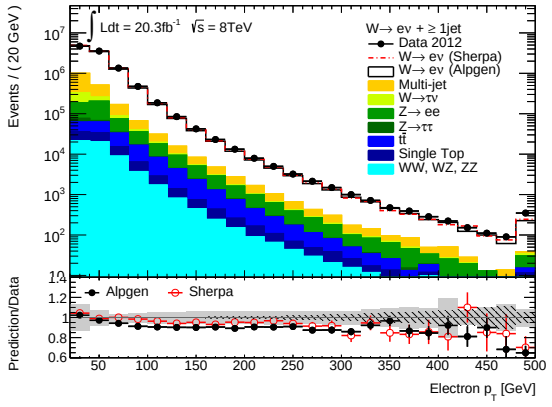
(b) Display in linear scale

Figure 5.7: Composition of signal and background in the W + jets signal region for the different inclusive jet multiplicities. For a better visualization, the same distribution is once shown with a logarithmic y-axis scale (a) and once with a linear y-axis scale (b). The latter has been zoomed in onto the interesting lower jet multiplicities. The lower panels show the ratio of the total prediction to data, where the hashed band indicates the fractional size of the data statistical uncertainties, the gray band the fractional size of the combined systematic uncertainties and the uncertainties in the ratio points result from the statistics in the MC simulation. For details on the considered systematic uncertainties see Chapter 7.

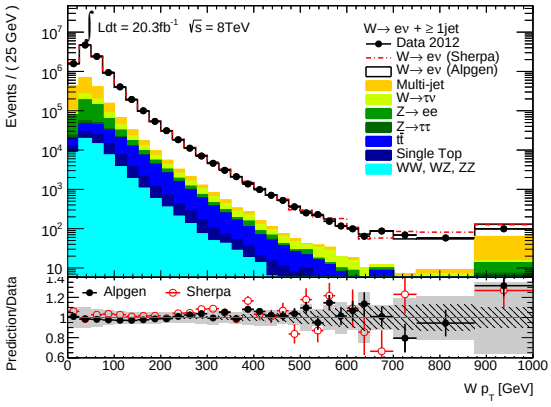
In general, a reasonable description of the data by the prediction is obtained which indicates that the background estimates are sound. Comparing data to the total prediction – with either ALPGEN or SHERPA used for the W + jets prediction –, it becomes clear that while ALPGEN mostly achieves good agreement with data, SHERPA has modelling issues in several distributions. In particular the higher jet multiplicities, the forward rapidity of jets and higher invariant masses of the leading two jets are not described by SHERPA. The electron p_{T} and the angular leading jets separation $\Delta R_{1,2}$ pose difficulties for both generators. However, overall ALPGEN performs much better and is thus used as nominal W+jets generator in this analysis.

Since the analysis presented in this thesis aims at measuring also differential W^+/W^- ratios, the same detector level comparisons are performed for separate W^+ and W^- selections. The corresponding Table and Figures are given in Appendix A.2. It is found that the composition in the signal region is in principle similar – as expected – but for the W^- selection larger background contributions and thus lower signal purity in the signal region are observed. For the most interesting ≥ 1 jet case, the signal contribution is reduced by approximately 4 % (absolute SR contribution difference) for W^- compared to W^+ and this difference increases towards the higher jet multiplicities. The signal region contributions shown in Table 5.6 for the charge-independent W selection thus implicitly correspond to cross section weighted averages of the W^+ and W^- selection compositions.

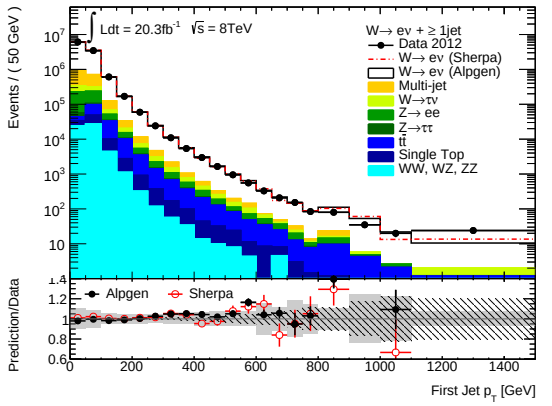
The general prediction modelling is investigated for W^+ and W^- separately for the same kinematic distributions as previously for the charge-independent W selection. As before, it does not hint at any problems in the background estimation. Similar features are observed as for the charge-independent W selection and mainly the electron p_{T} distribution (with small influences propagated through to derived distributions) shows a worsening of the data description by the prediction for W^- compared to W^+.



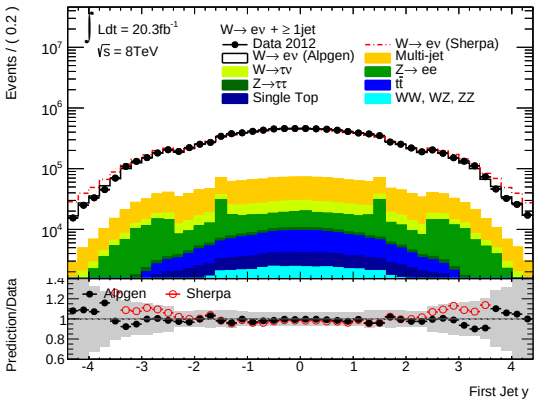
(a) Electron p_{T}



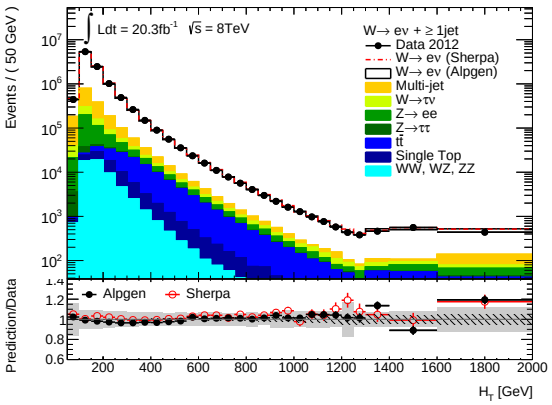
(b) W p_{T}



(c) Leading jet p_{T}

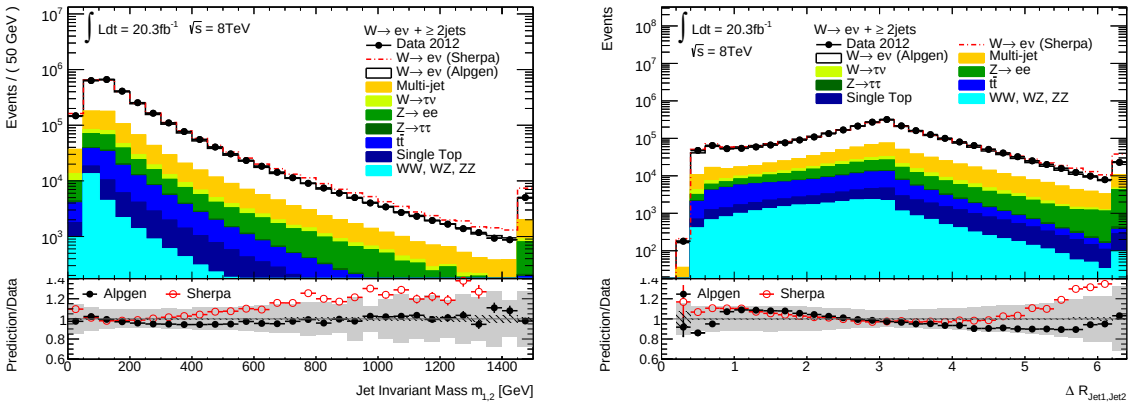


(d) Leading jet rapidity



(e) H_{T}

Figure 5.8: Kinematic distributions comparing data and the total prediction in the W + jets signal region for ≥ 1 jet: Electron p_{T} (a), W p_{T} (b), leading jet p_{T} (c) and rapidity (d) as well as H_{T} (e). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the W + jets signal. The gray and hashed bands are defined as explained in Figure 5.7.



(a) Invariant jet mass $m_{1,2}$

(b) Angular jet separation $\Delta R_{1,2}$

Figure 5.9: Kinematic distributions comparing data and the total prediction in the W + jets signal region for ≥ 2 jets: Invariant mass of the leading two jets $m_{1,2}$ (a) and their angular separation $\Delta R_{1,2}$ (b). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the W + jets signal. The gray and hashed bands are defined as explained in Figure 5.7.

CHAPTER 6

Unfolding

Measurements of physics processes, such as W + jets production, can be compared between different experiments or to new theory predictions independent of experimental details if detector specific effects are removed from the distributions. This means, the measured data counts need to be transformed into the physically more meaningful quantity of production cross sections.

The results in the previous chapter followed the opposite approach. Predictions for W+jets production were convoluted or *folded* with a model of the ATLAS detector and were scaled to the 2012 integrated luminosity. The inverse method is called *unfolding* and estimates the true kinematic spectrum from the observed distribution, recorded by the ATLAS detector.

The folding of, for example, a simulated W + jets event with detector effects is however probabilistic in nature. Assume that an observable is generated in the simulation in two bins and is reconstructed in either bin in 50 % of the cases – simply as a result of limited detector resolution. Then, the measurement of one event in bin 1 and one event in bin 2 does not allow exact conclusions in which bin the event was generated, i.e. to which bin it truely belongs. From the theoretical point of view, it is therefore important to note that the unfolding procedure cannot provide an exact determination of the true observable spectrum, but only an estimate of the likely true spectral values.

Several methods are available to perform unfolding of measured distributions. For this analysis, the iterative Bayesian unfolding method based on the technique described by D'Agostini [120, 121] has been chosen. This method accommodates the probabilistic nature of unfolding using Bayes' theorem as summarized in Section 6.1.

Distributions are unfolded to W + jets cross sections within a restricted – *fiducial* – phase space which is close to the measured one. The definition of the fiducial phase space used in this analysis is given in Section 6.2. Details on the implementation and tests of the iterative Bayesian unfolding in the W + jets analysis are described in Section 6.3 and Section 6.4, respectively. Unfolded data distributions are given in Section 6.5, while the measured W+jets cross sections at $\sqrt{s} = 8\,\text{TeV}$ are presented in the next chapter.

6.1 Iterative Bayesian unfolding

Bayesian unfolding estimates the true – *unfolded* – distribution from the measured data points by convoluting it with an unfolding matrix $\mathcal{M}(\mathbf{x_T}, \mathbf{x_D})$. $\mathbf{x_D} = \{x_{D,1}, x_{D,2}, ..., x_{D,m_D}\}$ here denotes the reconstructed observable values, split into m_D bins of the observable at the detector level, and $\mathbf{x_T} = \{x_{T,1}, x_{T,2}, ..., x_{T,m_T}\}$ specifies the same for the true observable

values, i.e. the observable at the *truth* level. The true bins $\mathbf{x_T}$ and the number of true bins m_T do not necessarily need to be the same as the respective quantities at the detector level, but are so in the W + jets analysis. The matrix relates $\mathbf{x_T}$ and $\mathbf{x_D}$ and comprises the information how probable a true observable value – the *cause* – is, when measuring the reconstructed observable value – the *effect*.

Denoting the number of measured events at the detector level as $N(\mathbf{x_D}) = \{N(x_{D,1}), ..., N(x_{D,m_D})\}$ and the number of estimated unfolded events similarly as $\hat{N}(\mathbf{x_T}) = \{\hat{N}(x_{T,1}), ..., \hat{N}(x_{T,m_T})\}$, the unfolding can be written as follows:

$$\hat{N}(\mathbf{x_T}) = \sum_{x_D} N(\mathbf{x_D}) \cdot \mathcal{M}(\mathbf{x_T}, \mathbf{x_D}) = \frac{1}{\varepsilon_{\mathbf{x_T}}} \sum_{x_D} N(\mathbf{x_D}) \cdot P(\mathbf{x_T}|\mathbf{x_D}) \,.$$

The unfolding matrix is thus given by the conditional probability that an observable has a true value of $\mathbf{x_T}$ given a reconstructed value of $\mathbf{x_D}$, $P(\mathbf{x_T}|\mathbf{x_D})$, and the efficiency, $\varepsilon_{\mathbf{x_T}}$, that a true value is actually reconstructed in any of the data bins. Viewed bin-wise, the probability $P(\mathbf{x_T}|\mathbf{x_D})$ specifies the matrix relating true and reconstucted bins $P(x_{T,i}|x_{D,j})$.

The probability $P(\mathbf{x_T}|\mathbf{x_D})$ can then be obtained by the application of Bayes' theorem:

$$P(\mathbf{x_T}|\mathbf{x_D}) = \frac{P(\mathbf{x_D}|\mathbf{x_T}) \cdot P(\mathbf{x_T})}{\sum_{\mathbf{x_T}} P(\mathbf{x_D}|\mathbf{x_T}) \cdot P(\mathbf{x_T})},$$

where $P(\mathbf{x_D}|\mathbf{x_T})$ is the conditional probability of reconstructing values $\mathbf{x_D}$ given the original true values $\mathbf{x_T}$. $P(\mathbf{x_T})$ indicates the *prior* which summarizes assumptions or prior knowlegde on the probability for a true value $\mathbf{x_T}$. The denominator just specifies the normalization and is otherwise not essential. $P(\mathbf{x_D}|\mathbf{x_T})$ – also denoted as response or smearing matrix – as well as $\varepsilon_{\mathbf{x_T}}$ – also called reconstruction efficiency – can be determined from MC simulation where true observable values can be generated and a model of the detector specifies if and how the observable is reconstructed.

The need to specify a prior is the weak point of Bayesian methods as they introduce an arbitrary element. This is resolved by the use of an *iterative* approach. Here, the unfolding process is repeated several times using the unfolded distribution from the previous iteration as new prior in the next iteration. The iterations thus signify a Bayesian unfolding with updating priors which diminishes the influence of the original prior usually after a few iterations. The convergence to the correct unfolded distribution is faster if the original prior is similar to the underlying true distribution. Therefore, the true distribution in the signal MC prediction is typically set as initial prior.

Usually, the observable values for the targeted signal process at the detector level $N(\mathbf{x_D})$ which are the input to unfolding, however, cannot directly be associated with the measured values in data. This is the result of the presence of backgrounds in the signal region. The number of signal events at the detector level can however be obtained by subtracting the number of estimated background events per observable bin $x_{D,j}$ from the number of events in data. Thus, $N(\mathbf{x_D}) = [N_{\text{data}} - N_{\text{bkg}}](\mathbf{x_D})$ which means, the unfolding can be written as:

$$\hat{N}(\mathbf{x_T}) = \sum_{x_D} [N_{\text{data}} - N_{\text{bkg}}](\mathbf{x_D}) \cdot \mathcal{M}(\mathbf{x_T}, \mathbf{x_D}) \,.$$

In order to relate reconstructed to true numbers of events, moreover, another effect has to be considered: the experimental issue of *fakes*. The fake fraction $f_{\mathbf{x_D}}$ takes into account that a certain fraction of the reconstructed signal events actually fall into their respective data bins only because e.g. pileup interactions modified the signature in the detector sufficiently. $f_{\mathbf{x_D}}$ is

either applied on $N(\mathbf{x_D})$ before the unfolding, or included into the definition of the unfolding matrix $\mathcal{M}(\mathbf{x_T}, \mathbf{x_D})$[1]. As a result of its origin, $f_{\mathbf{x_D}}$ must be obtained from a MC simulation of the considered signal process including the simulation of multiple proton interactions.

In total, three inputs from MC simulation are therefore needed for the unfolding: $P(\mathbf{x_D}|\mathbf{x_T})$, $\varepsilon_{\mathbf{x_T}}$ and $f_{\mathbf{x_D}}$. For this analysis, ALPGEN simulation samples which contain a simulation of overlayed minimum-bias events from PYTHIA8 are utilized to determine the three MC inputs. The iterative Bayesian unfolding is then performed using the ROOUNFOLD package [122].

Statistical uncertainties in the unfolded estimates $\hat{N}(\mathbf{x_T})$ are determined by ROOUNFOLD from the covariance matrix of the unfolded distribution:

$$V[\hat{N}(x_{T,i}), \hat{N}(x_{T,j})] = \sum_{k,l} \frac{\partial \hat{N}(x_{T,i})}{\partial N(x_{D,k})} V[N(x_{D,k}), N(x_{D,l})] \frac{\partial \hat{N}(x_{T,j})}{\partial N(x_{D,l})} .$$

This implies that the estimates for the bin contents in the unfolded distribution are correlated, even if the measured counts in data are independent between different bins, meaning the detector level covariance matrix $V[N(x_{D,k}), N(x_{D,l})]$ is diagonal. These correlations are the result of bin migrations described by the unfolding matrix as well as additional correlations introduced by the use of the unfolded distribution as prior in higher number of iterations, i.e. $\frac{\partial \hat{N}(x_{T,i})}{\partial N(x_{D,k})} = \mathcal{M}(x_{T,i}, x_{D,k}) + (\text{terms for iterations} > 1)$. For details see Ref. [122].

6.2 Fiducial phase space

The unfolding provides a way to estimate the shapes of the true kinematic distributions for the investigated signal process. For a measurement of the total W+jets cross section, the unfolding not only needs to correct for detector effects in the distributions, but also to extrapolate into regions of phase space without detector instrumentation. The latter largely relies on the proper modelling of this extrapolation in the MC simulation. Therefore, the measurement is unfolded to a *fiducial phase space* which is close to the detector level selection, and the fiducial cross section is measured instead of the total cross section.

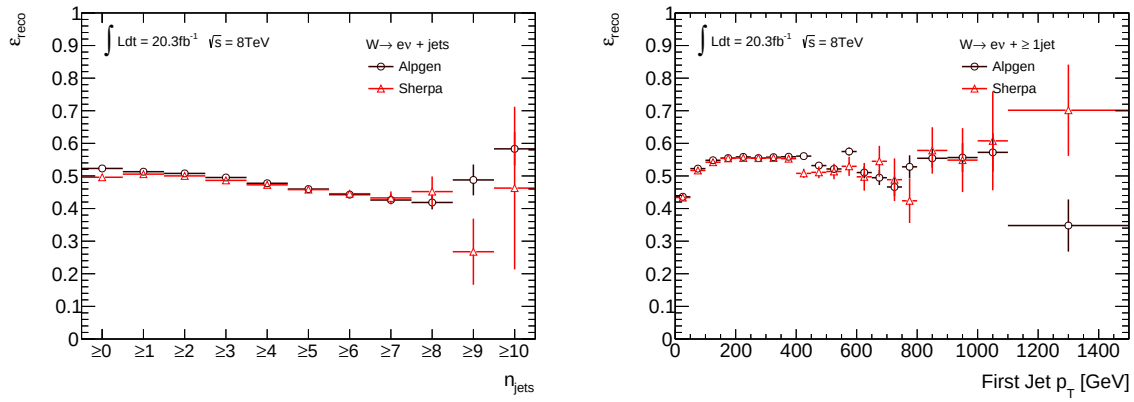
The fiducial phase space used in this analysis requires the presence of exactly one electron with p_{T} and $|\eta|$ requirements similar to the W+jets signal region, and $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} selections as at the detector level. The exact requirements on the fiducial phase space are listed in Table 6.1. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is obtained from the neutrino and both electron and neutrino originate from the W boson decay. To reflect the finite calorimeter resolution which measures small angle photon radiation as part of the electron cluster, the true electron is *dressed*, i.e. photons within $\Delta R < 0.1$ are added to the four-vector of the electron.

$W \to e\nu$ fiducial region	
Electron	Exactly 1 electron with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.5$
Neutrino	$E_{\mathrm{T}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$
W boson	$m_{\mathrm{T}} > 40\,\mathrm{GeV}$

Table 6.1: Fiducial phase space at particle level for the cross section measurement of $W \to e\nu$ +jets.

Similar to the detector level, events in the fiducial region are categorized according to the number of jets. Jets at the truth level are defined by the anti-k_t algorithm with a radius

[1] The latter approach is implemented in the used unfolding software and therefore employed in this analysis.



(a) n_{jets} (b) Leading jet p_{T}

Figure 6.1: Reconstruction efficiency as a function of n_{jets} (a) and the leading jet p_{T} for ≥ 1 jet (b) as determined from the two W + jets signal generators, ALPGEN and SHERPA.

parameter of $\Delta R = 0.4$, clustering truth particles with lifetimes longer than 30 ps (excluding muons and neutrinos - which are also not contained in jets at the detector level). The jets are required to have $p_{\text{T}} > 30\,\text{GeV}$ and $|\eta| < 4.4$ and must not overlap with the selected electron within $\Delta R < 0.5$.

6.2.1 Reconstruction efficiency

The efficiency to reconstruct a true event in the fiducial region in any observable bin at the detector level is an important correction needed for the unfolding. The reconstruction efficiency is defined as:

$$\varepsilon_{\text{reco}} = \frac{N_{\text{truth}}^{\text{reco}}}{N_{\text{truth}}^{\text{all,fid}}} \,.$$

So, it gives the fraction of truth events in the fiducial region which pass the selection of the W + jets signal region at the detector level. The reconstruction efficiency is measured in both ALPGEN and SHERPA MC simulation. Examples for the reconstruction efficiency as a function of n_{jets} and the leading jet p_{T} are given in Figure 6.1. Reconstruction efficiencies are found to be around 50 %, depending slightly on the exact observable value.

In turn, this means that approximately 50 % of the generated MC events in the fiducial region are lost in the reconstruction by failing one or more of the W + jets signal region requirements. In order to investigate the origin of the reconstruction losses, the event reduction from the different signal region requirements is computed for events passing the fiducial selection using the SHERPA W + jets simulation. The results are listed in Table 6.2 with and without the trigger requirement applied. In particular the used `EF_e24vhi_medium1` trigger item comprises, in addition to the electron p_{T} requirement of 24 GeV, electron identification and isolation criteria which influence the amount of event reduction from the corresponding requirements in the signal region selection. The numbers without the trigger requirement applied demonstrate that the largest loss in the reconstruction comes from the *tight* electron identification. It is in line with the *tight* electron identification efficiency measured by the ATLAS working group for electron performance and shown in Figure 4.1. The second largest contribution to the event loss arises from the electron isolation combined with the impact parameter requirements which do not have any correspondence in the fiducial phase space

definition. Together, the electron identification and isolation including the impact parameter criteria account for a loss of truth events in the reconstruction of approximately 38 %, i.e. the majority of the reconstruction loss. Additional influences from the electron η adjustment which is an acceptance correction and the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction are smaller, as expected.

Selection (Detector level)	Trigger required		Trigger not required	
	f_{rejected}	$f_{\mathrm{surviving}}$	f_{rejected}	$f_{\mathrm{surviving}}$
Trigger	23.1	76.9	–	–
Electron quality and η	2.2	97.8	4.4	95.6
Tight electron ID	13.4	86.6	26.3	73.7
Electron p_{T}	0.5	99.5	2.7	97.3
Electron isolation	3.7	96.3	8.9	91.1
Electron impact parameters	7.8	92.2	7.8	92.2
$E_{\mathrm{T}}^{\mathrm{miss}}$ validity	3.4	96.6	3.4	96.6
$E_{\mathrm{T}}^{\mathrm{miss}}$	9.1	90.9	9.1	90.9
m_{T}	0.4	99.6	0.4	99.6
b-Jet rejection	0.5	99.5	0.5	99.5
Total	49.9	50.1	49.9	50.1

Table 6.2: Rejected or surviving fractions of events in the fiducial region, f_{rejected} or $f_{\mathrm{surviving}}$, in %, if W+jets signal region requirements are applied. The total fraction of surviving events corresponds to the reconstruction efficiency of W + jets events in this analysis. The numbers have been obtained with SHERPA and are given with and without the trigger decision required, since the trigger items already contain electron identification and isolation criteria.

In total, the reconstruction efficiency of around 50 % results mainly from the stringent electron selection which is essential to suppress backgrounds, in particular the multi-jet background, and therefore unavoidable. The reconstruction efficiencies for the remaining unfolded distributions are given in Appendix A.3.

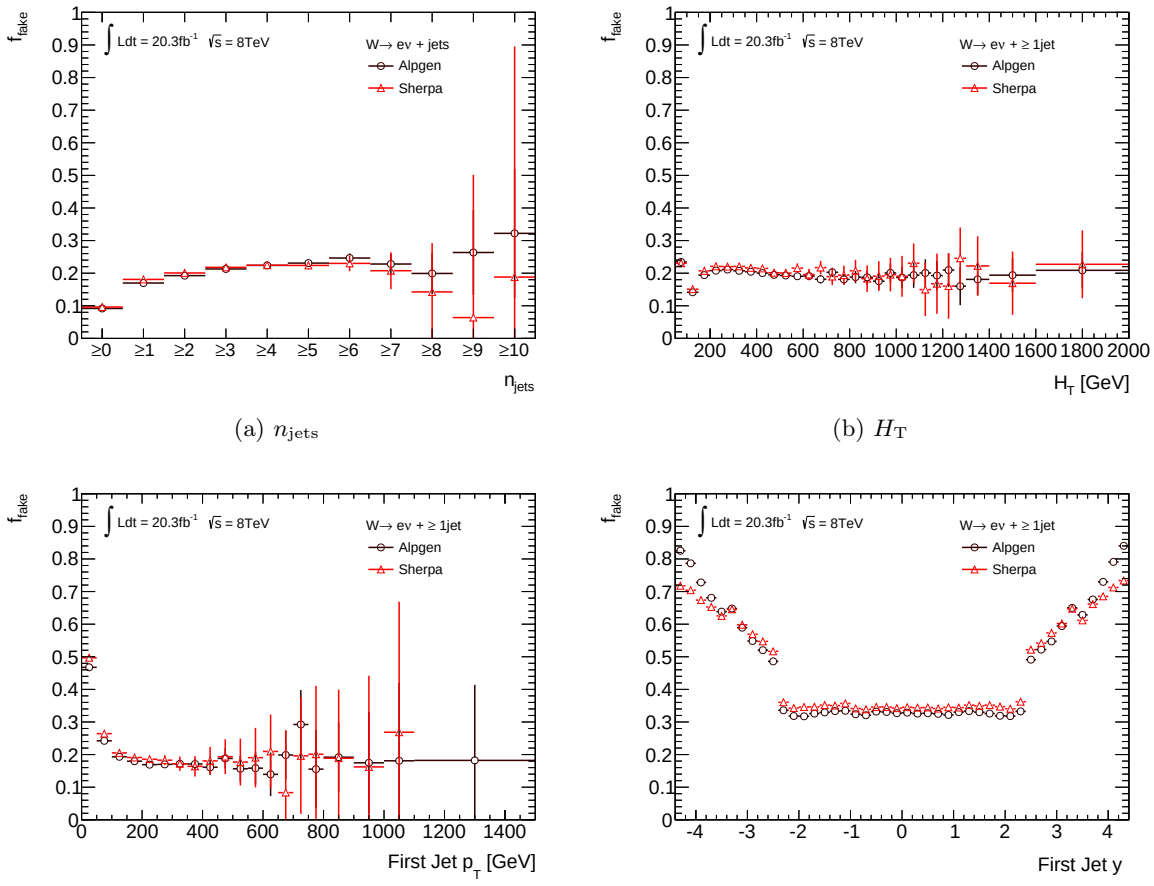
6.2.2 Fake fraction

The fraction of events reconstructed in the W + jets signal region, but not originating from the fiducial region is another important input to the unfolding procedure. This so-called *fake fraction* is defined as:

$$f_{\mathrm{fake}} = \frac{N_{\mathrm{reco}}^{\mathrm{not\ truth}}}{N_{\mathrm{reco}}^{\mathrm{all,\ SR}}} .$$

It is measured in ALPGEN and SHERPA MC simulation. The fake fraction as a function of n_{jets}, H_{T} as well as the leading jet p_{T} and rapidity is shown in Figure 6.2. The fake fractions for the other observables are given in Appendix A.3. In general, the fake fractions are around 20 %. For low p_{T} values of the leading jet, f_{fake} increases up to around 50 %. For central rapidity of the leading jet f_{fake} is about 35 %, but increases to about 70 % with small differences between the two W + jets generators.

The origin of the fakes, both of the general fake fraction of 20 % as well as the large increases towards low leading jet p_{T} and high rapidity, is investigated in the following. To trace the origin of the general fake fraction, the reduction of events in the W + jets signal region is studied, when the different criteria of the fiducial region are applied. The results are listed in Table 6.3. The largest reduction of reconstructed events and thus the main source of fakes in the jet-inclusive W + jets signal region is the $E_{\mathrm{T}}^{\mathrm{miss}}$. This is in line with observations



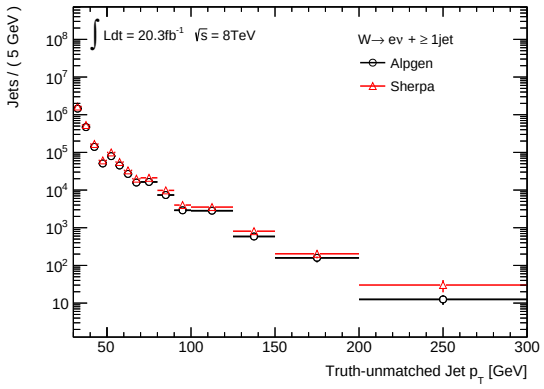
(a) n_{jets} (b) H_{T}

(c) Leading jet p_{T} (d) Leading jet rapidity

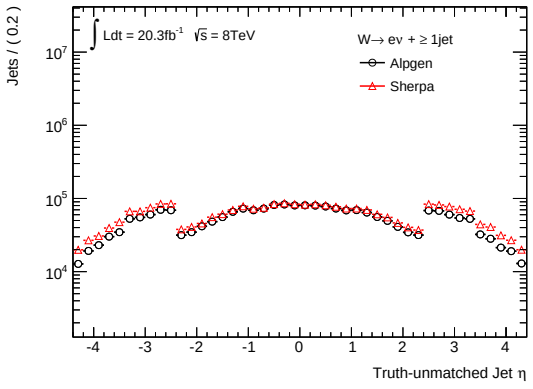
Figure 6.2: Fake fraction as a function of n_{jets} (a), H_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d) for ≥ 1 jet as determined from the two W + jets signal generators, ALPGEN and SHERPA.

Selection (Truth level)	f_{rejected}	$f_{\text{surviving}}$
Electron η	<0.1	>99.9
Electron p_{T}	0.3	99.7
$E_{\text{T}}^{\text{miss}}$	9.1	90.9
m_{T}	0.2	99.8
Total	9.5	90.5

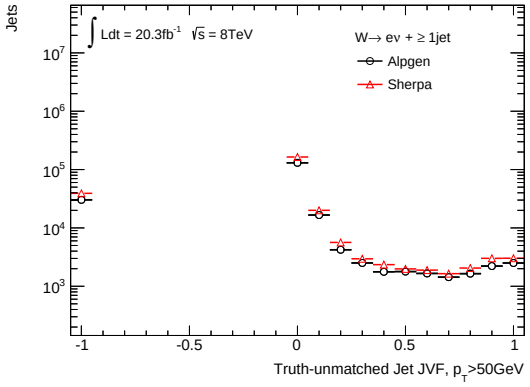
Table 6.3: Rejected or surviving fractions of events in the W + jets signal region, f_{rejected} or $f_{\text{surviving}}$, in %, if the fiducial region requirements are applied. Rejected events are classified as fake events. The numbers have been obtained using SHERPA W +jets signal prediction.



(a) Jet p_{T}



(b) Jet η

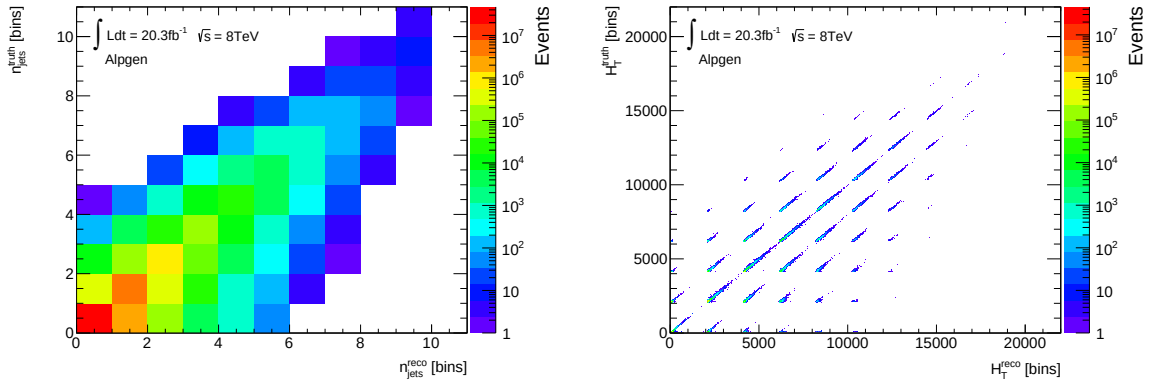


(c) JVF, if $p_{\mathrm{T}} > 50\,\mathrm{GeV}$

Figure 6.3: Kinematics of reconstructed jets which cannot be matched to a truth jet. Shown are the jet p_{T} (a), jet η (b) and JVF if $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ (c), determined both from ALPGEN and SHERPA W + jets simulation. Note that a JVF-value of -1 means that the JVF is not defined as e.g. for forward jets.

of a positive linearity bias in $W \to \mu\nu$ events for $E_{\mathrm{T}}^{\mathrm{miss,\,Truth}} < 40\,\mathrm{GeV}$ by the ATLAS performance group for $E_{\mathrm{T}}^{\mathrm{miss}}$-computation [80]. The linearity is defined as the mean value of the ratio $(E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,\,Truth}})/E_{\mathrm{T}}^{\mathrm{miss,\,Truth}}$ and a positive bias implies that the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ is overestimated compared to $E_{\mathrm{T}}^{\mathrm{miss,\,Truth}}$. Since the W + jets signal requirement lies in this low $E_{\mathrm{T}}^{\mathrm{miss}}$ region, the linearity bias leads to reconstructed events passing the W + jets selection, without fulfilling the fiducial requirements. This explains the fake fraction of about 10 % in the inclusive jet bin (≥ 0 jets) in Figure 6.2a. The increase of the fake fraction to 20 % occurs in the presence of at least 1 jet and is thus likely to result from the jet resolution at the detector level. This means that truth jets with p_{T} below 30 GeV are shifted in the reconstruction above the 30 GeV requirement and hence classify the event as W+≥ 1 jet.

The increase in the fake fraction for the leading jet at low p_{T} and high rapidities is investigated by matching reconstructed and truth jets according to their angular distance ($\Delta R < 0.4$) and examining reconstructed jets which could not be matched to a truth jet. Figure 6.3 displays the jet p_{T} and η of unmatched jets as well as the JVF of unmatched jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. Both the p_{T} and the η distributions show the boundaries of the pile-up jet suppression by the JVF requirement which is only applied for $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $|\eta| < 2.4$. The largest contribution of unmatched jets is found at very low jet p_{T}, i.e. just above the selection



(a) n_{jets} (b) H_{T}

Figure 6.4: Response matrices for the n_{jets} (a) and H_{T} (b) distributions, as determined from the ALPGEN W + jets generator. The response matrix for the H_{T} distribution contains the $H_{\text{T}}^{\text{truth}}$ vs. $H_{\text{T}}^{\text{reco}}$ correlation repeatedly for the different $n_{\text{jets}}^{\text{truth}}$ and $n_{\text{jets}}^{\text{reco}}$ values.

threshold of 30 GeV. This indicates again that the general fake fraction in the presence of jets is mainly the effect of migrations around the jet selection p_{T} threshold due to limited jet energy resolution. The forward jets and jets with $p_{\text{T}} > 50$ GeV and JVF-values close to zero are however most likely the result of pile-up. Pile-up is therefore the most probable explanation for the increase of the fake rates towards high rapidity. Further studies on the origin of the fakes are however needed to fully disentangle the pile-up and jet resolution contributions.

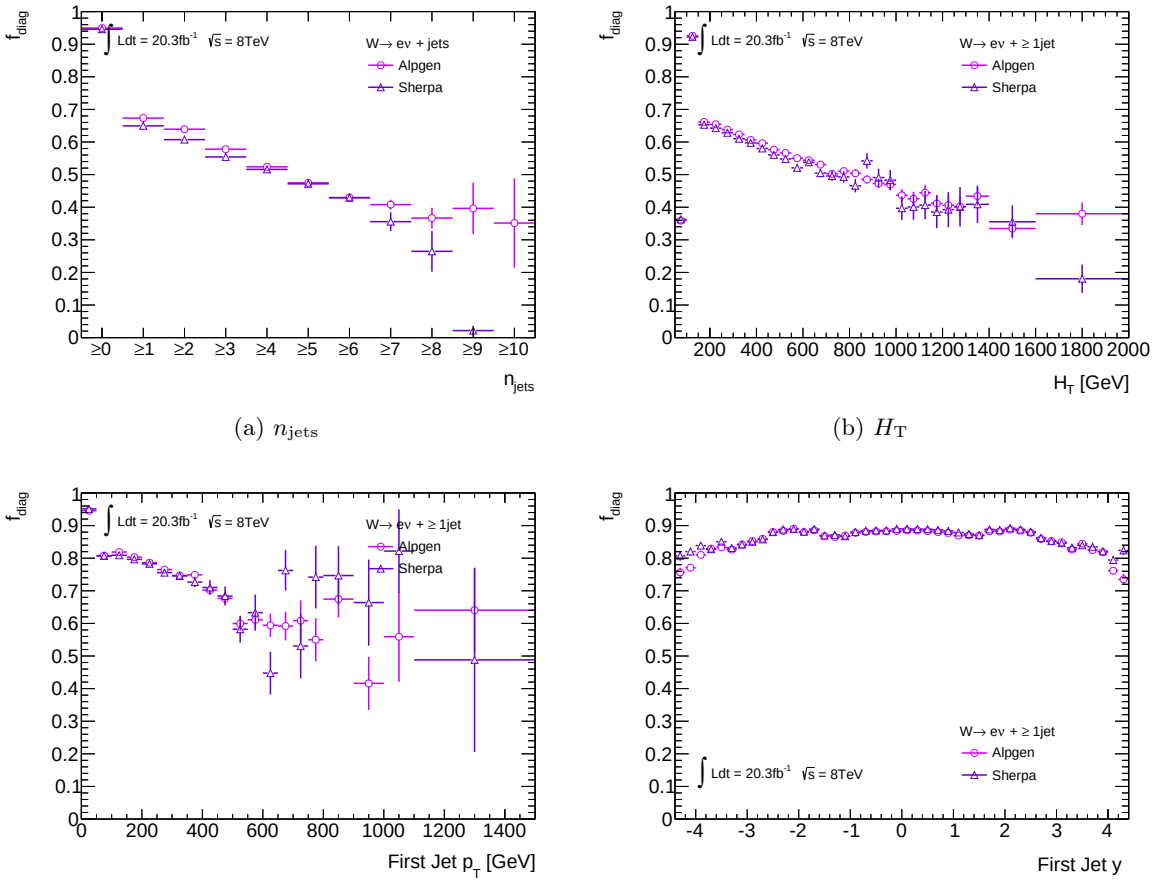
6.3 Extrapolation between reconstruction and particle level

The main ingredient for the Bayesian unfolding – the response matrix – contains the information on how reconstructed observables are related to the corresponding truth quantities. For each unfolded distribution one response matrix is determined from W + jets signal MC. It contains events which fulfil both the W + jets signal region requirements and the fiducial phase space selection. For all distributions (apart from the n_{jets} distribution), the unfolding is conducted in two-dimensions simultaneously. The first dimension is the jet multiplicity and second the respective observable. The response matrix thus contains the information on how events with e.g. $(n_{\text{jets}}^{\text{reco}}, H_{\text{T}}^{\text{reco}})$ are related to $(n_{\text{jets}}^{\text{truth}}, H_{\text{T}}^{\text{truth}})$. The response matrices which are used in the unfolding of data are determined with the ALPGEN W + jets simulation. The response matrices for the n_{jets} (1-dimensional unfolding) and, as an example, the H_{T} distribution (2-dimensional unfolding) are displayed in Figure 6.4. The response matrices for the other unfolded distributions are shown in Appendix A.4.

Events on the diagonal of the matrices are reconstructed in the same bin as they were generated. The number of these events compared to the total number of events in the matrix at same reconstructed observable value is defined as the *diagonal fraction*:

$$f_{\text{diag}} = \frac{N_{\text{matrix}}^{\text{reco=truth}}}{N_{\text{matrix}}^{\text{reco bin}}} .$$

For the 2-dimensional unfolding matrices, the denominator of f_{diag} contains only events where in addition $n_{\text{jets}}^{\text{reco}} = n_{\text{jets}}^{\text{truth}}$, i.e. only events from the sub-matrix sequence on the diagonal in Figure 6.4b are considered. The diagonal fraction provides an indication how large migrations



(a) n_{jets} (b) H_{T}

(c) Leading jet p_{T} (d) Leading jet y

Figure 6.5: Fraction of events on the diagonal of the response matrix as a function of n_{jets} (a) and H_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d) as determined from Alpgen and Sherpa. For the latter three, only events with ≥ 1 jet and $n_{\text{jets}}^{\text{reco}} = n_{\text{jets}}^{\text{truth}}$ are considered.

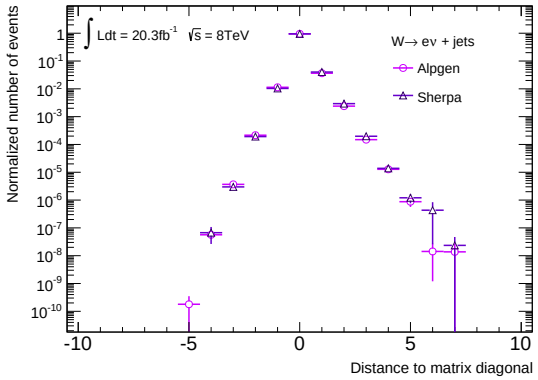


Figure 6.6: Spread of the n_{jets} response matrix with respect to the diagonal for both Alpgen and Sherpa.

between bins at the reconstructed level are, i.e. its reduction from one indicates what fraction of events is reconstructed in a different bin than generated.

For the n_{jets}, H_{T}, leading jet p_{T} and rapidity response matrices, f_{diag} is shown in Figure 6.5. For the n_{jets} and the H_{T} distribution, f_{diag} is observed to be mostly between 0.7 and 0.4. For n_{jets}, this reflects the spread in the jet multiplicity visible in the response matrix. The reduction of f_{diag} towards larger jet multiplicities furthermore indicates that jet multiplicity migrations increase at higher n_{jets}. For H_{T}, the f_{diag}-values demonstrate that bin migrations in H_{T} occur, in particular, towards higher H_{T}. This is understandable considering that H_{T} is an observable combining electron, $E_{\text{T}}^{\text{miss}}$ and jet momenta and is thus more prone to migrations between bins than observables based on individual particles. The diagonal fractions for the leading jet p_{T} and rapidity, consequently, are higher and lie mostly within 0.8-0.6 and 0.9-0.8, respectively. In particular for the leading jet rapidity, f_{diag} demonstrates that this observable can be reconstructed very well.

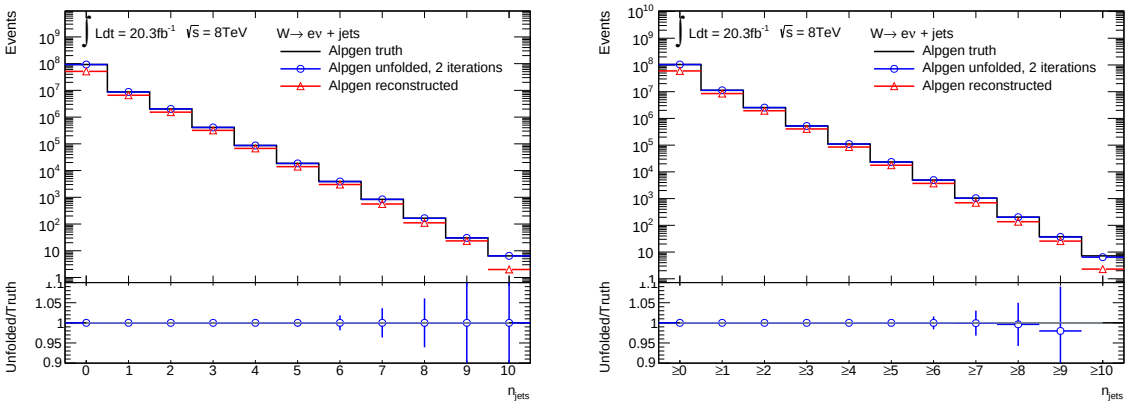
To investigate further the spread in the n_{jets} response matrix, the distance of entries to the diagonal is computed and displayed in Figure 6.6. The n_{jets}-spread is determined both from ALPGEN and SHERPA samples and is found to be in reasonable agreement between the two. The n_{jets} migrations are therefore probably the result of common simulation features at the low jet p_{T}-threshold such as jet response and pile-up. Both are simulated by separate software for the detector and the pile-up modelling (not ALPGEN or SHERPA) and are common in both the ALPGEN and SHERPA samples.

Bin migrations are thus observed mainly in the n_{jets} and H_{T} distributions and are taken into account by the unfolding procedure.

6.4 Tests of the unfolding method

The proper behaviour of the unfolding procedure as well as the validity of the unfolding inputs and the unfolded outputs are examined in a series of tests which are discussed in the following.

6.4.1 Closure test



(a) Exclusive n_{jets}

(b) Inclusive n_{jets}

Figure 6.7: Closure test of the 1-dimensional unfolding for the exclusive (a) and inclusive (b) n_{jets} distribution using the ALPGEN W + jets generator. The inclusive distribution is obtained by a summation of the respective jet multiplicities of the exclusive n_{jets} distribution.

The basic test of correct performance of the unfolding procedure assesses the closure of the sequence – generation of a true distribution, detector simulation, particle reconstruction and finally unfolding – and compares the unfolded distribution to the original true one. For the closure test, the reconstructed distribution in the MC simulation is unfolded using the response matrix and prior obtained from the same simulation. An almost 100 % agreement between the unfolded and the original true distribution is expected, where small deviations are possible only due to statistics losses in the reconstructed distribution and the response matrix.

Figure 6.7 shows the closure of the 1-dimensional unfolding of the n_{jets} distribution using the ALPGEN W + jets simulation. Note that the unfolding is performed based on exclusive jet multiplicities and inclusive numbers are obtained afterwards by adding the events above a certain jet multiplicity. The lower panel in Figure 6.7 displays the ratio of the unfolded to the original true distribution which agrees very well with unity. The unfolding thus recovers the correct distribution at the truth level from the reconstructed distribution. The reconstructed distribution is drawn in Figure 6.7 for comparison. The difference between reconstructed and unfolded distribution originates from the correction applied by the unfolding procedure in terms of normalization and shape according to the reconstruction efficiency and fake fraction.

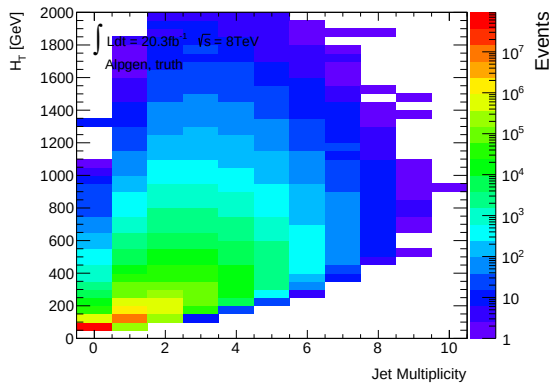
Figure 6.8 displays the original true, the reconstructed and the unfolded distribution obtained with the ALPGEN W + jets simulation exemplary for the H_{T} distribution. For the projections of individual jet multiplicities, e.g. 1 jet, bins in H_{T} with low statistics are combined, while the unfolding itself was performed with the equidistant binning of the migration matrix. The original equidistant binning corresponds to the bin size of the high statistics bins in the projections. The size of low-statistics bins was optimized by requiring at least about 10 events per bin for $n_{\text{jets}} \geq 1$, while still being reasonable also for the exclusive $n_{\text{jets}} = 1$ projection. Figure 6.9 presents the results of the closure test for the H_{T} distribution in the 1 and ≥ 1 jet bins. The lower panel shows the ratio of the unfolded distribution compared to the original true distribution and is found to agree well with unity. The smaller relative uncertainty at the higher H_{T} values in the unfolded distribution for inclusive ≥ 1 jet compared to exclusive 1 jet is the result of increased statistics at these H_{T} values due to contributions from higher jet multiplicities. The closure tests for the remaining unfolded distributions are shown in Appendix A.5. Good closure is observed there as well.

6.4.2 Definition of the number of iterations

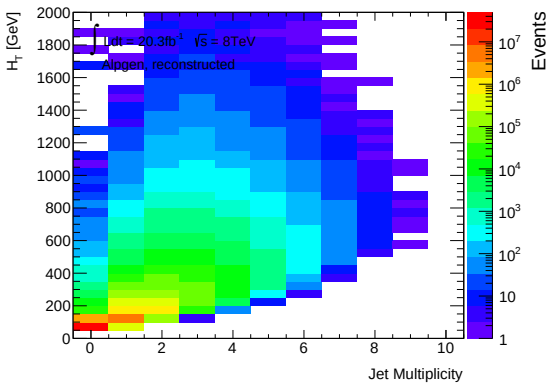
An important parameter of iterative Bayesian unfolding is the number of iterations in the unfolding procedure. Starting from an initial prior, the unfolding is performed repeatedly with the prior being updated to the estimate of the true, i.e. the unfolded distribution from the previous iteration. This reduces the influence of the initial prior, but at the same time statistical fluctuations in data are enhanced by the repeated use of data in the procedure. The chosen number of iterations therefore needs to balance the reduction of the influence from the initial prior and the increase of statistical uncertainties in the unfolded results.

Although the definition of the number of iterations is an issue for any iterative unfolding technique, there is no theoretically sound and experimentally approved single recipe or quantity which is generally accepted as the best optimization parameter. Several quantities are therefore investigated to obtain the number of iterations which balances the above criteria in the best possible way.

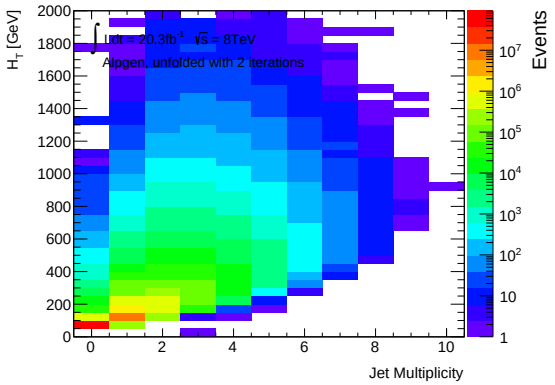
In order to check the influence of the number of iterations, the unfolding is therefore performed several times with increasing number of iterations, covering 1 to 10. The effect on the statistical uncertainties of the unfolded distributions is shown in Figure 6.10 for the



(a) Truth

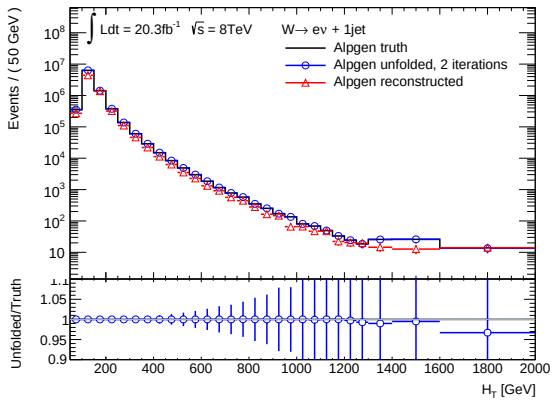


(b) Reconstructed

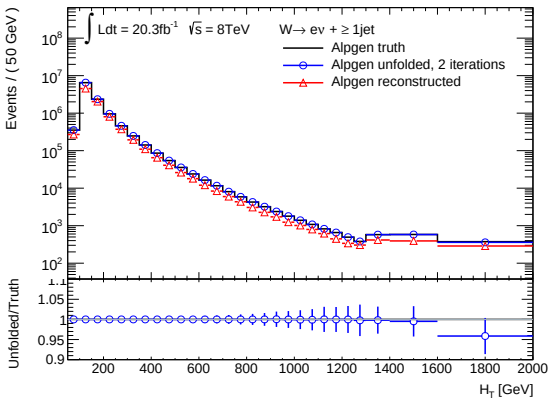


(c) Unfolded

Figure 6.8: H_{T} distribution in bins of jet multiplicity determined from Alpgen $W+$jets simulation at particle level (a), after reconstruction (b) and after unfolding (c). The unfolding was performed using the response matrix from Alpgen as well.

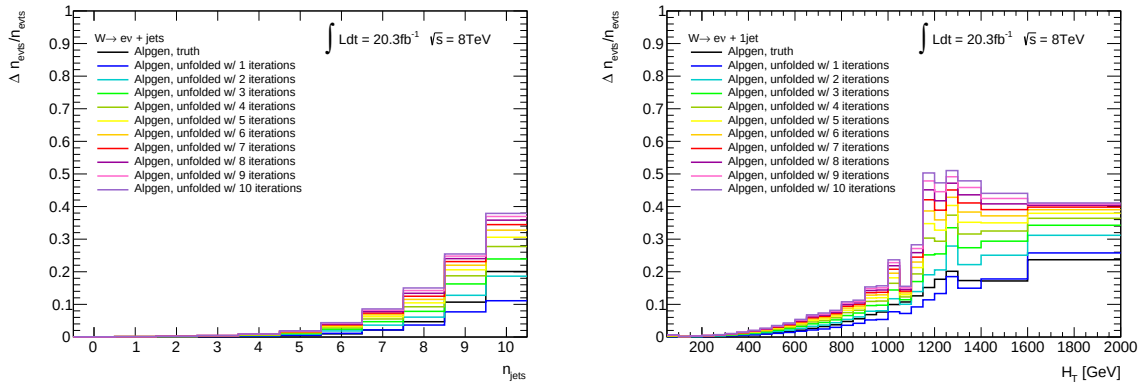


(a) 1 jet



(b) ≥ 1 jet

Figure 6.9: Closure test of the 2-dimensional unfolding of the H_{T} distribution in the case of 1 jet (a) and ≥ 1 jet (b) using the Alpgen $W+$jets generator. The distribution for ≥ 1 jet is obtained from a projection of the respective jet multiplicities in the 2-dimensional distributions as shown in Figure 6.8.



(a) n_{jets} (b) H_{T}

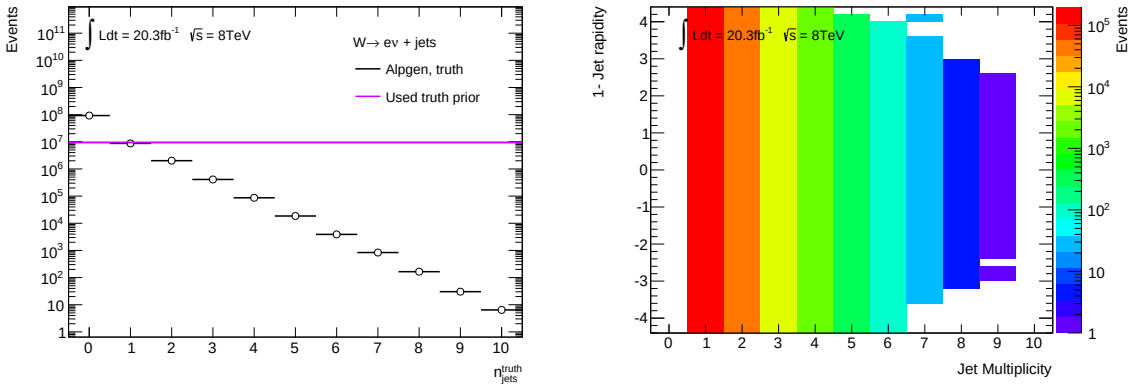
Figure 6.10: Relative statistical uncertainties in the unfolded results with increasing number of iterations and in the original truth distribution for n_{jets} (a) and H_{T} in the case of 1 jet (b). The unfolding here uses the same setup as for the closure test in Section 6.4.1.

n_{jets} and the H_{T} distribution for 1 jet. The unfolding is performed using the same setup as for the closure test, i.e. ALPGEN W + jets reconstructed distributions are unfolded using the response matrix obtained with ALPGEN. The statistical uncertainties in the unfolded result per bin correspond to the square root of the diagonal entries of the unfolding covariance matrix. Both the statistical uncertainties and the covariance matrix are discussed in more detail in Section 6.4.3 and Section 6.5.1, respectively. The relative size of these statistical uncertainties is observed to grow with decreasing statistics in the original distribution (i.e. towards a larger number of jets and higher H_{T}), but also with increasing number of iterations. For around 2 iterations, the statistical uncertainties are roughly the same size as the statistical uncertainties in the original true distribution.

The influence of the initial prior on the unfolded result is investigated by using a flat initial prior in the unfolding and comparing the behaviour of the unfolded distributions as a function of the number of iterations. The ROOUNFOLD package was modified for this purpose to allow the use of an external, here, flat prior. A flat initial prior, i.e. assuming the same initial probability for all truth bins, is no physically meaningful assumption, but a useful test case to see the dependence on the initial prior. It is however observed to be difficult for the Bayesian unfolding to obtain the correct normalization when the prior is flat as a function of the jet multiplicity where events counts differ by orders of magnitude. In the 2-dimensional unfolding, therefore, the priors are defined as flat within the individual jet multiplicities only. Figure 6.11 displays the flat test priors for the 1-dimensional unfolding of the n_{jets} distribution and the 2-dimensional unfolding of the leading jet rapidity as examples.

The unfolding is then performed using the reconstructed distributions as well as the response matrix from ALPGEN W + jets simulation, but with the defined flat initial priors. Figure 6.12 shows the unfolded distributions which are determined with growing number of iterations, compared to the ALPGEN truth distribution. While the influence of the prior is clearly visible for 1 iteration, it moves closer to the truth already for the second iteration. More than 2 iterations cause much smaller changes, slowly leading to a convergence of the unfolded distributions to the same value.

The difference of the unfolded to the true distribution as a function of the number of



(a) n_{jets} (b) Leading jet rapidity

Figure 6.11: Flat test priors for the 1-dimensional unfolding of the n_{jets} (a) and the 2-dimensional unfolding of the leading jet rapidity (b) distributions. The employed constant value corresponds to the average of non-zero bins within the distribution (a) or within a certain jet multiplicity (b). In the case of the n_{jets} distribution the true distribution which is normally used as prior is displayed for illustration purposes. The priors are normalized automatically within the RooUnfold package.

iterations is quantified by computing χ^2 as follows:

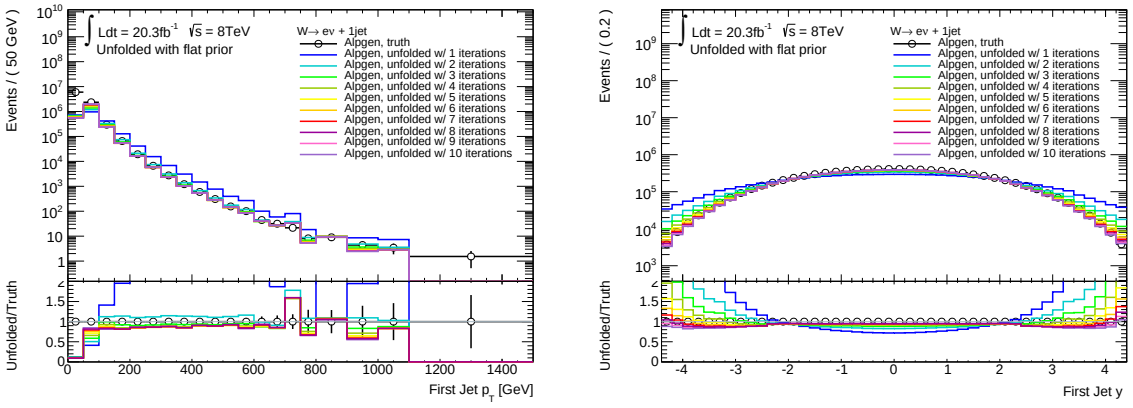
$$\chi^2 = \sum_i^{n_{\text{bins}}} \frac{\left(\hat{N}(x_{T,i}) - N(x_{T,i})\right)^2}{\Delta N(x_{T,i})^2},$$

with $\hat{N}(x_{T,i})$ being the unfolded number of events and $N(x_{T,i})$ and $\Delta N(x_{T,i})$ the number and uncertainty in the number of truth events in bin i.

Figure 6.13 presents the unfolded distributions with increasing number of iterations and the corresponding χ^2 values of the unfolded n_{jets} distribution. As mentioned above, the convergence of the n_{jets} distribution is found to be difficult when the prior is flat over n_{jets} and thus over large differences in event counts. Still, the unfolded result with 2 iterations is observed to agree roughly with the true distribution. With growing number of iterations, it converges however to a minimum different from the true distribution. This behaviour can also be deduced from the χ^2 distribution which has a clear minimum at 2 iterations which is visible both for Alpgen and Sherpa. The difference in the magnitude of the χ^2 values between Alpgen and Sherpa follows from the larger statistical uncertainties in Sherpa for which less MC events were generated.

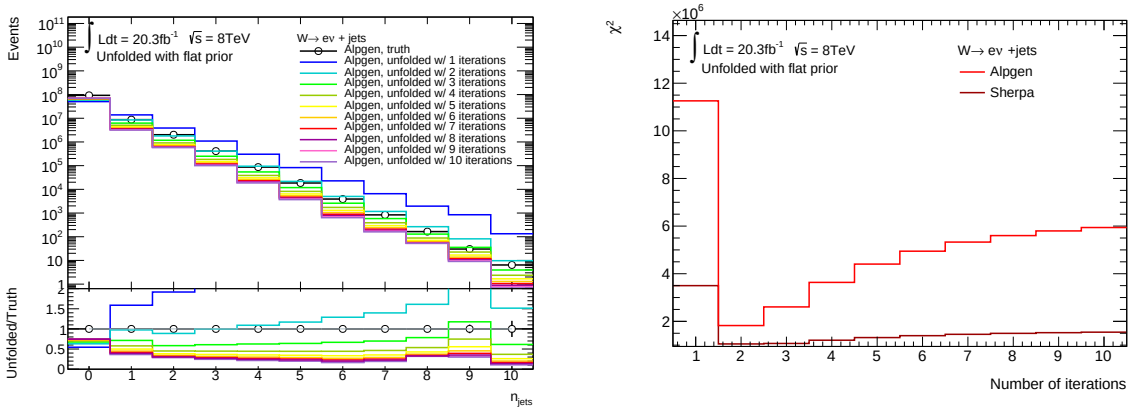
For the 2-dimensional unfolding, the χ^2 can be either computed using the 2-dimensional distributions or the 1-dimensional projections to the different jet multiplicities. Figure 6.14 displays the global χ^2 as well as the n_{jets}-dependent χ^2 values for the H_{T} distribution. The global χ^2 demonstrates a fast convergence already after the second iteration to the true distribution. This mainly results from the convergence behaviour of the 0-jet case which dominates the global χ^2 value, since it has the largest number of events with the smallest statistical uncertainties. For 1 or 2 jets, the convergence for higher number of iterations moves away from the truth, making lower number of iterations preferable.

Using a better assumption for the prior like the true distribution from a different MC simulation, e.g. Sherpa, the convergence of the unfolded distribution to the true distribution is very fast and observed for all unfolded distributions. In the unfolding of 2012 data, the true distribution from Alpgen is used as the prior.



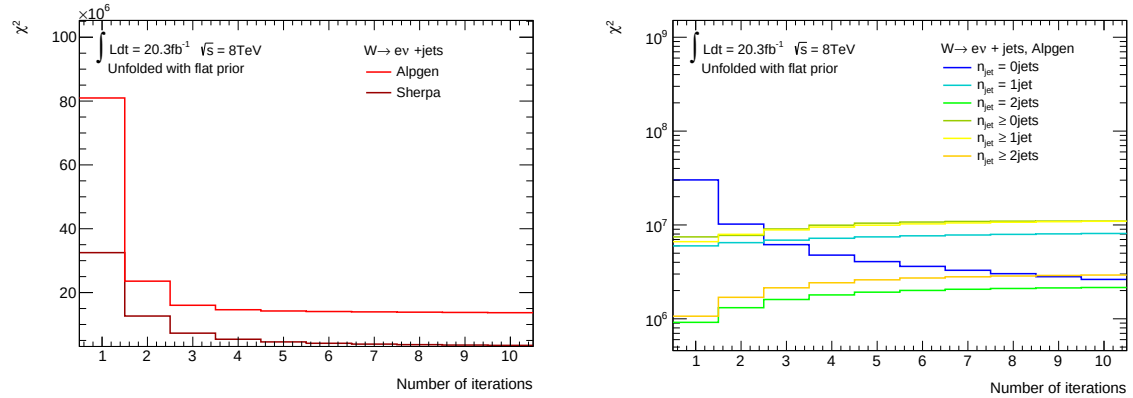
(a) Leading jet p_{T} (b) Leading jet rapidity

Figure 6.12: Unfolded results for the leading jet p_{T} (a) and rapidity (b) in the case of 1 jet as a function of the number of iterations and compared to the original true distribution. The reconstructed distribution and the response matrices are obtained with ALPGEN W + jets simulation and the unfolding starts from flat priors as shown in Figure 6.11.



(a) Unfolded distributions (b) χ^2

Figure 6.13: Impact of the number of iterations on the unfolded n_{jets} distribution, if a flat initial prior is used in the Bayesian unfolding. Shown are unfolded distributions after the different iterations compared to the true distribution for ALPGEN (a) and the χ^2 values when running the test with ALPGEN or SHERPA (b).



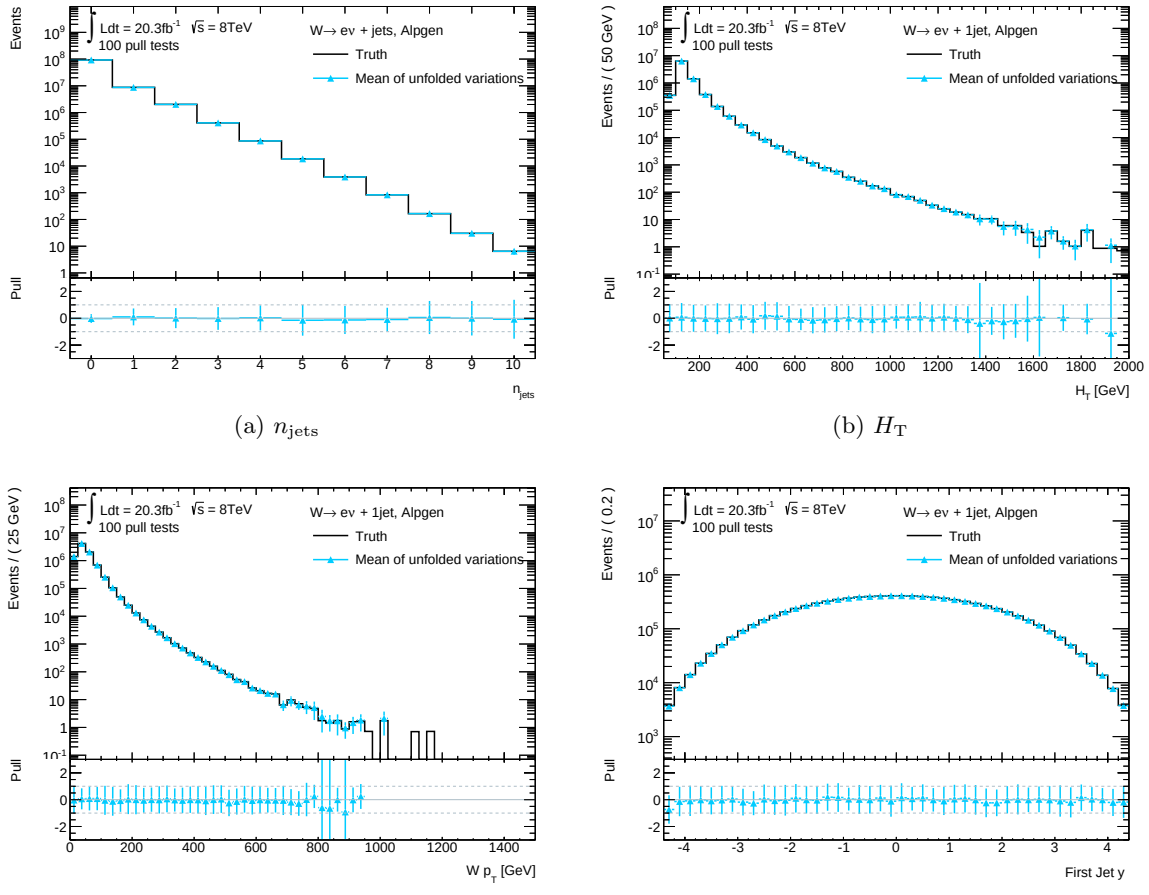
(a) Global χ^2 (b) n_{jets}-dependent χ^2

Figure 6.14: Difference between the unfolded H_{T} distribution and the true distribution as a function of the number of iterations for the global 2-dimensional χ^2 (a) and for different n_{jets} cases separately (b).

In general, a choice of 2 iterations is sufficient to reduce the influence from the prior on the unfolded result. After 2 iterations, most of the departure of the unfolded from the true distribution is resolved and further iterations only bring very small adjustments, but increase the statistical uncertainties. The statistical uncertainties after 2 iterations are close to the statistical uncertainties of the original true distribution and have not increased unreasonably due to the number of iterations yet. For these reasons, the number of iterations in the iterative Bayesian unfolding is set to 2 for the analysis presented in this thesis.

6.4.3 Statistical uncertainties

Estimates of the statistical uncertainties in the unfolded distributions correspond to the square root of the diagonal entries in the covariance matrix. It is important to assess whether these uncertainties reflect statistical fluctuations in the input data appropriately. One way to confirm this is a *pull test* as described in the following.

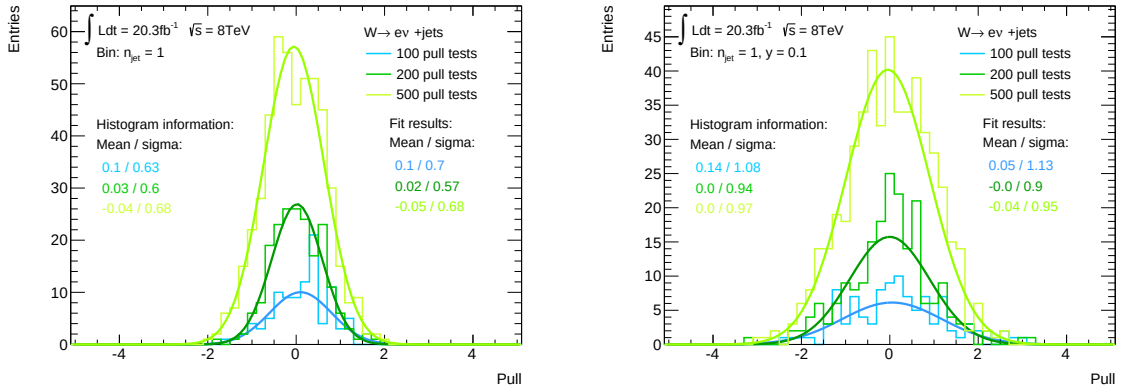


(a) n_{jets} (b) H_{T}

(c) W p_{T} (d) Leading jet rapidity

Figure 6.15: Unfolded results after Poisson variations of the reconstructed input distributions compared to the truth for the n_{jets} (a), H_{T} (b), W p_{T} (c) and leading jet rapidity (d) distributions for 1 jet, determined using Alpgen W + jets simulation. The lower panel gives the mean and spread of the bin-wise pull distributions.

For the pull test, reconstructed distributions from the Alpgen W + jets simulation are unfolded using the response matrix from Alpgen. In contrast to the setup for the closure test, however, the reconstructed distributions are bin-wise fluctuated before the unfolding. For



(a) n_{jets} (b) Leading jet rapidity

Figure 6.16: Pull distribution for the 1-jet-bin of the n_{jets} distribution (a) and the central bin of the leading jet rapidity distribution at 1 jet (b). The pull test is repeated 100, 200 or 500 times and the pull distributions are fitted with Gaussian functions.

each bin, the number of reconstructed events is varied randomly using a Poisson distribution with the original number of events as mean. In this way, a repetition of the measurement with statistical fluctuations is imitated.

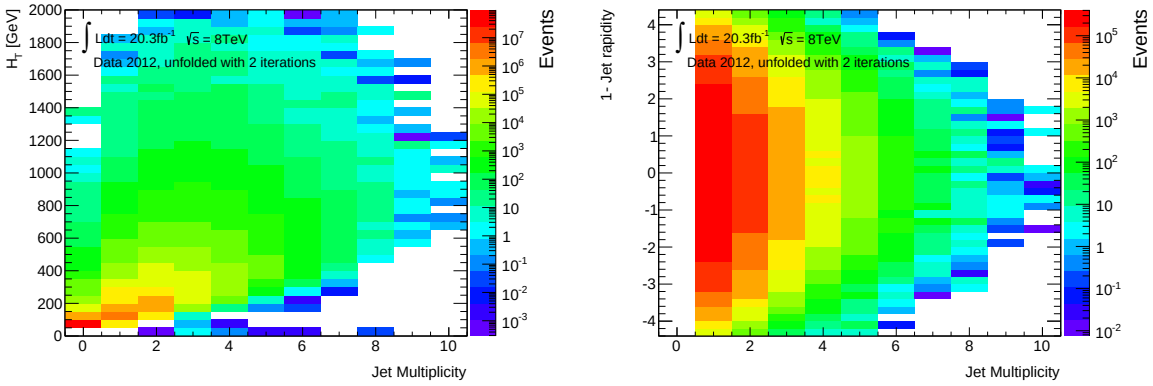
After unfolding of the modified reconstructed distribution, the pull per bin i is computed as:

$$p_i = \frac{\hat{N}(x_{T,i}) - N(x_{T,i})}{\Delta\hat{N}(x_{T,i})}$$

with $\hat{N}(x_{T,i})$ being the number and $\Delta\hat{N}(x_{T,i})$ the uncertainty estimate of the unfolded events and $N(x_{T,i})$ the number of true events in bin i. This procedure is repeated 100 times and a distribution of pulls is obtained per bin. If the uncertainties in the unfolded results correctly reflect statistical fluctuations in the reconstructed input, then the pulls should be Gaussian distributed with mean of zero and σ of 1.

Figure 6.15 presents the mean and spread of the repeatedly unfolded distributions for n_{jets}, H_{T}, W p_{T} and leading jet rapidity which almost exactly match the true distributions. The pulls per bin are displayed in the lower panel. Throughout, the pulls are centered at zero with a spread of 1 as indicated by the error bars on the pulls. Exceptions are only in low statistics bins at high H_{T} and W p_{T} as well as the 0 jet and to a minor extend the 1 jet bin in the n_{jets} distribution. While the two n_{jets} bins are still centered at zero, they exhibit smaller spreads of about 0.3 for $n_{\text{jets}} = 0$ and about 0.6 for $n_{\text{jets}} = 1$, which are relatively stable when increasing the number of pull tests to 200 or 500 (see discussion below). This indicates a slight overestimation of the statistical uncertainty in these bins, but since the general trend shows the expected valid behaviour, no correction to artificially modify the uncertainty in these bins is applied.

For two individual bins, the Gaussian shape of the pull distribution is validated by fitting with a Gaussian function, and the stability with respect to the number of pull tests, as noted briefly above, is verified by repeating the test 200 or 500 times. Figure 6.16 shows this for the 1-jet-bin of the n_{jets} distribution and the central bin of the leading jet rapidity at 1 jet. In both cases, the pull distribution is found to be described by the Gaussian fit. The fitted parameters for the mean and σ are similar to the corresponding histogram values and are stable with different numbers of tests.



(a) H_{T} (b) Leading jet rapidity

Figure 6.17: Result of the 2-dimensional unfolding of the measured data for the H_{T} (a) and the leading jet rapidity (b) distributions. For the inputs of the Bayesian unfolding, ALPGEN W + jets simulation is used.

The estimates of the statistical uncertainties, as determined from the unfolding covariance matrix are thus found to properly reflect potential statistical fluctuations in the input data and are therefore used as statistical uncertainties in the unfolded results and differential cross sections.

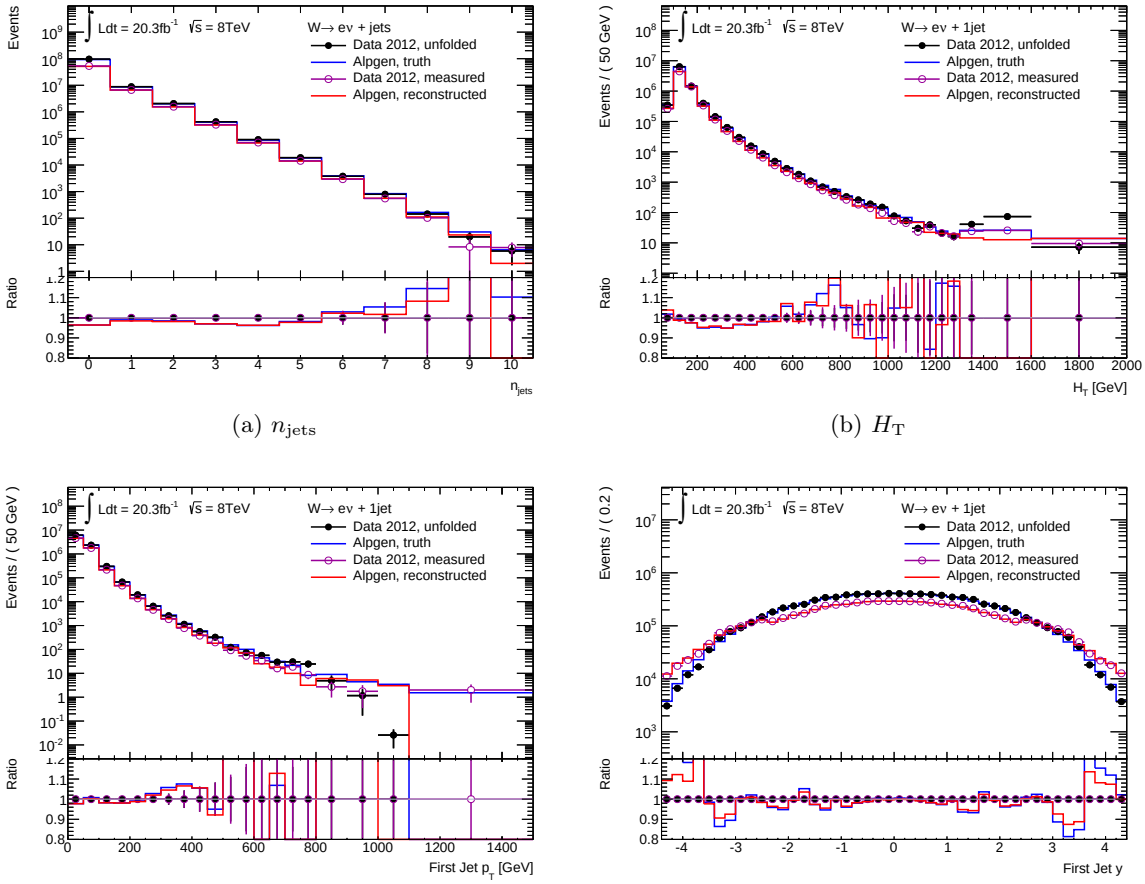
6.5 Unfolding of 2012 data

Seven distributions measured in data in 2012 are unfolded to particle level, using the iterative Bayesian unfolding with inputs from ALPGEN W + jets simulation. The estimated backgrounds are subtracted from data before unfolding. Negative values for data after background subtraction are truncated to zero.

The following distributions are unfolded:

- n_{jets}
- Leading jet p_{T} and rapidity
- Second leading jet p_{T} and rapidity
- W boson p_{T}
- H_{T}

Unfolded 2-dimensional data distributions are presented in Figure 6.17 for the H_{T} and the leading jet rapidity distributions, exemplary. Unfolded data for the other distributions with 2-dimensional unfolding are shown in Appendix A.6. The largest unfolded event counts are located at low energies and central rapidities and reduce towards higher jet multiplicities and larger energies. Leading and second leading jet distributions only contain entries for more than 1 or 2 jets, respectively, since these distributions obviously require the presence of at least 1 or 2 jets. In general, the unfolded distributions look reasonable.



(a) n_{jets} (b) H_{T}

(c) Leading jet p_{T} (d) Leading jet rapidity

Figure 6.18: Comparison of unfolded data with the ALPGEN truth prediction as well as measured data at the detector level with reconstructed ALPGEN for the n_{jets} (a), H_{T} (b) and leading jet p_{T} and rapidity (c) and (d) distributions. The latter three are for $n_{\text{jets}} = 1$. The lower panel presents the ratio of ALPGEN to the respective data, either measured or unfolded.

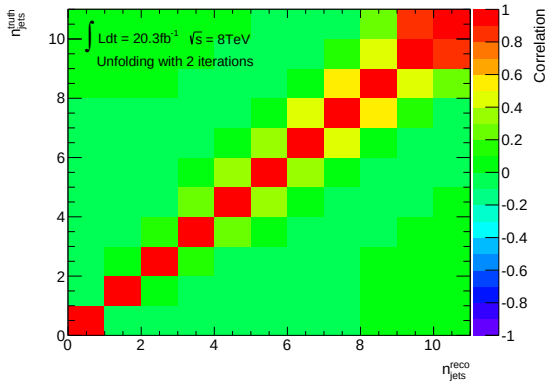


Figure 6.19: Correlation matrix of the unfolding of the n_{jets} distribution in data.

6.5.1 Consistency test and covariance matrix

Some unfolding techniques bias the results towards the MC simulation which is used as input in the unfolding procedure. To confirm that this does not occur in the iterative Bayesian method employed here, differences between W + jets MC predictions and data are investigated before and after the unfolding.

Figure 6.18 presents a comparison of unfolded data with the truth prediction in W + jets Alpgen simulation and of the measured data at the detector level with the reconstructed Alpgen prediction. Shown are the n_{jets}, H_{T}, leading jet p_{T} and rapidity distributions. In all four cases, differences which exist between reconstructed Alpgen and measured data are still visible between the unfolded data and Alpgen truth. The differences do not exhibit major changes in the magnitude caused by the unfolding, but only minor statistical fluctuations. The unfolding procedure adopted in this analysis therefore does not systematically distort the information contained in the measured data distribution.

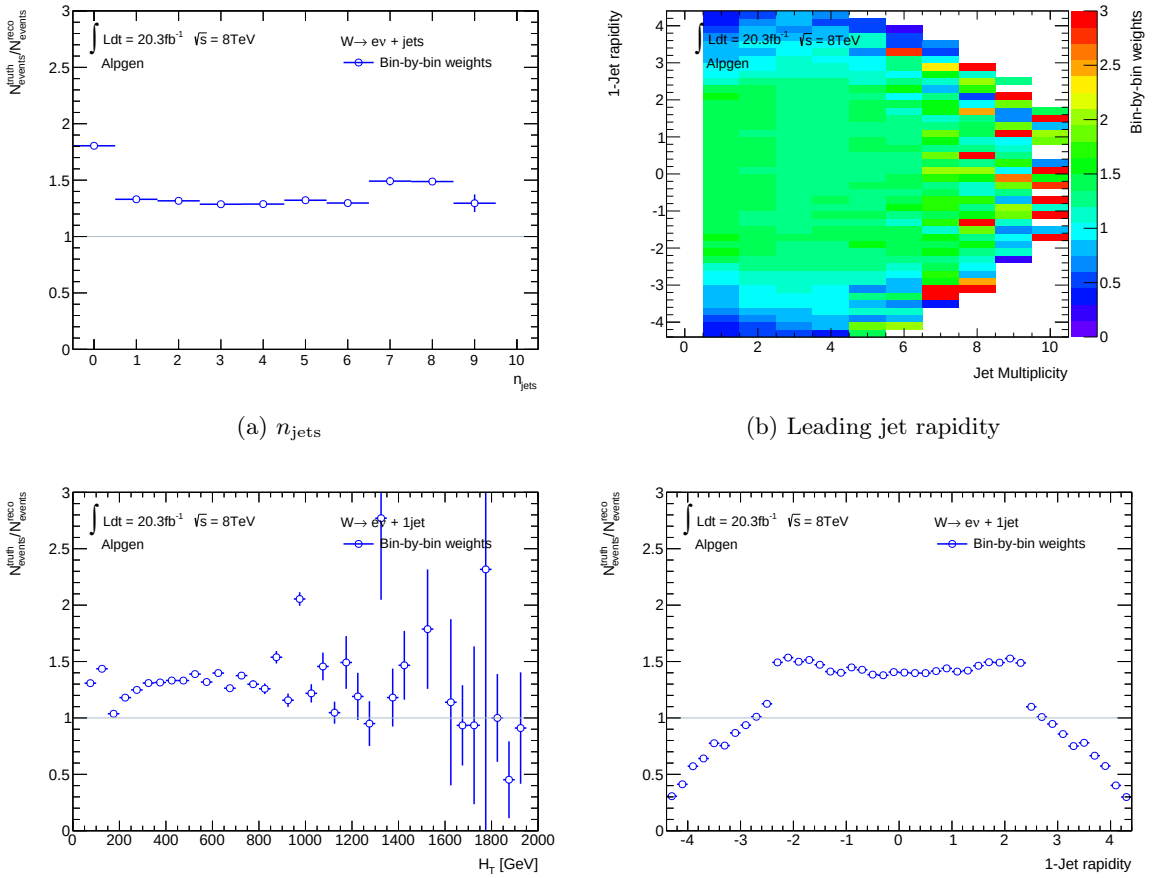
Bin-by-bin migrations which are treated properly in the Bayesian unfolding can introduce correlations between bins in the unfolded distribution. The unfolding covariance matrix provides this information and is converted into the correlation matrix for ease of judgement. Figure 6.19 displays the correlation matrix for the 1-dimensional unfolding of the n_{jets} distribution. For the low jets multiplicities, correlations in the n_{jets} unfolding are found to be very small. Significant correlations between neighbouring jet multiplicities are observed only for larger jet multiplicities. This demonstrates a good performance of the used unfolding method.

6.5.2 Bayesian vs. Bin-by-bin unfolding

In contrast to the rather involved iterative procedure, the most simple approach to unfolding is a bin-by-bin correction. Here, the number of measured events per bin is multiplied by a weight determined as the ratio between true and reconstructed distribution in the signal MC simulation. This approach relies on the truth-level shapes predicted by the MC simulation and is prone to a modelling bias. Furthermore, it does not consider migrations between reconstruction bins. Its simplicity therefore comes with major drawbacks and for these reasons, it is not used as nominal unfolding procedure in this analysis. It is however useful as cross check of the iterative Bayesian unfolding, in order to test basic agreement between two different unfolding procedures.

The weights for the bin-by-bin unfolding are determined using the Alpgen W + jets simulation and are shown in Figure 6.20 for the n_{jets}, H_{T} and leading jet rapidity distributions. For 2-dimensional distributions, the weighting is done in both dimensions, i.e. n_{jets} and observable, and weights for individual jet multiplicities are extracted for illustration purposes only. The weights implicitly contain the reconstruction efficiency and fake fraction corrections. The reconstruction efficiency of 50 % leads to the offset of the weights from one, while e.g. the large fake fraction at forward rapidity of the leading jet modifies the weights in these bins to below one. Since the true and reconstructed distributions used to calculate the weights partially include the same MC events, the uncertainties in the weights are computed using binomial errors.

The bin-by-bin unfolded data distributions are compared to the results from the iterative Bayesian unfolding in Figure 6.21 for the n_{jets}, H_{T}, leading jet p_{T} and leading jet rapidity distributions. The comparison for the other unfolded distributions is given in Appendix A.7. Within uncertainties, the results from both methods are found to agree very well. Small differences are observed in low statistics regions or e.g. a large jet rapidities where the increased fake fraction leads to bin-migrations and the bin-by-bin unfolding exhibits problems. The



(a) n_{jets}

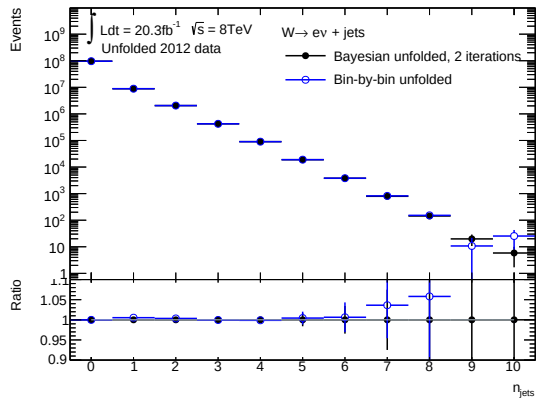
(b) Leading jet rapidity

(c) H_{T} at 1 jet

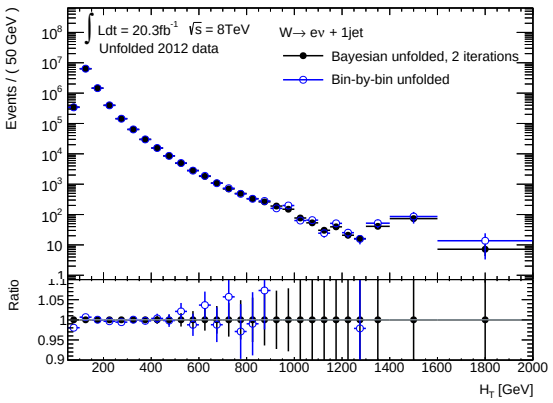
(d) Leading jet rapidity at 1 jet

Figure 6.20: Bin-by-bin weights for the n_{jets} (a) and leading jet rapidity (b) distributions as well as the projections of these weights for 1 jet for H_{T} (c) and the leading jet rapidity again (d). The weights are obtained from ALPGEN W + jets simulation as the ratio of the true over the reconstructed distributions.

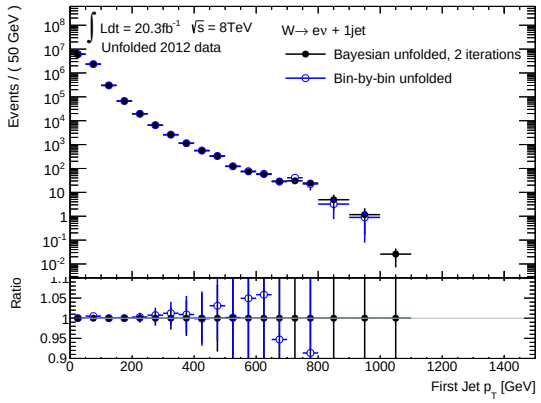
iterative Bayesian unfolding also appears slightly more stable towards low statistics regions, visible as smaller fluctuations in the Bayesian unfolded distribution compared to the bin-by-bin unfolded distribution at large H_{T} in Figure 6.21b or high W or second leading jet p_{T} in Appendix A.7. This further justifes the choice of the iterative Bayesian unfolding procedure as nominal unfolding method.



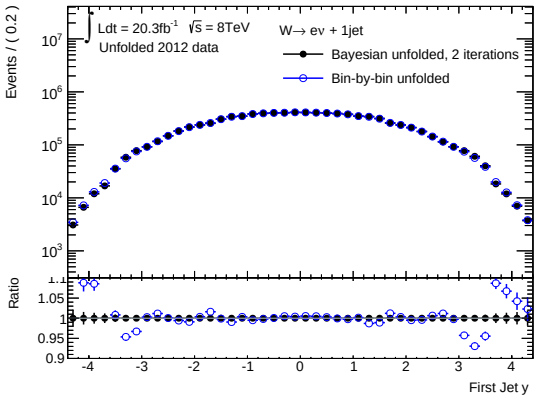
(a) n_{jets}



(b) H_{T}



(c) Leading jet p_{T}



(d) Leading jet rapidity

Figure 6.21: Comparison of the unfolded results obtained with iterative Bayesian unfolding or bin-by-bin unfolding for the n_{jets} (a), H_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d) distributions – the latter three for 1 jet. In both cases, the measured data distribution is unfolded using input from ALPGEN W + jets simulation.

CHAPTER 7

Estimate of uncertainties in differential cross sections

The derivation of the uncertainties is an important component of the cross section measurement of W + jets production. The size of the uncertainties determines to a large extend the competitiveness and the impact of the W+jets measurement on PDF fits. One of the main arguments for the measurement of the W^+/W^- ratio is the experimental fact that uncertainties from the same source which are present in a similar way for W^+ and W^- production cancel in the ratio. This concerns in particular uncertainties related to the jet energy which often constitute the or one of the dominant uncertainties, especially for measurements involving jets.

Uncertainties in this analysis are distinguished according to their origin into three major categories: systematic uncertainties at the detector level, systematic uncertainties in the unfolding method and statistical uncertainties in the unfolded cross sections.

The first class arises mainly from the calibration of energy scales or detector efficiencies for jets, electrons and $E_{\mathrm{T}}^{\mathrm{miss}}$ or from uncertainties related to the background estimation. These uncertainties are computed at the detector level as described in Section 7.1 and are propagated through the unfolding procedure as explained in Section 7.2. The second class of systematic uncertainties considers the impact of the chosen unfolding method on the unfolded results and is described in Section 7.3. The third category refers to the uncertainties in the unfolded results which arise from the statistical uncertainties in the data at the detector level. These have already been discussed in the previous chapter and are displayed in comparison to the total systematic uncertainties in Section 7.4 and Section 7.5 for the measured cross sections and the W^+/W^- ratio, respectively.

7.1 Systematic uncertainties at the detector level

The electron, $E_{\mathrm{T}}^{\mathrm{miss}}$ or jet energy calibration procedures aim at an energy determination with the best possible precision and have been described in Chapter 4. For electron and jet energy calibrations, an energy scale uncertainty of approximately 0.1 % and 1 %, respectively, were targeted [123]. The impact of these uncertainties on the W + jets measurement are evaluated in the following.

The systematic uncertainties at the detector level result from uncertainties in the reconstructed energy of the selected electron, the soft term contributions in the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation and the selected jets. Also uncertainties in different detector efficiencies (which are corrected in the MC simulation) contribute to the uncertainty in the W + jets measurement. Furthermore, the pile-up simulation in MC is tuned to match the average activity per event in data,

and the impact of the associated uncertainty on the W+jets measurement is determined. Last but not least, the background estimates are accompanied with uncertainties which influence the detector-level distributions presented in the last chapter and the unfolded W + jets cross section shown in the next chapter.

The impact of the different uncertainty sources on the W + jets measurement is evaluated by varying the corresponding quantity, e.g. the energy scale of the jets in the MC simulation, by one standard deviation and reapplying the W + jets selections. The variation is conducted separately for $+1\sigma$ and -1σ shifts and the difference between the nominal values and the variation is taken as the systematic uncertainty. The fit of the multi-jet background estimate (see Section 5.1) is repeated for each systematic variation and the effect from the shifted multi-jet fraction is taken into account. The results from the up- and down variations are averaged (absolute values), to slightly reduce the impact on the uncertainty from limited MC statistics, yielding a symmetric uncertainty. Several components of a particular uncertainty (all uncertainties associated to a particle type) are added in quadrature for the combined (particle) uncertainty.

The considered sources of uncertainty related to the electron, $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets are described in detail in Section 7.1.1, Section 7.1.2 and Section 7.1.3, respectively. Variations of electron or jet energies are also considered in the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation and the effect of the modified $E_{\mathrm{T}}^{\mathrm{miss}}$ is included in the respective electron and jet uncertainties. The uncertainties related to pile-up and background are discussed in Section 7.1.4 and Section 7.1.5, respectively.

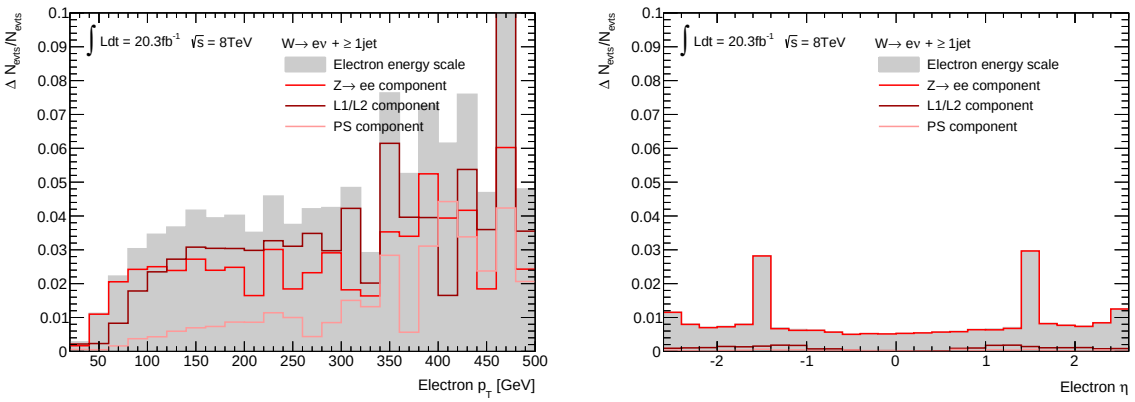
7.1.1 Electron uncertainties

The electron energy is calibrated as summarized in Section 4.1.1. Uncertainties in the calibration procedure arise mainly from the layer intercalibration and the energy scale measurement in $Z \to ee$ events. The measurement of the electron energies in $Z \to ee$ events defines the final energy scale and resolution within uncertainties of 0.03 % and 0.3 %, respectively, for central electrons. This measurement presupposes the relative calibration of the first and second electromagnetic calorimeter layers (L1/L2) and the calibration of the pre-sampler (PS) energy scale. The impact of the uncertainties in these measurements on the electron energy depend on the fractional contribution of energy deposited in these layers with respect to the full electron cluster energy. For central electrons, this translates into uncertainties in the energy scale of approximately 0.02 % at $p_{\mathrm{T}} \approx 25\,\mathrm{GeV}$, decreasing to a minimum at $p_{\mathrm{T}} \approx 40\,\mathrm{GeV}$ and increasing again to approximately 0.05 % at large electron p_{T}.

To estimate the impact of the electron energy scale (EES) uncertainties from the three different measurement sources, the electron energies are varied according to the electron cluster energy and the position in η. The impact of the EES uncertainty components on the total SM prediction at the detector level for the electron p_{T} and η distributions is shown in Figure 7.1. The EES component from the $Z \to ee$ measurement is the largest contribution at low electron p_{T} and roughly constant in η with peaks only at the barrel-endcap overlap of the electromagnetic calorimeter. For higher electron p_{T} ($\gtrsim 100\,\mathrm{GeV}$), the uncertainty from the L1/L2 intercalibration becomes roughly the same order of magnitude, while the uncertainty from the PS calibration remains sub-dominant.

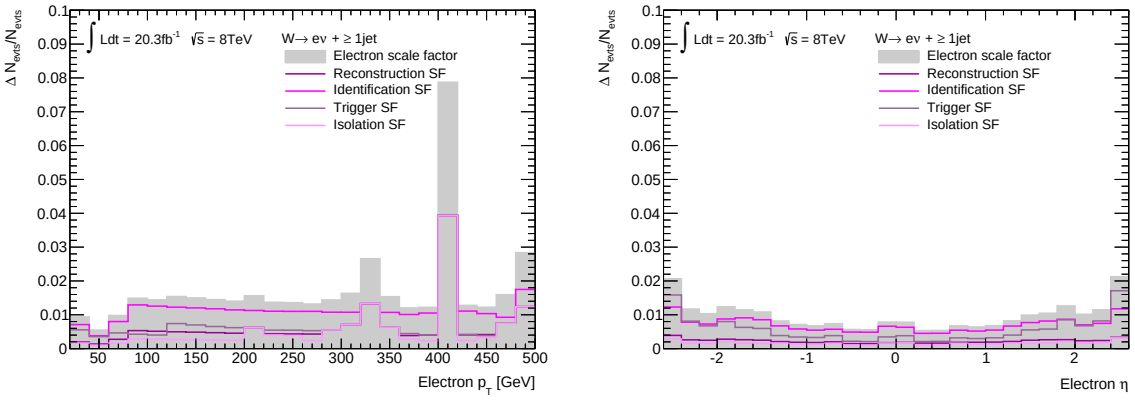
The electron energy is smeared in the MC simulation to match the resolution in data as mentioned in Section 4.1.1. The impact on the W + jets measurement is investigated by adjusting the smearing taking the uncertainty in the electron energy resolution (EER) measurement into account. The result for the total SM prediction is shown below together with the other electron uncertainties.

The electron triggering, reconstruction, identification and isolation efficiencies are corrected



(a) Electron p_{T} (b) Electron η

Figure 7.1: Relative electron energy scale systematic uncertainties in the total SM prediction at the detector level. Shown are the uncertainties in the electron p_{T} (a) and η (b) distributions for ≥ 1 jet.

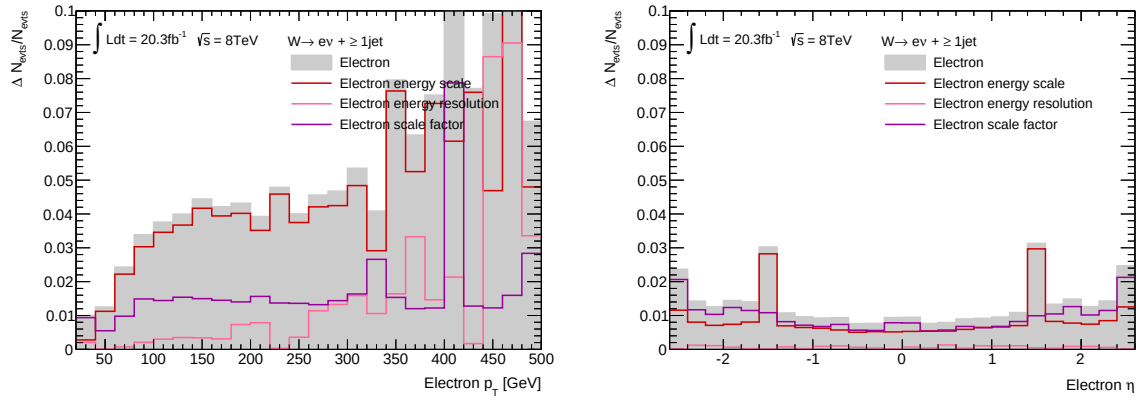


(a) Electron p_{T} (b) Electron η

Figure 7.2: Relative electron scale factor systematic uncertainties in the total SM prediction at the detector level. Shown are the uncertainties in the electron p_{T} (a) and η (b) distributions for ≥ 1 jet.

in the MC simulation to the values measured in data, using the scale factors (SF) discussed in Section 4.3. These are measured with $< 0.5\,\%$ and $1 - 2\,\%$ uncertainty in the reconstruction and identification efficiencies, respectively, for electrons with $E_{\mathrm{T}} > 25\,\mathrm{GeV}$ [78], for example. In the $W +$ jets analysis, the effect of these uncertainties is included by shifting the SFs by $\pm 1\sigma$ separately for each type of SF and re-measuring the $W +$ jets distributions. Figure 7.2 displays the relative uncertainty from electron SFs in the total SM prediction at the detector level for the electron p_{T} and η distributions. The largest observed impact arises from the electron identification SF, followed by the trigger SF, which is dominant at the highest η values. The impact from electron reconstruction and isolation SFs is small. This is in line with the very small corrections of $\lesssim 2\,\%$ to the electron reconstruction and isolation efficiencies in the simulation, while the identification and trigger efficiencies are corrected within $10\,\%$, reaching about $20\,\%$ in some cases.

The overview of the electron related uncertainties in the total SM prediction at the detector level is shown in Figure 7.3 for the electron p_{T} and η distributions. The total electron SF uncertainty is the largest source of uncertainty for electron p_{T} smaller than $40\,\mathrm{GeV}$, but for



(a) Electron p_{T} (b) Electron η

Figure 7.3: Relative electron-related systematic uncertainties in the total SM prediction at the detector level. Shown are the uncertainties in the electron p_{T} (a) and η (b) distributions for ≥ 1 jet.

larger electron p_{T} the EES uncertainty is dominating, reaching about 4.5 % at $p_{\mathrm{T}} \approx 300\,\mathrm{GeV}$.

7.1.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is split into the hard and soft terms as discussed in Section 4.1.2. The hard term defines an axis $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$ in the transverse plane and is expected to be balanced by the soft term in the absence of genuine $E_{\mathrm{T}}^{\mathrm{miss}}$. This balance method is used by the ATLAS $E_{\mathrm{T}}^{\mathrm{miss}}$ performance group to estimate the level of disagreement in the soft term calibration between data and MC simulation for $Z \to \mu\mu$ events in the absence of jets. Data-MC differences are then considered as systematic uncertainty. The uncertainties in the soft term are split into uncertainties in the soft term energy scale and the soft term energy resolution parallel and transverse to $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$. The largest variation of the soft term modelling in MC is found to result from differences between matrix element generators matched to parton shower models. For the soft term energy scale, the uncertainty is determined to be around 3 % of $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$ growing to approximately 6 % for the lowest $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$, while both the parallel and transverse soft term resolution uncertainties are around $5 - 8\,\%$ of $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$, increasing to $\gtrsim 30\,\%$ for the lowest $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$-values [80].

The impact of the soft term uncertainties on the W + jets measurement is investiated by varying the soft term energy scale once up and down by the uncertainty Δs:

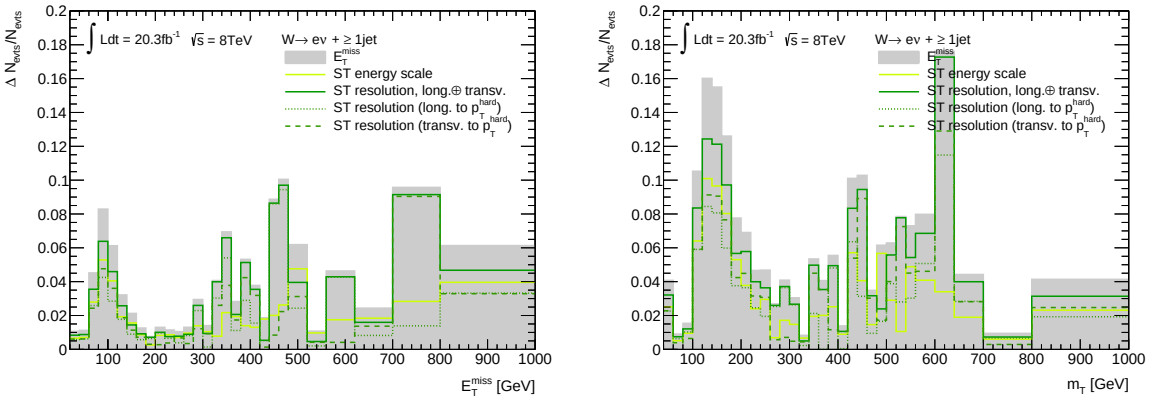
$$E_{\parallel,\mathrm{scale}\pm}^{\mathrm{miss,SoftTerm}} = E_{\parallel}^{\mathrm{miss,SoftTerm}} \pm \Delta s\,,$$

meaning only the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ component parallel to $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$ is modified. Similarly, the soft term resolution is varied by smearing the $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$-parallel or -transverse component separately within a Gaussian distribution with the resolution uncertainties $\sigma_{\parallel(\perp)}$ as widths:

$$E_{\parallel(\perp),\mathrm{reso}}^{\mathrm{miss,SoftTerm}} = E_{\parallel(\perp)}^{\mathrm{miss,SoftTerm}} + \mathrm{Gaus}(\Delta s, \sigma_{\parallel(\perp)})\,.$$

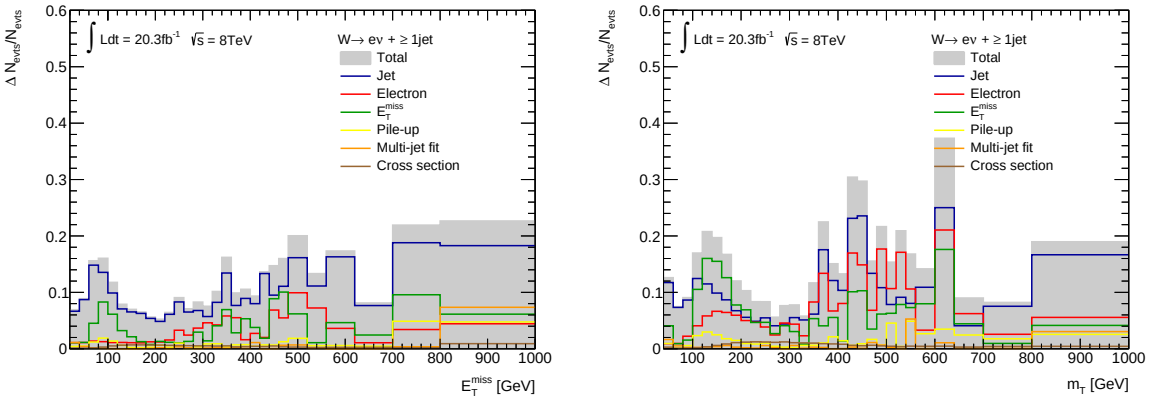
The uncertainty from both resolution variations is added in quadrature for the combined $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$-resolution uncertainty.

Figure 7.4 shows the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ fractional uncertainties in the total SM prediction at the detector level for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} distributions for ≥ 1 jet. While the contributions from the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainties are small in general, for these two distributions an increase



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

Figure 7.4: Relative $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss}}$ (a) and m_{T} (b) distributions for ≥ 1 jet. The soft term (ST) resolution uncertainty is the sum in quadrature from the smearing of the resolution longitudinal and transverse to $\vec{p}_{\mathrm{T}}^{\,\mathrm{hard}}$ (dashed lines).



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

Figure 7.5: Relative systematic uncertainty in the total SM prediction at the detector level from the considered uncertainties at the detector level. Shown are the uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss}}$ (a) and m_{T} (b) distributions for ≥ 1 jet.

of the combined $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainties up to approximately 8.5 % and 16 % is observed in the ranges $60 - 120\,\mathrm{GeV}$ for $E_{\mathrm{T}}^{\mathrm{miss}}$ and $100 - 200\,\mathrm{GeV}$ for m_{T}. For the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, the combined jet-related uncertainties (as described below) still yield the largest contribution in the total detector-level systematic uncertainty[1]. For the m_{T} distribution however the uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ energy scale and resolution are the dominant source of uncertainty in the mentioned m_{T} range. This can be seen in Figure 7.5, which displays the combined electron, $E_{\mathrm{T}}^{\mathrm{miss}}$ (soft term), jet, pile-up, multi-jet fit and cross section uncertainties at the detector level for the same $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} distributions. The mentioned $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} ranges correspond to the regions where both distributions at the cross section level heavily drop after their maximum from $W \to e\nu$ events and which are thus particularly prone to mismodelling effects from soft energy deposits or tracks contributing to the $E_{\mathrm{T}}^{\mathrm{miss}}$.

[1] In the absence of jets, the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainties are the dominant uncertainty contribution also for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the range of $60 - 120\,\mathrm{GeV}$.

7.1.3 Jet uncertainties

The energy calibration of jets is accompanied by a number of uncertainty sources from the applied corrections and calibrations (see Section 4.1.3). Uncertainties, both in the energy scale (JES) and the energy resolution (JER) of jets, are considered in the W + jets analysis. Furthermore, the jet-based selections of the b-jet veto and the pile-up jet suppression by the JVF requirement as described in Section 4.3.1 induce uncertainties which are considered in the W + jets analysis. Uncertainties in the b-jet veto are introduced by the jet b-tagging efficiency corrections applied in the MC simulation as mentioned in Section 4.3 and are in the following denoted as jet scale factor (SF) uncertainties.

Jet energy scale uncertainties

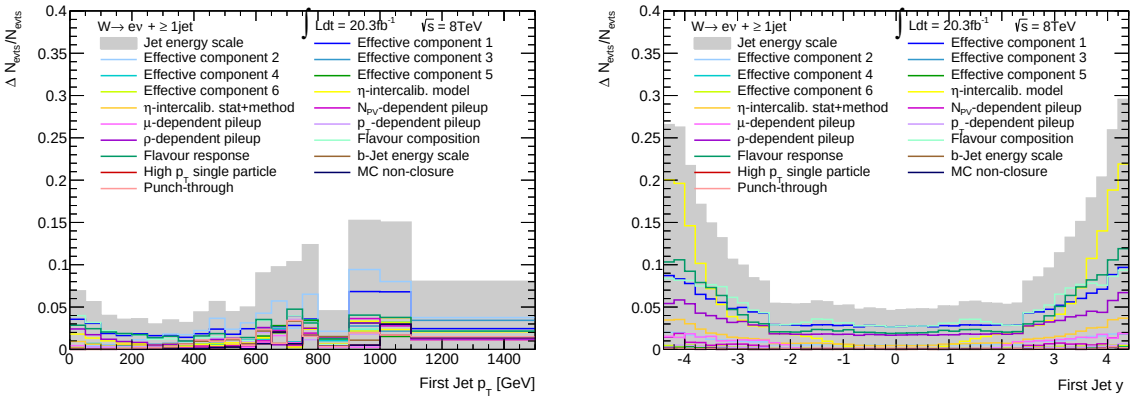
An overview of the JES calibration and its uncertainties in 2012 data is shown in Figure 4.2 in Section 4.1.3. The individual components which are considered in the W + jets analysis for the overall JES uncertainty are:

- 6 effective JES components from the in-situ determination of the jet energy scale in Z+jet, γ+jet and multi-jet balance measurements
- 2 η-intercalibration components – one contains the statistical and methodological uncertainties, while the other describes the modelling uncertainties
- 4 pile-up correction components, resulting from the $\langle\mu\rangle$, N_{PV}, p_{T} dependent pile-up corrections as well as the energy density ρ mismodelling in the pile-up correction
- 2 flavour components – one for the flavour composition assuming an unknown composition of quark and gluon-initiated jets (50:50 quark:gluon with 100 % uncertainty) and one for the flavour response considering the different detector responses for quark and gluon-initiated jets (only applied to non-b-jets)
- 1 b-jet energy scale component (only applied to b-jets)
- 1 single-hadron response component in the high p_{T} range used as extension of the in-situ measurements
- 1 MC closure component considering the calibration differences between MC generation versions for 8 TeV, LHC running conditions in 2012
- 1 punch-through component resulting from the punch-through correction in the GSC-step in the jet calibration

In total, these are 18 uncertainty sources[2]. For each component, the jet energies are varied by 1σ up and down, separately, yielding the systematic uncertainty in the W + jets distributions as the difference to the nominal. The combined JES uncertainty is the sum in quadrature from the individual components.

Figure 7.6 presents the fractional uncertainties from the individual JES components in the total SM prediction at the detector level for the leading jet p_{T} and rapidity distributions. The uncertainties demonstrate the expected behaviour, i.e. a decrease of the JES uncertainty from the lowest p_{T} to medium p_{T}, and an increase of the JES uncertainty from central to forward

[2]These correspond to the 15 uncertainty sources discussed mainly in Ref. [81] plus 3 additional ones discussed in Refs. [82–84, 90].



(a) Leading jet p_{T} (b) Leading jet y

Figure 7.6: Relative JES systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the leading jet p_{T} (a) and rapidity (b) distributions for ≥ 1 jet.

jets. Differences in the values of the JES uncertainties with respect to the offical values shown in Figure 4.2 come from the fact that the latter are at the jet level, while the ones presented here are at the cross section level. As a result of the steeply falling cross section of $W+$jets events with increasing jet p_{T}, a relatively small uncertainty of around 1 % in the jet energy as for $p_{\mathrm{T}} \gtrsim 100\,\mathrm{GeV}$ can translate into an uncertainty of approximately 4 % in the total SM prediction at the detector level.

Jet energy resolution uncertainty

The jet energy resolution agrees in 2012 data and MC simulation within the measurement uncertainties. Therefore, no adjustments to the jet energies in terms of the resolution are made for the nominal MC predictions. Uncertainties in this JER measurement are however considered as a systematic uncertainty.

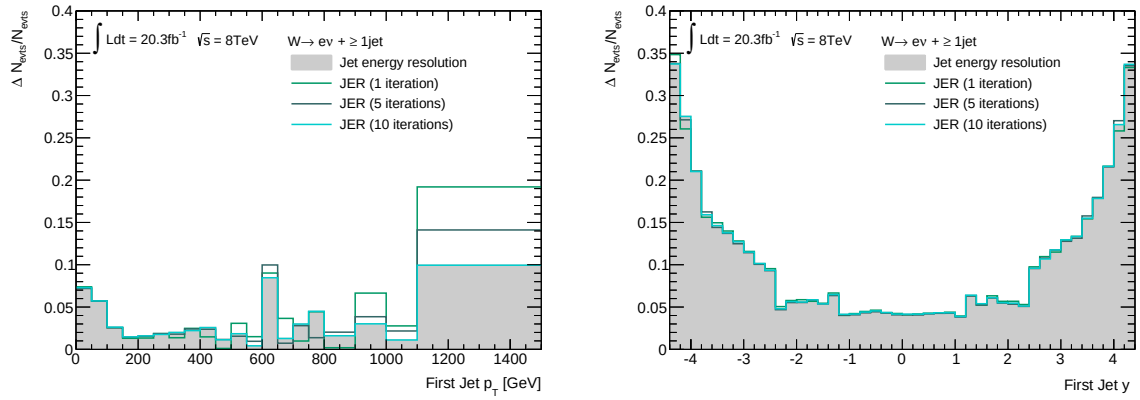
To estimate the impact of the JER measurement uncertainties in the $W+$jets measurement, the p_{T} of simulated jets is modified by random factors from a Gaussian distribution with mean of 1 and $\sigma = \sqrt{(\mathrm{JER}+\Delta\mathrm{JER})^2 - (\mathrm{JER})^2}$. The smearing procedure is repeated for each jet in the MC event 1, 5 or 10 times, in order to investigate the impact of statistical fluctuations on the variations.

Figure 7.7 displays the fractional uncertainty from the JER variations in the total SM prediction at the detector level for the leading jet p_{T} and rapidity. The main impact of the JER uncertainties is observed at low jet p_{T}, i.e. below about 100 GeV – as expected from Figure 4.3 – and at large jet rapidities. Very little difference is found between the different numbers of jet energy smearing, and the '10-smearings' uncertainty is used for the final JER uncertainty in the $W+$jets measurement.

Jet scale factor uncertainties

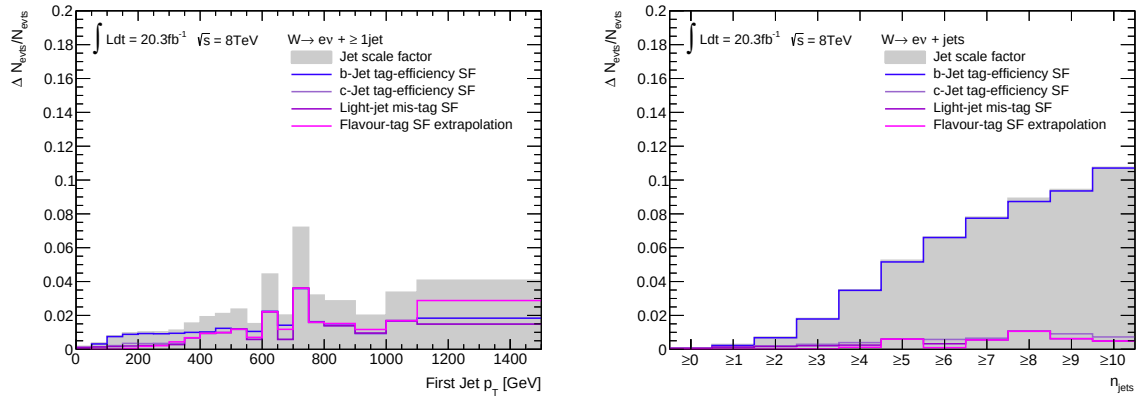
The measured b-jet tagging, c-jet tagging and light-jet mistagging efficiency corrections are applied to the simulation in form of jet-tagging SFs as discussed in Section 4.3. In the $W+$jets analysis, they are used in the $W+$jets signal region and the $t\bar{t}$ validation region.

Their impact is evaluated by modifying the b-tagging, c-tagging and light-jet-mistagging SFs independently up and down according to their uncertainties. Since the measurement of



(a) Leading jet p_{T} (b) Leading jet y

Figure 7.7: Relative JER systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the leading jet p_{T} (a) and rapidity (b) distributions for ≥ 1 jet, obtained with 1, 5 or 10 smearing iterations per jet. The final JER uncertainty corresponds to the version with 10 iterations.



(a) Leading jet p_{T} (b) n_{jets}

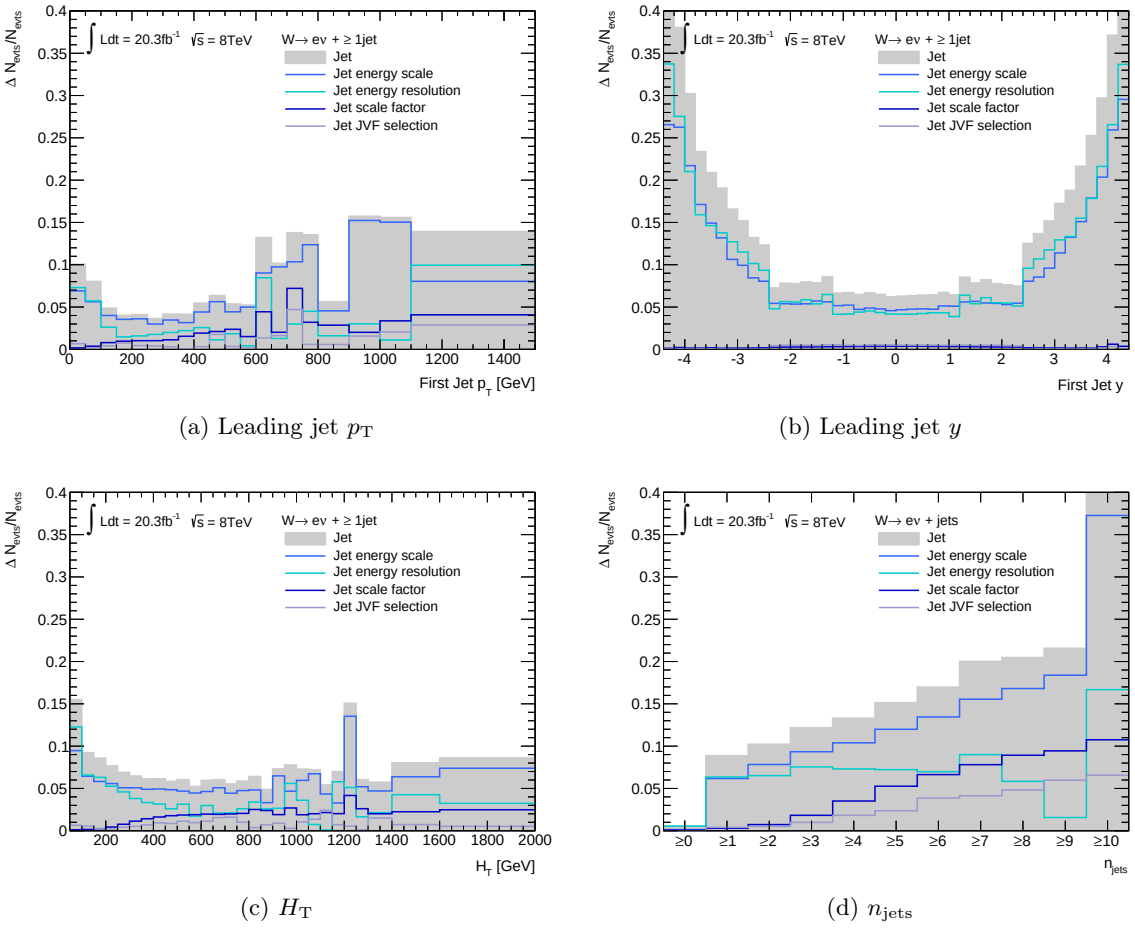
Figure 7.8: Relative jet SF systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the leading jet p_{T} distribution for ≥ 1 jet (a) and n_{jets} (b).

the tagging efficiencies is possible for b- and c-jets only up to 300 GeV and for light jets only up to 750 GeV, the SFs are extrapolated for jets with higher p_{T}. This p_{T}-extrapolation is associated with an additional uncertainty, considered for all jet flavours simultaneously.

The effect of the SF variations on the total SM prediction in the W + jets measurement at the detector level is shown in Figure 7.8 for the leading jet p_{T} distribution and the n_{jets} distribution. While the uncertainty from the jet SFs is mostly below 2 % even for the higher jet p_{T}, it increases with the number of jets up to around 10 %. The main contribution is the b-jet tagging uncertainty.

Jet pile-up suppression uncertainty

The JVF requirement, suppressing pile-up jets in the W + jets analysis (discussed in Section 4.1.3 and Section 4.3.1), is modified to investigate its impact on the W + jets distributions. The nominal jet vertex fraction of 50 % is varied for this purpose by ±3 % (absolute



(a) Leading jet p_{T}

(b) Leading jet y

(c) H_{T}

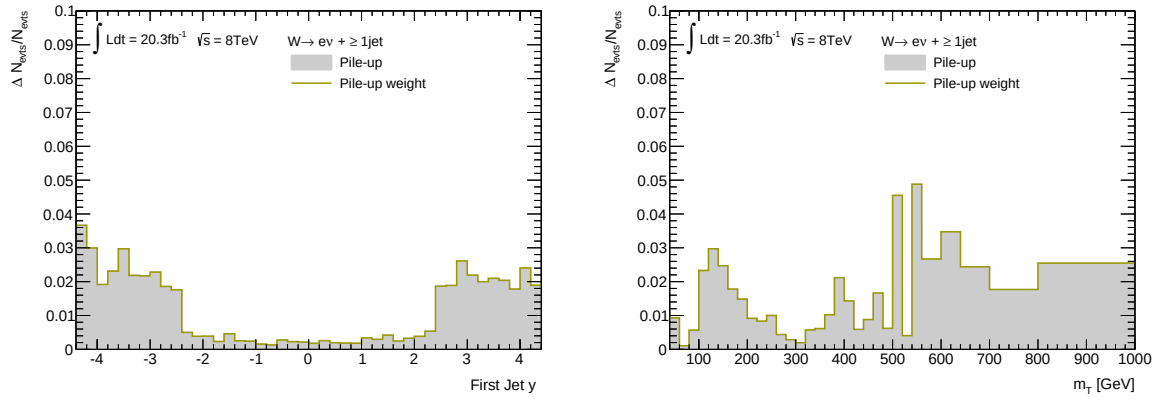
(d) n_{jets}

Figure 7.9: Relative jet-related systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the leading jet p_{T} (a), leading jet rapidity (b), H_{T} (c) distributions for ≥ 1 jet and n_{jets} (d).

value) [103]. The difference in the W + jets distributions due to the JVF variation is considered as systematic uncertainty and is found to be small compared to the other jet-related uncertainties (see next section).

Combined jet uncertainties

The combined jet uncertainties are determined by adding in quadrature the JES, JER, jet SF and JVF uncertainties. Figure 7.9 displays the combined jet uncertainties for the leading jet p_{T} and rapidity as well as the jet multiplicity and H_{T} distribution. The largest uncertainty contributions are found to result from the JES and JER uncertainty. In the presence of 1 jet, and at low jet p_{T} both uncertainties are roughly of the same order of magnitude, while for larger jet p_{T} and jet multiplicities the JES uncertainty dominates the combined jet uncertainty. The H_{T} distribution follows the trends from the leading jet p_{T}, but with a slightly larger influence of the JER at the lowest H_{T} value. For higher jet rapidity, both JES and JER uncertainties increase, but almost by the same amount. The jet SF and JVF uncertainties are sub-dominant everywhere, and only for larger jet p_{T}, H_{T} or n_{jets} the jet SF uncertainty becomes roughly of the same size as the there-subdominant JER uncertainty.



(a) Leading jet y (b) m_{T}

Figure 7.10: Relative pile-up systematic uncertainty in the total SM prediction at the detector level. Shown are the uncertainties in the leading jet rapidity (a) and m_{T} (b) distribution for ≥ 1 jet.

7.1.4 Pile-up uncertainty

The average activity in the detector per event follows from the number of pile-up interactions occurring simultaneous with the hard interaction. It is matched in the simulation to the one in data by adjusting $\langle\mu\rangle$. The impact of the $\langle\mu\rangle$-scaling on the W + jets measurement is estimated by modifying the nominal scaling factor of 1.09 by ± 0.07.

In general, the influence of the pile-up reweighting uncertainty is found to be small. Two examples where the pile-up uncertainty contribution to the total SM prediction at the detector level is non-negligible are shown in Figure 7.10: Leading jet rapidity and m_{T}. The influence of pile-up at large jet rapidities fits together with the increase in the fake fraction at large jet rapidities seen in the unfolding in Section 6.2.2. For the m_{T} distribution, the increase of the pile-up uncertainty in the range $100 - 200\,\mathrm{GeV}$ is in line with the observations for the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainties in this range (see Section 7.1.2), since this term is designed to capture the pile-up contributions in the $E_{\mathrm{T}}^{\mathrm{miss}}$-computation.

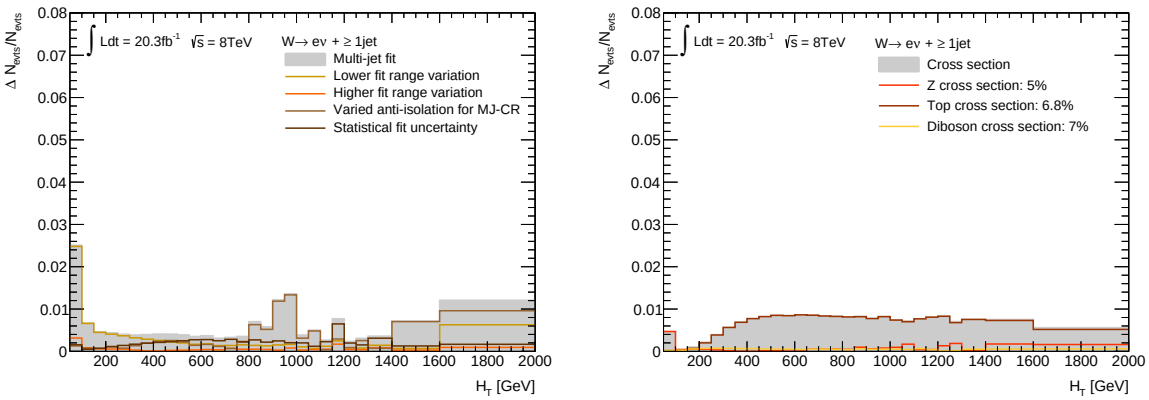
7.1.5 Background uncertainties

The systematic uncertainties in the background estimates arise from the data-driven methodology for the multi-jet background and from the cross section uncertainties in the MC predictions.

Systematic uncertainties in the multi-jet background estimates have been described in Section 5.1.2. Their influence on the W + jets predictions is estimated by adjusting the determined multi-jet fraction according to the systematic variation and re-evaluating the W + jets distributions. The impact on the total SM prediction depends on the relative contribution of the multi-jet background to the total prediction.

The impact of the cross section uncertainties of the other backgrounds is estimated by increasing or decreasing the corresponding background prediction by $\pm 1\sigma$ of the cross section uncertainty (including a repetition of the multi-jet background fit). The cross section uncertainties are the total theory uncertainties, i.e. the sum in quadrature of scale and PDF+α_{S} uncertainties.

For single Z boson, WW and ZZ diboson production, the cross section uncertainty is 5 %, while for WZ boson production the cross section uncertainty is 7 % [2]. Since the diboson



(a) Multi-jet fit

(b) MC cross section

Figure 7.11: Relative systematic uncertainty in the total SM prediction at the detector level from the multi-jet fit (a) and the MC background cross section uncertainties (b). Shown in both cases is the H_{T} distribution for ≥ 1 jet.

cross sections are assumed to be correlated and hence varied together, the more conservative value of 7 % is used in the $W+$jets analysis, although the largest contribution in the $W+$jets signal region is from WW production.

The cross section uncertainty of $t\bar{t}$ production is 6 % [124], while the cross section uncertainties of single top production are 3.9 % for both s- and t-channel production [125, 126] and 6.8 % for the Wt channel [127]. The $t\bar{t}$ and single top cross sections are varied in a correlated way and thus a conservative value of 6.8 % is used for the combined $t\bar{t}$ and single top cross section uncertainty.

Figure 7.11 presents the fractional uncertainties in the total SM prediction at the detector level from the multi-jet fit variations and the Z, top and diboson cross sections for the H_{T} distribution. The largest multi-jet uncertainty component at low H_{T} values is the variation of the lower fit range boundary. However, for higher H_{T} values, the modified anti-isolation of the multi-jet control region leads to a larger systematic uncertainty. This comes on the one hand from the re-fitted multi-jet fraction, but on the other hand from a changed H_{T} multi-jet template for this systematic uncertainty component using the modified multi-jet control region. The impact of the Z, top and diboson cross section uncertainties depends on the relative contribution of the corresponding background. This is largest for Z production at low energies and for top processes at higher energies. At the low jet multiplicities, the influence of the cross section uncertainties in the total SM prediction mostly remains below 1 %.

7.2 Propagation of systematic uncertainties through the unfolding procedure

Detector level systematic uncertainties are propagated to the final results through the unfolding procedure. Two types of uncertainties are distinguished in this case:

- Uncertainties in particle calibrations and MC description of data like electron, $E_{\mathrm{T}}^{\mathrm{miss}}$, jet and pile-up uncertainties,

- Uncertainties in the background estimates like cross section and multi-jet fit uncertainties.

Uncertainties in background estimates are propagated to the final results by modifying the background estimates by their uncertainties and repeating the unfolding. The systematically varied unfolded events, e.g. for the multi-jet fit range variation (MJ-FR) are given by:

$$\hat{N}^{\text{MJ-FR}}(\mathbf{x_T}) = \sum_{x_D} [N_{\text{data}}(\mathbf{x_D}) - N_{\text{bkg}}^{\text{MJ-FR}}(\mathbf{x_D})] \cdot \mathcal{M}(\mathbf{x_T}, \mathbf{x_D}) \,.$$

The systematic uncertainty $\Delta\hat{N}^{\text{MJ-FR}}(\mathbf{x_T})$ is then determined as the difference to the result from the nominal unfolding:

$$\Delta\hat{N}^{\text{MJ-FR}}(\mathbf{x_T}) = \hat{N}^{\text{MJ-FR}}(\mathbf{x_T}) - \hat{N}(\mathbf{x_T}) \,.$$

In case of the cross section uncertainties e.g. of the top background, the multi-jet estimate is repeated with the modified cross section and thus, both the top background and the multi-jet background are adjusted in the unfolding procedure.

Uncertainties in the electron, $E_{\text{T}}^{\text{miss}}$ and jet calibrations as well as the pile-up uncertainty are estimated in the MC simulation. They are propagated through the unfolding procedure by modifying the unfolding matrix, determined from $W+\text{jets}$ MC. For example, for the electron energy scale (EES), the systematically varied unfolded result is given by:

$$\hat{N}^{\text{EES}}(\mathbf{x_T}) = \sum_{x_D} [N_{\text{data}} - N_{\text{bkg}}](\mathbf{x_D}) \cdot \mathcal{M}^{\text{EES}}(\mathbf{x_T}, \mathbf{x_D}) \,,$$

and the systematic uncertainty $\Delta\hat{N}^{\text{EES}}(\mathbf{x_T})$ is computed as:

$$\Delta\hat{N}^{\text{EES}}(\mathbf{x_T}) = \hat{N}^{\text{EES}}(\mathbf{x_T}) - \hat{N}(\mathbf{x_T}) \,.$$

Since the systematic uncertainties at the detector level only affect the reconstructed, but not the true particles, only the reconstructed *axis* of the unfolding matrix changes as result of the systematic variation. Therefore, the systematically varied unfolding matrix $\mathcal{M}^{\text{EES}}(\mathbf{x_T}, \mathbf{x_D})$ is obtained by weighting the nominal matrix with the ratio of systematically varied to nominal reconstructed distributions in the $W+\text{jets}$ MC.

$$\mathcal{M}^{\text{EES}}(\mathbf{x_T}, \mathbf{x_D}) \rightarrow \frac{N_{W+\text{jets}}^{\text{EES}}(\mathbf{x_D})}{N_{W+\text{jets}}(\mathbf{x_D})} \cdot \mathcal{M}(\mathbf{x_T}, \mathbf{x_D})$$

This has the advantage of reducing the required computing resources in comparison to a direct determination of the unfolding matrix for each systematic uncertainty, in particular for the matrices used in the two-dimensional unfolding. For individual systematic uncertainties it has been checked that both approaches give similar unfolding matrices for the systematics variation. The weights $N_{W+\text{jets}}^{\text{syst}}(\mathbf{x_D})/N_{W+\text{jets}}(\mathbf{x_D})$ are determined from the Alpgen $W+\text{jets}$ simulation like the nominal unfolding matrix $\mathcal{M}(\mathbf{x_T}, \mathbf{x_D})$.

7.3 Systematic uncertainties in the unfolding method

In addition to the propagated detector-level systematic uncertainties, systematic uncertainties from the unfolding method itself are taken into account.

Two main MC-based inputs to the unfolding procedure affect the unfolded result: MC statistics and the MC-based reconstruction efficiency and fake fraction estimates. Estimates

of the impact from both sources are explained in the following and are treated as a systematic uncertainty. They are added in quadrature to the other uncertainties to obtain the total systematic uncertainty in the unfolded results.

Uncertainty due to limited Monte Carlo statistics

The number of generated MC events for the W + jets simulation determines the size of statistical fluctuations in the distribution predictions, in particular towards higher energies and larger jet multiplicities. In these regions, the differential production cross sections drop significantly. A sufficient population of the low cross section tails with predicted events is however needed in particular to fill the unfolding matrix.

The main sample for W +1 jet production in the ALPGEN simulation contains about 48 Mio events. Considering the reconstruction efficiency of 50 %, this is about 2-3 times the data statistics for W production with 1 jet.

To estimate the impact of the finite MC statistics in the unfolded result, the MC-based unfolding inputs are statistically varied and the unfolding is repeated. Variations are also applied for the *missed* and *fake* event distributions which encode the information of reconstruction efficiency and fake fraction. The variation is performed per bin within a Gaussian distribution with the bin entry as mean and the σ related to the statistical MC uncertainty for this bin – similar as for the template variations in the multi-jet background estimation (see Section 5.1.2).

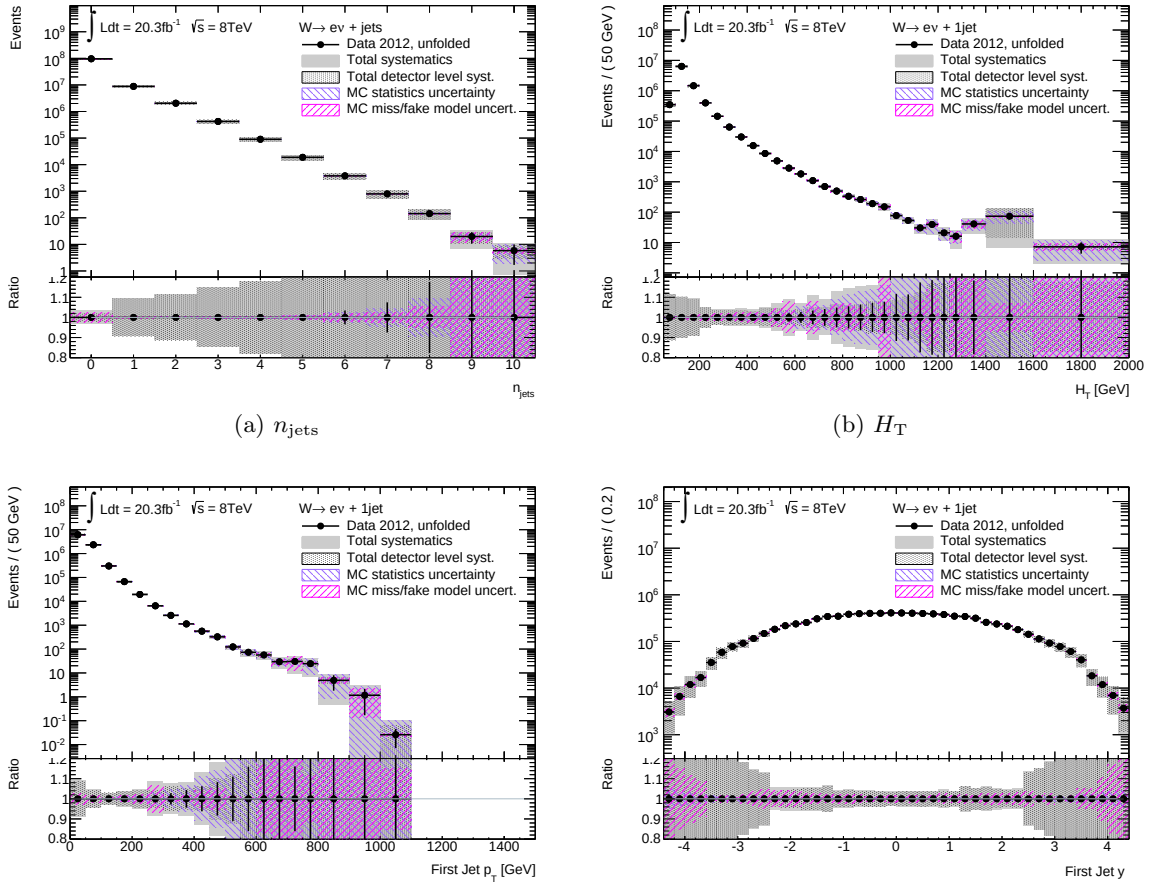
The variations of the MC-inputs and subsequent unfolding are performed 100 times and the spread of the unfolded results per bin is considered as systematic uncertainty due to limited MC statistics. The results for this systematic uncertainty are shown in Figure 7.12 together with the uncertainty from the reconstruction efficiency and fake fraction modelling in the MC simulation discussed next.

Uncertainty from reconstruction efficiency and fake fraction

The impact of the reconstruction efficiency and fake fraction mismodelling in the W+jets MC simulation is estimated by applying the corresponding values from a different MC generator – SHERPA. The difference to the unfolded results with reconstruction efficiency and fake fraction from ALPGEN is assigned as systematic uncertainty. The results for this systematic uncertainty are presented in Figure 7.12 discussed next.

Combination of unfolding method uncertainties

Figure 7.12 displays the unfolded data distributions together with the total systematic uncertainty, illustrating the contributions from the detector-level uncertainties and the two unfolding method uncertainties, separately. As expected, the systematic uncertainty from MC statistics increases strongly towards low statistics regions, i.e. at very large jet multiplicities, high H_{T} or increasing leading jet p_{T}, making it the dominant uncertainty in these regions, in particular for exclusive jet multiplicities. The systematic uncertainty due to reconstruction efficiency and fake fraction modelling is found to constitute the largest uncertainty component at 0 jets in the n_{jets} distribution and also creates a non-negligible contribution e.g. at forward rapidities of the leading jet rapidity distribution. Towards the low statistics regions, e.g. at high H_{T} or leading jet p_{T}, it exhibits sometimes strong fluctuations, as the result of the smaller MC statistics in SHERPA in these regions than in ALPGEN. In principle, the statistical fluctuations in reconstruction efficiency and fake fraction modelling uncertainty could



(a) n_{jets} (b) H_{T}

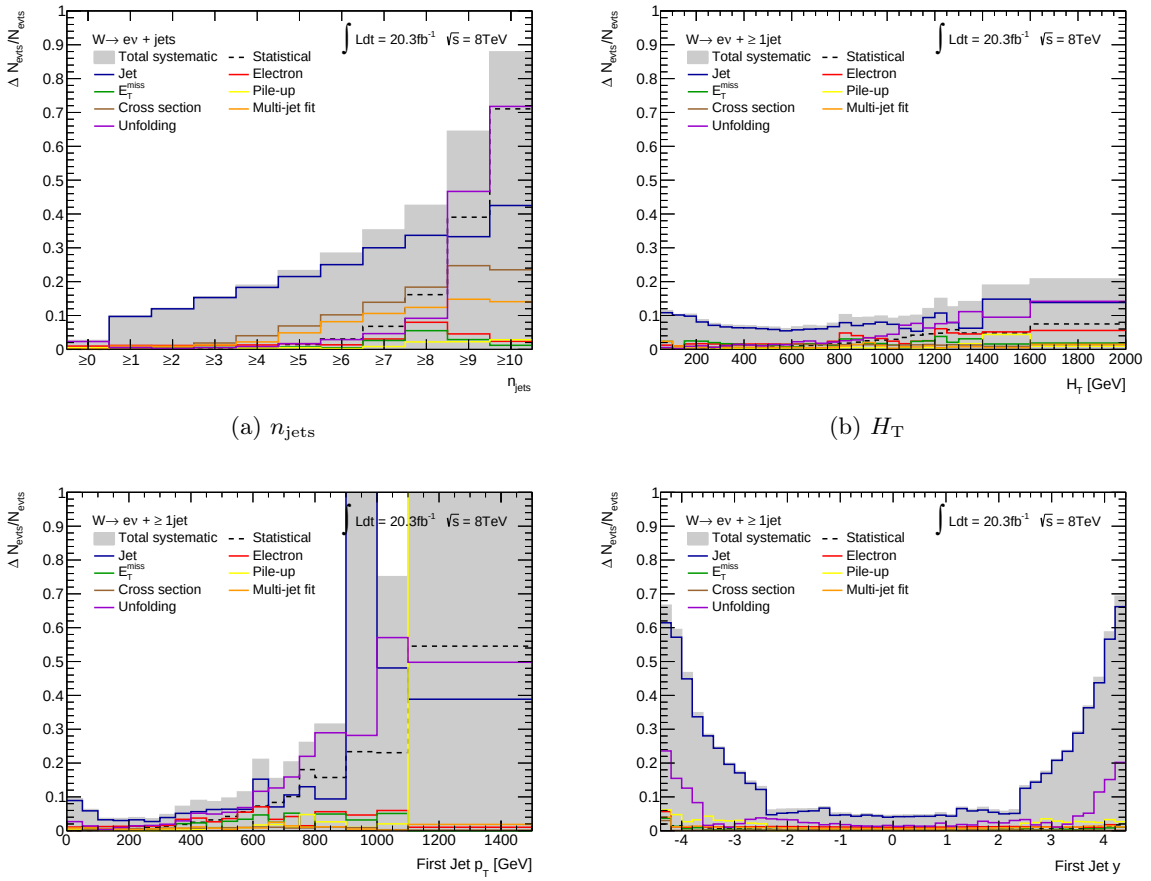
(c) Leading jet p_{T} (d) Leading jet y

Figure 7.12: Unfolded data with the total systematic uncertainty (gray) and the uncertainty contributions from the detector-level (black dotted), MC statistics (purple) and MC reconstruction efficiency and fake fraction modelling (miss/fake model) (pink) for the n_{jets} (a), H_{T} (b), leading jet p_{T} (c) and rapidity (d) distributions. The latter three are for 1 jet.

be reduced with a smoothing of the reconstruction efficiency and fake fraction distributions in SHERPA in the future, but this has not been applied here.

7.4 Total uncertainty in the unfolded cross sections

After unfolding, the W + jets cross sections are computed as explained in the next chapter. The relative size of the systematic and statistical uncertainties in the unfolded results is shown in Figures 7.13 - 7.14 for the differentially unfolded observables for ≥ 1 jet and inclusive jet multiplicities in the case of n_{jets}. At low jet energies, the total systematic uncertainty of around 10 % is dominated by the jet uncertainties, which come in roughly equal amounts from the jet energy scale and resolution uncertainties. With increasing energy (jet p_{T}, H_{T} or W p_{T}), the total uncertainty reduces to around $5-8$ %, depending on the distribution. For even higher energies, the systematic uncertainties increase strongly to values between about 20 % and above 100 %, depending on the observable. In the jet rapidity distributions, the total systematic uncertainty is about the same size for the central rapidities ($|y| < 2.4$) as in the medium p_{T}-range. For forward leading and second leading jet rapidities, the uncertainties



(a) n_{jets} (b) H_{T}

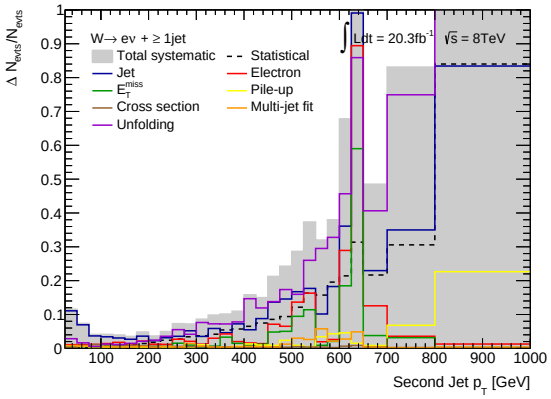
(c) Leading jet p_{T} (d) Leading jet y

Figure 7.13: Total systematic and statistical uncertainties in the unfolded results for the n_{jets} (a), H_{T} (b), as well as leading jet p_{T} (c) and rapidity (d) distributions – the latter three for ≥ 1 jet. The total systematic uncertainy corresponds to the sum in quadrature of the propagated detector level and the unfolding method uncertainties, while the statistical uncertainty is kept separate.

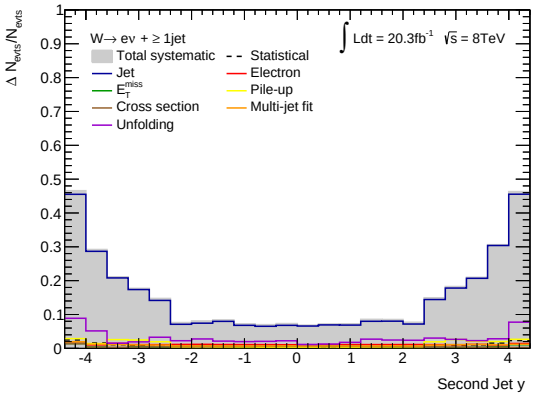
increase to around $47 - 70\,\%$. In the case of the rapidity distributions, the systematics increase is due to jet uncertainties and was already observed at the detector level. For the p_{T}-like distributions, the systematics increase comes to a large extend from the unfolding uncertainty, namely the uncertainty due to limited MC statistics, supplemented by the jet uncertainties, mainly the uncertainty in the jet energy scale.

For H_{T}, the uncertainty due to MC statistics in the unfolding grows to approximately the same size as the jet uncertainty, (partially) dominating the total systematic uncertainty, from around 1050 GeV onwards. For the leading and second leading jet, the uncertainty due to limited MC statistics dominates the total systematic uncertainty from approximately 600 GeV and 300 GeV, respectively, and for W p_{T} from about 400 GeV onwards. These H_{T} and p_{T}-values are rather low compared to the targeted measurement ranges, which suggests the generation of more MC statistics before the final publication of this analysis.

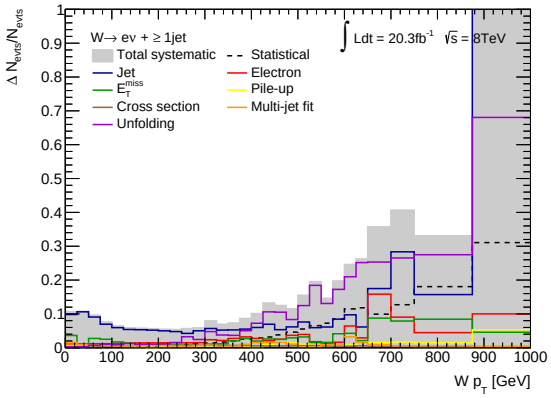
The statistical uncertainties in the unfolded results increase as expected with the H_{T} or p_{T}-values. Towards large jet or W p_{T}, they are approximately similar in size to the jet uncertainties. For large H_{T}, the statistical uncertainties are approximately a factor of 2 smaller than the two dominating systematic uncertainties, leaving this distribution systematics



(a) Second leading jet p_{T}



(b) Second leading jet y



(c) W p_{T}

Figure 7.14: Total systematic and statistical uncertainties in the unfolded results for the second leading jet p_{T} (a) and rapidity (b), as well as the W boson p_{T} distributions for ≥ 1 jet. The total systematic uncertainy corresponds to the sum in quadrature of the propagated detector level and the unfolding method uncertainties, while the statistical uncertainty is kept separate.

dominated even if the uncertainty from limited MC statistics could be reduced. In the rapidity distributions, statistical uncertainties are negligible.

Two further interesting observations concern the jet multiplicity and again the jet rapidity distributions. In the ≥ 0-jets bin of the n_{jets} distribution, the unfolding uncertainty dominates the total systematic uncertainty which amounts to about 2.6 %. In the case of the leading and second leading jet rapidity distributions, the unfolding systematics increase quite clearly towards forward rapidities, though here the jet uncertainties still dominate. In both cases, the source is however not the MC statistics, but the reconstruction efficiency and fake fraction modelling in Alpgen compared to Sherpa.

The systematic and the statistical uncertainties in the unfolded results per inclusive jet multiplicity are listed in Table 7.1. For the cross section measurement also the uncertainty due to the luminosity estimate as explained in the next chapter needs to be taken into account. This is a global value of 2.8 %, and not included in the distributions and numbers given here. The systematics plots as displayed above, but separately for W^+ and W^- selections, are given in Appendix A.8 and show similar conclusions apart from an even greater sensitivity of the W^- to fluctuations in the systematics as the result of limited MC statistics. For the

	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
Jet	0.37	9.71	11.97	15.29	18.30	21.51	25.03	30.01
Electron	0.96	1.16	1.14	1.17	1.29	1.68	1.34	3.06
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.41	0.39	0.49	0.56	0.44	0.72	0.47	2.62
Pile-up	0.49	0.55	0.55	0.51	0.42	0.50	0.23	0.83
Multi-jet bkg	0.10	0.96	0.93	1.29	2.19	4.88	8.17	10.60
Bkg cross sections	0.02	0.21	0.67	1.87	4.02	6.91	10.20	13.91
Unfolding	2.31	0.56	0.23	0.50	0.82	1.67	2.87	4.65
Total systematic	2.61	9.87	12.10	15.53	18.94	23.26	28.42	35.28
Statistical	0.02	0.05	0.12	0.28	0.63	1.42	3.06	6.80

Table 7.1: Total systematic and statistical uncertainties in the unfolded results in % for inclusive jet multiplicities, equivalently to what is shown in Figure 7.13a.

final publication of this analysis, a smoothing of the systematic uncertainties in low statistics regions could help to reduce fluctuations in the uncertainties, in particular for detector-level uncertainties, but this has not been applied here.

7.5 Uncertainty in the cross section ratios

Systematic uncertainties in the W^+/W^- cross section ratios shown in Chapter 9 are computed from the systematic uncertainties in the W^+ and W^- cross sections, differentiating between correlated and uncorrelated systematic uncertainties.

Systematic uncertainties like the jet energy scale components are correlated between W^+ and W^- cross sections which is reflected in a (partial) cancellation of these uncertainties in the W^+/W^- ratio. For correlated systematic uncertainties, the W^+/W^- ratio is computed with the systematic variations applied both in the numerator and the denominator and the uncertainty in the ratio r is then given, e.g. as $\Delta r^{\mathrm{JES}} = r^{\mathrm{JES}} - r$.

The statistical uncertainty in the unfolded results as well as a few systematic uncertainties of statistical nature are treated as uncorrelated between W^+ and W^- cross sections. These systematic uncertainties are the statistical uncertainty in the multi-jet fit result and the uncertainty in the unfolding method arising from limited MC statistics. For these uncorrelated uncertainties, the uncertainty in the W^+/W^- ratio is computed by adding in quadrature the relative uncertainties due to these sources:

$$\Delta r^{\mathrm{stat}} = r\sqrt{\left(\frac{\Delta\sigma^{\mathrm{stat}}_{W^+}}{\sigma_{W^+}}\right)^2 + \left(\frac{\Delta\sigma^{\mathrm{stat}}_{W^-}}{\sigma_{W^-}}\right)^2},$$

where σ_{W^+} (σ_{W^-}) are the cross sections for W^+ (W^-) production, explained in more detail in the next chapter.

The systematic uncertainties in the W^+/W^- ratio as a function of the jet multiplicity are shown in Figure 7.15. At the lowest jet multiplicities, a large reduction of the systematic uncertainties is visible compared to the uncertainties in the unfolded results discussed in the previous section. For example, at 1 jet, the total systematic uncertainty in the unfolded result is reduced from about 10 % for the absolute W^+ and W^- cross sections down to below 1 % for the ratio. From 3 jets onwards, the uncertainty in the ratio rises clearly, reaching slightly above 20 % at 6 jets. The uncertainty is here dominated by the uncertainty in the multi-jet fit, originating from the uncorrelated fit uncertainty. For even higher jet multiplicities,

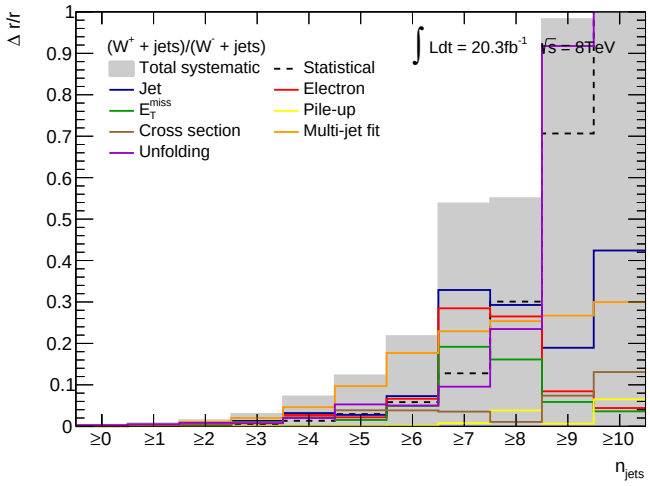


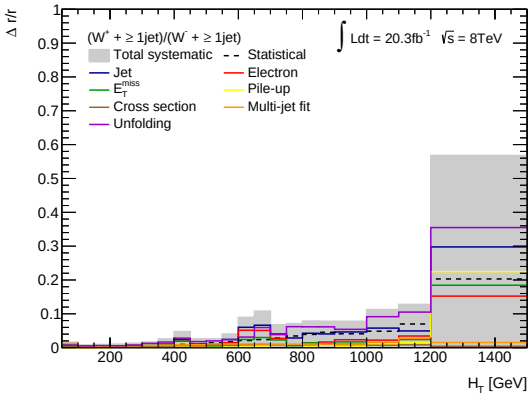
Figure 7.15: Total systematic and statistical uncertainties in the W^+/W^- cross section ratio in association with inclusive jet multiplicities.

in particular the uncertainty in the unfolding procedure grows strongly, until it eventually dominates. The large size of this uncertainty originates mainly from the uncorrelated MC statistics uncertainty which does not cancel in the ratio. The statistical and systematic uncertainties in the W^+/W^- ratio as a function of the inclusive jet multiplicity are listed again in Table 7.2.

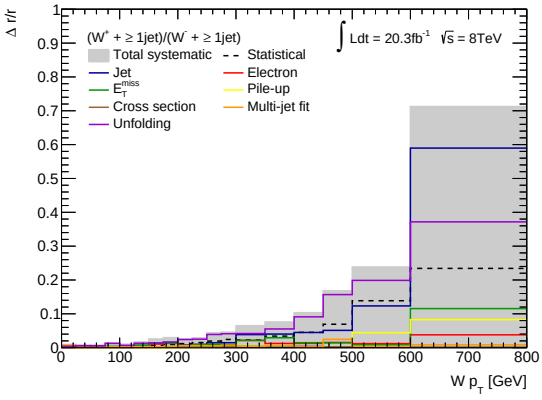
	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
Jet	0.05	0.38	0.38	1.29	3.18	2.75	7.28	32.89
Electron	0.09	0.10	0.19	0.97	2.72	2.17	6.64	28.47
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.09	0.16	0.36	1.01	2.02	1.53	4.93	19.19
Pile-up	0.01	0.02	0.05	0.07	0.04	0.05	0.15	0.78
Multi-jet bkg	0.15	0.39	1.06	2.00	4.61	9.69	17.68	22.93
Bkg cross sections	0.01	0.07	0.26	0.90	2.23	3.87	3.83	3.54
Unfolding	0.27	0.48	0.82	0.73	1.98	5.28	5.01	9.56
Total systematic	0.34	0.75	1.47	3.00	7.19	12.30	21.77	53.76
Statistical	0.03	0.11	0.26	0.58	1.33	2.94	5.83	12.78

Table 7.2: Total systematic and statistical uncertainties in the W^+/W^- cross section ratio in % for inclusive jet multiplicities, equivalent to what is shown in Figure 7.15.

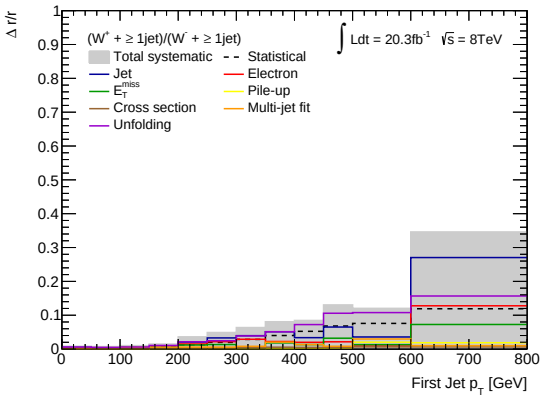
The statistical and systematic uncertainties in the W^+/W^- ratio in association with at least 1 jet are displayed in Figure 7.16 as function of the unfolded observables: H_{T}, W p_{T}, leading and second leading jet p_{T} and rapidity. At low energy scales and for central rapidities, a very good cancellation of systematic uncertainties is observed leading to the total systematic uncertainties mostly below 2 %. With increasing energy scale and for forward rapidities, however, the jet uncertainties and, in particular, the unfolding uncertainty from the limited MC statistics rise clearly. In most bins towards high energy scales, the uncertainty from limited MC statistics dominates the measurement uncertainty. The increase in the systematic uncertainty, but also in the statistical uncertainty is quite strong for the H_{T}, W and jet p_{T} distributions such that the measured range is limited to lower H_{T} and p_{T}-values than measured for the differential cross sections – also see the discussion in Section 9.2.2.



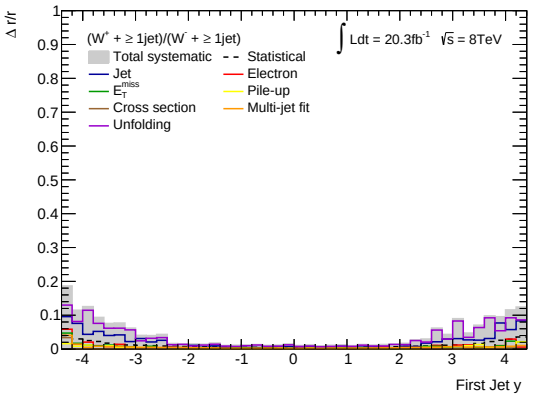
(a) H_{T}



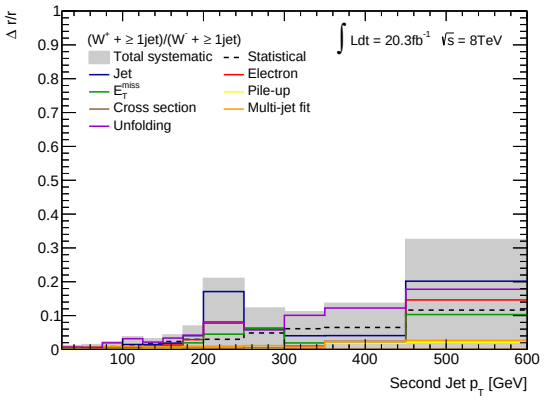
(b) W p_{T}



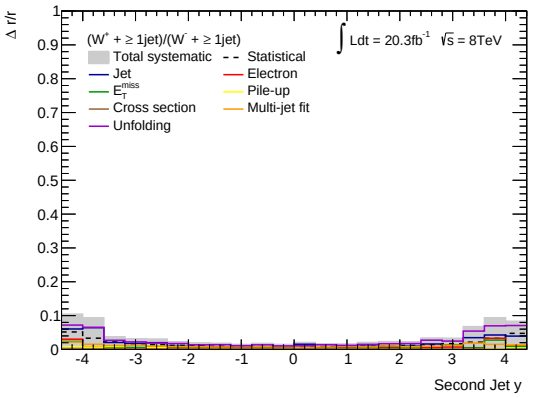
(c) Leading jet p_{T}



(d) Leading jet rapidity



(e) Second leading jet p_{T}



(f) Second leading jet rapidity

Figure 7.16: Total systematic and statistical uncertainties in the W^+/W^- cross section ratio in association with at least one jet as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} and rapidity (c) and (d) as well as second leading jet p_{T} and rapidity (e) and (f).

The statistical uncertainty in the data for the H_{T}, W p_{T} and jet p_{T} distributions is approximately of the same size as the jet uncertainties. The jet uncertainties originate mostly from imperfect cancellation of jet energy scale uncertainties with non-negligible contributions from the other jet uncertainty sources and are the second largest systematic component in most of the measured bins.

The uncertainties in the W^+/W^- ratio therefore demonstrate the cancellation of correlated systematic uncertainties – one of the main arguments for the measurement of a cross section ratio. At large energy scales, however, uncorrelated systematic uncertainty components increase the uncertainties in the W^+/W^- ratio. In particular, a limitation of the measurement precision due to limited MC statistics in the unfolding seems unfortunate, emphasizing the need to generate more MC statistics before the final publication of this analysis.

CHAPTER 8

W + jets cross section measurement

The measurements of production cross sections of SM processes provide stringent tests of SM predictions and challenges their theoretical calculation. In particular, the measurements of processes involving jets drive the need for more precise calculations of perturbative and non-perturbative QCD effects, increasing the order of the perturbative expansion in α_S.

Experimentally, the production cross section for a specific process is given as the ratio of the rate of produced events to the instantaneous luminosity, $\sigma = \dot{N}/\mathcal{L}$, where $\mathcal{L}$ is a measure for the number of particles per area and time which can undergo an interaction. At the LHC, it specifies the beam brightness (see Section 3.1). Instead of counting particles rates, it is however easier to count the total number of particles which were produced in a certain amount of time, $N = \int \dot{N} dt$. Equivalently, the integrated luminosity is then defined as $L_{\text{int}} = \int \mathcal{L} dt$. The cross section for a specific process – W + jets production – is then given as:

$$\sigma_{W+\text{jets}} = \frac{N^{\text{measured}}_{W+\text{jets}}}{A \cdot \epsilon \cdot L_{\text{int}}} .$$

The additional factors A and ϵ describe the acceptance and the efficiency of the detector used to measure $N^{\text{measured}}_{W+\text{jets}}$. In this analysis, the combined correction of A and ϵ is implemented by means of the unfolding procedure using $N^{\text{measured}}_{W+\text{jets}} = N^{\text{measured}}_{\text{data}} - N^{\text{estimated}}_{\text{bkg}}$ as input as described in detail in Chapter 6. The value of the integrated luminosity in 2012 data-taking is $(20.3 \pm 0.6)\,\text{fb}^{-1}$ [128] where the uncertainty in the luminosity is 2.8 %. For the results in this chapter and the following one, the uncertainty in the luminosity is kept separate from the other sources of systematic uncertainties discussed in Chapter 7.

After unfolding, the production cross section is determined as:

$$\sigma_{W+\text{jets}} = \frac{N^{\text{unfolded}}_{W+\text{jets}}}{L_{\text{int}}} .$$

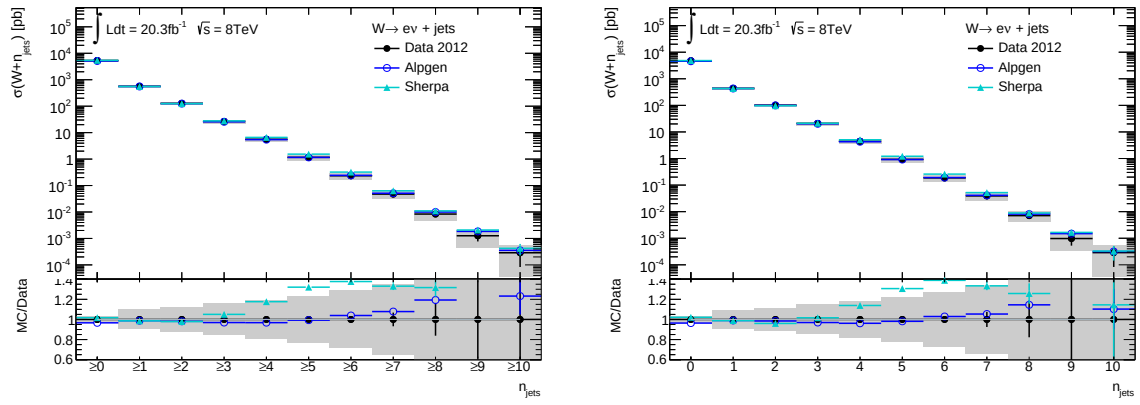
8.1 Cross section of W +n jets production

The cross section for W production in association with different numbers of jets is measured in this analysis. The measurement is conducted once for exclusive and once for inclusive jet multiplicities, i.e. in the former case exactly n jets have to be present, while in the latter at least n jets are required. The measurement is performed separately for the charge-independent W selection as well as for W^+ and W^- production. In all cases, the branching of the W boson as $W \to e\nu$ is considered to be part of the cross section. In order to limit the

amount of simulation-based extrapolations which are needed for measurement in a 4π phase space, the cross section is only measured in a fiducial phase space volume, denoted as σ_{fid}. The fiducial phase space is defined in the unfolding and was discussed in Section 6.2.

The measured fiducial W+n jets cross sections are given in Section 8.1.1. The measurement is compared to the theoretical prediction for $W+ \geq 1$ jet at NLO in pQCD, obtained with MCFM [59]. This is explained in Section 8.1.2.

8.1.1 Measurement of the $W + n$ jets cross sections



(a) Inclusive n_{jets} (b) Exclusive n_{jets}

Figure 8.1: Measured cross sections for W boson production in association with jets as a function of the inclusive number of jets (a) and the exclusive number of jets (b). The measured cross sections are compared to two LO multi-leg generators, Alpgen and Sherpa. The lower panel gives the ratio of MC to data. The error bars show the relative size of the statistical uncertainties (error bars) and the total systematic uncertainty (gray band). For the MC predictions only statistical uncertainties are shown.

The presented cross sections are the first measurement of W boson production cross sections as a function of n_{jets} at $\sqrt{s} = 8\,\text{TeV}$ within ATLAS. They are compared to the predictions at particle level from the LO multi-leg generators Alpgen and Sherpa. Figure 8.1 presents the cross sections for both inclusive and exclusive jet multiplicities. Theoretical uncertainties from the implicit jet vetos present for exclusive jet multicities, rejecting the presence of even more jets, are not well-known and only first ideas on their treatment exist so far as e.g. in Ref. [129]. Inclusive jet multiplicities are theoretically simpler and are therefore usually the target of NLO and NNLO pQCD calculations. Experimentally, they also provide higher statistics in the high p_{T} tails and are thus the focus of this measurement. The cross sections in Figure 8.1 are shown up to 10 jets, but the highest jet multiplicities (≥ 8 jets) are not considered reliable. This is due to the decreasing signal contribution at the highest jet multiplicities in data and the larger systematic uncertainties. The precise measurement of these jet multiplicities would require a more refined background estimate at high n_{jets} and the consideration of additional background uncertainties at these jet multiplicities. Cross section numbers are therefore only given up to 7 jets. Table 8.1 gives the measured cross sections for both inclusive and exclusive jet multiplicities, including the statistical, systematic and luminosity uncertainties. In order to show both statistical and systematic uncertainties with the same precision, the cross section and uncertainties are rounded to three significant digits on the systematic uncertainty.

The measured $W+ \geq n$ jets cross sections vary over five orders of magnitude from $(5.3 \pm 0.2)\,\text{nb}$ for inclusive W production down to $(48 \pm 17)\,\text{fb}$ for $W+ \geq 7$ jets. The experimental

n_{jets}	Inclusive n_{jets}							Exclusive n_{jets}						
	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$
0	5304	$\pm$	1	$\pm$	138	$\pm$	149	4741	$\pm$	1	$\pm$	155	$\pm$	133
1	562.8	$\pm$	0.3	$\pm$	55.6	$\pm$	15.8	435.3	$\pm$	0.3	$\pm$	40.4	$\pm$	12.2
2	127.6	$\pm$	0.2	$\pm$	15.4	$\pm$	3.6	101.1	$\pm$	0.1	$\pm$	11.3	$\pm$	2.8
3	26.49	$\pm$	0.07	$\pm$	4.11	$\pm$	0.74	20.88	$\pm$	0.07	$\pm$	3.06	$\pm$	0.58
4	5.61	$\pm$	0.04	$\pm$	1.06	$\pm$	0.16	4.449	$\pm$	0.031	$\pm$	0.798	$\pm$	0.125
5	1.165	$\pm$	0.017	$\pm$	0.271	$\pm$	0.033	0.931	$\pm$	0.015	$\pm$	0.206	$\pm$	0.026
6	0.2340	$\pm$	0.0072	$\pm$	0.0665	$\pm$	0.0066	0.1865	$\pm$	0.0064	$\pm$	0.0502	$\pm$	0.0052
7	0.0475	$\pm$	0.0032	$\pm$	0.0168	$\pm$	0.0013	0.0391	$\pm$	0.0029	$\pm$	0.0134	$\pm$	0.0011

Table 8.1: Measured fiducial cross sections σ_{fid} in pb for W+n jets production at $\sqrt{s} = 8\,\text{TeV}$ for inclusive and exclusive jet multiplicities separately and including the statistical, systematic and luminosity uncertainties, $\Delta\sigma_{\text{stat}}$, $\Delta\sigma_{\text{syst}}$ and $\Delta\sigma_{\text{lumi}}$, respectively.

uncertainty is dominated by the combined systematic uncertainty ($\Delta\sigma_{\text{syst}}$), except for the inclusive W production where the uncertainty from the luminosity estimate ($\Delta\sigma_{\text{lumi}}$) is largest ($\Delta\sigma_{\text{lumi}}$ is not contained in $\Delta\sigma_{\text{syst}}$).

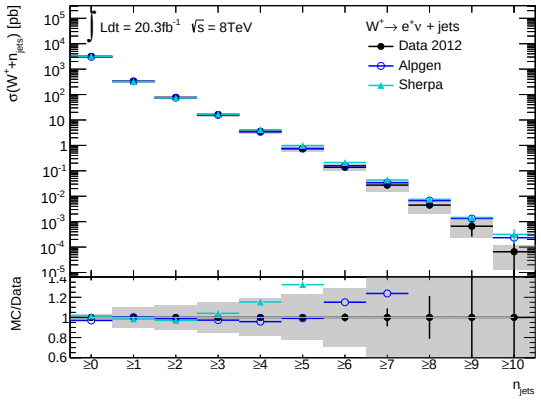
The cross sections for W^+ and W^- production are obtained with separate unfoldings for W^+ and W^-. The sum of the measured cross sections for W^+ and W^- should agree with the W cross section within uncertainties, but can differ numerically from the W cross section value because of the independent unfoldings.

Figure 8.2 displays the measured cross sections for W^+ and W^- production, in both cases for inclusive and exclusive jet multiplicities. The predictions are in reasonable agreement with data, but Sherpa mismodels the jet multiplicities from 4 jets onwards which has already been observed at the detector level. It is interesting that both Alpgen and Sherpa overpredict from 6 jets onwards for W^+ and underpredict for W^-. Both effects cancel approximately, and no clear over- or underprediction trend in Alpgen even at these high jet multiplicities remains for the charge-independent W selection (see Figure 8.1). The jet multiplicities of ≥ 6 jets are however outside of the range which Alpgen and Sherpa compute via matrix elements and jets there originate from the parton shower. Due to this and since the differences are still within uncertainties, this is not worrysome, but nevertheless interesting to note.

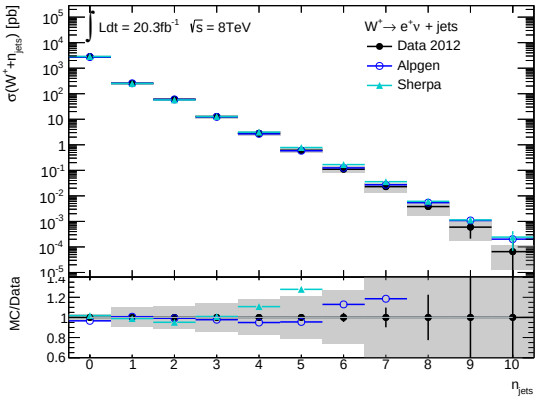
n_{jets}	Inclusive n_{jets}							Exclusive n_{jets}						
	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$
0	3146.5	$\pm$	0.6	$\pm$	85.5	$\pm$	88.1	2816.1	$\pm$	0.6	$\pm$	95.4	$\pm$	78.9
1	330.4	$\pm$	0.2	$\pm$	33.1	$\pm$	9.3	253.9	$\pm$	0.2	$\pm$	24.1	$\pm$	7.1
2	76.46	$\pm$	0.12	$\pm$	9.20	$\pm$	2.14	60.20	$\pm$	0.11	$\pm$	6.73	$\pm$	1.69
3	16.26	$\pm$	0.06	$\pm$	2.48	$\pm$	0.46	12.69	$\pm$	0.05	$\pm$	1.82	$\pm$	0.36
4	3.569	$\pm$	0.026	$\pm$	0.668	$\pm$	0.100	2.820	$\pm$	0.023	$\pm$	0.503	$\pm$	0.079
5	0.749	$\pm$	0.013	$\pm$	0.168	$\pm$	0.021	0.611	$\pm$	0.011	$\pm$	0.130	$\pm$	0.017
6	0.1379	$\pm$	0.0056	$\pm$	0.0398	$\pm$	0.0039	0.1104	$\pm$	0.0050	$\pm$	0.0296	$\pm$	0.0031
7	0.0274	$\pm$	0.0024	$\pm$	0.0119	$\pm$	0.0008	0.0230	$\pm$	0.0023	$\pm$	0.0096	$\pm$	0.0006

Table 8.2: Measured fiducial cross sections σ_{fid} in pb for W^++n jets production at $\sqrt{s} = 8\,\text{TeV}$ for inclusive and exclusive jet multiplicities separately and including the statistical, systematic and luminosity uncertainties, $\Delta\sigma_{\text{stat}}$, $\Delta\sigma_{\text{syst}}$ and $\Delta\sigma_{\text{lumi}}$, respectively.

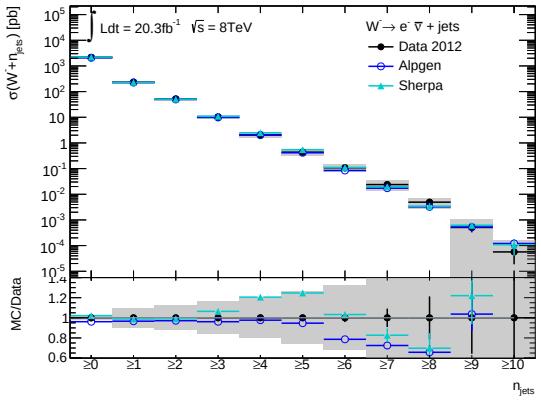
The measured cross sections for inclusive and exclusive jet multiplicities for $W^+ + n$ jets and $W^- + n$ jets production are listed in Table 8.2 and Table 8.3, respectively. Cross sections and uncertainties are rounded to three significant digits on the systematic uncertainty. Only for the exclusive 7 jets cross section, the precision has been reduced to two significant digits



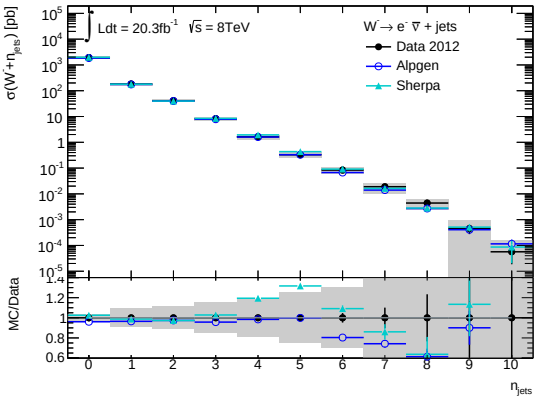
(a) W^+, inclusive n_{jets}



(b) W^+, exclusive n_{jets}



(c) W^-, inclusive n_{jets}



(d) W^-, exclusive n_{jets}

Figure 8.2: Measured cross sections for W^+ and W^- production in association with an inclusive number of jets (a) and (c) or an exclusive number of jets (b) and (d). The uncertainties and MC predictions follow the same format as in Figure 8.1.

due to the smallness of this cross section. The sum of W^+ and W^- cross sections is in good agreement with the numbers for the combined W cross sections, with small fluctuations within uncertainties, mainly for higher n_{jets}, coming from the reduction of the statistics in particular for the W^- measurement. The measured W^+ cross sections range from (3.1 ± 0.1) nb for inclusive W^+ production down to (27 ± 12) fb for $W^+ + \geq 7$ jets, while for W^- they cover values from (2.2 ± 0.1) nb down to (24 ± 11) fb, respectively. For inclusive W production, there is a clear excess of W^+ production compared to W^- production – as one would expect from the naive uud-proton content and the production mechanisms for W^+ and W^-. For ≥ 7 jets, however, the production cross sections move more towards roughly equal amounts of W^+ and W^-. More detailed studies of this are discussed in the next chapter.

8.1.2 Comparison to NLO theory prediction

The measured $W^{(+/-)}+ \geq 1$ jet cross sections are compared to the NLO prediction obtained with MCFM. MCFM computes the cross sections separately for W^+ and W^- and 100 000 events were generated per W charge for an acceptable MC precision in the cross section values. The predictions are calculated using the same phase space as the fiducial cross section

n_{jets}	Inclusive n_{jets}							Exclusive n_{jets}						
	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$	σ_{fid}	$\pm$	$\Delta\sigma_{\text{stat}}$	$\pm$	$\Delta\sigma_{\text{syst}}$	$\pm$	$\Delta\sigma_{\text{lumi}}$
0	2156.9	$\pm$	0.5	$\pm$	53.0	$\pm$	60.4	1924.4	$\pm$	0.5	$\pm$	59.8	$\pm$	53.9
1	232.5	$\pm$	0.2	$\pm$	22.6	$\pm$	6.5	181.4	$\pm$	0.2	$\pm$	16.4	$\pm$	5.1
2	51.11	$\pm$	0.10	$\pm$	6.30	$\pm$	1.43	40.85	$\pm$	0.09	$\pm$	4.66	$\pm$	1.14
3	10.26	$\pm$	0.05	$\pm$	1.66	$\pm$	0.29	8.19	$\pm$	0.04	$\pm$	1.25	$\pm$	0.23
4	2.064	$\pm$	0.023	$\pm$	0.416	$\pm$	0.058	1.627	$\pm$	0.020	$\pm$	0.305	$\pm$	0.046
5	0.437	$\pm$	0.011	$\pm$	0.115	$\pm$	0.012	0.3300	$\pm$	0.0096	$\pm$	0.0827	$\pm$	0.0092
6	0.1075	$\pm$	0.0045	$\pm$	0.0344	$\pm$	0.0030	0.0837	$\pm$	0.0040	$\pm$	0.0253	$\pm$	0.0023
7	0.0238	$\pm$	0.0022	$\pm$	0.0104	$\pm$	0.0007	0.0189	$\pm$	0.0019	$\pm$	0.0086	$\pm$	0.0005

Table 8.3: Measured fiducial cross sections σ_{fid} in pb for W^-+n jets production at $\sqrt{s} = 8\,\text{TeV}$ for inclusive and exclusive jet multiplicities separately and including the statistical, systematic and luminosity uncertainties, $\Delta\sigma_{\text{stat}}$, $\Delta\sigma_{\text{syst}}$ and $\Delta\sigma_{\text{lumi}}$, respectively.

measurement and include the decay of the W boson as $W \to e\nu$. The renormalization and factorization scales, μ_R and μ_F, were chosen to be fixed at the mass of the W boson. The CT10 NLO global PDF set [33] is employed as implemented via the LHAPDF library [130].

Uncertainties in the theoretical predictions are calculated using variations of μ_R and μ_F and the uncertainties provided by the used PDF set. μ_R and μ_F are varied by factors of $1/2$ and 2 separately and simultaneously. The maximum deviation from the μ_R and μ_F variations is denoted as *scale* uncertainty. Uncertainties from the PDF set are computed by MCFM using the *asymmetric Hessian* prescription [33] as required for the CT10 PDF set and are described in more detail in Section 8.1.2.

The obtained NLO predictions for the $W^+ + \geq 1$ jet and $W^- + \geq 1$ jet are:

$$\sigma^{\text{NLO}}_{W^+} = (339.3 \pm 3.0\ (\text{stat})\ ^{+8.2}_{-9.2}\ (\text{PDF})\ ^{+23.5}_{-21.9}\ (\text{scale}))\,\text{pb}$$
$$\sigma^{\text{NLO}}_{W^-} = (236.4 \pm 1.9\ (\text{stat})\ ^{+5.8}_{-6.5}\ (\text{PDF})\ ^{+15.9}_{-15.8}\ (\text{scale}))\,\text{pb}$$

with statistical uncertainties below 1 %, PDF uncertainties around 2.5 % and scale uncertainties around 7 %. The predictions given here agree very well within uncertainties with the measured values of $(330.4 \pm 34.4)\,\text{pb}$ for W^+ and $(232.5 \pm 23.5)\,\text{pb}$ for W^- production.

8.2 Differential cross sections for W + jet production

In addition to the fiducial $W + n$ jets cross sections, differential distributions provide further information on the dynamics of the production process. Higher orders in the α_S expansion are often needed to describe the distributions. For example, the W p_T distribution is only non-trivial at order $(\alpha_S)^1$, i.e. it requires the emission of an additional parton against which the W boson can recoil to acquire the transverse momentum. For the total W production cross section, $(\alpha_S)^1$ is already NLO precision, since the lowest order prediction – $(\alpha_S)^0$ – does not involve the emission of a parton. The NLO prediction for the W p_T distribution is of the order $(\alpha_S)^2$. This makes the computation of differential cross sections more complicated than total cross section predictions when aiming at the same precision of LO, NLO or even NNLO.

Six kinematic distributions are measured in the W + jets analysis. In the following, the inclusive distributions for ≥ 1 jet are shown and compared to theoretical predictions. In addition to the LO multi-leg generators Alpgen and Sherpa, the differential cross sections are compared to a NLO prediction from MCFM. The NLO predictions were computed at an earlier stage of the analysis and thus do not exactly correspond to the final fiducial phase space of the analysis. A global scale factor discussed in more detail in the next section is

applied to the NLO prediction to correct the offset in the cross sections resulting from the differences in the phase space. Residual differences however still occur at lower energies where the thresholds of the phase space requirements are particularly relevant.

The details of the NLO calculation are described in Section 8.2.1 and the measured cross sections are shown in Section 8.2.2 for W production (charge independent), in Section 8.2.3 for W^+ production and in Section 8.2.4 for W^- production.

8.2.1 Calculation of NLO predictions

The differential NLO predictions for $W^{(+/-)}+ \geq 1$ jet production are obtained using MCFM interfaced to APPLGRID [131]. While MCFM computes the NLO cross section, APPLGRID allows to keep track of the x-values of the initial partons as well as the energy scale Q^2 of the interaction. This permits the computation of theoretical uncertainties in the distributions after the actual running of the generator.

The following versions of the software were used: MCFM 6.8 and APPLGRID 1.4.70. The toolkit HOPPET [132], version 1.1.5, is employed for the DGLAP evolution. The CT10 NLO PDF set is used in the cross section computation and is managed via the LHAPDF interface, version 5.9.1.

MCFM/APPLGRID phase space		
Electron	p_{T}	$> \mathbf{20}$ GeV
	|rapidity|	< 2.5
Neutrino	$E_{\mathrm{T}}^{\mathrm{miss}}$	> 25 GeV
W boson	m_{T}	> 40 GeV
Jets	p_{T}	$> \mathbf{20}$ GeV
	$\|y\|$	$< \mathbf{4.5}$
	$\Delta R(\mathrm{el}, \mathrm{jet})$	$> \mathbf{0.3}$
Leading jet	p_{T}	$> \mathbf{25}$ GeV

Table 8.4: Phase space for NLO predictions of differential W cross sections obtained with MCFM interfaced to APPLGRID. Values which are different from the fiducial phase space of the W + jets cross section measurement for historical reasons are printed in bold and blue colour.

The phase space for the NLO calculation is similar to the fiducial region of the analysis, but some requirements are placed at slightly different values. An overview of the phase space selected in APPLGRID is given in Table 8.4. A correction factor in the fiducial cross section of ⅔ is determined from the MCFM prediction in the correct W + jets phase space and a second MCFM run with phase space requirements similar to Table 8.4. This is applied as the global scale factor, mentioned before, to the differential distributions predicted with MCFM/APPLGRID.

Approximately 10^{10} events per observable were generated with MCFM to fill the APPLGRID grids which store the x- and Q^2-values. The renormalization and factorization scales, μ_R and μ_F, were fixed at the mass of the W boson, as for the computation of the fiducial $W + n$ jets cross section.

Uncertainties in the theoretical predictions are calculated as the sum in quadrature of scale uncertainties, PDF uncertainties and variations of α_{S}. The scale uncertainty is the envelope of the effect of the μ_R and μ_F variations as in Section 8.1.2. Uncertainties in the PDF are determined according to the prescription for the CT10 PDF set, i.e. 52 PDF variations of 26 eigenvectors (up and down variations) are combined according to the *asymmetric Hessian*

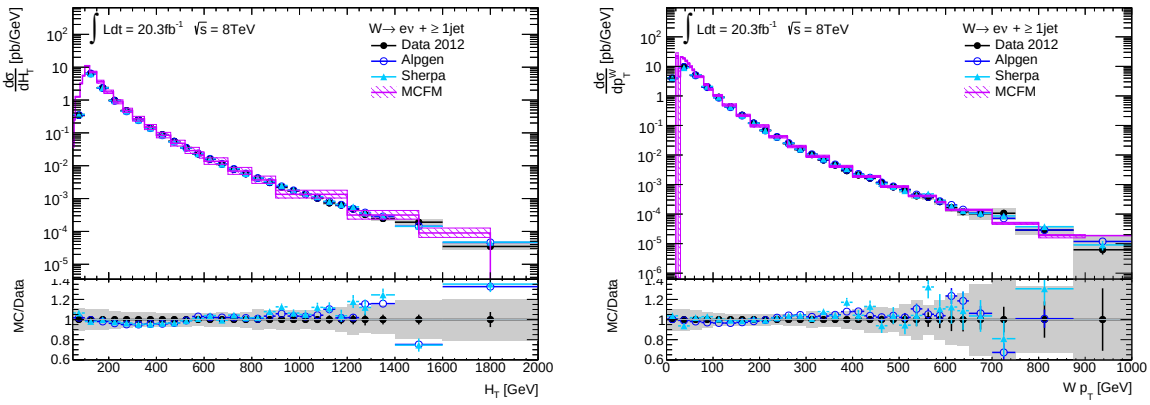
prescription. This prescription yields asymmetric uncertainties $\Delta\sigma_{\mathrm{PDF,+}}$ and $\Delta\sigma_{\mathrm{PDF,-}}$ on the nominal cross section prediction σ given as follows:

$$\Delta\sigma_{\mathrm{PDF,+}} = \sqrt{\sum_i \left(\max\left[\sigma^{(+)}_{\mathrm{PDF,i}} - \sigma, \sigma^{(-)}_{\mathrm{PDF,i}} - \sigma, 0\right]\right)^2}$$

$$\Delta\sigma_{\mathrm{PDF,-}} = \sqrt{\sum_i \left(\max\left[\sigma - \sigma^{(+)}_{\mathrm{PDF,i}}, \sigma - \sigma^{(-)}_{\mathrm{PDF,i}}, 0\right]\right)^2}$$

where $\sigma^{(+/-)}_{\mathrm{PDF,i}}$ indicates the cross section prediction obtained with the up (+) or down (−) variation of the error PDF i and the sum runs over all eigenvectors, i.e. error PDFs [33]. The error PDFs for the CT10 eigenvectors are given at 90 % CL, and are converted to 68 % CL, dividing by the factor 1.64485. The central value for α_{S} used in the CT10 PDF is $\alpha_{\mathrm{S}} = 0.1180$. According to the PDF4LHC recommendations [133], α_{S} uncertainties are to be estimated by varying the α_{S} value in the PDF by ± 0.0015 and use the envelope of the cross section prediction with the varied values of α_{S} as α_S-uncertainty. For CT10, only modified PDFs with α_{S} values of 0.120 and 0.116 exist, and the resulting $\Delta\sigma^{(+/-)}_{\alpha_{\mathrm{S}}}$ is scaled by the ratio 0.0015/0.0020, in line with the PDF4LHC prescription.

8.2.2 Differential W + jet cross section measurement



(a) H_{T} (b) W p_{T}

Figure 8.3: Measured differential $W+ \geq 1$ jet cross sections as a function of H_{T} (a) and W p_{T} (b). The data is compared to Alpgen and Sherpa predictions as well as NLO MCFM predictions. The lower panels show the size of the relative statistical and systematic uncertainties in the measurement as well as ratios to Alpgen and Sherpa predictions. A ratio to MCFM cannot be computed due to inconsistent binning.

The differential cross sections for $W+ \geq 1$ jet production as a function of H_{T} or the W boson p_{T} are shown in Figure 8.3. The data is compared to the multi-leg LO generators Alpgen and Sherpa, as well as the NLO MCFM calculation, explained in the previous section. Uncertainties in the LO predictions are statistical only, while the uncertainties in the NLO calculation correspond to the sum in quadrature of scale, PDF and α_{S}-uncertainties.

The measured distributions extend up to 2 TeV and 1 TeV for H_{T} and the W p_{T}, respectively. Very good agreement between multi-leg LO as well as NLO predictions with data is observed for both observables. Small differences between the NLO prediction and data are visible at low H_{T} or W p_{T} as a result of the slightly different phase space in the MCFM

prediction. The W p_{T} distribution at NLO is computed above 20 GeV. In the W p_{T} bin at $25-30$ GeV, the effect of imperfect cancellations between real emission and virtual amplitudes is observed which is typical for higher order calculations close to the phase space boundaries.

The H_{T} distribution is a theoretically difficult observable as it contains important contributions from higher jet multiplicities which require the modelling of additional hard parton emissions as well as subprocesses involving W emission in dijet production ($qq \to qq + W$). In particular, the latter can only be taken correctly into account at NNLO precision [134]. The H_{T} distribution has been measured several times, also up to 2 TeV e.g. in the ATLAS W + jets measurement based on 2011 data. There, the range $680 - 2000$ GeV is however measured in three bins, while the presented measurement provides 16 data points in the similar range $650 - 2000$ GeV.

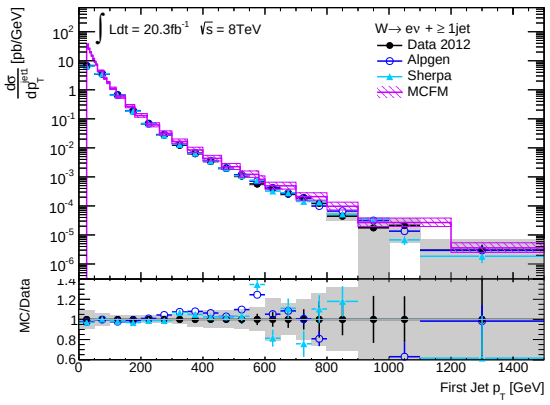
The W p_{T} distribution has been measured by ATLAS at the center-of-mass energy of $\sqrt{s} = 7$ TeV based on 31 pb^{-1} of data [135] and by CMS at $\sqrt{s} = 8$ TeV based on 18 pb^{-1} of data [136]. At low W p_{T} specific care needs to be taken to estimate the W p_{T} from the hadronic recoil at percent-level accuracy, i.e. much more precise than what is obtained in this measurement. The presented W p_{T} measurement therefore focusses on the high-p_{T} tail. The previous ATLAS measurement extends only up to 300 GeV, while the CMS measurement reaches up to 600 GeV with the last bin comprising the W p_{T} from 250 GeV onwards. The W + jets measurement presented here extends the existing measurements to much higher energy scales, providing an important benchmark for higher order predictions at large W p_{T}.

Figure 8.4 presents the measured differential cross sections for $W+ \geq 1$ jet production as a function of the leading and second leading jet p_{T} and rapidity. Alpgen and Sherpa mostly describe the measured cross sections within experimental uncertainties. Some discrepancies are observed for moderate second leading jet p_{T} and for the forward leading and second leading jet rapidities where in particular Sherpa fails to describe the distribution. The latter has already been observed at the detector level and was seen also in the ATLAS 2011 W + jets measurement, so it does not come as a surprise. For the NLO prediction from MCFM a stronger difference is observed, barely covered by the systematic uncertainty band for the leading jet p_{T}, but clearly distinct from data for the second leading jet p_{T}. The latter can be understood from the fact that MCFM provides a NLO prediction for $W + 1$ jet, which includes the second jet as real emission correction, but the second jet is then only predicted at the LO. In comparison to Alpgen and Sherpa, it does not include higher jet multiplicities which impact mainly the high p_{T} regions and thus leads to the non-optimal description of data in these regions. In the very low jet p_{T} regions, the effect of the non-ideal match of the phase spaces between data and MCFM is observed. The jet rapidity distributions are dominated by the low p_{T} jets. The slightly lower jet p_{T} requirement for MCFM leads to a higher p_{T}-integrated cross section as a function of the jet rapidity, resulting in the offset in Figure 8.4b. For the second leading jet rapidity, MCFM is therefore not shown.

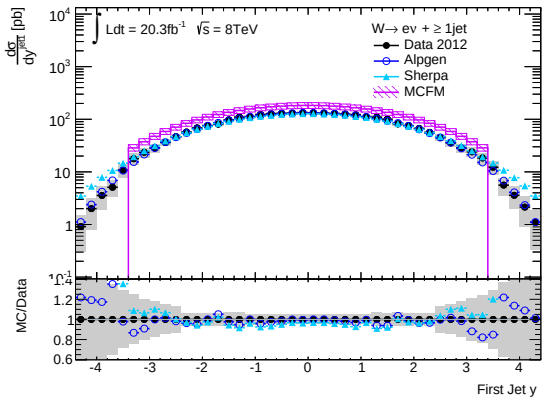
Overall, good agreement between the measured differential $W+ \geq 1$ jet cross sections and corresponding predictions is observed. Similar conclusions are obtained for the measurement of differential $W+ \geq 2$ jets cross sections shown in Appendix B.1.

8.2.3 Differential W^+ + jet cross section measurement

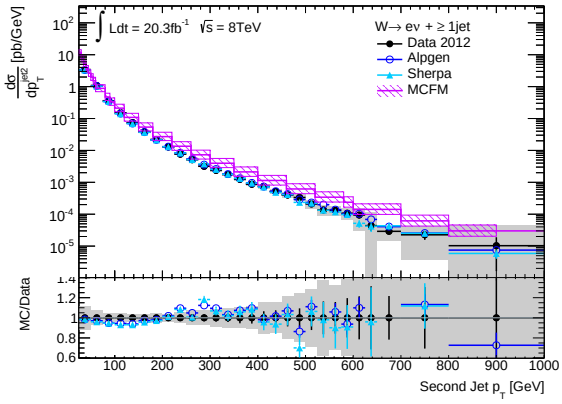
The differential $W^+ + \geq 1$ jet cross sections as a function of H_{T}, W p_{T}, as well as leading and second leading jet p_{T} and rapidity are displayed in Figure 8.5. The measured cross sections are compared to Alpgen, Sherpa and MCFM. In general, the conclusions are similar to those obtained for the differential W + jet cross sections. A small, but more pronounced discrepancy between Alpgen and data is observed for $W^+ + \geq 1$ jet at around $250 - 300$ GeV



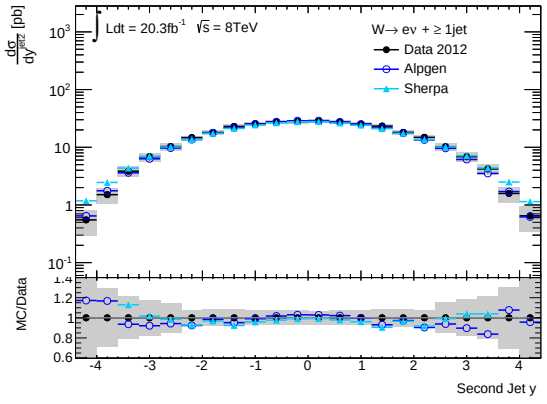
(a) Leading jet p_{T}



(b) Leading jet y



(c) Second leading jet p_{T}



(d) Second leading jet y

Figure 8.4: Measured differential $W+ \geq 1$ jet cross sections as a function of leading jet p_{T} (a) and rapidity (b) as well as second leading jet p_{T} (c) and rapidity (d). Comparisons of data to predictions are done similar to Figure 8.3. The prediction by MCFM for the leading jet rapidity is restricted to $|y| < 3.4$ for technical reasons. For the second leading jet rapidity, MCFM displays a similar offset as visible in the leading jet rapidity and is therefore not shown.

in the W p_{T} and slightly higher for the leading jet p_{T}.

Differential W^+ cross sections measured in association with at least 2 jets are presented in Appendix B.2. Conclusions from these are the same as obtained for the charge independent W production discussed above.

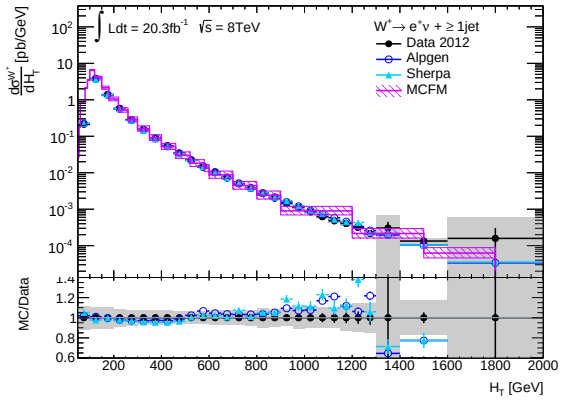
8.2.4 Differential W^- + jet cross section measurement

Differential $W^-+ \geq 1$ jet cross sections are presented in Figure 8.6 as function of H_{T}, W p_{T}, leading and second leading jet p_{T} as well as their rapidities. The measurements are compared to MC predictions using the three generators as before. In addition to similar general conclusions, both ALPGEN and SHERPA demonstrate larger discrepancies in the description of data at medium H_{T}, W p_{T} and slightly for the leading jet p_{T}. In the second leading jet p_{T} a clear difference is visible in the medium p_{T} range. In these ranges, where experimental uncertainties are smallest, both generators (ALPGEN even more than SHERPA) underpredict the data. This effect is almost not visible in the W^+ cross sections discussed before.

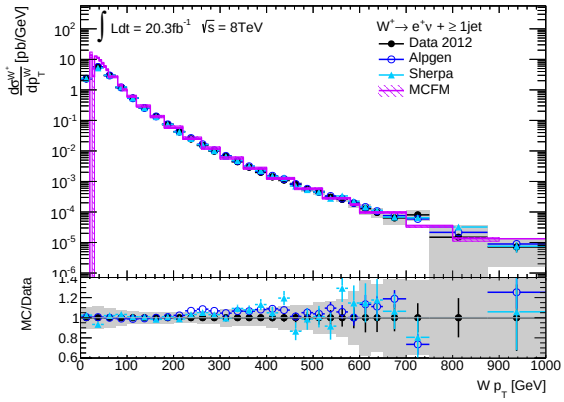
Comparing to $W^{(\pm)}$ production in association with exact jet counts, a strong underpredic-

tion of SHERPA, but also for ALPGEN is observed in the same range of the second leading jet p_{T} for exactly two jets in both W charges. For SHERPA, an underprediction for exactly two jets is visible also in the H_{T} distribution, but not for ALPGEN. ALPGEN underpredicts H_{T} and W p_{T} in terms of exclusive jet multiplicities only for W^- production in association with exactly one jet – with residual effect on the W prediction, of course. Figure 8.7 displays the differential W cross section with exactly two jets for the second leading jet p_{T} and Figure 8.8 the W^- cross section in association with exactly one jet for H_{T}. The observed underprediction for $W^-+ \geq 1$ jet production is probably a combination of both effects.

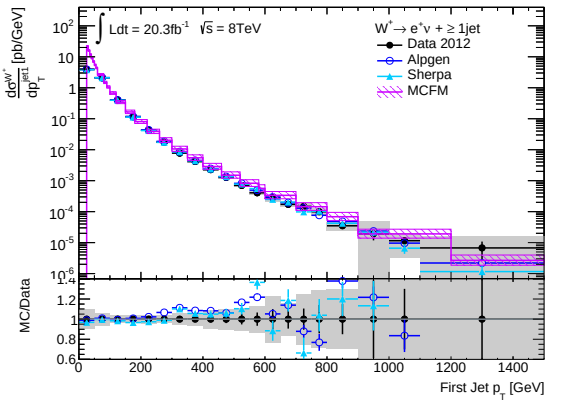
Differential cross sections for $W^-+ \geq 2$ jets are again given in Appendix B.2, with conclusions similar to what was discussed before.



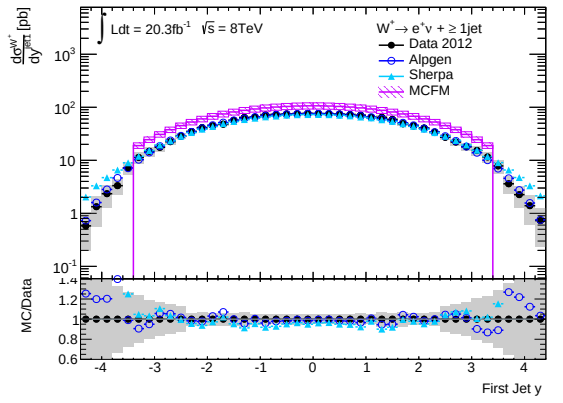
(a) H_{T}



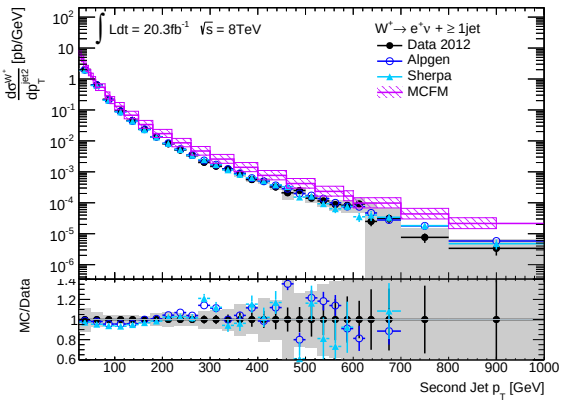
(b) W^+ p_{T}



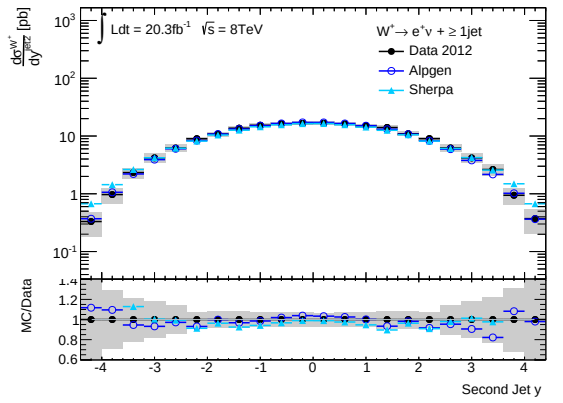
(c) Leading jet p_{T}



(d) Leading jet y

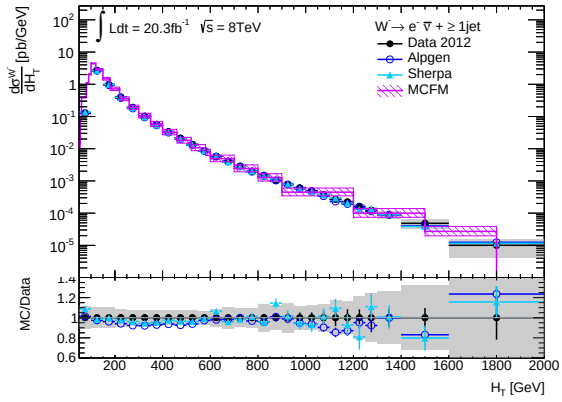


(e) Second leading jet p_{T}

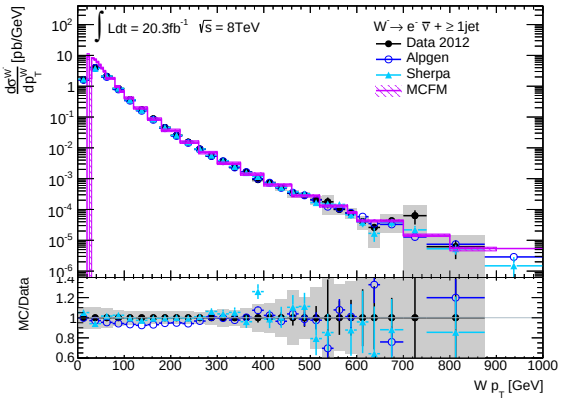


(f) Second leading jet y

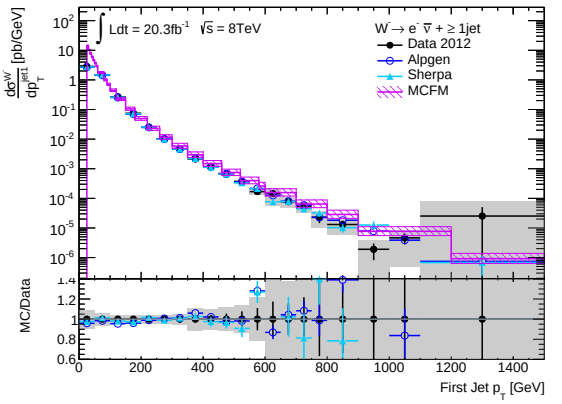
Figure 8.5: Measured differential $W^+ + \geq 1$ jet cross sections as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and rapidity (d) as well as second leading jet p_{T} (e) and rapidity (f). Comparisons of data to predictions are done similar to Figure 8.3 and Figure 8.4.



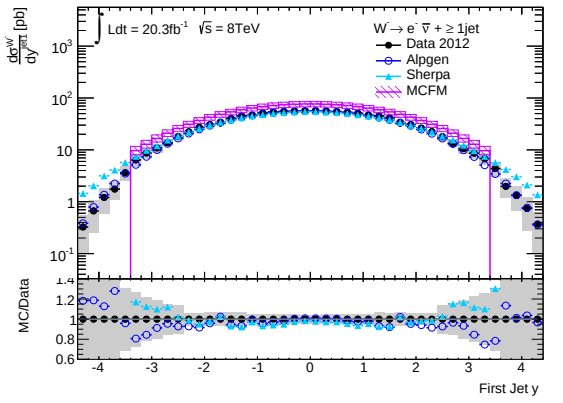
(a) H_{T}



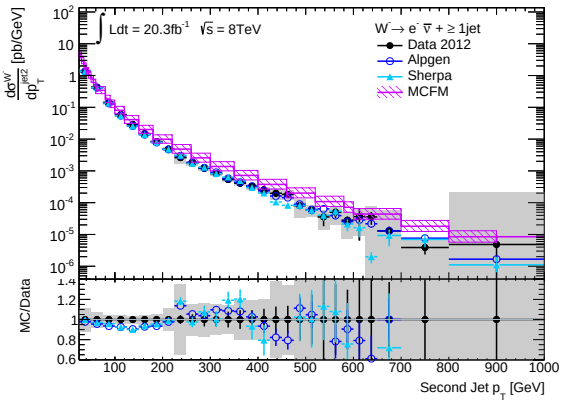
(b) W^- p_{T}



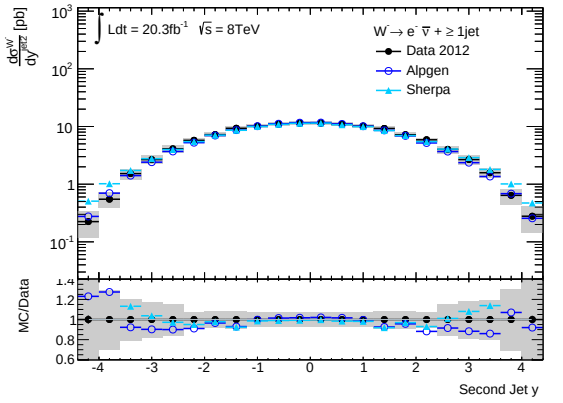
(c) Leading jet p_{T}



(d) Leading jet y



(e) Second leading jet p_{T}



(f) Second leading jet y

Figure 8.6: Measured differential W^- + ≥ 1 jet cross sections as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and rapidity (d) as well as second leading jet p_{T} (e) and rapidity (f). Comparisons of data to predictions are done similar to Figure 8.3 and Figure 8.4.

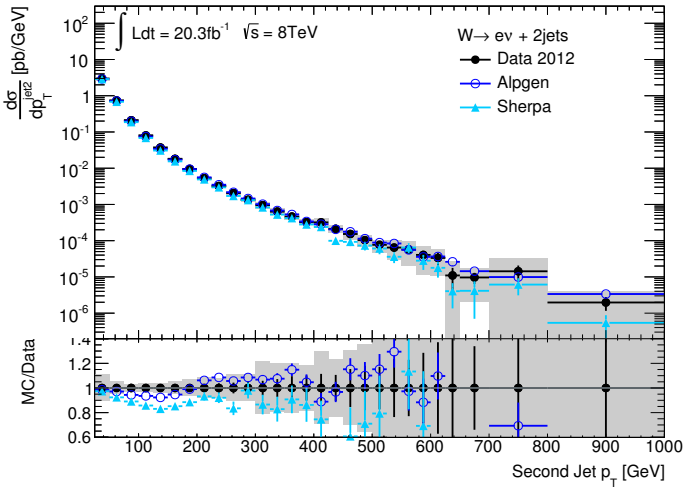


Figure 8.7: Differential cross sections for W production in association with exactly two jets as a function of the second leading jet p_{T}. Comparisons of data to Alpgen and Sherpa are done similar to Figure 8.3.

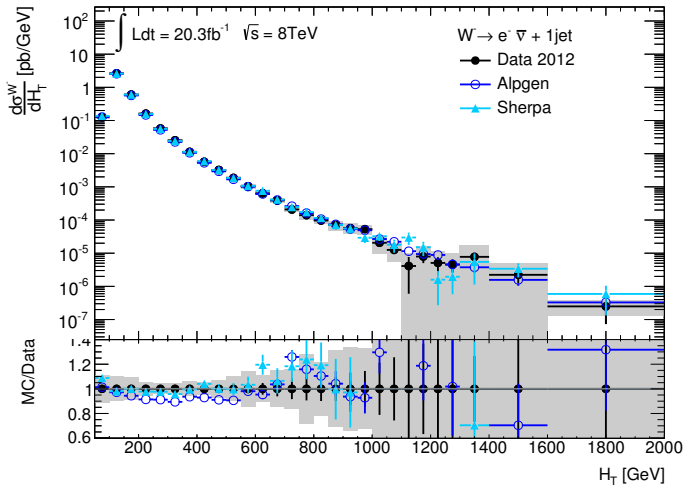


Figure 8.8: Differential cross sections for W^- production in association with exactly one jet as a function of H_{T}. Comparisons of data to Alpgen and Sherpa are done similar to Figure 8.3.

CHAPTER 9

Measurement of W^+/W^- ratios

The measurement of W^+ and W^- production cross sections with correlated systematic uncertainties, presented as W^+/W^- ratio, is expected to provide important information on the valence quark PDFs in the proton. Measurements of the kinematic distributions of $W^\pm$ production in association with at least 1 jet thereby yield complementary information to what is typically probed by W asymmetry measurements by ATLAS or CMS.

Studies of the potential sensitivity of observables to PDFs are a useful tool for designing the analysis in view of a later use in PDF fits. The inclusion of PDF sensitive measurements in global PDF fits is then done by the PDF collaborations themselves, provided that complete information on correlated and uncorrelated uncertainties is published together with the measured cross sections.

In this analysis, studies of PDF sensitivity are conducted using MCFM interfaced to APPLGRID and are presented in Section 9.1. Results from these studies are implemented in the analysis by the choice of measured differential cross sections. The W^+/W^- ratios are measured in $\sqrt{s} = 8\,\text{TeV}$ data from the ATLAS experiment and are presented in Section 9.2, including a comparison to LO and NLO predictions.

9.1 Theory studies on W^+/W^- ratios

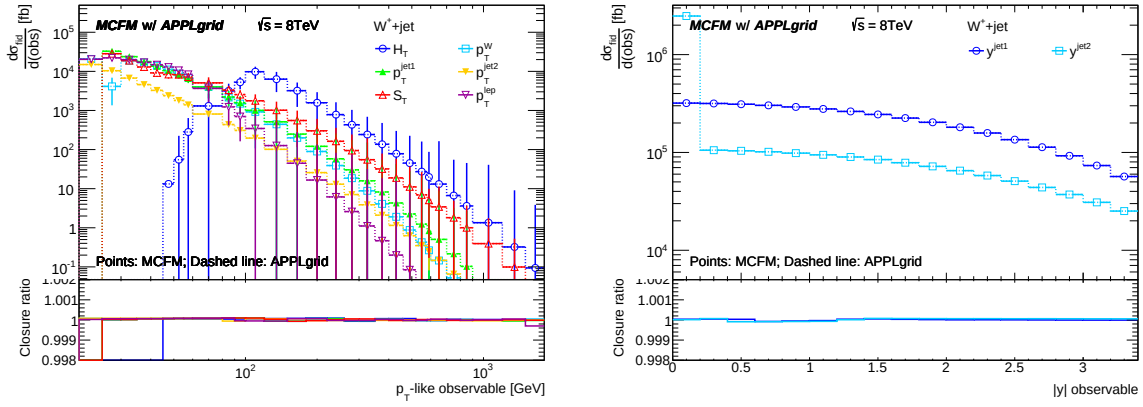
The relation between cross section predictions and PDFs is possible to study in NLO predictions for $W^+ + \text{jet}$ and $W^- + \text{jet}$ production obtained with MCFM interfaced to APPLGRID. APPLGRID allows to store the perturbative coefficients of NLO QCD calculations in look-up tables and thus enables the a-posteriori inclusion of parameter modifications for cross section predictions. Possible are e.g. variations of PDF sets, including their uncertainties and the value of α_S, variations of renormalization and factorization scales μ_R and μ_F or the center-of-mass energy. Also access to the contributing subprocesses, i.e. the simulated initial state configurations, is provided.

In one main run of the generator MCFM, APPLGRID stores *weights* in the three-dimensional space of the momentum fractions of the initial partons x_1 and x_2 and the energy scale Q^2, per order of α_S, per bin of the investigated kinematic observable and per subprocess. For a cross section calculation at NLO, the α_S orders correspond to the LO and NLO contributions. Depending mainly on the number of bins of the observable distribution as well as the bins in x and Q^2, the APPLGRID files, called *grids*, can become quite large and their determination poses requirements on the available computing and memory capacities.

In the W + jets analysis, APPLGRID grids were determined for the following eight observables both for W^+ + jet and W^- + jet cross sections at NLO with MCFM:

- H_T
- S_T
- W p_T
- Electron p_T
- Leading jet p_T
- Leading jet y
- Second leading jet p_T
- Second leading jet y

The observable S_T is defined as the scalar sum of the jet p_T's only, in contrast to H_T which also contains electron p_T and E_T^{miss} in the scalar sum. The determined grids have 40 bins in x and 15 bins related to Q^2.



(a) p_T-like observables

(b) y observables

Figure 9.1: Differential W^+ + jet APPLGRID cross section predictions in fb (dashed lines) in comparison to the reference prediction from MCFM (points) for p_T-like observables (a) and rapidities (b). The lower panel displays the ratio of the APPLGRID prediction to the reference. In some distributions of the p_T-like observables, the phase space requirements (e.g. leading jet $p_T > 25\,\text{GeV}$) are visible at the start of the distribution. The lowest p_T bin starts at 20 GeV. For the second jet rapidity, the y-value was set to zero if no second jet was found in the event, leading to the peak in that distribution at zero.

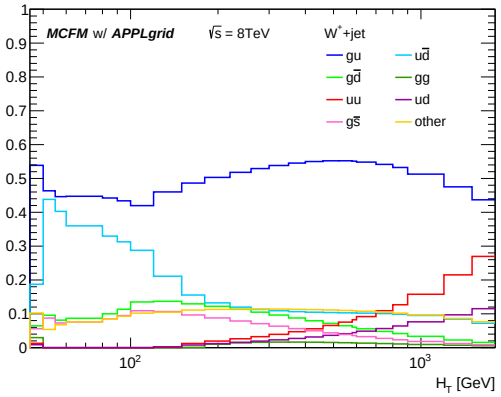
The agreement of the cross sections computed from the APPLGRID weights with the reference from MCFM is examined in the W + jets analysis and confirmed, as shown in Figure 9.1 for W^+ + jet. The results from W^- + jet look similar. Details on the computation of MCFM predictions interfaced to APPLGRID were given in Section 8.2.1.

Using the APPLGRID files, studies of the subprocess contributions, the correlation between W^+/W^- predictions and PDFs as well as studies of the theoretical uncertainties in the W^+/W^- predictions were performed. They are discussed in Section 9.1.1, Section 9.1.2 and Section 9.1.3, respectively.

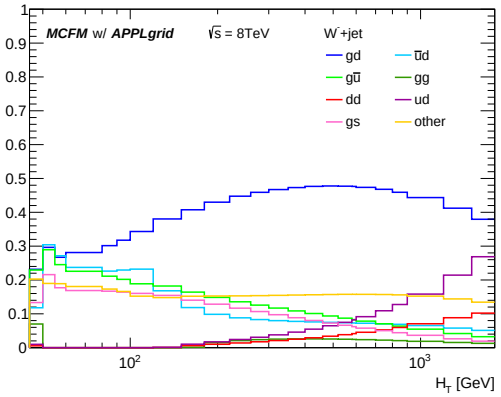
9.1.1 Contributions from different subprocesses

Contributions of different subprocesses to the cross section predictions are obtained from APPLGRID by convoluting the PDFs for individual parton flavours with the weights from APPLGRID. This allows to examine the fractional contribution from specific parton flavour combinations in the initial state as a function of the investigated observable.

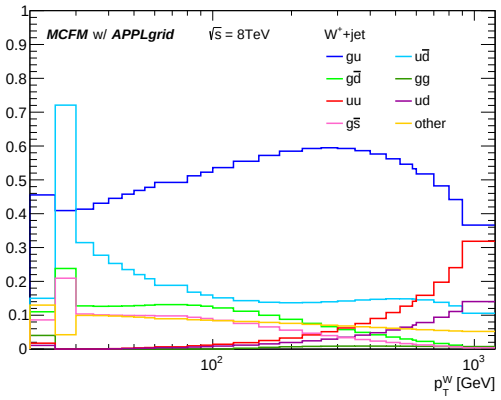
For the simulation of W^+ + jet and W^- + jet with MCFM, subprocess contributions are computed using the CT10 NLO PDF set. Results are shown in Figure 9.2 as a function of H_T, W p_T and leading jet p_T.



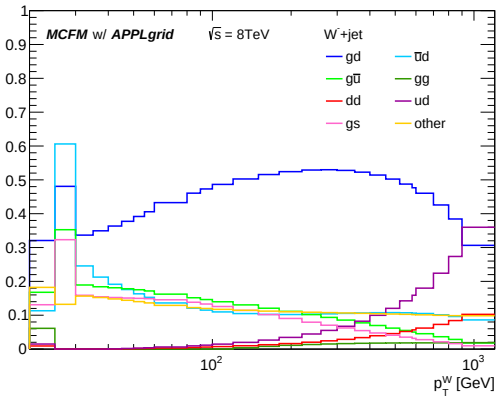
(a) W^+, H_{T}



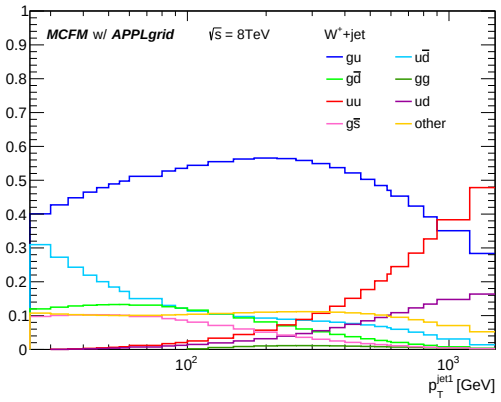
(b) W^-, H_{T}



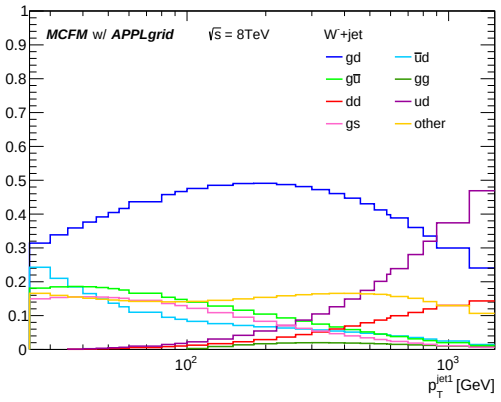
(c) W^+ p_{T}



(d) W^- p_{T}

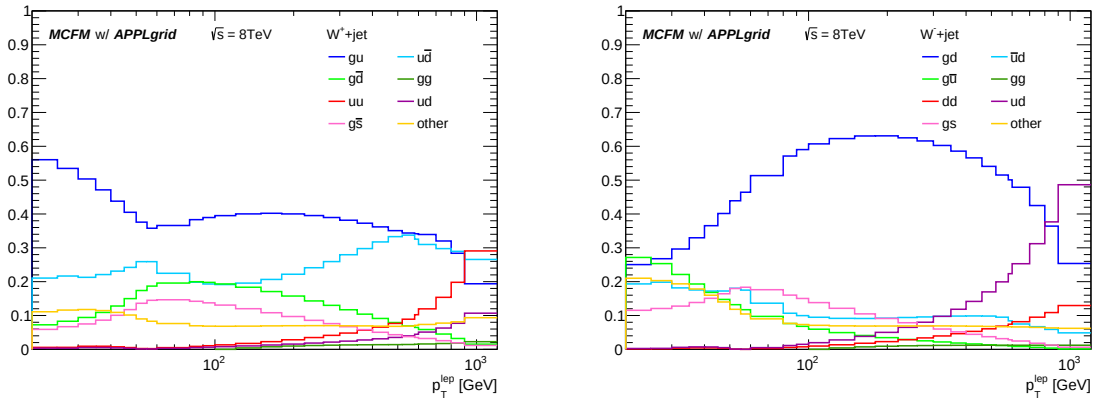


(e) W^+, leading jet p_{T}



(f) W^-, leading jet p_{T}

Figure 9.2: Fractional contribution of specific subprocesses to the differential cross section of $W^+ +$ jet or $W^- +$ jet production as a function of H_{T} (a) and (b), W p_{T} (c) and (d) and the leading jet p_{T} (e) and (f). The NLO CT10 PDF set is used in the subprocess determination.



(a) W^+, electron p_{T} (b) W^-, electron p_{T}

Figure 9.3: Fractional contribution of specific subprocesses to the differential cross section as a function of the electron p_{T} for $W^+ + \mathrm{jet}$ (a) and $W^- + \mathrm{jet}$ (b). The NLO CT10 PDF set is used in the subprocess determination.

The dominating subprocess, constituting around 50 % of the total cross section for most of the H_{T} and p_{T} ranges, corresponds to the initial state configuration gu and gd for $W^+ + \mathrm{jet}$ and $W^- + \mathrm{jet}$, respectively. Its fractional contribution is slightly higher for W^+ than for W^- in almost all distributions. At low H_{T} and p_{T} values, the initial state configurations $u\bar{d}$ and $\bar{u}d$ account for around $30 - 40$ % of the cross section with a spike at the W p_{T} of $25 - 30$ GeV where the cross section calculation has the instability in the LO/NLO amplitude cancellation discussed in Section 8.2.2. Towards higher H_{T} and p_{T} values, uu and ud or ud and dd configurations for $W^+ + \mathrm{jet}$ or $W^- + \mathrm{jet}$, respectively, become significant or even dominant. The magnitude of these contributions is roughly a factor of 2 larger for uu compared to ud in W^+, and again larger for ud compared to dd in W^-– in accordance with the naive expectation to find the ratio 2:1 of u:d valence quarks in the proton. These subprocesses basically correspond to dijet production with subsequent W emission, mentioned briefly in Section 8.2.2. The initial state u and d valence quarks provide the high momentum fraction x needed for the large H_{T} or p_{T} values.

The subprocess composition as a function of the electron p_{T} is shown in Figure 9.3. For $W^- + \mathrm{jet}$ production, the split-up from the different subprocesses appears similar to the distributions shown previously. For the positron p_{T} from the W^+ decay, however, a lower fraction from the gu initial state and a larger fraction from $u\bar{d}$ is observed at moderate to large p_{T}. Also the contribution from $g\bar{u}$ rises quite significantly at p_{T} around 100 GeV, compared to W^- or the other kinematic distributions where it contributes mainly at low values. The reason for the different behaviour of the electron p_{T} distribution is a large left-handed polarization of both W^+ and W^- bosons produced at the LHC, combined with decay kinematics in the presence of polarizations [137]. For the gu and gd subprocesses, both W^+ and W^- bosons are produced with up to 80 % left-handed polarization for a scattering by 90°, corresponding to asymptotically large W p_{T}. Although other subprocesses dilute the polarization fraction, it still remains large – around 50 % for W $p_{\mathrm{T}} \gtrsim 50$ GeV, rising to about 70 % at large W p_{T} ($\gtrsim 400 - 600$ GeV). Due to the left-handed nature of the W coupling, the positron is emitted in the decay of a left-handed W^+ backwards to the W flight direction, while in the decay of the left-handed W^- the electron is emitted forward. In the laboratory frame, this leads to a strong reduction of the positron p_{T}, while the electron p_{T} follows the

W^- p_{T}. Consequently, large values of the positron p_{T} in the W^+ decay arise either from unpolarized or right-handed W^+ bosons as obtained in the $u\bar{d}$ or the $g\bar{u}$ configurations. For more details on the polarization of W bosons at the LHC see e.g. Ref. [137].

These studies therefore demonstrate the dominance of the gu and gd initial state configurations over a large range of the observable values which is favourable for a sensitivity to the valence quark ratio. Only for the differential cross section as a function of the electron p_{T}, important differences in the subprocess contributions are observed, indicating potentially different PDF sensitivities.

9.1.2 Correlations between W^+/W^- and $u_{\mathrm{v}}/d_{\mathrm{v}}$

The impact of PDFs on differential cross section predictions can be estimated by studying the correlation ρ between the cross section and the PDF. As a function of x and for example H_{T}, the correlation is defined for the parton flavour p as follows:

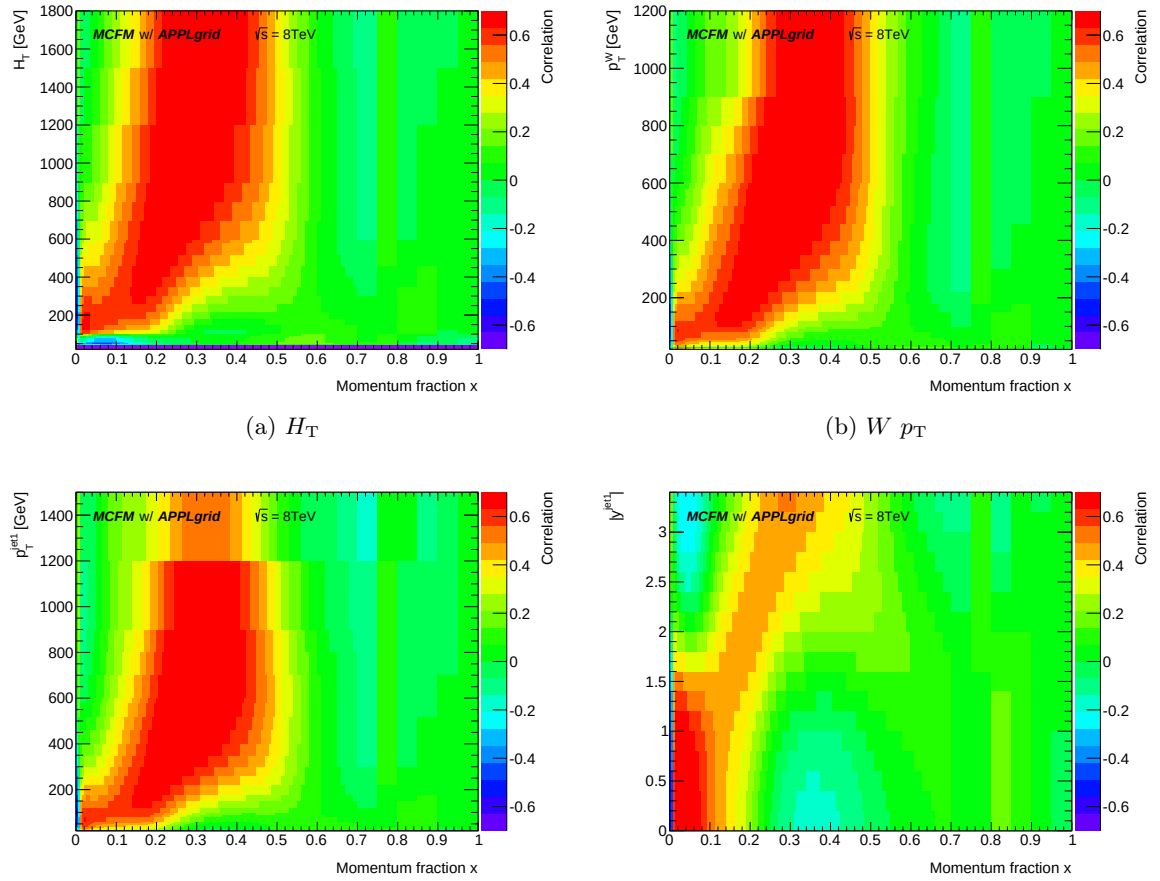
$$\rho_p(x, H_{\mathrm{T}}) = \frac{\langle \sigma_{\mathrm{fid}}(H_{\mathrm{T}}) \cdot x f_p(x) \rangle - \langle \sigma_{\mathrm{fid}}(H_{\mathrm{T}}) \rangle \cdot \langle x f_p(x) \rangle}{\sigma_{\sigma_{\mathrm{fid}}} \cdot \sigma_{x f_p}} ,$$

where $\langle ... \rangle$ denotes a sample mean and $\sigma_{...}$ the standard deviation of the sample variations. σ_{fid} is given as a function of H_{T} ($\sigma_{\mathrm{fid}}(H_{\mathrm{T}})$) and the PDF as a function of x ($x f_p(x)$). This definition implies a sample of PDFs which are also used to obtain a sample of cross sections. The PDF sets from the NNPDF collaboration are based on PDF fits to statistical replicas of a chosen set of data where the replicas are drawn from multi-gaussian probability distributions constrained by statistical and systematic uncertainties in data [138]. This yields a set of PDF replicas which are provided by the NNPDF collaboration as error PDFs and can be used to calculate the sample correlation coefficient ρ defined above.

For the W+jets analysis, the PDF set NNPDF 2.3 [40] with 100 replicas is used to estimate the correlations between cross section and PDFs. For the correlation of the W^+/W^- ratio to the $u_{\mathrm{v}}/d_{\mathrm{v}}$ valence quark ratio, $\sigma_{\mathrm{fid}}(H_{\mathrm{T}})$ and $x f_p(x)$ are replaced by the respective ratios $(\sigma_{\mathrm{fid}}^{W^+}/\sigma_{\mathrm{fid}}^{W^-})(H_{\mathrm{T}})$ and $(x f_{u_{\mathrm{v}}}/x f_{d_{\mathrm{v}}})(x)$. In the following, the notation W^+/W^- implies that both W^+ and W^- are produced in association with at least 1 jet, i.e. correctly $(W^+ + \text{jet})/(W^- + \text{jet})$, if not explicitly mentioned otherwise.

The correlation of the W^+/W^- ratio to the $u_{\mathrm{v}}/d_{\mathrm{v}}$ ratio as a function of H_{T}, W p_{T}, leading jet p_{T} and leading jet rapidity is shown in Figure 9.4. Correlations of about 60 % are observed for momentum fractions $x \sim 0.15 - 0.45$ above approximately 300 GeV/400 GeV in H_{T}, W p_{T} or leading jet p_{T}, depending slightly on the observable. For lower energy scales, the x value with high correlation also moves to lower values, reaching down to $x \sim 0.02$. For the leading jet rapidity, a different correlation pattern is observed than for the p_{T}-like variables. Central rapidities exhibit around 60 % correlation to the $u_{\mathrm{v}}/d_{\mathrm{v}}$ ratio similar to the p_{T}-observables, but in the lower x-range of about $0.02 - 0.1$. This is in line with the probed x-range of lepton asymmetry measurements at ATLAS or CMS, discussed in Section 2.3. The jet rapidity can therefore be seen as an interesting cross check to these measurements, but does not cover the main x-range targeted in this analysis particularly well. Only some reduced correlation in the forward region is observed, demonstrating that higher rapidities probe larger x-values. This is however most likely not significant enough for tight constraints on PDFs.

The strength of this measurement therefore lies in the kinematic variables related to the W p_{T}, i.e. H_{T}, W p_{T}, leading jet p_{T}, etc. To get an idea of the approximate magnitude of the correlations per observable and to be able to quantify the sensitivity of different observables, the correlation per x is averaged over the entire observable range. In a few cases, the averaged



(a) H_T (b) W p_T

(c) Leading jet p_T (d) Leading jet y

Figure 9.4: Correlation of the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio to the u_v/d_v valence quark ratio as a function of the momentum fraction x as well as H_T (a), W p_T (b), leading jet p_T (c) and leading jet rapidity (d).

observable range is limited to above a certain value, in order not to average over anti-correlated and correlated regions as visible e.g. for H_T below approximately 100 GeV. The averaged correlations are displayed in Figure 9.5. Similar correlations are observed for most observables except the leading jet rapidity in the forward region – as expected from Figure 9.4d. The second leading jet rapidity also demonstrates reasonable correlation between W^+/W^- and u_v/d_v comparable to the p_T-like observables. The largest correlations are achieved for the electron p_T and H_T. The S_T which is very similar to H_T, but focusses only on jets, does not provide additional gain in the correlation compared to H_T and has therefore not been measured in this thesis. W p_T and leading jet p_T are rather similar, while the second leading jet p_T shows the second largest correlation after the electron p_T in the range $x \sim 0.25 - 0.45$. Since the second leading jet is however simulated by MCFM only at LO, this is not considered too conclusive.

Two other interesting correlations are furthermore observed for the fiducial cross sections for $W^+ + \text{jet}$ and $W^- + \text{jet}$ production directly.

Figure 9.6 shows the averaged correlations of the $W^+ + \text{jet}$ and $W^- + \text{jet}$ cross sections to the gluon PDF. For $W^+ + \text{jet}$ production, very large, i.e. above 70 % correlation to the gluon PDF is observed for the W p_T distribution, and still about 60 % correlation for the leading and second leading jet p_T distributions. Interesting here is also the x-range which

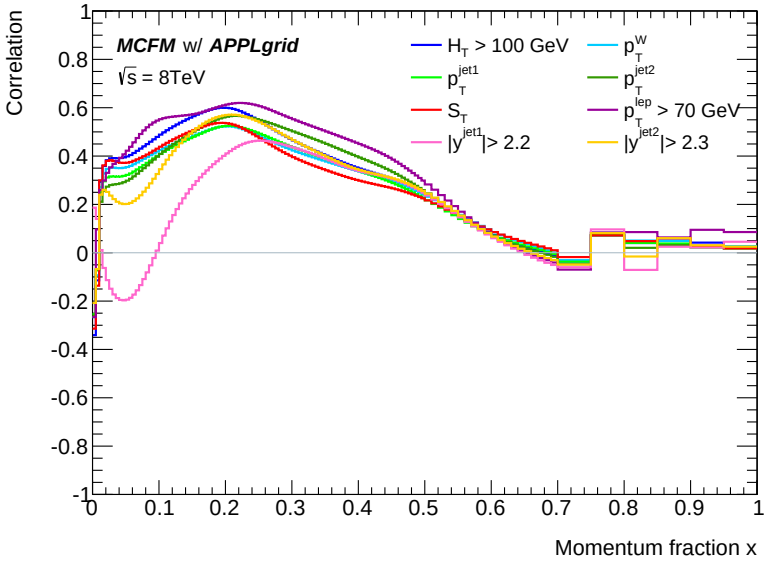


Figure 9.5: Correlation of the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio to the u_v/d_v valence quark ratio as a function of the momentum fraction x, averaged over observable values as indicated in the legend.

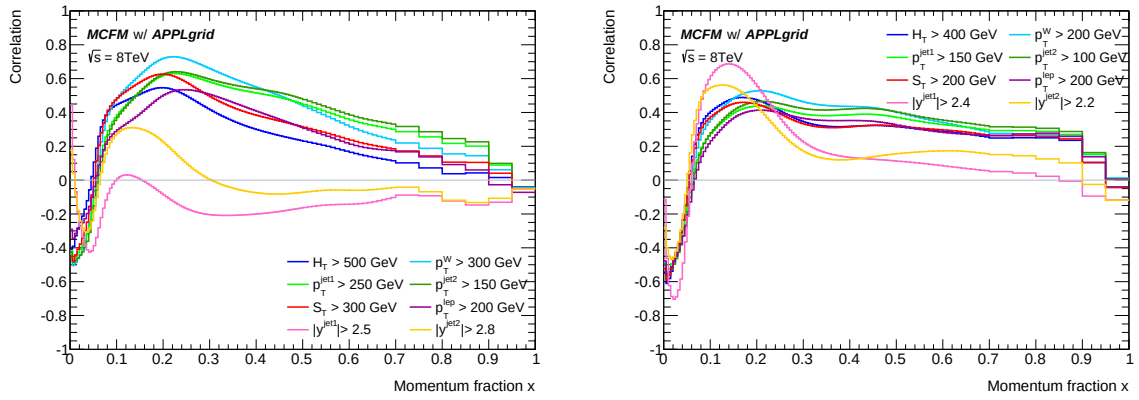
reaches from about 0.1 to high x values of around $0.5 - 0.6$ with reasonable $(40 - 50\,\%)$ correlation. In case of $W^- + \text{jet}$ production, the largest correlation of slightly below 70 % is observed for forward leading jet rapidities, but more within $x \sim 0.1 - 0.2$. For the p_T-like observables, the correlation of $W^- + \text{jet}$ production to the gluon PDF is reduced compared to $W^+ + \text{jet}$ production and only reaches around 50 % for the W p_T. In the W^+/W^- ratio, these correlations however mostly cancel.

The measurement of the absolute production cross section has the disadvantage that larger systematic uncertainties might limit the impact on the PDFs, but the relatively high x-values covered could still make this interesting for the extraction of the gluon PDF in PDF fits.

The other interesting feature is observed for the electron p_T distribution. As discussed in the previous section, polarization effects here lead to a very different behaviour of $W^+ + \text{jet}$ and $W^- + \text{jet}$ production as a function of the electron p_T. In terms of the correlations, this yields significant differences between W^+ and W^- with respect to the sea quark PDF. The sea quark PDF is defined as $\bar{u} + \bar{d} + \bar{s}$, and, due to subdominant contributions from subprocesses involving sea quarks, no major sensitivity of $W + \text{jets}$ production to this PDF is expected. Figure 9.7 presents the correlation of W^+ and W^- cross sections to the sea quark PDF as a function of the electron p_T and the momentum fraction x. As a result of the polarization effects, correlations of $W^+ + \text{jet}$ production to the sea quark PDF above 60 % are observed with the related x-range rising with increasing electron p_T. In a core area of $100 - 600\,\text{GeV}$ in electron p_T, the covered x-range is approximately 0.15 to 0.3. For $W^- + \text{jet}$ production, however, only a very localized small anti-correlation is observed, more in accordance with the naive expectation of no correlation. In the W^+/W^- ratio, the correlation then follows the pattern observed for $W^+ + \text{jet}$ production with the advantage of the cancellation of correlated uncertainties[1].

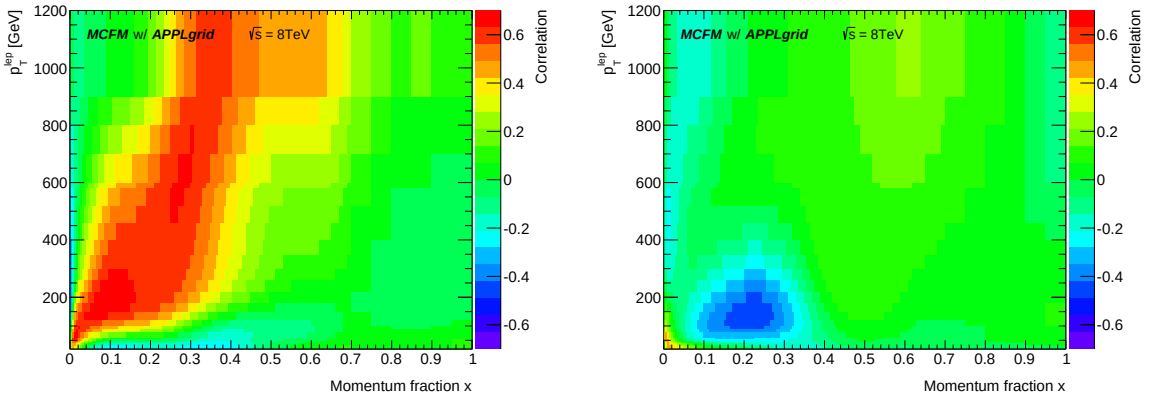
In general, good correlation of the W^+/W^- ratio to the valence quark ratio u_v/d_v, but also correlations to other PDFs are observed as a function of several observables in the targeted

[1]For technical reasons, the cross section as a function of electron p_T was not measured in this thesis, but will most certainly be included in the final publication of this analysis as a result of these studies.



(a) $W^+ + \text{jet}$ (b) $W^- + \text{jet}$

Figure 9.6: Correlation of $W^+ + \text{jet}$ (a) and $W^- + \text{jet}$ (b) production to the gluon PDF as a function of the momentum fraction x, averaged over observable values as indicated in the legend.



(a) $W^+ + \text{jet}$ (b) $W^- + \text{jet}$

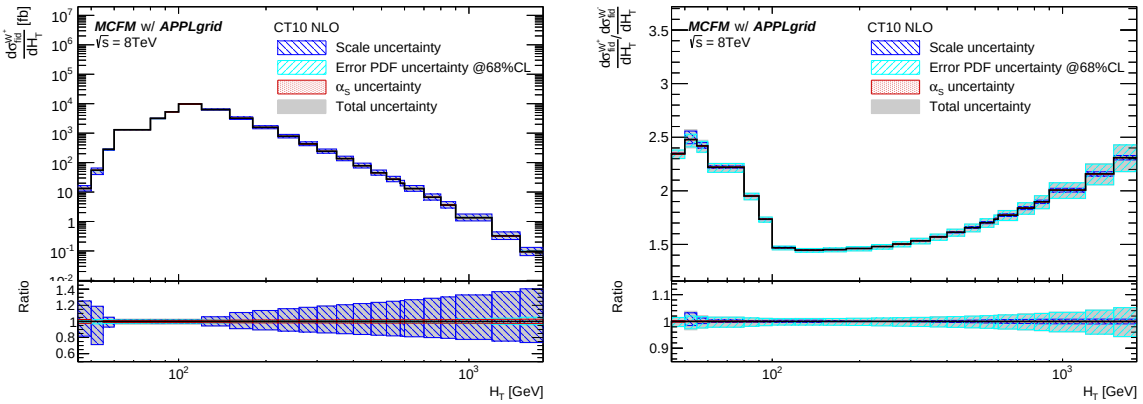
Figure 9.7: Correlation of $W^+ + \text{jet}$ (a) and $W^- + \text{jet}$ (b) production to the sea quark PDF as a function of the electron p_{T} and the momentum fraction x.

range of the momentum fraction of $x \sim 0.1-0.5$. As discussed in Section 2.3, the actual impact of the $W+\text{jets}$ measurement however also depends on the precision of this measurement and the precision and coverage of this x-range by other measurements. The shown correlations nevertheless demonstrate that the prerequisite of a potential impact is fulfilled.

9.1.3 Theoretical uncertainty in the W^+/W^- ratio

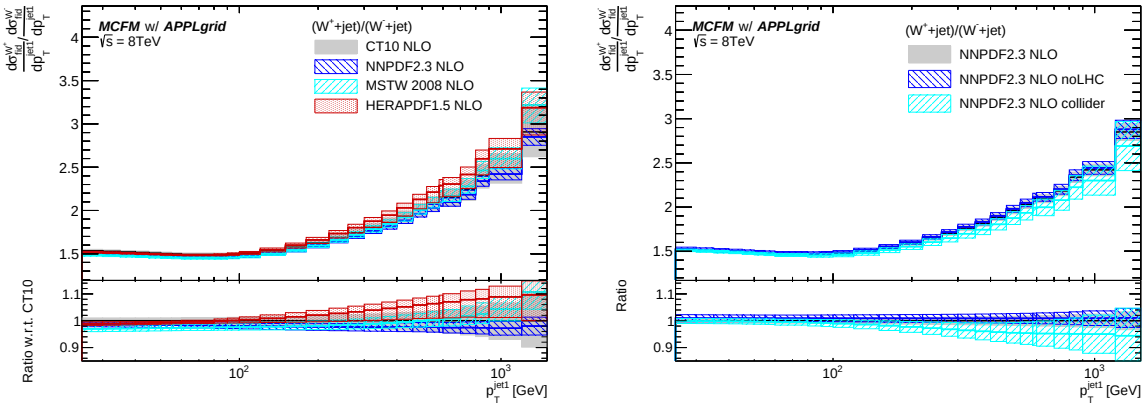
Theoretical uncertainties in the MCFM predictions coming from the different uncertainty components are investigated, in order to evaluate their relative contributions and the total size as a function of the studied observables.

The total theoretical uncertainty is calculated as the sum in quadrature of the scale, PDF and α_{S} uncertainties, described in Section 8.2.1. Figure 9.8 presents the fiducial cross section prediction for $W^+ + \text{jet}$ production and the W^+/W^- prediction as a function of H_{T}, using the CT10 NLO PDF set. The individual uncertainty components are displayed separately and are compared to the total theoretical uncertainty. The scale uncertainty clearly dominates



(a) $W^+ + \text{jet}$ (b) $(W^+ + \text{jet})/(W^- + \text{jet})$

Figure 9.8: Theoretical uncertainty in the $W^+ + \text{jet}$ cross section (a) and the W^+/W^- ratio (b) as a function of H_T, computed with MCFM and the CT10 NLO PDF set. The total uncertainty is the sum in quadrature of scale, PDF and α_S uncertainties.



(a) Different PDF sets (b) Different input data for NNPDF 2.3

Figure 9.9: Predictions of the W^+/W^- ratio as a function of the leading jet p_T for different PDF sets (a) and different versions of the NNPDF 2.3 PDF set (b).

the uncertainty in the cross section prediction. It grows up to $30 - 40\,\%$ at the highest H_T values. The PDF uncertainty at high H_T amounts to around $10\,\%$, and is roughly of the same size as the α_S uncertainty. The large scale uncertainty certainly indicates the need for the NNLO calculation of the differential cross section. It however also demonstrates that the used choice of scales: $\mu_R = \mu_F = m_W$ is probably not the best energy scale for W production with significant recoil from the emitted jet. For the W^+/W^- ratio, however, the scale uncertainties cancel almost completely, leaving the PDF uncertainties of about $5\,\%$ at the highest H_T as dominant uncertainty. The α_S uncertainty is almost negligible in the W^+/W^- ratio. Thus, the measurement of the W^+/W^- ratio provides input for a prediction limited by PDF uncertainties.

In addition to the split-up of the theoretical uncertainties, it is worth comparing W^+/W^- predictions obtained with different PDF sets and their respective uncertainties. This is shown in Figure 9.9 as a function of the leading jet p_T. Figure 9.9a compares the CT10 PDF set to predictions from the NNPDF 2.3 [40], the MSTW 2008 [53] and the HERAPDF 1.5

[139] NLO PDF sets. Uncertainties given here are the sum in quadrature of PDF and α_S uncertainties only. For MSTW 2008 and HERAPDF 1.5, the PDF uncertainty is computed using the asymmetric Hessian prescription with up and down variations of 20 and 10 eigenvectors, respectively, as explained for CT10 in Section 8.2.1. For the NNPDF 2.3 set, the standard deviation of the 100 PDF replicas is used as PDF uncertainty in accordance with the prescription for NNPDF uncertainties. The different PDF sets agree well within uncertainties with a few percent difference in the nominal value at low H_T and up to 10 % difference at high H_T. The largest difference of the tested PDF sets with respect to the CT10 PDF set is observed for HERAPDF 1.5 which uses only HERA data as input to the PDF fit. While this is a very consistent data set, it is only a subset of the data used by other PDF sets and differences to other PDF sets and larger PDF uncertainties are expected here to some extend. Interesting is furthermore the small uncertainty of the NNPDF 2.3 set. This PDF set is the only one among the displayed ones which includes some W, Z^0 and jet data from the LHC measurements using 2010 and early 2011 data[2]. The NNPDF set is furthermore the only one which does not assume an explicit parametrization of the x-dependence of the PDFs, but instead relies on neural networks as interpolating functions in the PDF fit.

The nominal NNPDF 2.3 set is moreover compared to two alternatives in Figure 9.9b where for the first no LHC data and for the second only collider data is considered in the PDF fit. The impact of the early LHC data is thus reflected in the difference of the *no-LHC* PDF to the nominal PDF set which yields a small offset over the entire jet p_T range, but approximately the same uncertainty. The *collider-only* version however clearly shows a trend to lower values of the W^+/W^- ratio at high leading jet p_T with much increased uncertainties. For the collider-only version, data is restricted to HERA, Tevatron and LHC results, leaving only 1212 data points for the NLO PDF fit, instead of 3482 as in the nominal fit [40]. For the updated NNPDF 3.0 set, a full collider-only PDF fit has not been re-attempted.

Further measurements at colliders, mainly at the LHC, are therefore needed to make the collider-only PDF competitive in performance with the nominal PDF, and permit the omission of fixed-target DIS data with not-up-to-date uncertainty treatment from global PDF fits.

9.2 Measurement of W^+/W^- ratios

Ratios of inclusive W^+ vs. W^- production have been measured by ATLAS at $\sqrt{s} = 7\,\text{TeV}$ [142] and 13 TeV [143], and by CMS at 7 TeV [144], 8 TeV [145] and 13 TeV [146]. As a function of the jet multiplicity, the W asymmetry has been measured by CMS based on $36\,\text{pb}^{-1}$ of data at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$, reaching up to 3 jets [9]. The measurement of W^+/W^- ratios in association with jets presented in this thesis combines both aproaches and is the first measurement of this kind at the LHC. The inclusive W^+/W^- ratio (≥ 0 jets) is moreover the first measurement of this value at 8 TeV with the ATLAS detector.

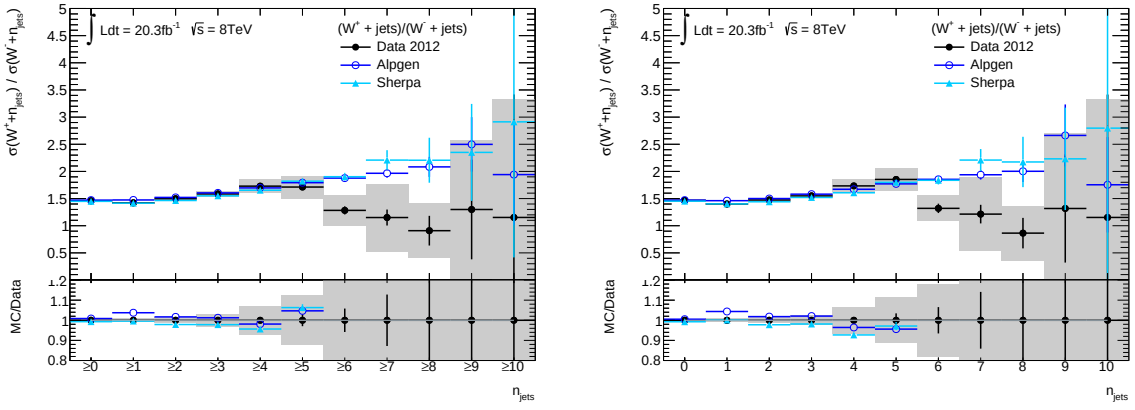
The W^+/W^- ratios are computed based on the measured fiducial cross sections presented in Chapter 8. The treatment of correlated and uncorrelated uncertainties in the W^+/W^- ratio is explained in Section 7.5. The uncertainty in the cross section from the luminosity measurement is identical for W^+ and W^- and cancels in the ratio exactly. The measured ratios are compared to the LO and NLO predictions: Alpgen, Sherpa and MCFM. Theoretical uncertainties in the NLO MCFM prediction are treated as correlated in the W^+/W^- ratio.

The measurement of the W^+/W^- ratio as a function of the jet multiplicity, $W^+ + (\geq)n$ jets vs. $W^- + (\geq)n$ jets, is presented in Section 9.2.1. Differential $(W^+ + \text{jet})/(W^- + \text{jet})$ ratios

[2]By now, updates of the shown PDF sets exist: CT14 [37], NNPDF 3.0 [140] and MMHT 2014 [141] including LHC data and HERAPDF 2.0 [139] based on the combined HERAI+II data.

are measured as a function of six observables and are shown in Section 9.2.2.

9.2.1 Measurement of W^+/W^- in association with n jets



(a) Inclusive n_{jets} (b) Exclusive n_{jets}

Figure 9.10: Measured ratios of W^+ to W^- cross sections in association with an inclusive (a) or an exclusive (b) number of jets. The measured W^+/W^- ratios are compared to predictions from Alpgen and Sherpa. The lower panel shows the relative size of the statistical uncertainties (black error bars) and the total systematic uncertainty (gray band) as well as the ratio of MC to data. Uncertainties in the LO MC predictions are statistical only.

The n_{jets}-dependent W^+/W^- ratio measurements are performed for inclusive as well as exclusive jet multiplicities separately. Figure 9.10 shows the ratios in comparison to predictions from the LO multi-leg generators Alpgen and Sherpa. The W^+/W^- ratio is measured to be approximately 1.5 at the low jet multiplicities, growing to about 1.7/1.8 at 5 jets and then dropping considerably for even higher jet multiplicities albeit with strongly increasing uncertainties. The trend at low jet multiplicities ($\leq$ 5 jets) is approximately described by Alpgen and Sherpa, but the drop above 5 jets in data is not followed by the MC generators who agree among themselves. At low jet multiplicities, very good cancellation of systematic uncertainties is observed. The total systematic uncertainty of approximately 10 % for W^+ and W^- cross section in association with 1 jet reduces to below 1 % in the ratio. The strongest disagreement between data and the predictions at low multiplicities is observed for Alpgen in association with 1 jet. It overpredicts the measured ratio in this n_{jets} bin by about 4 % – similar for inclusive and exclusive jet multiplicities.

The measured W^+/W^- ratios are listed in Table 9.1, rounded to two significant digits in the systematic uncertainty. Systematic uncertainties dominate the uncertainty in the measured W^+/W^- ratios.

For the ratio in association with at least one jet, the measured value is compared to the cross section ratio computed at NLO with MCFM.

$$r^{\text{NLO}} = (1.4354 \pm 0.0173\ (\text{stat})\ ^{+0.0004}_{-0.0009}\ (\text{PDF})\ ^{+0.0315}_{-0.0186}\ (\text{scale}))\,,$$

where the MC statistical uncertainty is treated as uncorrelated in the ratio, but the PDF and scale uncertainties are treated as correlated. For the PDF uncertainties, the CT10 NLO PDF set is used, considering only the combined PDF uncertainty values for $W^+ + \text{jet}$ or $W^- + \text{jet}$ in the ratio, not the individual eigenvector variations.

n_{jets}	Inclusive n_{jets}				Exclusive n_{jets}			
	r	$\pm$ Δr_{stat}	$\pm$	Δr_{syst}	r	$\pm$ Δr_{stat}	$\pm$	Δr_{syst}
0	1.4588	$\pm$ 0.0005	$\pm$	0.0050	1.4634	$\pm$ 0.0005	$\pm$	0.0061
1	1.421	$\pm$ 0.002	$\pm$	0.011	1.400	$\pm$ 0.002	$\pm$	0.011
2	1.496	$\pm$ 0.004	$\pm$	0.022	1.474	$\pm$ 0.004	$\pm$	0.020
3	1.585	$\pm$ 0.009	$\pm$	0.047	1.549	$\pm$ 0.010	$\pm$	0.036
4	1.73	$\pm$ 0.02	$\pm$	0.12	1.73	$\pm$ 0.03	$\pm$	0.11
5	1.71	$\pm$ 0.05	$\pm$	0.21	1.85	$\pm$ 0.06	$\pm$	0.21
6	1.28	$\pm$ 0.07	$\pm$	0.28	1.32	$\pm$ 0.09	$\pm$	0.24
7	1.15	$\pm$ 0.15	$\pm$	0.62	1.21	$\pm$ 0.17	$\pm$	0.67

Table 9.1: Measured cross section ratios, $r = (W^+ + n \text{ jets})/(W^- + n \text{ jets})$, at $\sqrt{s} = 8\,\text{TeV}$ for inclusive and exclusive jet multiplicities separately and including the statistical and systematic uncertainties, Δr_{stat} and Δr_{syst}, respectively. Luminosity uncertainties are correlated between W^+ and W^- and cancel in the ratio.

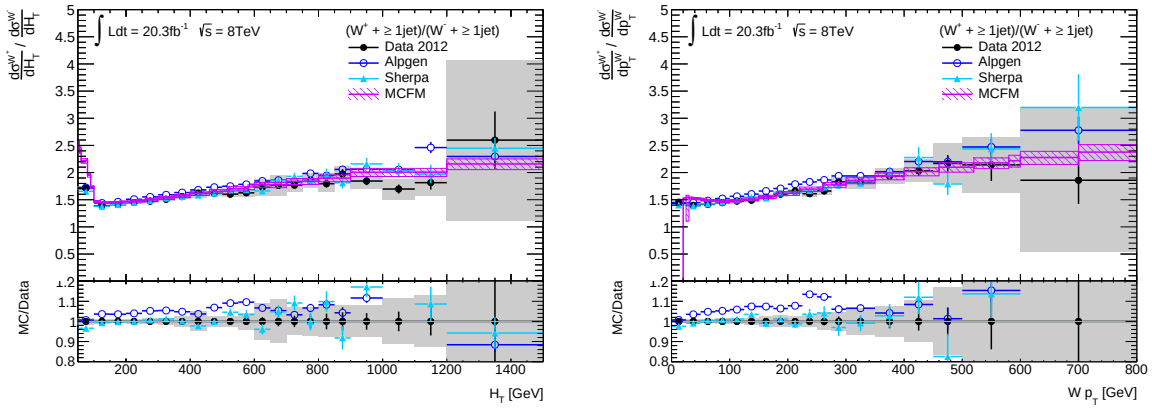
The NLO prediction for the ratio agrees within its uncertainties with the experimentally measured value of $r^{\text{exp}}_{\geq 1 \text{ jet}} = (1.421 \pm 0.011)$. The largest theoretical uncertainty is the uncertainty due to the scales μ_R and μ_F and is with about 2 % larger than the total measurement uncertainty. The PDF uncertainties cancel very well in the ratio, but are expected to grow according to the results in Section 9.1 with higher energy scale of the W^+ and W^- production, when also the required momentum fraction x for the W production increases.

9.2.2 Measurement of differential $(W^+ + \text{jet})/(W^- + \text{jet})$ ratios

Differential cross section ratios for W^++jet vs. W^-+jet production have been measured as a function of H_{T}, W p_{T}, leading and second leading jet p_{T} as well as leading and second leading jet rapidities. Systematic uncertainties in the ratio are treated as explained in Section 7.5. The measured W^+/W^- ratios are compared to LO multi-leg predictions from Alpgen and Sherpa as well as the NLO prediction from MCFM interfaced to Applgrid, as described in Section 8.2.1. The theoretical uncertainties in the MCFM prediction include scale, PDF and α_{S} uncertainties whose respective contributions were discussed in Section 9.1.3. As a result of the loss in statistics towards higher energy scales, the p_{T}-like distributions use a coarser binning in the high p_{T} tail than displayed in the cross section measurement in Section 8.2.3 and Section 8.2.4. Also the reach is limited to lower values than shown in the cross section measurement. In addition to the reduction of data statistics in particular for the W^- distributions, also the large increase in the systematic uncertainties in the high p_{T} tails makes these modifications necessary.

The measured W^+/W^- ratio in association with at least one jet as a function of H_{T} or the W p_{T} is presented in Figure 9.11. The ratio is about 1.5 at low H_{T} (>50 GeV) and W p_{T} and rises to around 2 for large H_{T} or slightly above 2 for high W p_{T}. The measured ratio agrees well with the predictions from Sherpa and MCFM (apart from the very low H_{T} and W p_{T} region where differences in phase space impact the MCFM prediction). In comparison with Alpgen, the global offset observed already for the W^+/W^- ratio as function of n_{jets} in the 1 jet bin is visible here as well, apart from the very first bin in H_{T} and W p_{T} where data and Alpgen agree. As a result of the statistics reduction at high H_{T} and W p_{T}, the measured range is reduced from 2 TeV to 1.5 TeV and from 1 TeV to 800 GeV, respectively.

The ratios as a function of leading and second leading jet p_{T} and rapidity are displayed in Figure 9.12. Again an increase of the ratio from approximately 1.5 to around 2 is observed



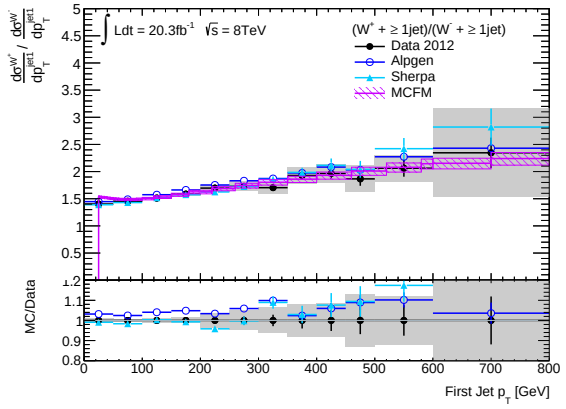
(a) H_{T} (b) W p_{T}

Figure 9.11: Measured differential W^+/W^- ratios in association with at least one jet as a function of H_{T} (a) and W p_{T} (b). The data is compared the LO predictions from Alpgen and Sherpa and the NLO prediction from MCFM. The lower panels show the relative size of statistical (black error bars) and systematic (gray band) uncertainties in the data as well as the ratio of MC to data.

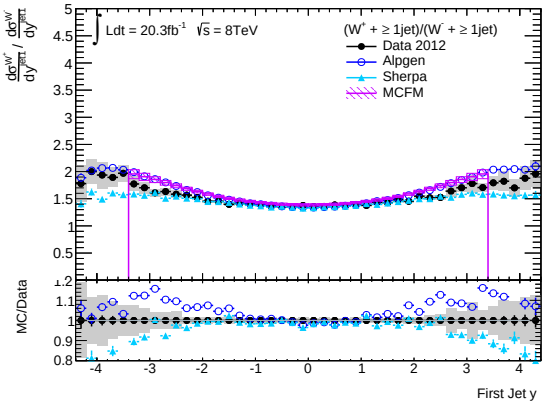
over the measured leading and second leading jet p_{T} ranges. In terms of rapidity, this change in the W^+/W^- ratio is observed when going from central to forward rapidities, although with reduced size in the case of the second leading jet rapidity. Regarding the jet p_{T}'s, Sherpa and MCFM predictions mostly agree with data, while Alpgen displays the already known offset. In terms of the leading jet rapidity, however, both Alpgen and Sherpa describe the W^+/W^- ratio at central rapidities, but at forward rapidities Alpgen overpredicts, while Sherpa underpredicts. MCFM follows the trend from Alpgen and overpredicts data for $|y| \gtrsim 2$ up to $|y| = 3.4$ to which the MCFM prediction was restricted. For the second leading jet rapidity, the data-MC agreement is improved in the forward region with a small overprediction from MCFM and Alpgen and a small underprediction from Sherpa. Due to reduced statistics and the increase of the systematic uncertainty, the measured range for the jet p_{T}'s was also lowered from 1.5 TeV to 800 GeV and from 1 TeV to 600 GeV for the leading and second leading jet p_{T}, respectively.

Measurements of the W^+/W^- ratio in association with at least two jets are performed in addition and are given in Appendix B.3. Reasonable agreement of data with Sherpa predictions is observed there as well. Alpgen displays a small slope in the ratio to data for the H_{T}, W p_{T} and leading jet p_{T} distributions, but the global offset is not present anymore – in line with the agreement between Alpgen and data in the W^+/W^- ratio as a function of n jets in the 2-jet bin.

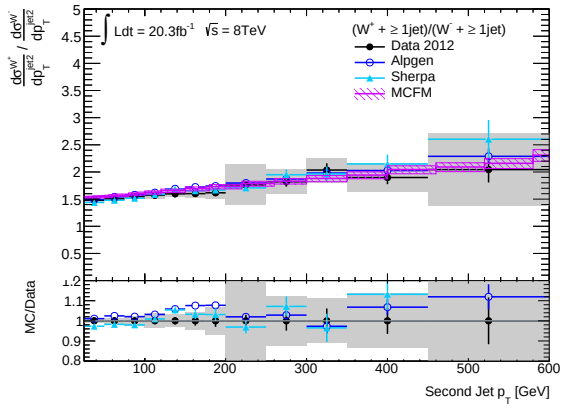
In general, a reasonable agreement of the predictions with the measured differential $W^+ +$ jet vs. $W^- +$ jet cross section ratios is observed. For higher energy scales, the systematic uncertainties in the measurement increase far above the theoretical NLO uncertainties, limiting the experimental impact of the high-p_{T} tails on pQCD predictions significantly.



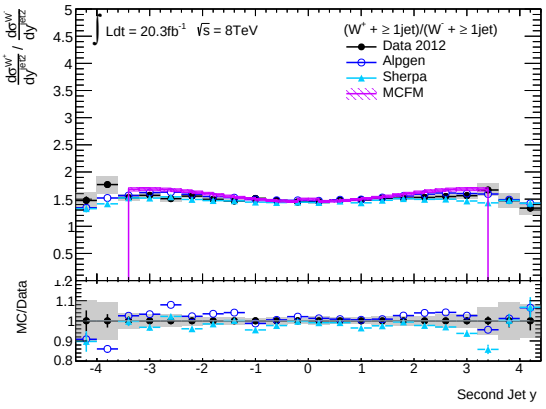
(a) Leading jet p_{T}



(b) Leading jet y



(c) Second leading jet p_{T}



(d) Second leading jet y

Figure 9.12: Measured differential W^+/W^- ratios in association with at least one jet as a function of the leading jet p_{T} and rapidity (a) and (b) as well as the second leading jet p_{T} and rapidity (c) and (d). Comparisons of data to predictions as well as the lower panel information are as in Figure 9.11.

CHAPTER 10

Summary

The production of W bosons is one of the most abundant processes in proton-proton (pp) collisions at the LHC. In association with jets, it serves as a precision test of perturbative quantum chromodynamics (pQCD), requiring higher order calculations as the energy scales probed in the measurement increase. In the charge-asymmetric initial state of pp-collisions at the LHC, the relative cross sections of W^+ and W^- production provide access to the valence quark composition of the proton, testing higher values of Bjorken x in the presence of jets than inclusive W asymmetry measurements at the LHC.

The production of W bosons in association with jets has been measured in this thesis using $20.3\,\text{fb}^{-1}$ of pp-collision data at a center-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. The measurement is performed in the decay channel $W \to e\nu$ and presents the first measurement of W + jets production at 8 TeV with the ATLAS detector.

Backgrounds to W + jets production arise mainly from multi-jet production at low jet multiplicities and from $t\bar{t}$ production at high jet multiplicities ($\geq$ 4 jets). The multi-jet background is estimated using a data-driven method, while the $t\bar{t}$ background and further sub-dominant electroweak backgrounds are determined from Monte Carlo simulation.

The background subtracted W + jets distributions in data are unfolded to particle level using an iterative Bayesian unfolding procedure. Differential distributions are unfolded in two dimensions: an observable and the jet multiplicity. Extensive tests are conducted to ensure the proper performance of the unfolding procedure.

The cross sections are measured as a function of the jet multiplicity for the charge-independent W production as well as W^+ and W^- production separately for up to 7 jets. Differential cross sections are presented in this thesis for six observables: H_T, W p_T, leading and second leading jet p_T and rapidity for W, W^+ and W^- production in association with at least one jet and at least two jets. The measured cross sections are compared to leading-order multi-leg and next-to-leading order predictions from Alpgen, Sherpa and MCFM. Good agreement to data is observed in most distributions, but a few modelling issues are seen in certain regions of phase space depending on the generator.

The relative measurement of W^++jets and W^-+jets cross sections considering correlations in the systematic uncertainties is presented as the cross section ratio $(W^+ + \text{jets})/(W^- + \text{jets})$. In the presence of at least one jet, this cross section ratio is sensitive to the ratio of the parton distribution functions (PDFs) of the valence quarks in the proton in a range of the momentum fraction $x \sim 0.1 - 0.5$. This range is inaccessible by W asymmetry measurements from ATLAS and CMS and so far mainly constrained by fixed-target deep inelastic scattering experiments and W asymmetry measurements at the Tevatron. Additional data as the $(W^+ + \text{jet})/(W^- + \text{jet})$ ratio measurement presented in this thesis provide essential input to global PDF fits.

Bibliography

[1] *ATLAS public results on SM measurements*, 2016, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/StandardModelPublicResults.

[2] J Butterworth et al., *Single Boson and Diboson Production Cross Sections in pp Collisions at* $\sqrt{s} = 7\,\text{TeV}$, tech. rep. ATL-COM-PHYS-2010-695, Geneva: CERN, Aug. 2010, URL: https://cds.cern.ch/record/1287902.

[3] CDF Collaboration, *Measurement of jet multiplicity in W events produced in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.8$ *TeV*, *Phys. Rev. Lett.* 70 (1993) 4042–4046, DOI: 10.1103/PhysRevLett.70.4042.

[4] CDF Collaboration, *Measurement of the cross section for W boson production in association with jets in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Rev.* D77 (2008) 011108, DOI: 10.1103/PhysRevD.77.011108, arXiv: 0711.4044 [hep-ex].

[5] DØ Collaboration, *Studies of W boson plus jets production in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Rev.* D88.9 (2013) 092001, DOI: 10.1103/PhysRevD.88.092001, arXiv: 1302.6508 [hep-ex].

[6] DØ Collaboration, *Measurements of inclusive W+jets production rates as a function of jet transverse momentum in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Lett.* B705 (2011) 200–207, DOI: 10.1016/j.physletb.2011.10.011, arXiv: 1106.1457 [hep-ex].

[7] ATLAS Collaboration, *Study of jets produced in association with a W boson in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, *Phys. Rev. D* 85 (2012) 092002, DOI: 10.1103/PhysRevD.85.092002, arXiv: 1201.1276 [hep-ex].

[8] ATLAS Collaboration, *Measurements of the W production cross sections in association with jets with the ATLAS detector*, *Eur. Phys. J. C* 75 (2015) 82, DOI: 10.1140/epjc/s10052-015-3262-7, arXiv: 1409.8639 [hep-ex].

[9] CMS Collaboration, *Jet Production Rates in Association with W and Z Bosons in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* 01 (2012) 010, DOI: 10.1007/JHEP01(2012)010, arXiv: 1110.3226 [hep-ex].

[10] CMS Collaboration, *Differential cross section measurements of W bosons produced in association with jets in proton-proton collisions at* $\sqrt{s} = 8$ TeV, tech. rep. CMS-PAS-SMP-14-023, Geneva: CERN, 2016, URL: https://cds.cern.ch/record/2147989.

[11] ATLAS Collaboration, *Search for the direct production of charginos, neutralinos and staus in final states with at least two hadronically decaying taus and missing transverse momentum in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* 10 (2014) 096, DOI: 10.1007/JHEP10(2014)096, arXiv: 1407.0350 [hep-ex].

[12] V. Lang and R. Stamen, *Precision Synchronization of the ATLAS Level-1 Calorimeter Trigger with Collision Data in 2010 and 2011*, tech. rep. ATL-DAQ-PUB-2012-001, Geneva: CERN, June 2012, URL: https://cds.cern.ch/record/1454684.

[13] V. Lang and R. Stamen, *A Fit Method for Input Signals of the ATLAS Level-1 Calorimeter Trigger*, tech. rep. ATL-DAQ-INT-2011-004, Geneva: CERN, Nov. 2011, URL: https://cds.cern.ch/record/1397021.

[14] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}$= 540 GeV*, *Phys. Lett.* B122 (1983), [,611(1983)] 103–116, DOI: 10.1016/0370-2693(83)91177-2.

[15] UA2 Collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider*, *Phys. Lett.* B122 (1983) 476–485, DOI: 10.1016/0370-2693(83)91605-2.

[16] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider*, *Phys. Lett.* B126 (1983) 398–410, DOI: 10.1016/0370-2693(83)90188-0.

[17] UA2 Collaboration, *Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider*, *Phys. Lett.* B129 (1983) 130–140, DOI: 10.1016/0370-2693(83)90744-X.

[18] CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* 74 (1995) 2626–2631, DOI: 10.1103/PhysRevLett.74.2626, arXiv: hep-ex/9503002 [hep-ex].

[19] DØ Collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* 74 (1995) 2632–2637, DOI: 10.1103/PhysRevLett.74.2632, arXiv: hep-ex/9503003 [hep-ex].

[20] DONUT Collaboration, *Observation of tau neutrino interactions*, *Phys. Lett.* B504 (2001) 218–224, DOI: 10.1016/S0370-2693(01)00307-0, arXiv: hep-ex/0012035 [hep-ex].

[21] 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, DOI: 10.1016/j.physletb.2012.08.020, arXiv: 1207.7214 [hep-ex].

[22] 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, DOI: 10.1016/j.physletb.2012.08.021, arXiv: 1207.7235 [hep-ex].

[23] Royal Swedish Academy of Science, *Press release from 8 October*, 2013, URL: http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/press.pdf.

[24] MissMJ, *Standard Model content*, Particle Data Group Author, PBS NOVA, Fermilab, Office of Science, United States Department of Energy, 2008, URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[25] K. A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* C38 (2014) 090001, DOI: 10.1088/1674-1137/38/9/090001.

[26] F. Halzen and A. D. Martin, *Quarks and leptons*, 1st ed., John Wiley and Sons, 1984, ISBN: 0471887412.

[27] M. Thomson, *Modern Particle Physics*, 1st ed., Cambridge University Press, 2013, ISBN: 9781107034266.

[28] C. S. Wu et al., *Experimental Test of Parity Conservation in Beta Decay*, *Phys. Rev.* 105 (4 Feb. 1957) 1413–1415, DOI: 10.1103/PhysRev.105.1413, URL: http://link.aps.org/doi/10.1103/PhysRev.105.1413.

[29] Bayes et al., *Experimental Constraints on Left-Right Symmetric Models from Muon Decay*, *Phys. Rev. Lett.* 106 (4 Jan. 2011) 041804, DOI: 10.1103/PhysRevLett.106.041804, URL: http://link.aps.org/doi/10.1103/PhysRevLett.106.041804.

[30] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl.Phys.* B126 (1977) 298, DOI: 10.1016/0550-3213(77)90384-4.

[31] 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.

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

[33] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev.* D82 (2010) 074024, DOI: 10.1103/PhysRevD.82.074024, arXiv: 1007.2241 [hep-ph].

[34] R. Boughezal et al., *W-Boson Production in Association with a Jet at Next-to-Next-to-Leading Order in Perturbative QCD*, *Phys. Rev. Lett.* 115 (6 Aug. 2015) 062002, DOI: 10.1103/PhysRevLett.115.062002, URL: http://link.aps.org/doi/10.1103/PhysRevLett.115.062002.

[35] C. F. Berger et al., *An Automated Implementation of On-Shell Methods for One-Loop Amplitudes*, *Phys. Rev.* D78 (2008) 036003, DOI: 10.1103/PhysRevD.78.036003, arXiv: 0803.4180 [hep-ph].

[36] H. Ita et al., *Precise Predictions for Z + 4 Jets at Hadron Colliders*, *Phys. Rev.* D85 (2012) 031501, DOI: 10.1103/PhysRevD.85.031501, arXiv: 1108.2229 [hep-ph].

[37] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* D93.3 (2016) 033006, DOI: 10.1103/PhysRevD.93.033006, arXiv: 1506.07443 [hep-ph].

[38] S. A. Malik and G. Watt, *Ratios of W and Z cross sections at large boson p_T as a constraint on PDFs and background to new physics*, *JHEP* 02 (2014) 025, DOI: 10.1007/JHEP02(2014)025, arXiv: 1304.2424 [hep-ph].

[39] C.-H. Kom and W. J. Stirling, *Charge asymmetry in W + jets production at the LHC*, *Eur. Phys. J.* C69 (2010) 67–73, DOI: 10.1140/epjc/s10052-010-1353-z, arXiv: 1004.3404 [hep-ph].

[40] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* B867 (2013) 244–289, DOI: 10.1016/j.nuclphysb.2012.10.003, arXiv: 1207.1303 [hep-ph].

[41] CMS Collaboration, *Measurement of the differential cross section and charge asymmetry for inclusive pp to W + X production at $\sqrt{s}$ = 8 TeV* (2016), arXiv: 1603.01803 [hep-ex].

[42] S. Catani et al., *QCD matrix elements + parton showers*, *JHEP* 11 (2001) 063, DOI: 10.1088/1126-6708/2001/11/063, arXiv: hep-ph/0109231 [hep-ph].

[43] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* 08 (2002) 015, DOI: 10.1088/1126-6708/2002/08/015, arXiv: hep-ph/0205283 [hep-ph].

[44] M. L. Mangano, M. Moretti, and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study*, *Nucl. Phys.* B632 (2002) 343–362, DOI: 10.1016/S0550-3213(02)00249-3, arXiv: hep-ph/0108069 [hep-ph].

[45] B. Andersson, *The Lund Model*, Part of Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 2005, ISBN: 9780521017343.

[46] T. D. Gottschalk, *An Improved Description of Hadronization in the QCD Cluster Model for e^+e^- Annihilation*, *Nucl. Phys.* B239 (1984) 349–381, DOI: 10.1016/0550-3213(84)90253-0.

[47] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, *Nucl. Phys.* B238 (1984) 492–528, DOI: 10.1016/0550-3213(84)90333-X.

[48] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* 504 (2011) 145–233, DOI: 10.1016/j.physrep.2011.03.005, arXiv: 1101.2599 [hep-ph].

[49] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* C74 (2014) 2916, DOI: 10.1140/epjc/s10052-014-2916-1, arXiv: 1405.1160 [hep-ex].

[50] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* 07 (2003) 001, DOI: 10.1088/1126-6708/2003/07/001, arXiv: hep-ph/0206293 [hep-ph].

[51] T. Sjostrand et al., *High-energy physics event generation with PYTHIA 6.1*, *Comput. Phys. Commun.* 135 (2001) 238–259, DOI: 10.1016/S0010-4655(00)00236-8, arXiv: hep-ph/0010017 [hep-ph].

[52] S. Catani et al., *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* 103 (2009) 082001, DOI: 10.1103/PhysRevLett.103.082001, arXiv: 0903.2120 [hep-ph].

[53] A. D. Martin et al., *Parton distributions for the LHC*, *Eur. Phys. J.* C63 (2009) 189–285, DOI: 10.1140/epjc/s10052-009-1072-5, arXiv: 0901.0002 [hep-ph].

[54] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* 0902 (2009) 007, DOI: 10.1088/1126-6708/2009/02/007, arXiv: 0811.4622 [hep-ph].

[55] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* 0812 (2008) 039, DOI: 10.1088/1126-6708/2008/12/039, arXiv: 0808.3674 [hep-ph].

[56] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* 0803 (2008) 038, DOI: 10.1088/1126-6708/2008/03/038, arXiv: 0709.1027 [hep-ph].

[57] S. Höche et al., *QCD matrix elements and truncated showers*, *JHEP* 0905 (2009) 053, DOI: 10.1088/1126-6708/2009/05/053, arXiv: 0903.1219 [hep-ph].

[58] J.-C. Winter, F. Krauss, and G. Soff, *A Modified cluster hadronization model*, *Eur. Phys. J.* C36 (2004) 381–395, DOI: 10.1140/epjc/s2004-01960-8, arXiv: hep-ph/0311085 [hep-ph].

[59] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC*, *Nuclear Physics B - Proceedings Supplements* 205-206 (2010), Loops and Legs in Quantum Field Theory, Proceedings of the 10th DESY Workshop on Elementary Particle Theory 10–15, ISSN: 0920-5632, DOI: 10.1016/j.nuclphysbps.2010.08.011.

[60] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* 1006 (2010) 043, DOI: 10.1007/JHEP06(2010)043, arXiv: 1002.2581 [hep-ph].

[61] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* 0411 (2004) 040, DOI: 10.1088/1126-6708/2004/11/040, arXiv: hep-ph/0409146 [hep-ph].

[62] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* 0711 (2007) 070, DOI: 10.1088/1126-6708/2007/11/070, arXiv: 0709.2092 [hep-ph].

[63] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* 185 (2014) 2930, DOI: 10.1016/j.cpc.2014.06.021, arXiv: 1112.5675 [hep-ph].

[64] G. Corcella et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* 0101 (2001) 010, DOI: 10.1088/1126-6708/2001/01/010, arXiv: hep-ph/0011363 [hep-ph].

[65] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* 70 (2010) 823, DOI: 10.1140/epjc/s10052-010-1429-9, arXiv: 1005.4568 [hep-ex].

[66] S. Agostinelli et al., *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth. A* 506 (2003) 250–303, DOI: 10.1016/S0168-9002(03)01368-8.

[67] E. Rutherford, *The Scattering of α and β Particles by Matter and the Structure of the Atom*, *Philosophical Magazine. Series 6* 21 (1911) 669–688.

[68] O. S. Brüning et al., *LHC Design Report*, Geneva: CERN, 2004, URL: https://cds.cern.ch/record/782076.

[69] *ATLAS Collaboration Webpage*, 2016, URL: http://www.atlas.ch/.

[70] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* 3 (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[71] T. Schörner-Sadenius, ed., *The Large Hadron Collider - Harvest of Run 1*, 1st ed., Springer International Publishing, 2015, ISBN: 9783319150017.

[72] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC*, ATLAS-PERF-2015-01-001, To be published 2016, URL: https://cds.cern.ch/record/2140276.

[73] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, *JINST* 9 (2014) P08009, DOI: 10.1088/1748-0221/9/08/P08009, arXiv: 1404.7473 [hep-ex].

[74] M. Lamont, *The LHC's first long run*, CERN Courier, 2016, URL: http://cerncourier.com/cws/article/cern/54381.

[75] E. Eisenhandler et al., *The ATLAS Level-1 Calorimeter Trigger*, *JINST **3** P03001* (2008).

[76] V. Lang, *Performance and Improvements of the ATLAS Jet Trigger System*, *Proceedings of the IEEE 2012 conference, Anaheim, USA, 2012, N14-230*, 2012 1406–1409, DOI: 10.1109/NSSMIC.2012.6551342.

[77] ATLAS Collaboration, *ATLAS Computing: technical design report*, Technical Design Report ATLAS, Geneva: CERN, 2005, URL: https://cds.cern.ch/record/837738.

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

[79] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* 74 (2014) 3071, DOI: 10.1140/epjc/s10052-014-3071-4, arXiv: 1407.5063 [hep-ex].

[80] ATLAS Collaboration, *Reconstruction and Performance of Missing Transverse Momentum in the ATLAS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV*, ATLAS-PERF-2014-04-002, To be published 2016, URL: https://cds.cern.ch/record/2144372.

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

[82] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-017, 2015, URL: http://cdsweb.cern.ch/record/2008678.

[83] ATLAS Collaboration, *Determination of the jet energy scale and resolution at ATLAS using Z/γ-jet events in data at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-057, 2015, URL: http://cdsweb.cern.ch/record/2059846.

[84] ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-002, 2015, URL: http://cdsweb.cern.ch/record/2001682.

[85] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, Geneva: CERN, Apr. 2008, URL: https://cds.cern.ch/record/1099735.

[86] *Expected electron performance in the ATLAS experiment*, tech. rep. ATL-PHYS-PUB-2011-006, Geneva: CERN, Apr. 2011, URL: https://cds.cern.ch/record/1345327.

[87] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton–proton collision data*, *Eur. Phys. J. C* 72 (2012) 1909, DOI: 10.1140/epjc/s10052-012-1909-1, arXiv: 1110.3174 [hep-ex].

[88] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data*, *Eur. Phys. J. C* 74 (2014) 2941, DOI: 10.1140/epjc/s10052-014-2941-0, arXiv: 1404.2240 [hep-ex].

[89] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* 73 (2013) 2304, DOI: 10.1140/epjc/s10052-013-2304-2, arXiv: 1112.6426 [hep-ex].

[90] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J. C* 75 (2015) 17, DOI: 10.1140/epjc/s10052-014-3190-y, arXiv: 1406.0076 [hep-ex].

[91] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k_t jet clustering algorithm*, *JHEP* 04 (2008) 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv: 0802.1189 [hep-ph].

[92] G. P. Salam, *Towards Jetography*, *Eur. Phys. J.* C67 (2010) 637–686, DOI: 10.1140/epjc/s10052-010-1314-6, arXiv: 0906.1833 [hep-ph].

[93] S. Catani et al., *New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation*, *Phys. Lett.* B269 (1991) 432–438, DOI: 10.1016/0370-2693(91)90196-W.

[94] S. Catani et al., *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* B406 (1993) 187–224, DOI: 10.1016/0550-3213(93)90166-M.

[95] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* D48 (1993) 3160–3166, DOI: 10.1103/PhysRevD.48.3160, arXiv: hep-ph/9305266 [hep-ph].

[96] Y. L. Dokshitzer et al., *Better jet clustering algorithms*, *JHEP* 08 (1997) 001, DOI: 10.1088/1126-6708/1997/08/001, arXiv: hep-ph/9707323 [hep-ph].

[97] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998, arXiv: hep-ph/9907280 [hep-ph].

[98] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett.* B641 (2006) 57–61, DOI: 10.1016/j.physletb.2006.08.037, arXiv: hep-ph/0512210 [hep-ph].

[99] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* C72 (2012) 1896, DOI: 10.1140/epjc/s10052-012-1896-2, arXiv: 1111.6097 [hep-ph].

[100] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* B659 (2008) 119–126, DOI: 10.1016/j.physletb.2007.09.077, arXiv: 0707.1378 [hep-ph].

[101] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC*, *Eur. Phys. J.* C73.3 (2013) 2305, DOI: 10.1140/epjc/s10052-013-2305-1, arXiv: 1203.1302 [hep-ex].

[102] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, *JHEP* 04 (2008) 005, DOI: 10.1088/1126-6708/2008/04/005, arXiv: 0802.1188 [hep-ph].

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

[104] *ATLAS public results on the total data-taking luminosities*, 2016, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[105] M Baak et al., *Data Quality Status Flags and Good Run Lists for Physics Analysis in ATLAS*, tech. rep. ATL-COM-GEN-2009-015, Geneva: CERN, Mar. 2009, URL: https://cds.cern.ch/record/1168026.

[106] *Good Run List for full detector operation in 2012*, 2016, URL: https://atlasdqm.web.cern.ch/atlasdqm/grlgen/All_Good/data12_8TeV.periodAllYear_DetStatus-v61-pro14-02_DQDefects-00-01-00_PHYS_StandardGRL_All_Good.xml.

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

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

[109] ATLAS Collaboration, *Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton–Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2010-038, 2010, URL: http://cdsweb.cern.ch/record/1277678.

[110] S. Laplace and J. B. de Vivie, *Calorimeter isolation and pile-up*, tech. rep. ATL-COM-PHYS-2012-467, Geneva: CERN, May 2012, URL: https://cds.cern.ch/record/1444890.

[111] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, 2014, URL: http://cdsweb.cern.ch/record/1664335.

[112] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, 2014, URL: http://cdsweb.cern.ch/record/1741020.

[113] *ATLAS Luminosity Calculator*, 2016, URL: https://atlas-lumicalc.cern.ch/.

[114] L. Lyons, *Statistics for nuclear and particle physicists*, Reprinted edition 1992, Cambridge: Cambridge University Press, 1986.

[115] G. Cowan, *Statistical Data Analysis*, New York: Oxford University Press, 1998.

[116] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, tech. rep. physics/0306116, Stanford, CA: SLAC, June 2003, URL: https://cds.cern.ch/record/622147.

[117] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, *Computer Physics Communications* 10.6 (1975) 343–367, ISSN: 0010-4655, DOI: 10.1016/0010-4655(75)90039-9.

[118] *ROOT Data Analysis Framework*, CERN, 2016, URL: https://root.cern.ch/.

[119] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV*, tech. rep. ATL-PHYS-PUB-2015-002, Geneva: CERN, Jan. 2015, URL: http://cds.cern.ch/record/1981319.

[120] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, *Nucl. Instr. Meth.* A362 (1995) 487–498.

[121] G. D'Agostini, *Improved iterative Bayesian unfolding* (2010), arXiv: 1010.0632v1 [physics.data-an].

[122] T. Adye, *Unfolding algorithms and tests using RooUnfold*, *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318*, 2011 313–318, arXiv: 1105.1160 [physics.data-an].

[123] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, Technical Design Report ATLAS, Geneva: CERN, 1999, URL: https://cds.cern.ch/record/391176.

[124] J. Ferrando and D. Wendland, *Reference $t\bar{t}$ production cross sections for use in ATLAS analyses*, tech. rep. ATL-COM-PHYS-2014-112, Geneva: CERN, Feb. 2014, URL: https://cds.cern.ch/record/1662536.

[125] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, *Phys. Rev.* D81 (2010) 054028, DOI: 10.1103/PhysRevD.81.054028, arXiv: 1001.5034 [hep-ph].

[126] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* D83 (2011) 091503, DOI: 10.1103/PhysRevD.83.091503, arXiv: 1103.2792 [hep-ph].

[127] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, *Phys. Rev.* D82 (2010) 054018, DOI: 10.1103/PhysRevD.82.054018, arXiv: 1005.4451 [hep-ph].

[128] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8TeV$ using the ATLAS detector at the LHC*, ATLAS-DAPR-2013-01-001, To be published 2016, URL: https://cds.cern.ch/record/2133046.

[129] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, *Phys. Rev.* D85 (2012) 034011, DOI: 10.1103/PhysRevD.85.034011, arXiv: 1107.2117 [hep-ph].

[130] M. R. Whalley, D. Bourilkov, and R. C. Group, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, *HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B*, 2005, arXiv: hep-ph/0508110 [hep-ph].

[131] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur. Phys. J.* C66 (2010) 503–524, DOI: 10.1140/epjc/s10052-010-1255-0, arXiv: 0911.2985 [hep-ph].

[132] G. P. Salam and J. Rojo, *A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)*, *Comput. Phys. Commun.* 180 (2009) 120–156, DOI: 10.1016/j.cpc.2008.08.010, arXiv: 0804.3755 [hep-ph].

[133] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* G43 (2016) 023001, DOI: 10.1088/0954-3899/43/2/023001, arXiv: 1510.03865 [hep-ph].

[134] R. Boughezal, X. Liu, and F. Petriello, *A comparison of NNLO QCD predictions with 7 TeV ATLAS and CMS data for V+jet processes* (2016), arXiv: 1602.05612 [hep-ph].

[135] ATLAS Collaboration, *Measurement of the transverse momentum distribution of W bosons in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev. D* 85 (1 Jan. 2012) 012005, DOI: 10.1103/PhysRevD.85.012005.

[136] CMS Collaboration, *Measurement of the transverse momentum distribution of* W *bosons in* pp *collisions at $\sqrt{s} = 8$ TeV*, CMS-PAS-SMP-13-006 (2014), URL: https://cds.cern.ch/record/1951372.

[137] Z. Bern et al., *Left-Handed W Bosons at the LHC*, *Phys. Rev.* D84 (2011) 034008, DOI: 10.1103/PhysRevD.84.034008, arXiv: 1103.5445 [hep-ph].

[138] R. D. Ball et al., *A Determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* B809 (2009), [Erratum: Nucl. Phys.B816,293(2009)] 1–63, DOI: 10.1016/j.nuclphysb.2008.09.037,10.1016/j.nuclphysb.2009.02.027, arXiv: 0808.1231 [hep-ph].

[139] H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J.* C75.12 (2015) 580, DOI: 10.1140/epjc/s10052-015-3710-4, arXiv: 1506.06042 [hep-ex].

[140] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* 04 (2015) 040, DOI: 10.1007/JHEP04(2015)040, arXiv: 1410.8849 [hep-ph].

[141] L. A. Harland-Lang et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* C75.5 (2015) 204, DOI: 10.1140/epjc/s10052-015-3397-6, arXiv: 1412.3989 [hep-ph].

[142] ATLAS Collaboration, *Measurement of the inclusive $W^{\pm}$ and Z/γ^{*} cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev.* D85 (2012) 072004, DOI: 10.1103/PhysRevD.85.072004, arXiv: 1109.5141 [hep-ex].

[143] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector* (2016), arXiv: 1603.09222 [hep-ex].

[144] CMS Collaboration, *Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* 10 (2011) 132, DOI: 10.1007/JHEP10(2011)132, arXiv: 1107.4789 [hep-ex].

[145] CMS Collaboration, *Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev. Lett.* 112 (2014) 191802, DOI: 10.1103/PhysRevLett.112.191802, arXiv: 1402.0923 [hep-ex].

[146] CMS Collaboration, *Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV*, CMS-PAS-SMP-15-004 (2015), URL: https://cds.cern.ch/record/2093537.

APPENDIX A

Supplementary material

A.1 Multi-jet estimate for W^+ and W^- selections

The fitted multi-jet fractions extrapolated to the W + jets signal regions are listed in Table A.1 for events with positively charged electron (positron) and in Table A.2 for events with negatively charged electron.

n_{jets}	$f_{\text{MJ,SR}}$	$\Delta f^{\text{stat}}_{\text{MJ,SR}}$	$\Delta f^{\text{syst}}_{\text{MJ,SR}}$	$\frac{\Delta f^{\text{stat}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$[%]	$\frac{\Delta f^{\text{syst}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$[%]
0	2.71	0.02	0.07	0.63	2.60
1	7.50	0.05	0.59	0.69	7.81
2	13.01	0.15	0.49	1.12	3.78
3	13.99	0.32	0.56	2.30	3.98
4	13.21	0.66	0.83	4.98	6.32
5	12.43	1.35	0.84	10.88	6.79
≥6	16.28	2.91	3.26	17.87	20.05

Table A.1: Fraction of multi-jet events, $f_{\text{MJ,SR}}$, in % per fitted jet multiplicity in the signal region for the W^+ selection. The combined statistical and systematic uncertainty, $\Delta f^{\text{stat}}_{\text{MJ,SR}}$ and $\Delta f^{\text{syst}}_{\text{MJ,SR}}$, as well as their fractional size with respect to the determined multi-jet fraction are listed in addition.

n_{jets}	$f_{\text{MJ,SR}}$	$\Delta f^{\text{stat}}_{\text{MJ,SR}}$	$\Delta f^{\text{syst}}_{\text{MJ,SR}}$	$\frac{\Delta f^{\text{stat}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$[%]	$\frac{\Delta f^{\text{syst}}_{\text{MJ,SR}}}{f_{\text{MJ,SR}}}$[%]
0	4.19	0.02	0.12	0.48	2.86
1	9.41	0.07	0.78	0.71	8.33
2	16.67	0.20	0.60	1.18	3.59
3	17.88	0.46	0.71	2.58	3.95
4	19.60	0.93	0.56	4.75	2.88
5	20.00	2.25	1.75	11.25	8.77
≥6	8.95	3.24	1.73	36.23	19.30

Table A.2: Fraction of multi-jet events, $f_{\text{MJ,SR}}$, in % per fitted jet multiplicity in the signal region for the W^- selection. The combined statistical and systematic uncertainty, $\Delta f^{\text{stat}}_{\text{MJ,SR}}$ and $\Delta f^{\text{syst}}_{\text{MJ,SR}}$, as well as their fractional size with respect to the determined multi-jet fraction are listed in addition.

A.2 Detector level agreement for W^+ and W^- selections

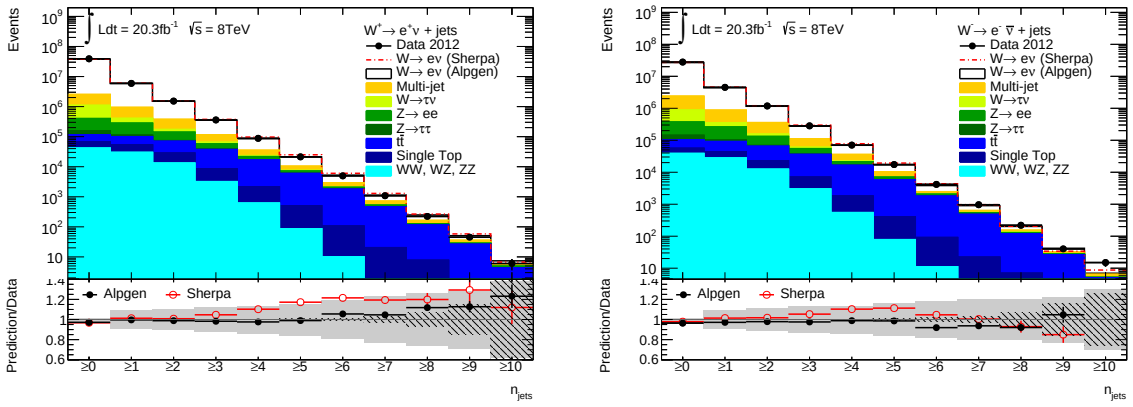
The signal and background composition for the W^+ and W^- selections in the W + jets SR are listed in Table A.3 and Table A.4, respectively, for inclusive jet multiplicities. The jet multiplicity distribution is shown in Figure A.1 for both W^+ and W^- selections, while Figure A.2 and Figure A.3 present several kinematic distributions for W^+ and W^- for ≥ 1 jet and Figure A.4 and Figure A.5 for ≥ 2 jets.

	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
WW, WZ, ZZ	0.1	0.5	0.9	0.9	0.7	0.4	0.2	0.1
Single t	0.1	0.4	1.0	1.5	1.8	2.0	1.8	1.7
$t\bar{t}$	0.1	0.8	2.7	8.1	16.8	25.6	31.7	38.6
$Z \to \tau\tau$	0.1	0.2	0.3	0.3	0.3	0.3	0.2	0.3
$Z \to ee$	0.6	2.8	4.4	5.3	5.5	5.4	4.9	5.1
$W \to \tau\nu$	1.8	2.1	2.1	2.1	2.0	1.8	1.5	1.4
Multi-jet	3.8	9.0	13.4	14.0	13.5	13.4	15.3	15.4
$W \to e\nu$	93.4	84.2	75.3	67.7	59.4	51.1	44.3	37.4

Table A.3: Composition of signal and background processes in the W +jets signal region for the W^+ selection, as expected from MC simulation and the data-driven multi-jet estimate. The numbers are in % with respect to the total SM prediction and are listed for different inclusive jet multiplicities. For the signal numbers, the Alpgen prediction has been used.

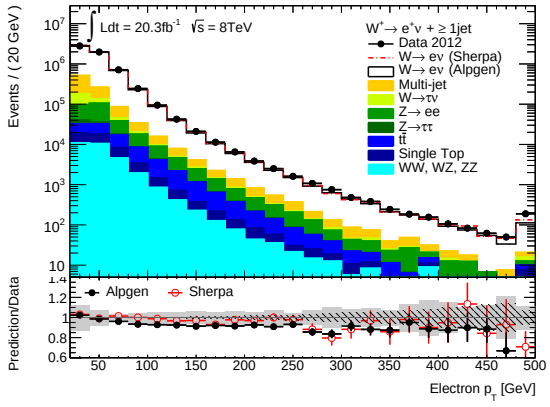
	Inclusive	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets	≥ 5 jets	≥ 6 jets	≥ 7 jets
WW, WZ, ZZ	0.1	0.6	1.1	1.1	0.8	0.5	0.3	0.2
Single t	0.1	0.4	0.9	1.5	1.8	1.9	2.1	2.2
$t\bar{t}$	0.2	1.1	3.6	10.4	20.7	31.4	43.6	50.2
$Z \to \tau\tau$	0.1	0.3	0.3	0.4	0.3	0.4	0.4	0.3
$Z \to ee$	0.8	3.6	5.4	6.4	6.6	6.4	6.7	6.2
$W \to \tau\nu$	1.9	2.0	2.0	1.9	1.7	1.5	1.3	1.1
Multi-jet	5.6	11.7	17.4	18.6	19.2	17.5	9.7	9.5
$W \to e\nu$	91.2	80.3	69.3	59.8	48.9	40.4	35.9	30.2

Table A.4: Composition of signal and background processes in the W +jets signal region for the W^- selection, as expected from MC simulation and the data-driven multi-jet estimate. The numbers are in % with respect to the total SM prediction and are listed for different inclusive jet multiplicities. For the signal numbers, the Alpgen prediction has been used.

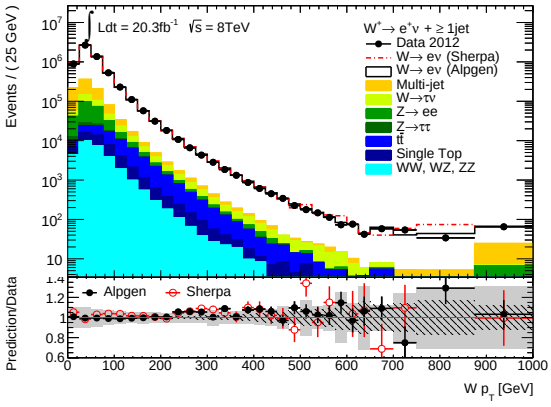


(a) W^+ (b) W^-

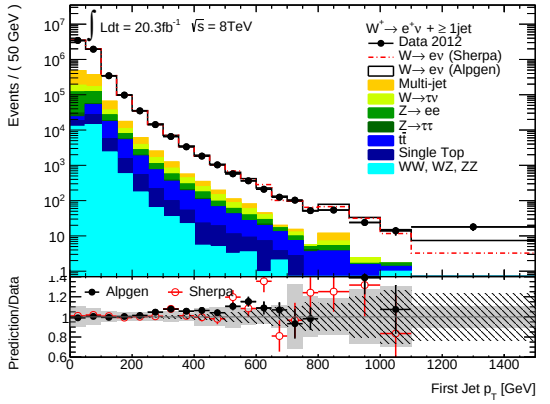
Figure A.1: Composition of signal and background in the W + jets signal region for the different inclusive jet multiplicities for the W^+ (a) and W^- (b) selections separately. The lower panels show the ratio of the total prediction to data, where the total prediction has been computed using two different MC generators for the W + jets signal. The gray and hashed bands are defined as explained in Figure 5.7.



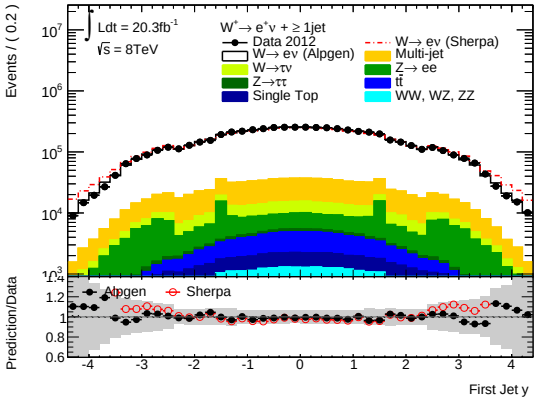
(a) Electron p_{T}



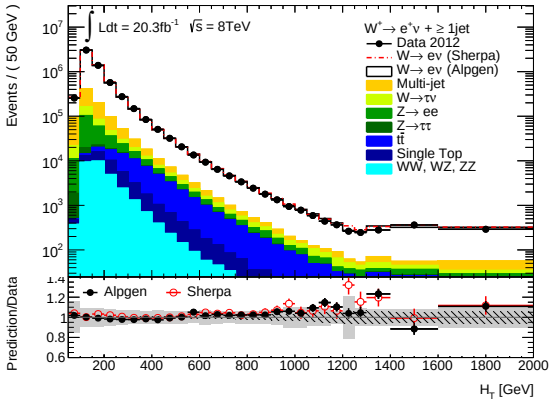
(b) W p_{T}



(c) Leading jet p_{T}

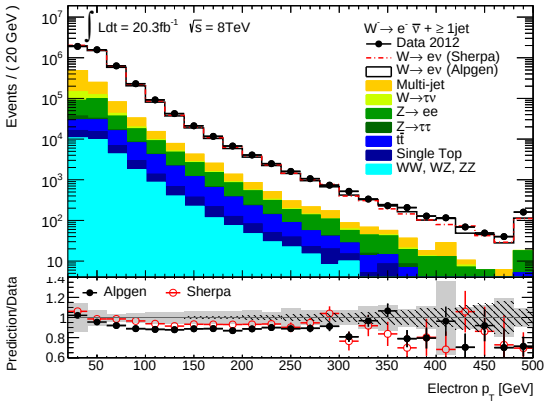


(d) Leading jet rapidity

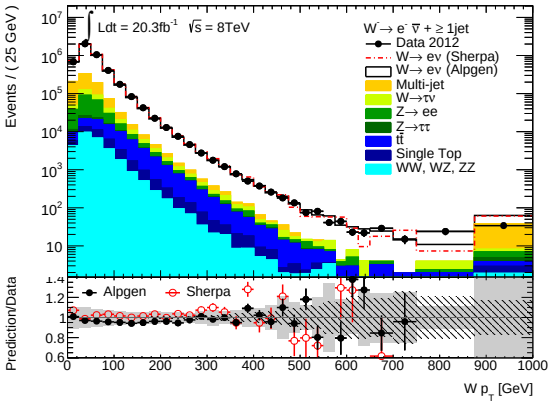


(e) H_{T}

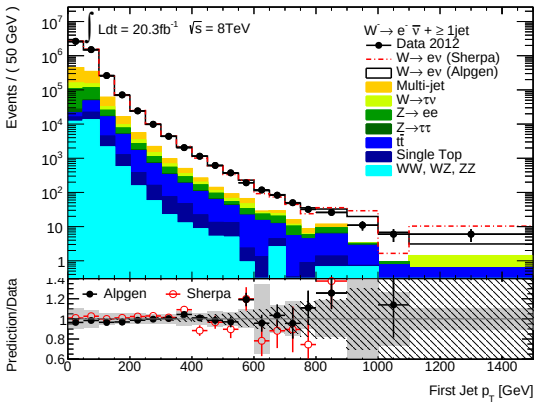
Figure A.2: Kinematic distributions comparing data and the total prediction for the W^+ selection in the $W+$jets signal region for ≥ 1 jet: Electron p_{T} (a), W p_{T} (b), leading jet p_{T} (c) and rapidity (d) as well as H_{T} (e). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the $W+$jets signal. The gray and hashed bands are defined as explained in Figure 5.7.



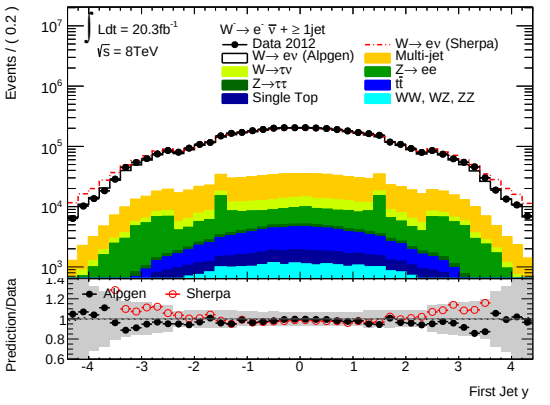
(a) Electron p_{T}



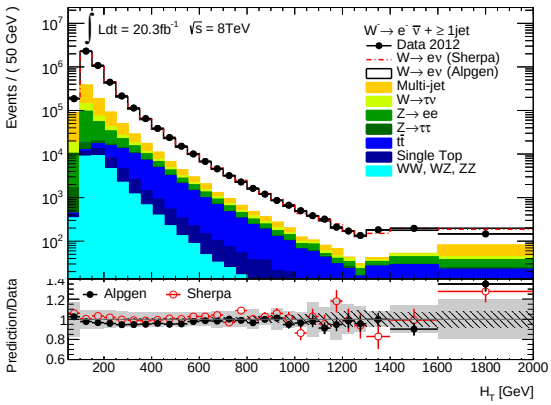
(b) W p_{T}



(c) Leading jet p_{T}

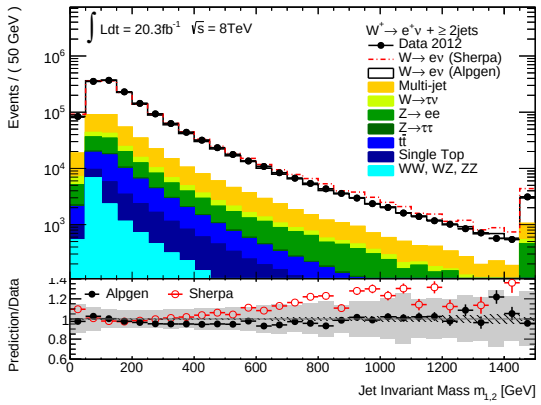


(d) Leading jet rapidity

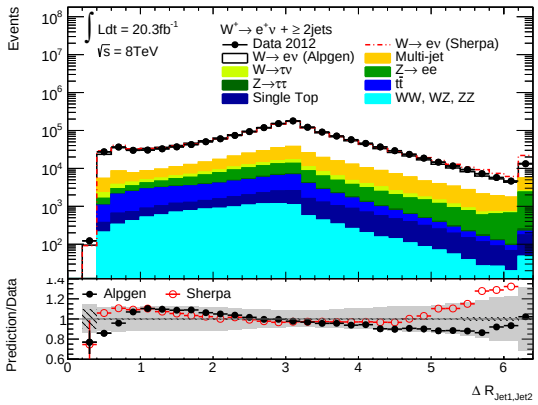


(e) H_{T}

Figure A.3: Kinematic distributions comparing data and the total prediction for the W^- selection in the W + jets signal region for ≥ 1 jet: Electron p_{T} (a), W p_{T} (b), leading jet p_{T} (c) and rapidity (d) as well as H_{T} (e). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the W + jets signal. The gray and hashed bands are defined as explained in Figure 5.7.

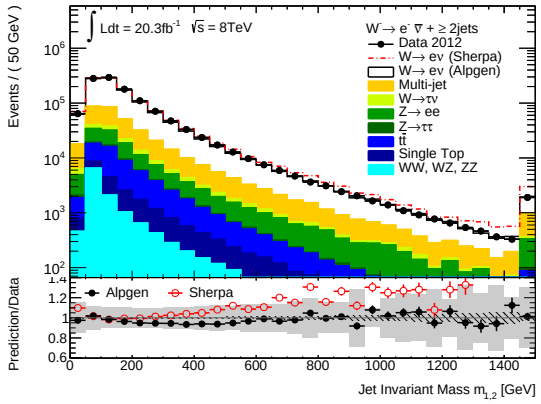


(a) Invariant jet mass $m_{1,2}$

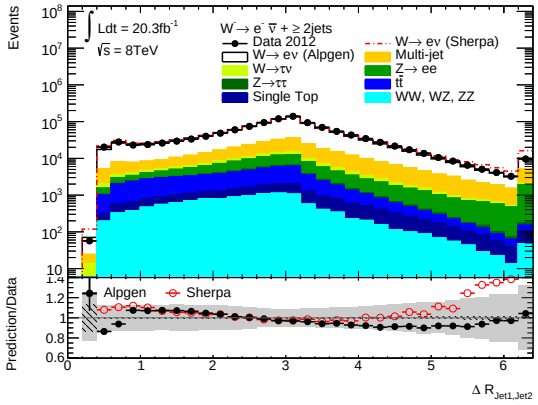


(b) Angular jet separation $\Delta R_{1,2}$

Figure A.4: Kinematic distributions comparing data and the total prediction for the W^+ selection in the $W+$jets signal region for ≥ 2 jets: Invariant mass of the leading two jets $m_{1,2}$ (a) and their angular separation $\Delta R_{1,2}$ (b). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the $W+$jets signal. The gray and hashed bands are defined as explained in Figure 5.7.



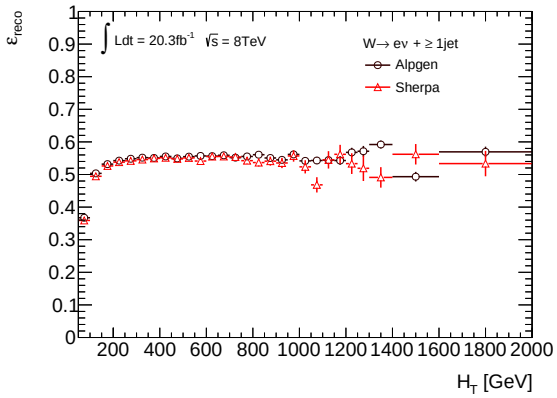
(a) Invariant jet mass $m_{1,2}$



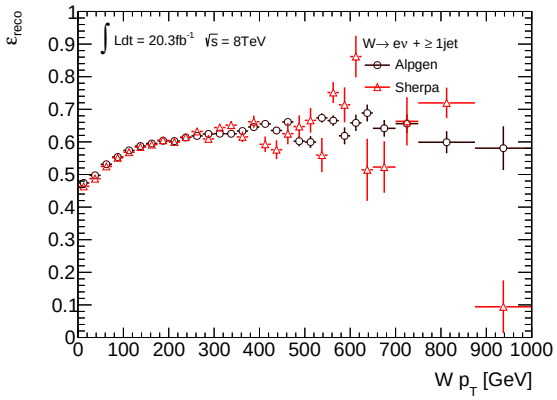
(b) Angular jet separation $\Delta R_{1,2}$

Figure A.5: Kinematic distributions comparing data and the total prediction for the W^- selection in the $W+$jets signal region for ≥ 2 jets: Invariant mass of the leading two jets $m_{1,2}$ (a) and their angular separation $\Delta R_{1,2}$ (b). The lower panels show the ratio of the total signal+background prediction to data, where the total prediction has been computed using two different MC generators for the $W+$jets signal. The gray and hashed bands are defined as explained in Figure 5.7.

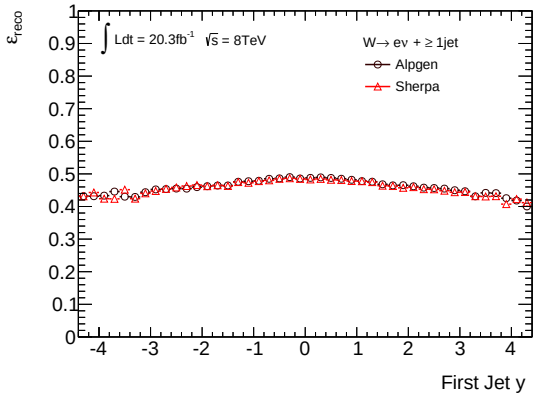
A.3 Further reconstruction efficiencies and fake rates



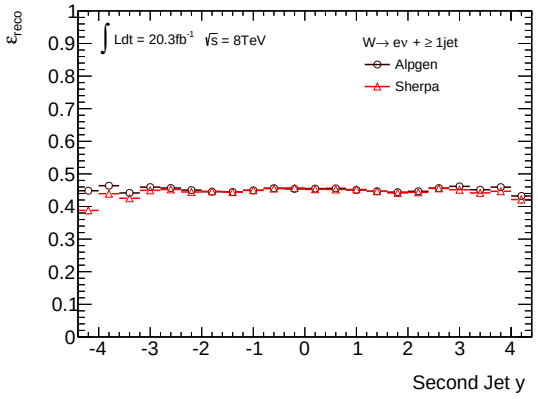
(a) H_{T}



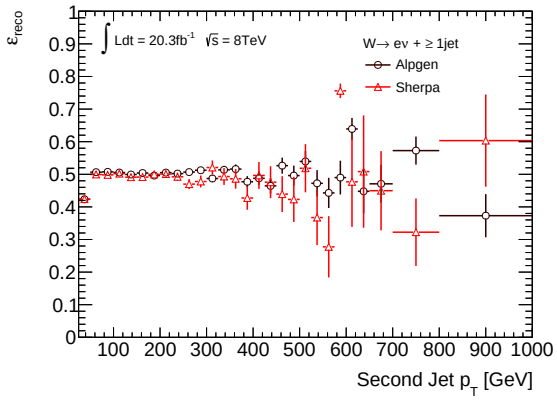
(b) W p_{T}



(c) Leading jet rapidity

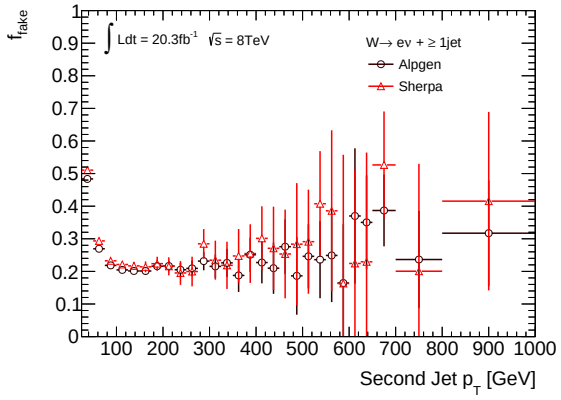


(d) Second leading jet rapidity

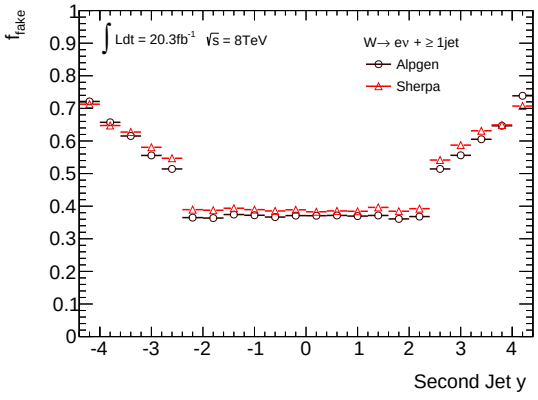


(e) Second leading jet p_{T}

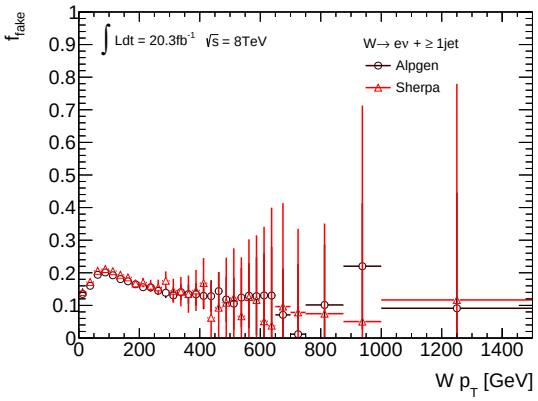
Figure A.6: Reconstruction efficiency as a function of H_{T} (a), W p_{T} (b), leading and second leading jet rapidities (c) and (d) as well as the second leading jet p_{T} (e) for ≥ 1 jet as determined from ALPGEN and SHERPA W + jets simulation.



(a) Second leading jet p_{T}



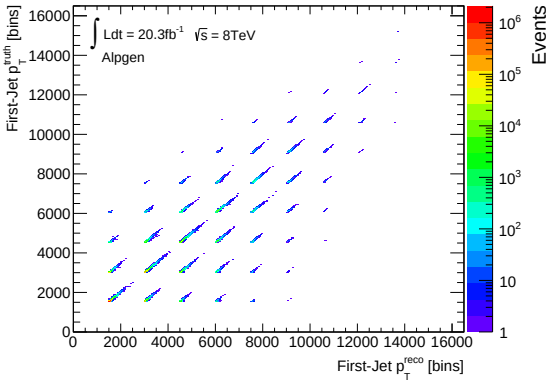
(b) Second leading jet rapidity



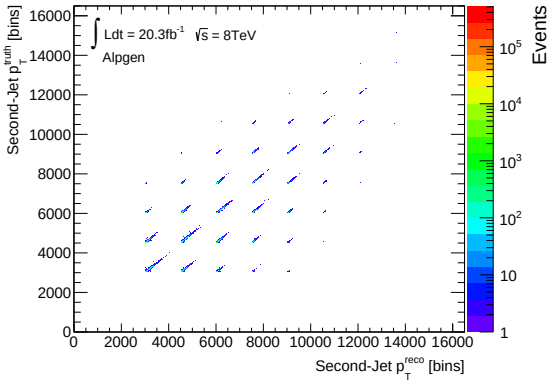
(c) W p_{T}

Figure A.7: Fake fraction as a function of second leading jet p_{T} (a) and rapidity (b) as well as W p_{T} for ≥ 1 jet as determined from Alpgen and Sherpa W + jets simulation.

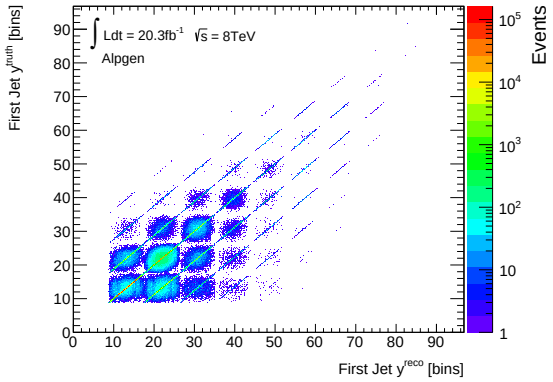
A.4 Further response matrices



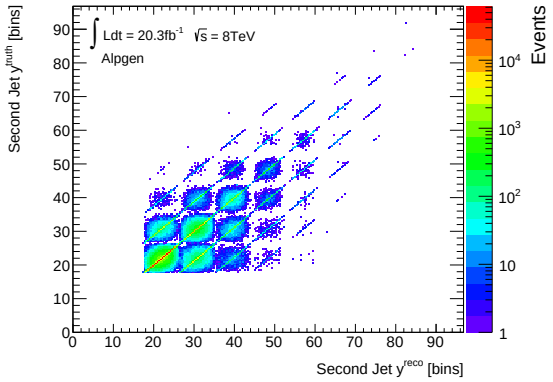
(a) Leading jet p_{T}



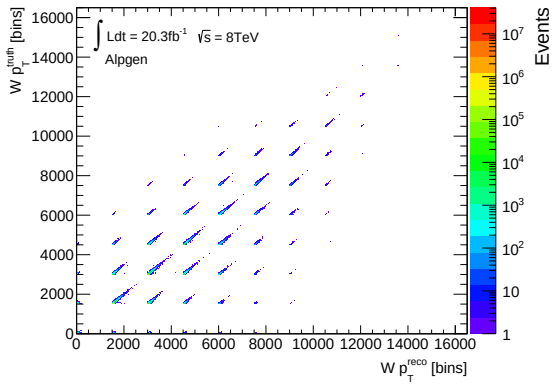
(b) Second leading jet p_{T}



(c) Leading jet rapidity



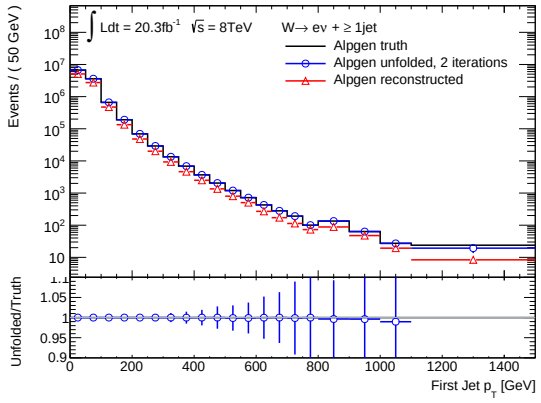
(d) Second leading jet rapidity



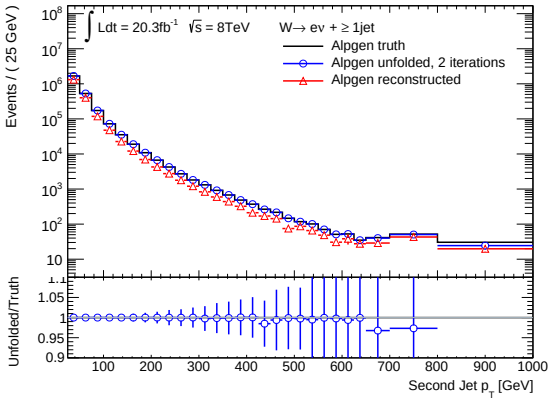
(e) W p_{T}

Figure A.8: Response matrices for the leading and second leading jet p_{T} (a) and (b), the leading and second leading jet rapidity (c) and (d) as well as the W p_{T} (e) as determined from ALPGEN $W+$jets simulation. The response matrices contain the truth vs. reco correlations repeatedly for the different $n_{\mathrm{jets}}^{\mathrm{truth}}$ and $n_{\mathrm{jets}}^{\mathrm{reco}}$ values.

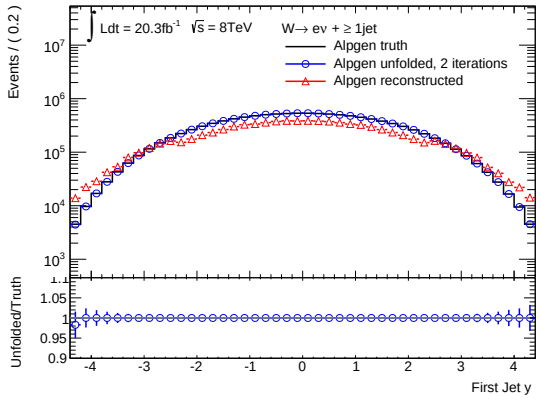
A.5 Unfolding closure tests



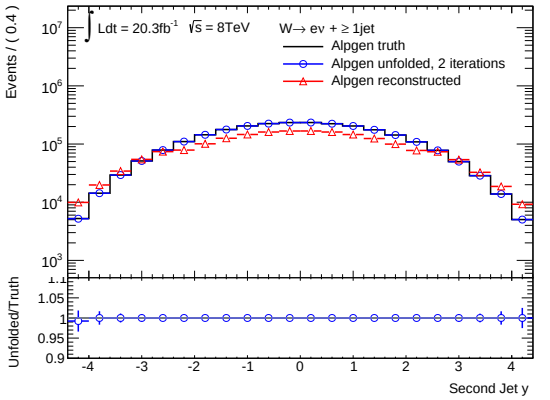
(a) Leading jet p_{T}



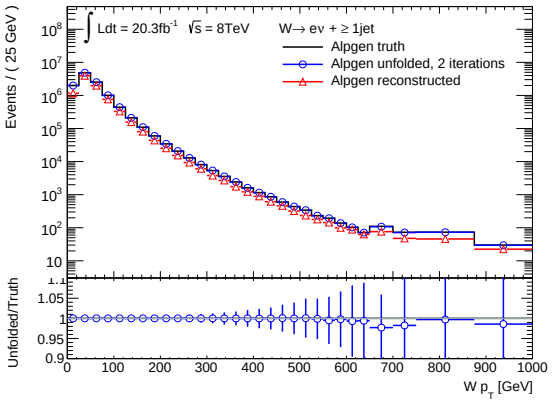
(b) Second leading jet p_{T}



(c) Leading jet rapidity



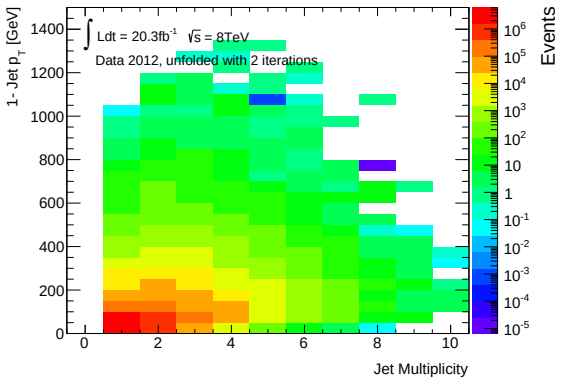
(d) Second leading jet rapidity



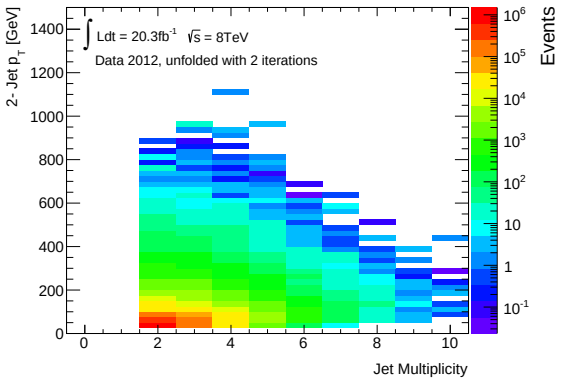
(e) W p_{T}

Figure A.9: Closure test of the 2-dimensional unfolding for the leading and second leading jet p_{T} (a) and (b), the leading and second leading jet rapidity (c) and (d) as well as the W p_{T} (e) distributions for ≥ 1 jet. The distributions for ≥ 1 jet are obtained as projection of the respective jet multiplicities in the 2-dimensional true, reconstructed and unfolded distributions as shown e.g. in Figure 6.8

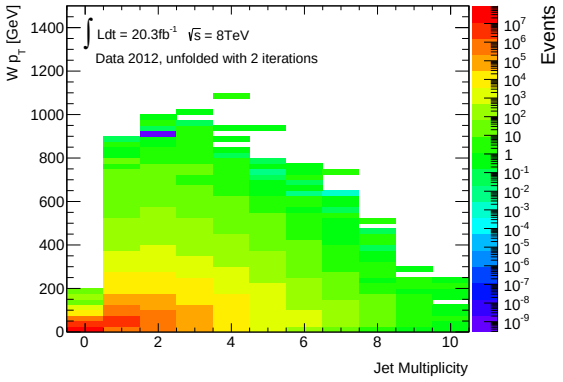
A.6 Further unfolded data distributions



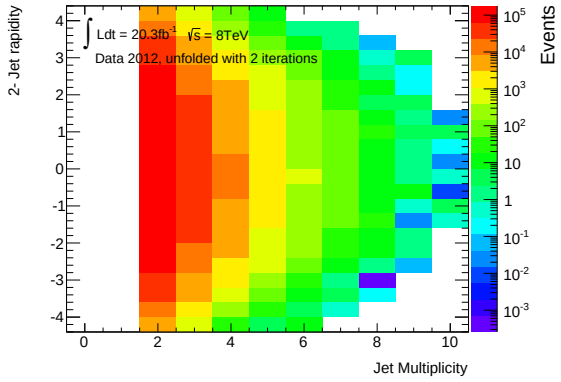
(a) Leading jet p_{T}



(b) Second leading jet p_{T}



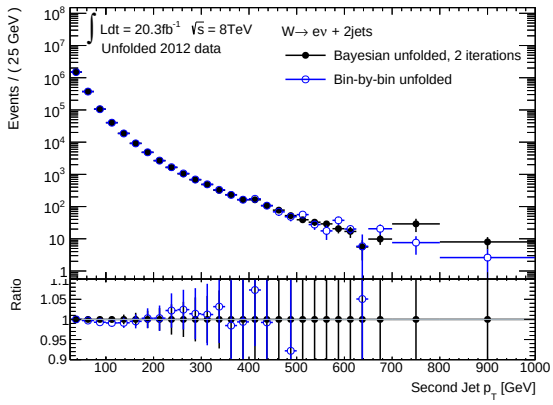
(c) W p_{T}



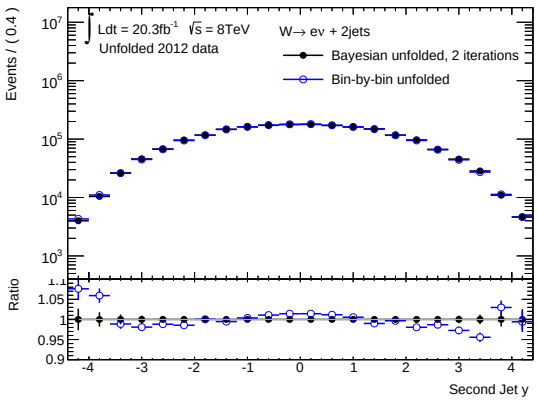
(d) Second leading jet rapidity

Figure A.10: Result of the 2-dimensional unfolding of the measured data for the leading and second leading jet p_{T} (a) and (b), the W p_{T} (c) and the second leading jet rapidity (d) distributions.

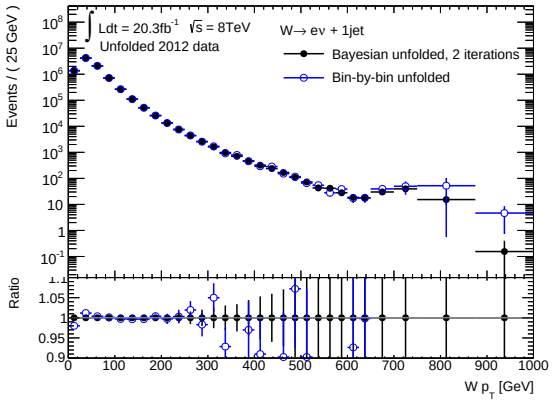
A.7 Comparison to bin-by-bin unfolding



(a) Second leading jet p_{T}



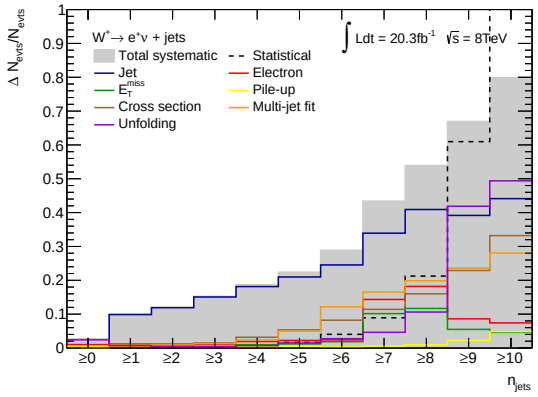
(b) Second leading jet rapidity



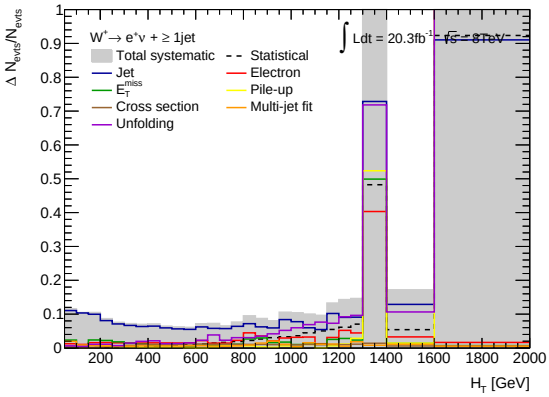
(c) W p_{T}

Figure A.11: Comparison of the unfolded results obtained with the iterative Bayesian unfolding or the bin-by-bin unfolding for the second leading jet p_{T} and rapidity for 2 jets (a) and (b) and the W boson p_{T} for 1 jet (c).

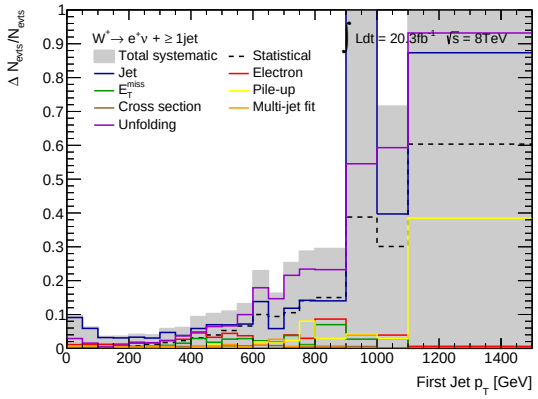
A.8 Total systematic and statistical uncertainties in the unfolded W^+ and W^- distributions



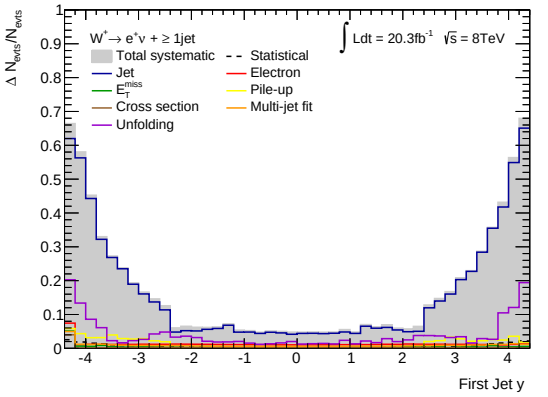
(a) n_{jets}



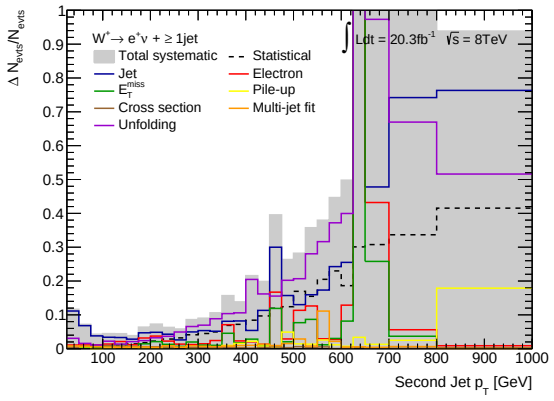
(b) H_{T}



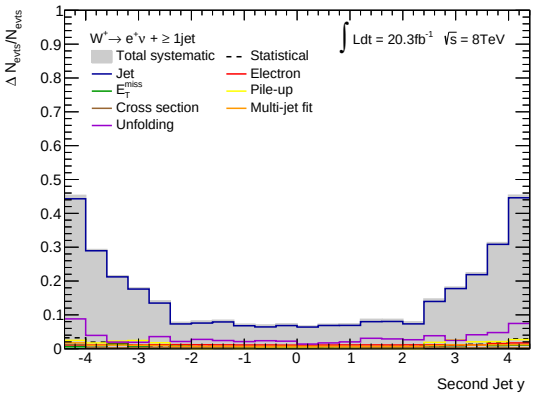
(c) Leading jet p_{T}



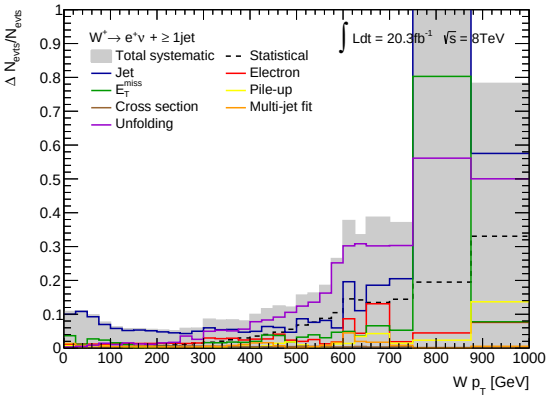
(d) Leading jet y



(e) Second leading jet p_{T}

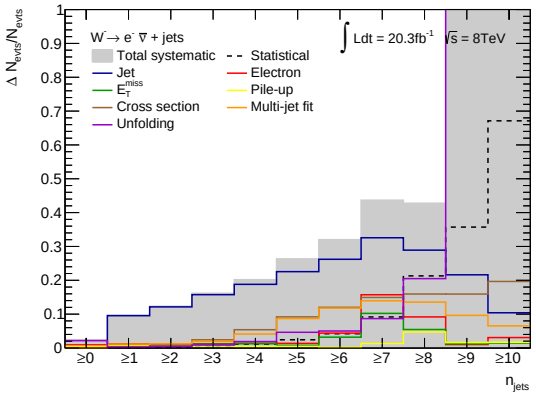


(f) Second leading jet y

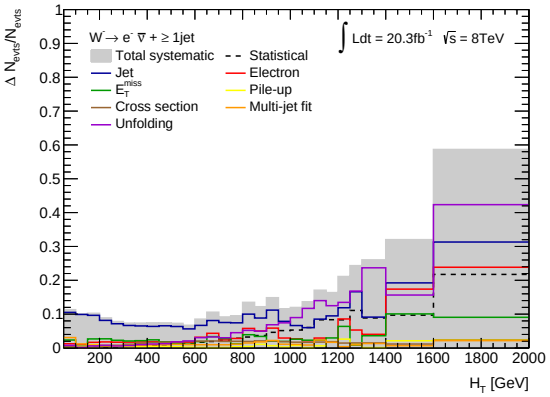


(g) W p_{T}

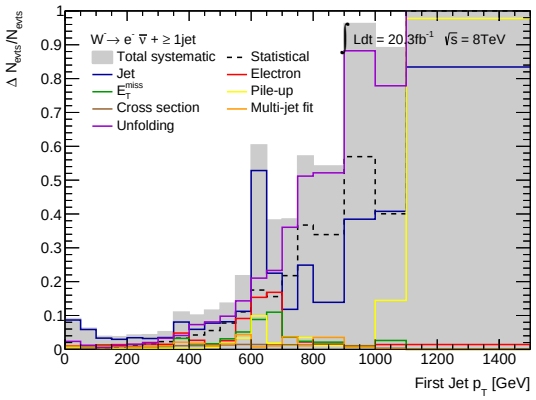
Figure A.12: Total systematic and statistical uncertainties in the unfolded results for W^+ production for the n_{jets} (a), H_{T} (b), leading jet p_{T} (c) and rapidity (d), second leading jet p_{T} (e) and rapidity (f) as well as W p_{T} distributions – the latter six for ≥ 1 jet.



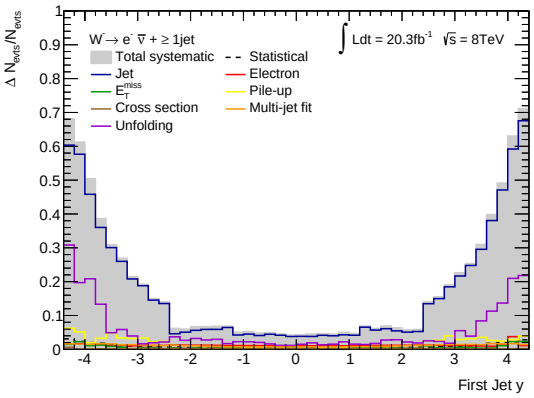
(a) n_{jets}



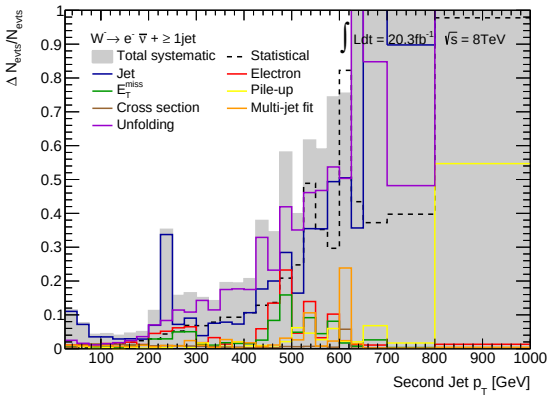
(b) H_{T}



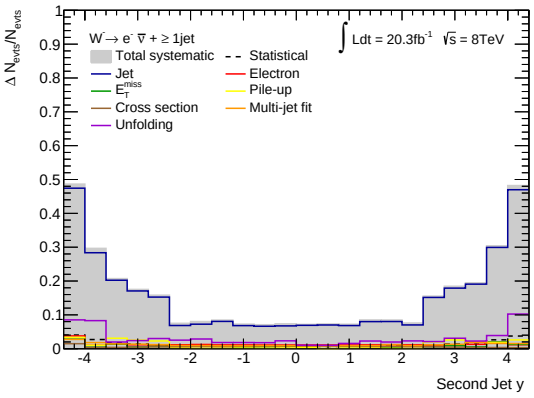
(c) Leading jet p_{T}



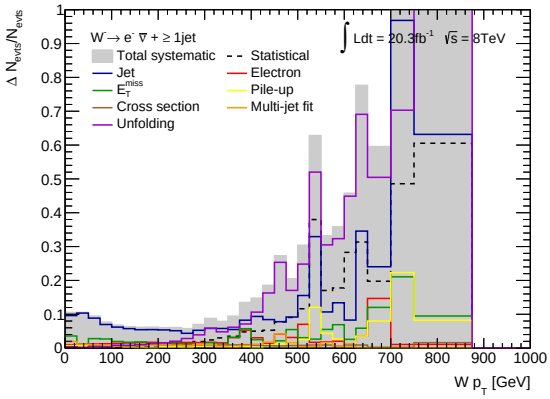
(d) Leading jet y



(e) Second leading jet p_{T}



(f) Second leading jet y



(g) W p_{T}

Figure A.13: Total systematic and statistical uncertainties in the unfolded results for W^- production for the n_{jets} (a), H_{T} (b), leading jet p_{T} (c) and rapidity (d), second leading jet p_{T} (e) and rapidity (f) as well as W p_{T} distributions – the latter six for ≥ 1 jet.

APPENDIX B

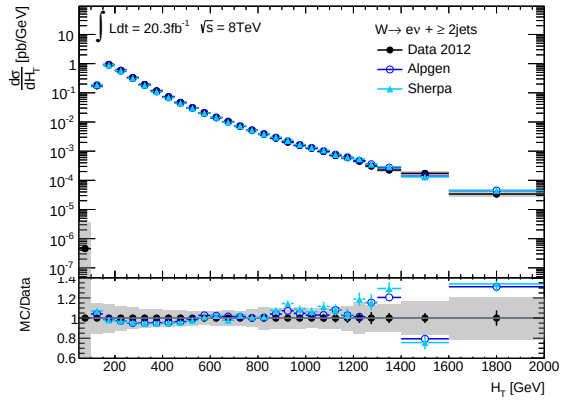
Additional results

B.1 Differential cross section measurement for $W+ \geq 2$ jets

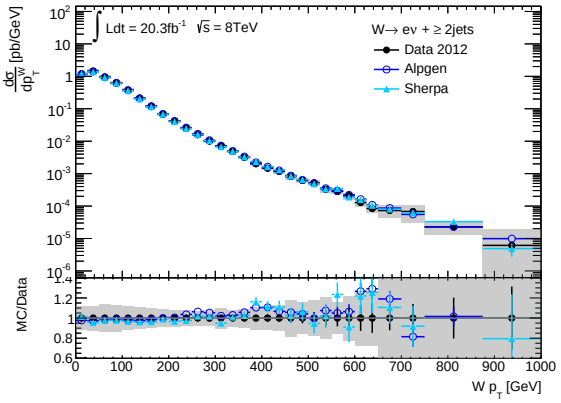
Differential cross section measurements for $W+ \geq 2$ jets are shown in Figure B.1 as a function of H_{T}, W p_{T} as well as leading jet p_{T} and leading jet rapidity. Distributions for the second leading jet are identical to the ones for $W+ \geq 1$ jet, since the second jet distributions obviously result from events with 2 jets, and are therefore identically contained in $W+ \geq 1$ jet and $W+ \geq 2$ jets. They are thus not shown again.

B.2 Differential cross section measurement for $W^{+/-}+ \geq 2$ jets

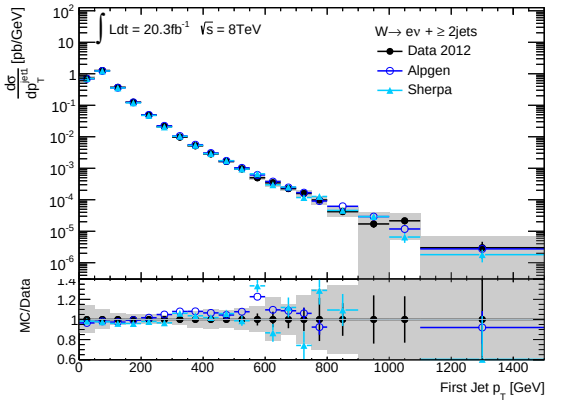
Measured differential cross sections similar to the ones shown above are also obtained for W^+ and W^- separately and are displayed in Figure B.2 and Figure B.3, respectively.



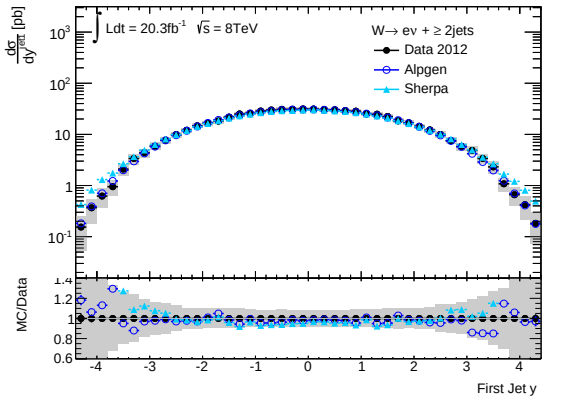
(a) H_{T}



(b) W p_{T}

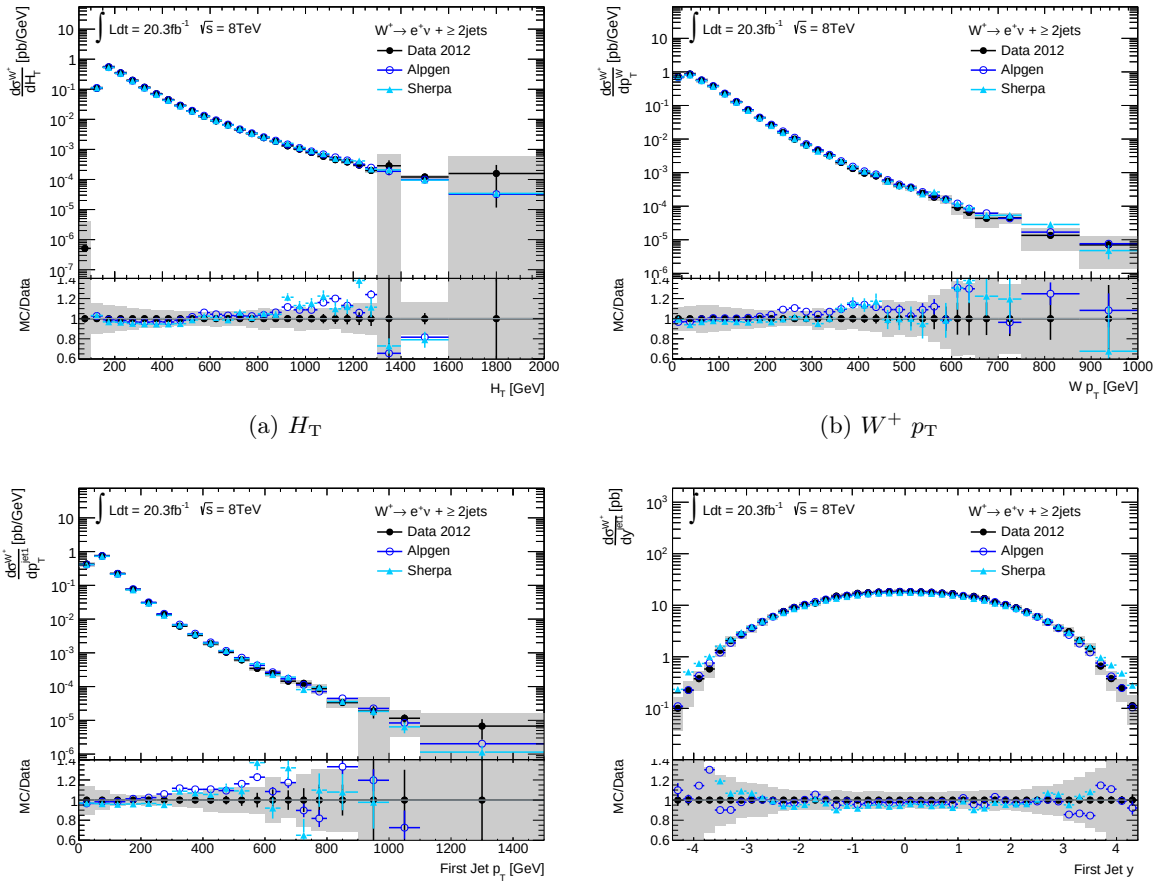


(c) Leading jet p_{T}



(d) Leading jet y

Figure B.1: Measured differential $W+ \geq 2$ jets cross sections as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d). The data is compared to Alpgen and Sherpa predictions. The lower panels show the size of the relative statistical and systematic uncertainties on the measurement as well as ratios of Alpgen and Sherpa predictions to data.



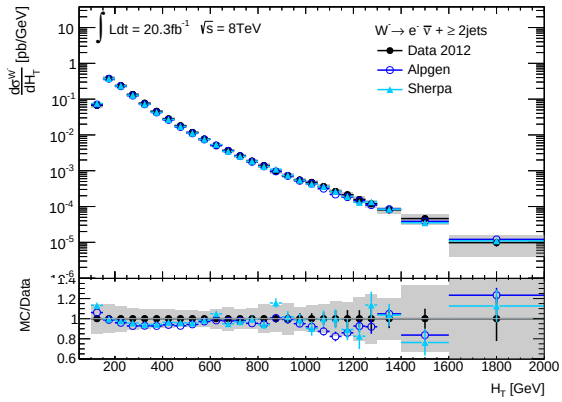
(a) H_{T}

(b) W^+ p_{T}

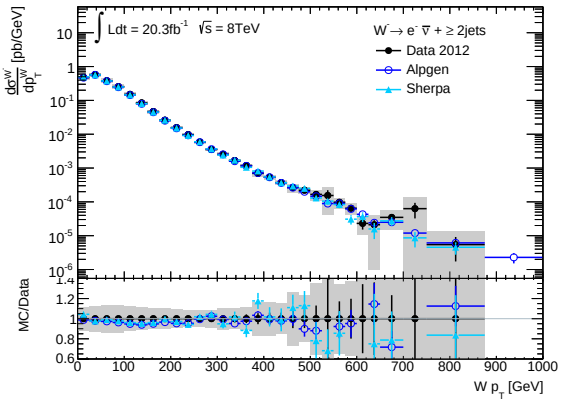
(c) Leading jet p_{T}

(d) Leading jet y

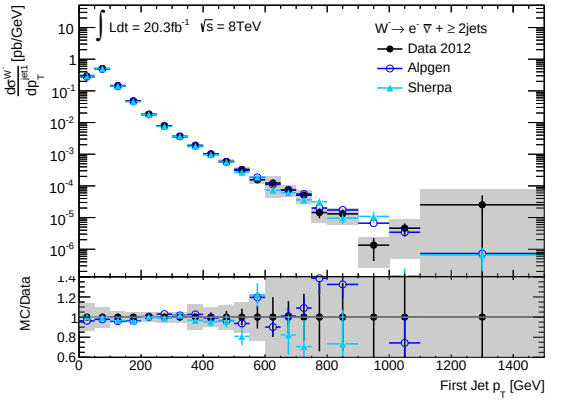
Figure B.2: Measured differential $W^+ + \geq 2$ jets cross sections as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d). Comparisons of data to predictions are done in the same way as in Figure B.1.



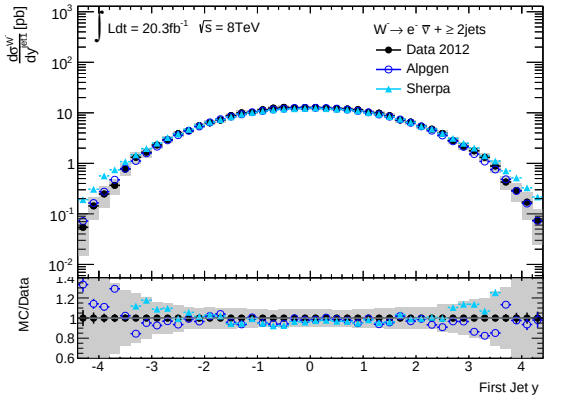
(a) H_{T}



(b) W^- p_{T}



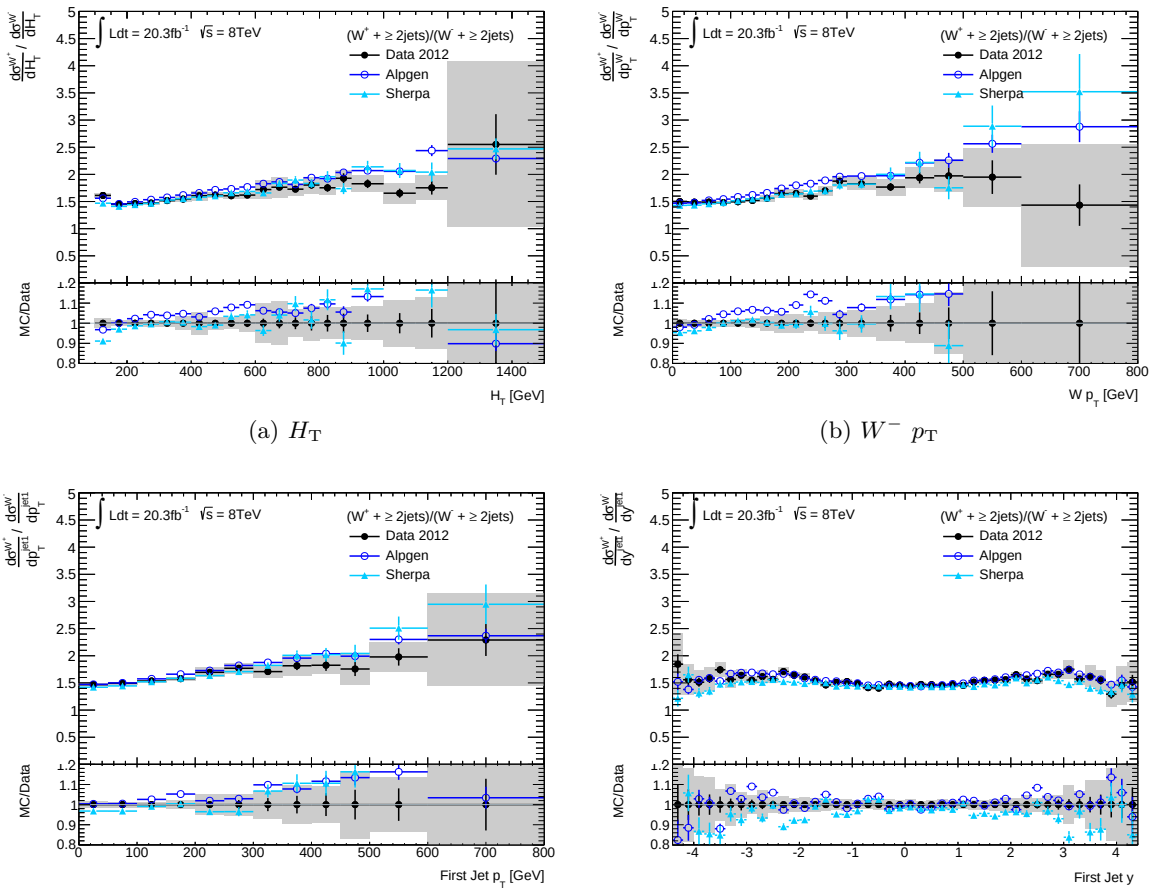
(c) Leading jet p_{T}



(d) Leading jet y

Figure B.3: Measured differential $W^-+ \geq 2$ jets cross sections as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d). Comparisons of data to predictions are done in the same way as in Figure B.1.

B.3 Measurement of differential W^+/W^- ratios with ≥ 2 jets



(a) H_{T}

(b) W^- p_{T}

(c) Leading jet p_{T}

(d) Leading jet y

Figure B.4: Measured differential W^+/W^- ratios in association with at least 2 jets as a function of H_{T} (a), W p_{T} (b), leading jet p_{T} (c) and leading jet rapidity (d). Comparisons of data to predictions are done in the same way as in Figure 9.11, except for MCFM.

Acknowledgements

This thesis and all the work leading to it was only possible with the help and support from many people.

First of all, I would like to thank Prof. Dr. Hans-Christian Schultz-Coulon for the possibility to explore the world of particle physics within the ATLAS collaboration and work on the research topic of my choice. I am grateful for the chance to experience two different types of analysis – a search and a measurement – and for his encouragement when it was much needed.

I would also like to thank Prof. Dr. Johanna Stachel who kindly agreed to be the second referee of this thesis. I am thankful for her time invested in reading this thesis and for a few inspiring statements of hers which guided me in this thesis and still make me think.

A huge thanks goes to Rainer, Monica and Pavel for their supervision, their advice and their thorough proof-reading of this thesis. I especially like to thank Monica for her motivation and support in physics, politics and life and the chance to take on this study of W + jets. Working in the ATLAS group at the Kirchhoff-Institute for Physics, I appreciate the enjoyable atmosphere in our combined ATLAS+ILC group and the many wonderful colleagues. I feel proud that we are a group with one of the largest shares of female physicists at the institute which makes work and conversations over lunch much more diverse. In particular, I like to thank my colleagues Julia, Alessandra and Merve who have become good friends and my current and former office mates, especially Nadia, Hanno and Sebastian who were often willing to go for a coffee when I needed a break from writing.

At CERN, I would like to thank the many interesting people and colleagues I have met in the past years who always make a visit to CERN worthwhile. Special thanks goes to Alessandro, Kristin and my friend Sarah for their help and input to the analysis, particularly in the last months when I needed answers to my questions fast. I am also thankful to everyone I have forgotten now as it is already rather late at night while I am writing this.

Last but not least, I want to thank my family for their understanding and support and their never-waving trust that I could do this. I also like to thank my friends, especially Selina and Nadia who helped me a lot in a very difficult time and asked for a sign of life when I got too absorbed in my work, especially in the last months. Yes, I am still alive and I did it!