Measurement of the cross-section for W boson production in association with b-jets in pp collisions at √s=7 TeV with the ATLAS detector


DOI
10.1007/JHEP06(2013)084

Publication date
2013

Document Version
Final published version

Published in
The Journal of High Energy Physics

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Measurement of the cross-section for $W$ boson production in association with $b$-jets in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: This paper reports a measurement of the $W+b$-jets ($W+b+X$ and $W+b\bar{b}+X$) production cross-section in proton-proton collisions at a centre-of-mass energy of 7 TeV at the LHC. These results are based on data corresponding to an integrated luminosity of 4.6 fb$^{-1}$, collected with the ATLAS detector. Cross-sections are presented as a function of jet multiplicity and of the transverse momentum of the leading $b$-jet for both the muon and electron decay modes of the $W$ boson. The $W+b$-jets cross-section, corrected for all known detector effects, is quoted in a limited kinematic range. Combining the muon and electron channels, the fiducial cross-section for $W+b$-jets is measured to be $7.1 \pm 0.5$ (stat) $\pm 1.4$ (syst) pb, consistent with the next-to-leading order QCD prediction, corrected for non-perturbative and double-parton interactions (DPI) contributions, of $4.70 \pm 0.09$ (stat) $^{+0.60}_{-0.49}$ (scale) $\pm 0.06$ (PDF) $\pm 0.16$ (non-pert) $^{+0.52}_{-0.38}$ (DPI) pb.

KEYWORDS: Hadron-Hadron Scattering

doi:10.1007/JHEP06(2013)084
1 Introduction

This paper reports a measurement of the cross-section for W boson production in association with b-quark jets in proton-proton (pp) collisions at \( \sqrt{s} = 7 \) TeV. The measurement, performed differentially in the b-jet transverse momentum \( (p_T) \) for the 1-jet and 2-jet final states, provides an important test of perturbative quantum chromodynamics (QCD) in the presence of heavy quarks.

Next-to-leading-order (NLO) perturbative QCD calculations of the W+b-jets process have recently become available in Monte Carlo (MC) simulations, both at the parton level [1] and enhanced with parton shower models [2–5]. Several processes contribute to W+b-jets production at NLO. In the four-flavour number scheme (4FNS), where only u, d, c, s are considered as initial-state quarks, these are \( q\bar{q} \rightarrow Wb\bar{b}(g) \) and \( gq \rightarrow Wb\bar{b}q \). When considering the presence of b-quarks in the initial state (five-flavour number scheme, or 5FNS), the single b-quark processes \( bq \rightarrow Wbq(g) \) and \( b\bar{q} \rightarrow Wbq\bar{q} \) also play a significant role at LHC energies [6–9]. In addition, double-parton interactions (DPI), in which a W
boson and $b$-jets are produced from different parton-parton interactions within the same $pp$
collision, are also expected to contribute to the total observed $W+b$-jets cross-section [10].

The production of $W+b$-jets events via top-quark decay (from single or pair-produced
top-quarks) is not included in the signal definition for the primary measurement. An additional set of measurements is performed including the contribution from single top-quark production. These are of particular relevance for the differential cross-section, since single top-quark and $W+b$-jets events are difficult to separate and their relative contribution depends strongly on the $b$-jet transverse momentum.

The $W+b$-jets process is an important background to the Higgs boson associated-production process $WH$ with $H \rightarrow b\bar{b}$ decays [11]. The associated-production measurements are a substantial ingredient in determining the coupling of the Higgs boson to fermions, through the decay $H \rightarrow b\bar{b}$, and searches in this channel have been included in studies of the Higgs-like boson [12]. The $W+b$-jets process is also an irreducible background in some searches for physics beyond the Standard Model [13], and in measurements of single top-quark properties [14], due to the dominating branching fraction of the $t \rightarrow Wb$ decay.

Measurements of the $W+b$-jets fiducial cross-section in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV have been reported by the CDF Collaboration [15] and more recently also by the D0 Collaboration [16]. The ATLAS Collaboration reported a previous measurement based on 36 pb$^{-1}$ of data collected in $pp$ collisions at $\sqrt{s} = 7$ TeV [17]. The CDF measurement of $2.74 \pm 0.27$ (stat) $\pm 0.42$ (syst) pb, and the ATLAS measurement of $10.2 \pm 1.9$ (stat) $\pm 2.6$ (syst) pb are both found to be larger than the corresponding theoretical cross-sections calculated at NLO $(1.22 \pm 0.14$ pb and $4.8 \pm 1.3$ pb) by 2.8 and 1.5 standard deviations, respectively. The D0 measurement of $1.05 \pm 0.12$ pb is found to be lower than the theoretical prediction of $1.34^{+0.41}_{-0.34}$ pb [1], but in agreement within theoretical uncertainties.

In this paper, the $W+b$-jets cross-section is measured using the ATLAS detector$^1$ in a restricted fiducial region defined at the particle level$^2$ and given in table 1. For the first time, the $W+b$-jets differential cross-section is also measured as a function of the $b$-jet $p_T$. To enter the fiducial region, events at the generator level are required to contain an electron or muon and a neutrino originating from a $W$ boson decay, and one or two particle-level jets. At least one of the jets is required to be a $b$-jet, defined by the presence of a weakly decaying $b$-hadron with $p_T > 5$ GeV and within $\Delta R = 0.3$ of the jet axis. The fiducial region is divided, according to the number of particle-level jets, in the 1-jet and 2-jet fiducial regions.

---

$^1$The ATLAS experiment uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the $z$-axis along the beam pipe. The $x$-axis points from the IP to the centre of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln(\tan(\theta/2))$ and the rapidity is defined as $y = \ln[(E + p_z)/(E - p_z)]/2$. The distance $\Delta R$ in $\eta-\phi$ space is defined as $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.

$^2$For leptons, the particle level corresponds to ‘dressed’ muons and electrons, calculated using final-state leptons from the $W$ decay for which collinear radiation in a cone of $\Delta R < 0.1$ is added to the lepton four-momentum. Particle-level jets are clustered from all final-state particles (with a proper lifetime longer than 10 ps), excluding the $W$ decay products. Given this definition, muons and neutrinos from decaying hadrons are included in particle-level jets.
### Table 1. Definition of the phase space for the fiducial region. The \( W \) transverse mass is defined as
\[
m_T(W) = \sqrt{2p_T(\ell)p_T(\nu)(1 - \cos(\phi_\ell - \phi_\nu))}.
\]

At the reconstruction level, events are required to be consistent with the decay of a \( W \) boson to the \( \ell \nu \) (\( \ell = \mu, e \)) final state, and to contain either one or two jets. Events are selected if exactly one jet passes the \( b \)-tagging requirements. A \( b \)-tagged jet can be a \( b \)-jet, but also a mis-tagged \( c \)-jet or light-jet. Events with two or more \( b \)-tagged jets are rejected, as are events with three or more jets, to reduce the top-quark background. Given the fiducial region definition and the reconstruction-level selection, the measurement is performed using reconstructed events containing a single \( b \)-tagged jet. However, it is unfolded to the fiducial region with one or more \( b \)-jets. The measurement is performed separately in the \( W \to \mu \nu \) and \( W \to e \nu \) decay channels and in the exclusive 1-jet and 2-jet final states.

## 2 The ATLAS detector

The ATLAS detector \cite{18} is a multi-purpose particle physics detector operating at one of the interaction points of the LHC. It consists of an inner detector tracking system (ID) within a 2 T axial magnetic field provided by a superconducting solenoid, surrounded by electromagnetic and hadronic calorimeters, and by a muon spectrometer (MS) embedded in the magnetic field provided by three air-core superconducting toroidal magnets.

The ID consists of pixel and silicon microstrip detectors surrounded by a transition radiation tracker, and it provides measurements of charged-particle tracks within \(|\eta| < 2.5\). The calorimeters provide three-dimensional reconstruction of particle showers in the region of \(|\eta| < 4.9\), with a finely segmented inner layer used for electron identification in \(|\eta| < 2.5\). They are based on liquid-argon (LAr) sampling technologies, except for the barrel region (\(|\eta| < 1.7\)) of the hadronic calorimeter where scintillator tiles are used as the active media. The MS consists of three layers of high-precision tracking chambers (monitored drift-tubes and cathode strips) in the region \(|\eta| < 2.7\), and resistive-plate or thin-gap chambers providing trigger signals in the region \(|\eta| < 2.4\).
3 Simulated event samples

Monte Carlo simulated samples are used to model the reconstructed $W+b$-jets signal and most of its background contributions, as well as to extract a fiducial cross-section from the measured $W+b$-jets yield.

The processes of $W$ boson production in association with $b$-jets, $c$-jets and light-jets are simulated separately using the ALPGEN 2.13 [19] generator, interfaced to HERWIG 6.510 [20, 21] for parton showers and hadronization, and JIMMY 4.31 [22] for the underlying-event simulation, using the AUET2-CTEQ6L1 [23] tune. Exclusive samples with zero to four additional partons and an inclusive sample with five or more additional partons are used. The MLM [24] matching scheme, as implemented in ALPGEN, is used to remove overlaps between samples with the same parton multiplicity originating from the matrix element (ME) and the parton shower (PS). In addition, overlap between samples with heavy-flavour quarks originating from the ME and from the PS is removed. Large samples of dijet events simulated using PYTHIA 6.423 [25] with the AUET2b [26] tune are used to model the light and heavy-flavour jet properties relevant to the $b$-jet identification in the $W$+jets sample.

The $Z$+jets background is simulated with ALPGEN interfaced to HERWIG and JIMMY, using the same configuration as for $W$+jets. The diboson ($WW, WZ, ZZ$) background is simulated with HERWIG. The $t$-channel, $s$-channel and $Wt$-channel single-top processes are simulated with AcerMC 3.7 [27] interfaced to PYTHIA. The $t\bar{t}$ background is simulated with POWHEG [2] interfaced to PYTHIA.

The total cross-sections of the $W$+jets and $Z$+jets samples are normalized to the inclusive NNLO predictions [28], while other backgrounds are normalized to NLO predictions [29–31]. The $t\bar{t}$ yields in the 1-jet and 2-jet analysis regions, and the single-top yield in the 2-jet analysis region, are estimated from data.

For all the processes modelled, multiple interactions per bunch crossing (pile-up), with an average of nine interactions per crossing, are accounted for by overlaying minimum-bias events simulated with PYTHIA onto the generated hard process. The detector simulation [32] is based on the GEANT4 program [33].

4 Data sample and event selection

The analysis considers data recorded in the year 2011 during periods with stable $pp$ collisions at $\sqrt{s} = 7$ TeV, and where all relevant parts of the detector were operating normally. The resulting data set corresponds to 4.6 fb$^{-1}$ of integrated luminosity, with an uncertainty of 3.9% [34, 35]. Events were collected using single-muon or single-electron triggers. The $p_T$ threshold of the muon trigger was 18 GeV, while the transverse energy ($E_T$) threshold used for the electron trigger was initially 20 GeV and was later raised to 22 GeV to cope with the increasing LHC instantaneous luminosity.

Candidate $W+b$-jets events are required to have exactly one high-$p_T$ electron or muon, as well as missing transverse momentum ($E_T^{\text{miss}}$) consistent with a neutrino from a $W$ boson, and one or two reconstructed jets, exactly one of which must be $b$-tagged. All events must have at least one reconstructed vertex formed by the intersection of at least three tracks.
with $p_T > 400$ MeV. In events with multiple vertices, the vertex with the largest sum of squared $p_T$ of the associated tracks is taken to be the primary hard-scatter vertex (PV).

Requiring events to have exactly one $b$-tagged jet significantly reduces the top-quark background contribution in the 2-jet analysis region. $W+b$-jets events with a second $b$-jet satisfying the fiducial selection represent 10% of the 2-jet fiducial region. Most of these events have a single $b$-tagged jet, and they are included in the 2-jet region at the reconstruction level.

Electron candidates are formed by matching clusters found in the electromagnetic calorimeter to tracks reconstructed in the ID in the region of $|\eta| < 2.47$ and are required to have $E_T > 25$ GeV. To ensure good containment of electromagnetic showers in the calorimeter, the transition region $1.37 < |\eta| < 1.52$ between the barrel and the endcaps is excluded. The lateral and transverse shapes of the clusters must be consistent with those of an electromagnetic shower [36]. Muon candidates, reconstructed by combining tracks reconstructed in the ID and the MS, are selected in the region $|\eta| < 2.4$, and are required to have $p_T > 25$ GeV. Both the electron $E_T$ and the muon $p_T$ requirements are chosen to be on the efficiency plateau for the respective triggers. The selection efficiency of electrons and muons in simulated events, as well as their energy and momentum scale and resolution, are adjusted to reproduce those observed in $Z \rightarrow \ell\ell$ events in data [36–38].

In order to reduce the large background from multijet production, lepton candidates are required to be isolated from neighbouring tracks within $\Delta R = 0.4$ of their direction, as well as from other calorimeter energy depositions, corrected for pile-up contributions, within $\Delta R = 0.2$. In the muon case, the sum of transverse momenta of neighbouring tracks must be less than 2 GeV, while the sum of the calorimeter transverse energies must be less than 1 GeV. In the electron case, these requirements range between 1.35 GeV and 3.15 GeV depending on $p_T$ and $\eta$ in order to yield a constant efficiency across momentum ranges and detector regions. Additionally, leptons are required to be consistent with originating from the PV. Their longitudinal impact parameter ($|z_0|$) with respect to the PV must be smaller than 10 mm, and the ratio of the transverse impact parameter $d_0$ to its uncertainty ($d_0$ significance) must be smaller than 3 for muons, and 10 for electrons.

Jets are reconstructed from calorimeter energy topological clusters (topoclusters) [39] using the anti-$k_T$ algorithm [40] with a radius parameter $R = 0.4$. They are required to have a transverse momentum greater than 25 GeV, and a rapidity $|y| < 2.1$ in order for the entire jet to be reconstructed within the tracking region. Jets originating in pile-up interactions are suppressed by requiring that at least 75% of the total transverse momentum of tracks associated with each jet point to the PV. Jets within a distance $\Delta R = 0.5$ of the lepton candidate are removed, and jets arising from detector noise or cosmic rays are also rejected [41].

The jet energy is calibrated to account for the different response of the calorimeters to electrons and hadrons, for energy losses in un-instrumented regions, and for the energy offset introduced by pile-up, by applying jet calibration factors dependent on $p_T$, $\eta$ [42–44], and pile-up conditions [45]. A residual calibration derived from in-situ techniques is applied to the data to reduce differences between data and Monte Carlo simulation [46, 47].
Jets originating from $b$-quarks are identified using the combination of two $b$-tagging algorithms. The first one exploits the topology of weak $b$- and $c$-hadron decays inside the jet to reconstruct their decay vertices along a common line originating from the PV. The second one uses the impact parameter significance of each track within the jet to determine the likelihood that the jet originates from a $b$-quark. The properties measured by these two taggers are combined using an artificial neural network to determine a single discriminant variable (CombNN) \[48\].

The CombNN variable is used both to select a sample enriched with $b$-jets, as well as to discriminate between $b$-jets, $c$-jets and light-jets within the enriched sample. The working point used for the selection (CombNN > 2.2) corresponds to a $b$-tagging efficiency of about 40\% at low $p_T$, increasing to a plateau of 57\% for $b$-jets of $p_T$ above 60 GeV, with rejection rates of about 10 for $c$-jets and 1000 for light-jets. In order to reproduce the $b$-jet, $c$-jet and light-jet tagging efficiencies measured in data \[49–51\], event weights in Monte Carlo simulation are scaled as a function of the number of tagged and untagged jets of each flavour and the corresponding $p_T$ and $\eta$. The CombNN distribution in the $b$-tagged sample is then used to separate statistically the remaining $c$-jet and light-jet contributions from the $W+b$-jets signal.

The measurement of $E_T^{\text{miss}}$ in each event is based on an algorithm \[52\] which performs the vector sum of transverse energies of high-$p_T$ objects such as electrons, muons and jets. All topological energy clusters \[39\] not associated with any high-$p_T$ physics objects are also added to the vector sum, after being individually calibrated \[53\]. To be consistent with a $W$ boson decay, and to reduce the multijet background, the $E_T^{\text{miss}}$ is required to be larger than 25 GeV, and the $W$ boson transverse mass $m_T(W) = \sqrt{2p_T^\ell p_T^\nu (1 - \cos(\phi^\ell - \phi^\nu))}$ is required to be larger than 60 GeV.

5 Signal and background estimation

Several processes contribute to the overall background for the $W+b$-jets signal, accounting for more than 85\% of the selected sample. Some of the backgrounds, such as single-top, $t\bar{t}$ and multijet, are characterized by rather different kinematics from $W+b$-jets, but they have real $b$-jets in their final state and show similar $b$-tagging response. Others, $W+c$-jets and $W$+light-jets, have kinematic properties similar to the signal, but they can be statistically separated by studying the characteristics of $b$-tagged jets. The remaining backgrounds, diboson ($WW$ and $WZ$) and $Z$+jets, contribute less than 5\% of the selected sample.

In most cases, backgrounds are estimated directly from data in order to reduce the theoretical uncertainties on their normalization. A sequence of binned maximum likelihood (ML) fits is performed, in which a distribution measured in data is described by a linear combination of templates representing each contributing process. In each fit, the normalization of the process of interest is allowed to float freely, while some processes are constrained by Gaussian terms in the likelihood.\footnote{Using log-normal terms to ensure positive normalization yields consistent results with those obtained using Gaussian terms.} For a constrained process, the mean of the corresponding Gaussian constraint is fixed to the expected number of events, while
Figure 1. Overlay of the $W+b$-jets, $W+c$-jets and $W+\text{light}$-jets CombNN distributions in MC simulation for $b$-tagged (CombNN > 2.2) jets in the muon 1-jet analysis region.

The width is fixed to the associated uncertainty. This uncertainty is derived either from the results of a previous ML fit or from theoretical uncertainties, depending on the process considered. Pseudo-experiments are used to validate the behaviour and properties of each fit.

The single-top, $t\bar{t}$ and multijet contributions are estimated either in background-enriched control regions or using kinematic distributions directly in the signal regions. The $W+b$-jets, $W+c$-jets and $W+\text{light}$-jets contributions are then statistically separated, and the number of $W+b$-jets events is extracted, by fitting the CombNN weight distribution of $b$-tagged jets observed in data in each analysis region. Example templates for the muon 1-jet sample are shown in figure 1.

5.1 Multijet background

Multijet events from QCD production processes in which one of the jets is either identified as a lepton, or contains a real lepton originating from a heavy-quark decay, can occasionally enter the selected sample. Specific criteria are used to reduce this contamination, such as the lepton identification, isolation and impact parameter, and the $E_T^{\text{miss}}$ and $m_T(W)$ requirements mentioned in section 4. To estimate the remaining multijet contribution, complementary data samples highly enriched in multijet events are created by requiring that some of these criteria are not fulfilled. The normalization of these samples is then obtained by fitting the $E_T^{\text{miss}}$ distributions in data.

Specifically, the multijet background shape for each distribution is obtained in the muon channel by inverting the tracking isolation requirement, and in the electron channel by inverting part of the identification selection and waiving the calorimeter isolation requirement. The selection used to form the multijet template from data was studied using dijet Monte Carlo simulations to minimize kinematic biases with respect to the standard signal selection, while maintaining the large number of events required to obtain smooth templates.
The normalization of the multijet template is then assessed, in each analysis region, by performing a fit to the $E_T^{\text{miss}}$ distribution in data after relaxing the $m_T(W)$ requirement from 60 GeV to 40 GeV, and removing the $E_T^{\text{miss}} > 25$ GeV requirement. The templates used in this fit for the $W/Z+\text{jets}$, $t\bar{t}$, single-top and diboson processes are based on Monte Carlo simulation. The multijet and the $W+\text{jets}$ template normalizations are free parameters of the fit to the $E_T^{\text{miss}}$ distribution, while those of the other components are fixed to their expected cross-sections. The $E_T^{\text{miss}}$ distributions, normalized to the results of the fit, are presented in figure 2 for the 1- and 2-jet regions in the muon and electron channels.

To estimate the uncertainty of the multijet normalization, the $E_T^{\text{miss}}$ fit results are compared in each analysis region to the results obtained by fitting the alternative distributions of $m_T(W)$ and lepton $p_T$. As a result of the comparison, an uncertainty of 50%, applied as a Gaussian constraint in subsequent ML fits, is assigned to the multijet normalization in the 1-jet and 2-jet regions. In the differential measurement, the multijet background normalization is extrapolated from the inclusive estimates, and the same 50% uncertainty is applied as an independent Gaussian constraint in each $p_T^{\text{jet}}$ bin. In the 4-jet region used to estimate the $t\bar{t}$ background, this uncertainty is estimated from the different fit results to be 100%.
5.2 $t\bar{t}$ background

The $t\bar{t}$ background is estimated in data by selecting events with at least four jets and exactly one $b$-tag. A binned ML fit to the CombNN distribution is performed in this control region to extract the $t\bar{t}$ yield. The $t\bar{t}$ Monte Carlo simulation is then used to extrapolate the measured yield into the 1- and 2-jet analysis regions.

In this fit, the sum of the $W+b$-jets, $W+c$-jets and $W+\text{light-jets}$ MC templates is normalized to the NNLO $W$ inclusive cross-section. Their relative contributions are taken from the ALPGEN Monte Carlo prediction and a Gaussian normalization uncertainty constraint of 100% is applied to each. Similarly, the single-top template is assigned a 50% constraint that reflects the maximum uncertainty on the single-top normalization discussed in section 5.3. The multijet background is estimated using the technique described in the previous section and assigned a normalization uncertainty constraint of 100%, based on the fits to the alternative distributions. The $Z+\text{jets}$ contribution is assigned a 10% normalization uncertainty constraint based on theoretical calculations and previous measurements [54, 55]. Finally, the diboson contribution is assigned a 10% normalization uncertainty constraint, which is twice the uncertainty of the corresponding NLO predictions [56, 57].

The correction factors to the $t\bar{t}$ Monte Carlo normalization estimated by the fit in the “at least 4 jets 1-$b$-tag” region are $1.09 \pm 0.06$ for the muon channel and $1.08 \pm 0.07$ for the electron channel. These factors are in good agreement with those resulting from alternative fits to the “3-jet 1-$b$-tag” region, and with those resulting from a fit of the number-of-jets distribution in the “at least 1 jet 1-$b$-tag” and the “at least 4 jets 1-$b$-tag” regions. The fit projections corresponding to the CombNN distribution in the “at least 4 jets 1-$b$-tag” region and to the number-of-jets distribution in the “at least 1 jet 1-$b$-tag” are shown in figure 3. As a result, a 10% $t\bar{t}$ normalization uncertainty is applied as a Gaussian constraint in subsequent fits.

5.3 Single-top background

Single-top events containing a $W$ boson and at least one $b$-jet are, like $t\bar{t}$ events, an irreducible background for the $W+b$-jets signal. In the 2-jet region, where the single-top and $W+b$-jets contributions are comparable, kinematic observables are used to estimate the single-top normalization in data. The invariant mass of the combined $W$ boson and $b$-tagged jet system is computed for each event,\footnote{The $p_z$ of the neutrino is computed by setting the $W$ mass equal to the world average value of 80.399 GeV [58]. In case of complex solutions, only the real part is considered.} and the resulting $m(Wb)$ distribution, where single-top appears as a relatively narrow peak, is fitted. Example templates for the muon 2-jet sample are shown in figure 4.

The $W+\text{jets}$ and single-top template normalizations are free parameters of the fit. Gaussian constraints are applied to the multijet, $t\bar{t}$, $Z+\text{jets}$, and diboson backgrounds as described above. The single-top correction factors estimated by the fit in the electron and muon channel are, respectively, $1.13 \pm 0.15$ and $1.09 \pm 0.13$, and the corresponding fit projections are shown in figure 5.
Figure 3. CombNN distributions for the $b$-tagged (CombNN > 2.2) jet in the “at least 4 jets 1-$b$-tag” control region (top) and number-of-jets distributions in the “at least 1 jet 1-$b$-tag” region (bottom) in data and MC simulation. The muon (electron) channel is shown on the left (right). MC samples are normalized to the results of the respective ML fits.

Figure 4. Overlay of the $W$+jets, $t\bar{t}$ and single-top $m(Wb)$ distributions in MC simulation in the muon 2-jet analysis region.

These estimates of the single-top background contribution are verified with a fit to the $H_T = p_T^\ell + E_T^{\text{miss}} + \sum_{i=1}^{n} p_T^{\text{jet}}$ distribution, where single-top events are expected, on average, to be harder than $W$+jets events. The corresponding single-top correction factors (1.17 ± 0.17 and 1.12 ± 0.11 for the electron and muon channels, respectively) are consistent with those obtained from the $m(Wb)$ fit. The $m(Wb)$-derived factors are therefore used
to scale the ACERMC single-top prediction in the 2-jet region, and a 20% uncertainty is assigned to its normalization and applied as a Gaussian constraint in subsequent ML fits.

In the 1-jet region, where the expected single-top contribution is approximately half the size of the expected $W+b$-jets signal, the ACERMC prediction is used, and a large normalization uncertainty (50%) is assigned. This uncertainty reflects the difference observed in the single-top contribution estimated by the fit to the $m(Wb)$ distribution and the fit to the $H_T$ distribution in this analysis region.

### 5.4 $W+$jets backgrounds and signal

The different response of $b$-jets, $c$-jets, and light-jets to the CombNN $b$-tagging algorithm is used to separate statistically the $W+b$-jets component from the $W+c$-jets and $W+light$-jets ones in each analysis region. The $W+b$-jets, $W+c$-jets and $W+light$-jets normalizations are free parameters of the fit, while Gaussian constraints are applied to all other processes. In addition to the uncertainties discussed above for the multijet, $t\bar{t}$ and single-top backgrounds, 10% Gaussian constraints are assigned to the diboson and $Z+jets$ backgrounds as discussed previously.

The CombNN templates for the multijet component are extracted from data, while those from the other non-$W$ processes are extracted from the respective MC samples. For $W+b$-jets, $W+c$-jets and $W+light$-jets, the corresponding templates are prepared in each analysis region using large Pythia-generated samples.

The CombNN distributions normalized to the fit results are shown in figure 6, and the number of $W+b$-jets and background events estimated by the fits, along with their statistical uncertainties, are summarized in table 2. Table 3 shows the correction factors estimated by the fit to the data compared to the prediction for each process. While the electron and the muon sample backgrounds are treated as completely uncorrelated, the estimated background levels are found to be in good agreement across the channels. The behavior observed in the $Z+jets$ background prediction in table 2, when comparing the electron and muon channels in the 1-jet and 2-jet regions, is due to the different properties of $Z \to ee$ and $Z \to \mu\mu$ events in which one lepton is not reconstructed. In particular, $Z \to ee$ events tend to have a higher number of jets (from the missing electron), while $Z \to \mu\mu$
Figure 6. CombNN distributions for the $b$-tagged jet in data and MC simulation, where the MC samples are normalized to the results of the ML fit, for the 1-jet (top) and 2-jet (bottom) analysis regions, in the muon (left) and electron (right) channels.

Events tend to have higher $E_{\text{T}}^{\text{miss}}$ (from the missing muon). The $W+$light-jets estimate does not strongly affect the measured $W+b$-jets cross-section, as the $W+$light-jets and $W+b$-jets estimators have a 15% correlation. Instead, the $W+$light-jets and $W+c$-jets estimators are 90% anticorrelated, and a modest change in the dominant $W+c$-jets contribution can change the $W+$light-jets estimate significantly.

In the differential $p_T^{b\text{-jet}}$ measurement, a separate fit to the CombNN distribution is performed in each analysis region in four intervals of $b$-tagged jet $p_T$: 25–30 GeV, 30–40 GeV, 40–60 GeV and 60–140 GeV. The background contributions are extrapolated to each $p_T$ interval from the inclusive measurements, and the same Gaussian constraints as those of the inclusive fits are used. For the multijet background, this extrapolation is based on the $b$-tagged jet $p_T$ spectrum found in the multijet templates extracted from data. For all other backgrounds, the extrapolation is based on Monte Carlo simulation.

6 Cross-section extraction

The $W+b$-jets yields obtained from the CombNN fits are converted to a fiducial cross-section for $W+b$-jets times the branching ratio for each $W \to \ell\nu$ decay channel ($\ell = e, \mu$) using Monte Carlo simulation. The unfolding procedure is defined with respect to the fiducial region introduced in table 1. It accounts for trigger and object reconstruction efficiencies (including the $b$-jet identification efficiency) after applying corrections for all
Table 2. Estimated event yields for the eight contributions to the four analysis regions, including the statistical uncertainty from the binned ML fit.

<table>
<thead>
<tr>
<th>Process</th>
<th>$\mu$ 1-jet</th>
<th>$e$ 1-jet</th>
<th>$\mu$ 2-jet</th>
<th>$e$ 2-jet</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W+b$-jets</td>
<td>5300 ± 400</td>
<td>4800 ± 400</td>
<td>3000 ± 260</td>
<td>2220 ± 250</td>
</tr>
<tr>
<td>$W+c$-jets</td>
<td>15600 ± 600</td>
<td>13300 ± 500</td>
<td>4600 ± 400</td>
<td>4000 ± 400</td>
</tr>
<tr>
<td>$W$+light-jets</td>
<td>1600 ± 500</td>
<td>500 ± 500</td>
<td>1170 ± 330</td>
<td>490 ± 320</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>1230 ± 120</td>
<td>1100 ± 110</td>
<td>4300 ± 400</td>
<td>3690 ± 350</td>
</tr>
<tr>
<td>Single-top</td>
<td>1700 ± 500</td>
<td>1400 ± 500</td>
<td>2300 ± 400</td>
<td>1810 ± 350</td>
</tr>
<tr>
<td>Diboson</td>
<td>181 ± 18</td>
<td>139 ± 14</td>
<td>185 ± 18</td>
<td>155 ± 15</td>
</tr>
<tr>
<td>$Z$+jets</td>
<td>770 ± 70</td>
<td>258 ± 26</td>
<td>397 ± 40</td>
<td>365 ± 37</td>
</tr>
<tr>
<td>Multijet</td>
<td>780 ± 330</td>
<td>1000 ± 500</td>
<td>210 ± 150</td>
<td>1220 ± 290</td>
</tr>
</tbody>
</table>

Table 3. Correction factors estimated by the binned ML fit to the CombNN distribution for each process in the four analysis regions, including the statistical uncertainty. The multijet, $t\bar{t}$ and 2-jet single-top factors are given with respect to their estimate in data. The remaining factors are given with respect to the Monte Carlo expectations normalized to the NLO (single-top, diboson) and inclusive NNLO ($W/Z$+jets) cross-sections.

<table>
<thead>
<tr>
<th>Process</th>
<th>$\mu$ 1-jet</th>
<th>$e$ 1-jet</th>
<th>$\mu$ 2-jet</th>
<th>$e$ 2-jet</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W+b$-jets</td>
<td>1.68 ± 0.14</td>
<td>1.98 ± 0.16</td>
<td>1.14 ± 0.10</td>
<td>1.16 ± 0.13</td>
</tr>
<tr>
<td>$W+c$-jets</td>
<td>1.22 ± 0.04</td>
<td>1.30 ± 0.05</td>
<td>1.04 ± 0.09</td>
<td>1.10 ± 0.10</td>
</tr>
<tr>
<td>$W$+light-jets</td>
<td>0.70 ± 0.22</td>
<td>0.28 ± 0.25</td>
<td>1.15 ± 0.33</td>
<td>0.67 ± 0.44</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
<td>1.02 ± 0.10</td>
<td>1.01 ± 0.10</td>
</tr>
<tr>
<td>Single-top</td>
<td>1.07 ± 0.34</td>
<td>1.02 ± 0.36</td>
<td>1.08 ± 0.19</td>
<td>1.01 ± 0.19</td>
</tr>
<tr>
<td>Diboson</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
</tr>
<tr>
<td>$Z$+jets</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
<td>1.00 ± 0.10</td>
</tr>
<tr>
<td>Multijet</td>
<td>1.12 ± 0.47</td>
<td>0.80 ± 0.40</td>
<td>0.67 ± 0.49</td>
<td>1.79 ± 0.42</td>
</tr>
</tbody>
</table>

known detector effects. The small contribution (less than 5%) from $W \to \tau \nu$, where the $\tau$ decays to an electron or a muon, is not included in the fiducial region.

The Alpgen Monte Carlo simulation is used to produce correction factors to account for two effects: events passing the fiducial selection which fail the reconstructed-level selection, and events which pass the reconstructed-level selection but originate from outside the fiducial region. These factors are applied to the inclusive $W+b$-jets yield in each analysis region to obtain a fiducial cross-section.

The differential $W+b$-jets cross-section is also extracted, in the 1-jet and 2-jet regions, as a function of the transverse momentum of the leading $b$-jet, $p_{T}^{b}$, using the same bins as the CombNN differential fits. The measured quantity is therefore $d\sigma_{\text{fiducial}}/dp_{T}^{b}$. For this measurement, the correction factors mentioned above are produced in each $p_{T}^{b}$ bin.
ALPGEN Monte Carlo events which pass both reconstruction and fiducial selections are used to generate a response matrix to account for bin-to-bin migration effects between the reconstructed and generator-level distributions of $p_T^{b\text{-jet}}$. This response matrix is applied through an iterative Bayesian technique \cite{59}, in which the MC prediction is used as the initial prior, and three successive iterations are performed to remove the bias from the initial distribution.

The stability of the unfolding procedure is tested by comparing the unfolded spectra after three iterations with those obtained using two and four iterations, yielding consistent results. The bias introduced by the choice of prior is tested by creating an alternative sample, and unfolding it using the nominal response matrix; after three iterations, the unfolded distribution is significantly different from the nominal initial prior, and reproduces correctly the alternative generated distribution.

Different jet bins and lepton flavour channels are combined to yield more precise measurements of the $W+b\text{-jets}$ cross-section. In order not to introduce new assumptions on the background normalizations, the $W+b\text{-jets}$ yields are added after the CombNN fit, and their sum is unfolded using correction factors and response matrices obtained from Monte Carlo simulated events in the combined channels. This procedure is performed for each systematic variation and, in order to take into account the correlation of systematic uncertainties, correlated uncertainties are varied simultaneously in the samples being combined.

### 7 Systematic uncertainties

Several sources of systematic uncertainties on the measured $W+b\text{-jets}$ cross-section are considered. Each source may affect the background estimation in the control regions, the results of the CombNN fits, and the unfolding factors and response matrices. The strategy described here is used in all the jet multiplicity regions and $b\text{-jet}$ $p_T$ intervals.

The effect of each systematic source on the estimated number of $W+b\text{-jets}$ events is quantified using pseudo-experiments. For a given systematic variation, new sets of signal and background templates are prepared which may differ in both shape and normalization from the reference set used in the fit to data. The modified templates are used to generate pseudo-data samples that are fitted using the reference templates. In these pseudo-experiments, the same background constraints as those used in the fit to data are applied. Finally, the quoted fractional systematic uncertainty associated with a given source is defined as the ratio $(\bar{N}_{\text{fit}} - \mu_{\text{gen}}) / \mu_{\text{gen}}$, where $\bar{N}_{\text{fit}}$ is the mean of the estimator of the number of $W+b\text{-jets}$ events and $\mu_{\text{gen}}$ is the number of $W+b\text{-jets}$ events used in the pseudo-data generation for that particular systematic source.

The full analysis procedure is repeated for each systematic variation: the multijet, $t\bar{t}$ and single-top contributions are estimated in the corresponding control sample or distribution, and the CombNN distribution is fitted in the analysis regions after propagating the new background estimates.

Systematic uncertainties in the unfolding process are accounted for by using each systematically varied signal Monte Carlo sample to generate an alternative response matrix and set of unfolding factors. The difference in fiducial cross-section obtained when using the alternative Monte Carlo in place of the default one is quoted as the systematic uncertainty.
in the measurement. For systematic uncertainties that are split into an upwards and downwards variation, the unfolding is performed twice, and only the largest of the two resulting variations is taken as a symmetric uncertainty.

Most of the systematic effects considered here influence both the fitting and unfolding steps. In these cases, the systematic effects are propagated coherently and for a given systematic source the corresponding estimated \(W+b\)-jets yields are unfolded using the corresponding response matrix.

Background normalizations (multijet, \(t\bar{t}\), single-top, diboson and \(Z+jets\)) are treated as nuisance parameters of the ML fit to the CombNN distribution. As such, background normalization uncertainties are accounted for in the uncertainty on the number of \(W+b\)-jets events estimated by the fit. In the unfolding, the statistical uncertainty of the fiducial cross-section is evaluated using pseudo-experiments based on the uncertainty on the number of \(W+b\)-jets events estimated by the fit. For reference, if the background normalizations are fixed and the corresponding nuisance parameters are removed, the uncertainty on the number of \(W+b\)-jets events decreases by almost a factor of two.

The following effects are found to be non-negligible for the cross-section measurements:

**Jet energy scale and resolution.** The uncertainty on the jet energy scale (JES) is derived from data and from Monte Carlo simulation [42–47], and varies between 3% and 14% depending on the jet \(p_T\) and pseudorapidity. This uncertainty includes effects arising from the dependence of the jet response on the pile-up. It also accounts for differences between the calorimeter responses to light-quark-, gluon-, and heavy-quark-initiated jets, and for additional low-momentum jets found within \(\Delta R = 0.8\) of each jet considered. Uncertainties related to the jet energy resolution (JER) are derived from the jet response asymmetry measured in dijet events in data [42, 44].

The effects of the JES and JER uncertainties are quantified using alternative signal and background Monte Carlo templates in which the jet energy is modified by ±1\(\sigma\) or smeared, respectively. They represent the dominant sources of systematic uncertainties on the measured \(W+b\)-jets fiducial cross-sections and are found to be in the range 10–50%, depending on the jet multiplicity and \(p_T\) interval considered.

**Initial-state and final-state radiation (ISR/FSR).** Uncertainties on ISR and FSR affect the extrapolation of the \(t\bar{t}\) contribution in the analysis regions, as well the single-top Monte Carlo expectation in the 1-jet region and the data-driven single-top estimate in the 2-jet region. These effects are evaluated using the ACERMC generator interfaced to PYTHIA, and by varying the parameters controlling ISR and FSR in a range consistent with experimental data [60]. Their effect on the final cross-section measurements depends strongly on the jet \(p_T\) interval and varies between 2% and 30%.

**b-tagging efficiency calibration.** The calibration of the \(b\)-tagging efficiency is performed using control samples in data [49]. Uncertainties on these calibrations are estimated separately for light-jets, c-jets and \(b\)-jets as a function of the \(p_T\) and \(\eta\) of the jet [50, 51]. These uncertainties affect both the \(W+b\)-jets selection efficiency and the shape of the CombNN templates. The corresponding impact on the measured cross-section is estimated independently for \(b\)-jets, c-jets and light-jets to be in the range 1–8%.
Monte Carlo modelling. The uncertainty related to the ALPGEN $W+b$-jets fiducial acceptance modelling is estimated using alternative $W+b$-jets samples, generated using different settings. Specifically, the functional form of the factorization scale is varied; the set of parton distribution functions (PDF) is changed from CTEQ6L1 [61] to MRST2002LO [62]; the renormalization and factorization scales are halved and doubled; finally, the minimum jet $p_T$ used in the MLM matching is decreased (increased) to 15 (25) GeV, from the reference value of 20 GeV.

Because of the mild dependence of the CombNN template shape on the jet $p_T$, an additional systematic uncertainty due to the $c$-jets and $b$-jets $p_T$ modelling is quoted. This uncertainty is estimated using the full HERWIG parton shower sample of $c$-jets and $b$-jets in place of the ALPGEN matrix element ones. The full HERWIG parton shower $c$-jets and $b$-jets spectra are found to be softer than the corresponding spectra produced by the ALPGEN matrix element and the difference between the two is larger than any differences observed with the alternative ALPGEN samples mentioned above.

The systematic uncertainties on the measured cross-section related to the Monte Carlo modelling are in the range 2–8%.

Topological cluster energy scale and pile-up modelling. These account for the contribution to the $E_T^{\text{miss}}$ uncertainty due to uncertainties on the energy measurement of low-momentum jets and calorimeter cells that are not associated with electrons, muons or jets, as well as the uncertainty on the modelling of pile-up [63]. Their effect is estimated to be in the range 2–6%.

CombNN weight templates shape. Uncertainties on the CombNN shape of $b$-jets, $c$-jets and light-jets are quantified independently. These uncertainties affect the measured cross-section by changing the results of the CombNN fits, but they do not affect the unfolding process.

The systematic uncertainty associated with the $b$-jet CombNN shape is estimated using data. Events with at least four jets, two of which must be $b$-tagged, are selected. These events form a sample of $t\bar{t}$ candidates whose leading jet is a real $b$-jet approximately 95% of the time, as estimated in Monte Carlo simulation. This clean sample of $b$-jet candidates is used to compute the ratio of the corresponding CombNN distribution in data to that in MC simulation. This ratio is applied to all the $b$-jet templates used in the CombNN fit and the new set of templates is used to assess the corresponding systematic uncertainty.

The systematic uncertainty associated with the $c$-jet CombNN shape is estimated by preparing alternative CombNN templates for $c$-jets. Both the $c$-jet and $b$-jet CombNN shapes are sensitive to the description of the $b$-hadron and $c$-hadron branching ratios and in particular to the number of charged particles produced in their decay vertices. While it is possible to define a clean and almost unbiased control sample of $b$-jets in the data to check any possible discrepancy with the Monte Carlo simulation, the same is not feasible for $c$-jets. The alternative $c$-jet templates are then obtained by varying artificially the relative contribution of events with different track multiplicity associated with secondary vertices. Variations of 10% are considered in each track multiplicity bin, consistent with the relevant uncertainties in the $c$-hadron branching ratios.
Finally, a light-jet shape extracted from *Herwig* is used in place of the one obtained from *Pythia* to assess the systematic uncertainties on the light-jet CombNN shape. The effects of the variations of $b$-jet and $c$-jet CombNN shapes on the final cross-section measurements range between 2% and 8%, while the effect of the light-jet CombNN shape variation is negligible.

**Multijet background CombNN shape.** The systematic uncertainty on the multijet template shape is assessed using a control region defined by $E_{\text{T}}^{\text{miss}} < 25 \text{ GeV}$ and $m_{T}(W) < 40 \text{ GeV}$. Any mismodelling observed in this region is transported to the signal region and used to generate alternative multijet shapes, both in the electron and muon samples. The corresponding effect on the measured cross-section is larger in the electron sample where it is in the range 1–10% depending on the jet multiplicity and $p_{T}$ interval considered.

**Others.** Uncertainties related to the lepton trigger and reconstruction efficiencies are evaluated using tag-and-probe measurements in $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ events [36, 38]. Similarly, the $Z$-mass peak is used to determine the lepton momentum scales and resolutions and the corresponding uncertainties [36, 37]. The effect of these sources of uncertainties on the $W+b$-jets cross-section is between 1% and 2%. A 3.9% uncertainty on the integrated luminosity is also included [35].

8 Results

The unfolded result for the fiducial $W+b$-jets cross-section is presented in figure 7, while the measured differential $d\sigma/dp_{T}^{b}$ distributions are shown in figure 8. The numerical values corresponding to the combination of the electron and muon channels are shown in tables 4–6, where details of the systematic uncertainties and correlation matrices for the statistical and systematic uncertainties are also presented. The measured cross-sections for the 1-jet, 2-jet and 1+2-jet fiducial regions are:

$$
\sigma_{\text{fid}} (1 \text{ jet}) = 5.0 \pm 0.5 \text{ (stat)} \pm 1.2 \text{ (syst)} \text{ pb},$

$$
\sigma_{\text{fid}} (2 \text{ jet}) = 2.2 \pm 0.2 \text{ (stat)} \pm 0.5 \text{ (syst)} \text{ pb},$

$$
\sigma_{\text{fid}} (1+2 \text{ jet}) = 7.1 \pm 0.5 \text{ (stat)} \pm 1.4 \text{ (syst)} \text{ pb}.
$$

The results are compared to the NLO predictions of MCFM and Powheg, and to the *Alpgen* predictions scaled by the NNLO normalization factor for the inclusive $W$ cross-section [28]. Both the *Alpgen* and Powheg predictions implement a 4-flavour number scheme (4FNS) calculation, while the MCFM prediction, following the calculation described in ref. [9], includes terms which use the 5-flavour number scheme (5FNS) to account for the presence of $b$-quarks in the initial state originating from parton distribution functions.

The NLO predictions of MCFM and Powheg are evaluated using the MSTW2008 [64] NLO PDF, and the following dynamic renormalization and factorization scales ($\mu_R$ and
Figure 7. Measured fiducial cross-sections with the statistical (inner error bar) and statistical plus systematic (outer error bar) uncertainties in the electron, muon, and combined electron and muon channels. The cross-sections are given in the 1-jet, 2-jet, and 1+2-jet fiducial regions. The measurements are compared with NLO predictions calculated with MCFM [9] and corrected for hadronization and double-parton interaction (DPI) effects. The yellow bands represent the total uncertainty on the prediction. It is obtained by combining in quadrature the uncertainties resulting from variations of the renormalization and factorization scales, the PDF set, the DPI model and non-perturbative corrections. The NLO prediction from Powheg interfaced to Pythia, corrected for DPI effects, and the prediction from ALPGEN interfaced to HERWIG and Jimmy and scaled by the NNLO inclusive W normalization factor are also shown.

Figure 8. Measured differential W+b-jets cross-sections with the statistical plus systematic uncertainties as a function of $p_T^{b\text{-jet}}$ in the 1-jet (left) and 2-jet (right) fiducial regions, obtained by combining the muon and electron channel results. The measurements are compared to the MCFM predictions and to the ALPGEN predictions interfaced to HERWIG and Jimmy and scaled by the NNLO inclusive W normalization factor. The ratios between measured and predicted cross-sections are also shown.
Table 4. Measured fiducial $W+b$-jets cross-sections for the combination of the electron and muon channels with statistical and systematic uncertainties and breakdown of relative systematic uncertainties per jet multiplicity, and combined across jet bins.

\[ \mu_F = \mu_{F_R} = m_T^2 + p_T^2(t\nu) + \frac{m_b^2 + p_T^2(b)}{2} + \frac{m_{\bar{b}}^2 + p_T^2(\bar{b})}{2}. \] (8.1)

The PDF uncertainty is calculated with the MSTW2008 eigenvectors using the Hessian procedure [64]. The dependence of the result on the choice of scale, which dominates the theoretical uncertainty, is evaluated by varying the scale conservatively between a quarter and four times the value in equation (8.1), as in ref. [9]. These variations are used to

---

Footnote 5: In the 5-flavour number scheme, the production of one $b$-jet in the final state with an associated light jet can occur. In those cases, one of the two last terms in eq. (8.1) is omitted.
Fiducial cross-section, 1 jet

<table>
<thead>
<tr>
<th>$p_T^b$-jet [GeV]</th>
<th>[25, 30]</th>
<th>[30, 40]</th>
<th>[40, 60]</th>
<th>[60, 140]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d\sigma/dp_T^b$-jet [nb/GeV]</td>
<td>259</td>
<td>143</td>
<td>65</td>
<td>10.3</td>
</tr>
<tr>
<td>Statistical Uncertainty (%)</td>
<td>9</td>
<td>6</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Systematic Uncertainty (%)</td>
<td>24</td>
<td>19</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>Correlation coefficients of statistical uncertainties</td>
<td>1</td>
<td>0.415</td>
<td>$-0.38$</td>
<td>$-0.02$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$-0.01$</td>
<td>$-0.17$</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>Correlation coefficients of systematic uncertainties</td>
<td>1</td>
<td>0.893</td>
<td>0.740</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.887</td>
<td>0.750</td>
<td>0.875</td>
</tr>
</tbody>
</table>

Table 5. Measured fiducial $W+b$-jets cross-section in the 1-jet region with statistical and systematic uncertainties and their correlations in bins of $p_T^b$.

Fiducial cross-section, 2 jets

<table>
<thead>
<tr>
<th>$p_T^b$-jet [GeV]</th>
<th>[25, 30]</th>
<th>[30, 40]</th>
<th>[40, 60]</th>
<th>[60, 140]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d\sigma/dp_T^b$-jet [nb/GeV]</td>
<td>73</td>
<td>58</td>
<td>38</td>
<td>9.3</td>
</tr>
<tr>
<td>Statistical Uncertainty (%)</td>
<td>12</td>
<td>8</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Systematic Uncertainty (%)</td>
<td>26</td>
<td>22</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>Correlation coefficients of statistical uncertainties</td>
<td>1</td>
<td>0.585</td>
<td>$-0.45$</td>
<td>$-0.08$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.069</td>
<td>$-0.29$</td>
<td>$-0.20$</td>
</tr>
<tr>
<td>Correlation coefficients of systematic uncertainties</td>
<td>1</td>
<td>0.900</td>
<td>0.550</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.795</td>
<td>0.719</td>
<td>0.775</td>
</tr>
</tbody>
</table>

Table 6. Measured fiducial $W+b$-jets cross-section in the 2-jet region with statistical and systematic uncertainties and their correlations in bins of $p_T^b$.

calculate an asymmetric uncertainty before applying vetoes on additional jets. The effect of jet vetoes is then taken into account following the procedure outlined in ref. [65].

To compare the NLO calculations with data, the impact of non-perturbative effects and double-parton interactions has to be considered. The MCFM predictions are only available at the parton level, while the POWHEG predictions are interfaced with PYTHIA to model the non-perturbative effects of hadronization and the underlying event. A multiplicative correction derived from the POWHEG sample is therefore applied to the MCFM calculation.
to account for these non-perturbative effects. The uncertainty on the hadronization component of this correction is estimated by comparing the PYTHIA and HERWIG parton showers, while the uncertainty on the underlying event component is estimated using the alternative Perugia2011 [60] tune instead of the AUET2B [26] one. The effect of double-parton interactions, where a W boson and heavy-flavour jet are produced from different parton-parton interactions within the same proton, also has to be considered. Neither the MCFM nor the POWHEG calculations include this contribution, therefore an additive correction derived from the ALPGEN simulation interfaced to HERWIG and JIMMY has been applied to both calculations. This correction represents a 25% effect on the total cross-section, concentrated in the lowest momentum bins of the 1-jet region. The DPI contribution to ALPGEN has been shown to agree at the detector level with the ATLAS measurement of $\sigma_{\text{eff}}$ in the $W+2$-jet sample [66]. Based on this measurement, a $^{+39}_{-28}$% uncertainty is assigned to the DPI correction. The non-perturbative and DPI corrections for the 1-jet and 2-jets regions are presented in table 7. The fully corrected MCFM predictions are presented in table 8 for the 1-jet, 2-jet and 1+2-jet fiducial regions.

### 9 Results without single-top subtraction

The $W+b$-jets cross-section is also measured including the contribution of the single-top process. These measurements provide a complementary perspective on the $W+b$-tagged-jet sample, and they have a higher statistical precision than the single-top subtracted ones, especially at high $p_T^{b}$.

For each analysis region and $p_T^{b}$ bin, the same ML fit as for the $W+b$-jets measurement is used, as well as the same estimates and constraints for the multijet, $t\bar{t}$, Z+jets and

<table>
<thead>
<tr>
<th>Correction</th>
<th>1 jet</th>
<th>2 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-perturbative</td>
<td>$0.92 \pm 0.02$ (had.) $\pm 0.03$ (UE)</td>
<td>$0.96 \pm 0.05$ (had.) $\pm 0.03$ (UE)</td>
</tr>
<tr>
<td>DPI [pb]</td>
<td>$1.02 \pm 0.05$ (stat) $^{+0.40}_{-0.29}$ (syst)</td>
<td>$0.32 \pm 0.02$ (stat) $^{+0.12}_{-0.09}$ (syst)</td>
</tr>
</tbody>
</table>

**Table 7.** Multiplicative correction factors for non-perturbative effects and additive corrections for double-parton interactions, derived from the ALPGEN simulation, applied to the MCFM and POWHEG predictions for the comparisons with unfolded results. The non-perturbative uncertainties include the hadronization and underlying event modelling, while the DPI uncertainties are based on the ATLAS measurement of $\sigma_{\text{eff}}$ [66].

<table>
<thead>
<tr>
<th>MCFM NLO prediction [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 jet</td>
</tr>
<tr>
<td>2 jets</td>
</tr>
<tr>
<td>1+2 jets</td>
</tr>
</tbody>
</table>

**Table 8.** Theoretical NLO predictions for the $W+b$-jets fiducial cross-section for one lepton flavour calculated with the MCFM program, corrected for non-perturbative effects and DPI contributions.
diboson backgrounds. In the fit to the CombNN distribution, the $W+b$-jets and single-top templates are merged accounting for their respective predicted cross-sections, and they form a single template whose normalization is estimated. As a consequence of the single-top process being considered as part of the signal, the number of nuisance parameters in the fit is reduced, thereby increasing its statistical precision. After the CombNN fit, the number of estimated $W+b$-jets and single-top events is unfolded to a common fiducial region, identical to the $W+b$-jets fiducial region, using correction factors and a response matrix built from the sum of the two Monte Carlo samples.

The systematic uncertainties from the fit and unfolding steps are accounted for using the same methods as for the single-top subtracted measurement. An additional uncertainty is introduced to account for the relative normalization of $W+b$-jets and single-top. Alternative samples, in which the amounts of $W+b$-jets and single-top are doubled in turn, are used to perform the unfolding. The largest deviation obtained with respect to the nominal result, approximately 5%, is then quoted as a separate systematic uncertainty.

The resulting fiducial cross-sections for $W+b$-jets plus single-top, combining the electron and muon channels, are:

$$\sigma_{\text{fid}} (1 \text{ jet}) = 5.9 \pm 0.2 \text{ (stat)} \pm 1.3 \text{ (syst)} \text{ pb},$$

$$\sigma_{\text{fid}} (2 \text{ jet}) = 3.7 \pm 0.1 \text{ (stat)} \pm 0.8 \text{ (syst)} \text{ pb},$$

$$\sigma_{\text{fid}} (1+2 \text{ jet}) = 9.6 \pm 0.2 \text{ (stat)} \pm 1.7 \text{ (syst)} \text{ pb}.$$

The corresponding expected cross-sections, calculated for the $W+b$-jets process using ALPGEN interfaced to HERWIG and JIMMY and scaled by the NNLO inclusive $W$ normalization factor and for the single-top processes using ACERMC interfaced to PYTHIA and scaled to NLO, are 3.6 pb, 3.0 pb and 6.6 pb, respectively. The differential results as a function of $p_T^{b\text{-jet}}$ are presented in figure 9 and tables 9 and 10.

10 Conclusions

A measurement of the cross-section of $W$ boson production in association with $b$-jets at $\sqrt{s} = 7$ TeV is presented, based on data corresponding to an integrated luminosity of 4.6 fb$^{-1}$ collected with the ATLAS detector at the LHC. The measurement is performed with a single $b$-tagged jet requirement in the $W+1$-jet and $W+2$-jets samples. $W+b$-jets yields are estimated separately in the electron and muon decay channel and unfolded to a common fiducial region. Good agreement is found between the results in the electron and muon channels and combined measurements are provided.

In the 1-jet region, the measured fiducial cross-section is $5.0 \pm 0.5 \text{ (stat)} \pm 1.2 \text{ (syst)} \text{ pb}$, consistent within 1.5$\sigma$ with NLO predictions. In the 2-jet region, the measured fiducial cross-section is $2.2 \pm 0.2 \text{ (stat)} \pm 0.5 \text{ (syst)} \text{ pb}$, in good agreement with the theoretical calculations. As a result, the combined 1+2-jet measurement, yielding a cross-section of $7.1 \pm 0.5 \text{ (stat)} \pm 1.4 \text{ (syst)} \text{ pb}$, is found to be consistent within 1.5$\sigma$ with the MCFM NLO prediction, corrected for hadronization and DPI effects, of $4.70 \pm 0.09 \text{ (stat)}^{+0.60}_{-0.49} \text{ (scale)} \pm 0.06 \text{ (PDF)} \pm 0.16 \text{ (non-pert)}^{+0.52}_{-0.38} \text{ (DPI)} \text{ pb}$.
Figure 9. Measured differential $W+b$-jets cross-section without single-top subtraction as a function of $p_T^{b\text{-jet}}$ in the 1-jet (left) and 2-jet (right) samples, obtained by combining the electron and muon channels. The measurements are compared to the $W+b$-jets plus single-top predictions obtained using ALPGEN interfaced to HERWIG and JIMMY and scaled by the NNLO inclusive $W$ normalization factor plus ACERMC interfaced to PYTHIA and scaled to the NLO single-top cross-section. The ratios between measured and predicted cross-sections are also shown.

<table>
<thead>
<tr>
<th>$p_T^{b\text{-jet}}$ [GeV]</th>
<th>[25, 30]</th>
<th>[30, 40]</th>
<th>[40, 60]</th>
<th>[60, 140]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d\sigma/dp_T^{b\text{-jet}}$ [nb/GeV]</td>
<td>278</td>
<td>156</td>
<td>80</td>
<td>15.7</td>
</tr>
<tr>
<td>Statistical Uncertainty (%)</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Systematic Uncertainty (%)</td>
<td>23</td>
<td>15</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Correlation coefficients of statistical uncertainties</td>
<td>1</td>
<td>0.401</td>
<td>-0.31</td>
<td>-0.03</td>
</tr>
<tr>
<td>Correlation coefficients of systematic uncertainties</td>
<td>1</td>
<td>0.840</td>
<td>0.682</td>
<td>0.866</td>
</tr>
</tbody>
</table>

Table 9. Measured fiducial $W+b$-jets cross-section without single-top subtraction in the 1-jet region, with statistical and systematic uncertainties and their correlations in bins of $p_T^{b\text{-jet}}$.

A differential cross-section measurement as a function of the leading $b$-jet $p_T$ is also presented, for jets in the $p_T$ range between 25 GeV and 140 GeV. In the 1-jet fiducial region, the measured cross-section is larger than the NLO predictions, but compatible within the theoretical and experimental uncertainties. The same measurement in the 2-jet fiducial region is found to be in agreement with the theoretical predictions.
A second set of measurements, including the single-top contribution, is also presented. In the 1-jet fiducial region, the measured $W+b$-jets plus single-top cross-section is $5.9 \pm 0.2 \text{ (stat)} \pm 1.3 \text{ (syst)}$ pb, while in the 2-jet fiducial region it is $3.7 \pm 0.1 \text{ (stat)} \pm 0.8 \text{ (syst)}$ pb. The combined 1+2-jet fiducial cross-section is measured to be $9.6 \pm 0.2 \text{ (stat)} \pm 1.7 \text{ (syst)}$ pb. The corresponding $b$-jet $p_T$ differential cross-sections have significantly reduced uncertainties with respect to the single-top subtracted measurements and can be compared to combined single top-quark and $W+b$-jets calculations in the future.

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNSRC and Lundbeck Foundation, Denmark; EPLANET, ERC and NSRF, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, DFG, HGF, MPG and AvH Foundation, Germany; GSRT and NSRF, Greece; ISF, MINERVA, GIF, DIP and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; BRF and RCN, Norway; MNiSW, Poland; GRICES and FCT, Portugal; MERSYS (MECTS), Romania; MES of Russia and ROSATOM, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARRS and MVZT, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC,
the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[46] ATLAS collaboration, Probing the measurement of jet energies with the ATLAS detector using Z + jet events from proton-proton collisions at \( \sqrt{s} = 7 \) TeV, ATLAS-CONF-2012-053 (2012).

[47] ATLAS collaboration, Probing the measurement of jet energies with the ATLAS detector using photon+jet events in proton-proton collisions at \( \sqrt{s} = 7 \) TeV, ATLAS-CONF-2012-063 (2012).


[52] ATLAS collaboration, b\(^{-}\)-jet tagging calibration on c\(^{+}\)-jets containing D\(^{*}\)\(^{+}\) mesons, ATLAS-CONF-2012-039 (2012).


Also at Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia

Also at Nevis Laboratory, Columbia University, Irvington NY, United States of America

Also at Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom

Also at Department of Physics, Oxford University, Oxford, United Kingdom

Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America

Also at Discipline of Physics, University of KwaZulu-Natal, Durban, South Africa

* Deceased