Measurement of the cross section for the production of a W boson in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector


DOI
10.1016/j.physletb.2011.12.046

Publication date
2012

Document Version
Final published version

Published in
Physics Letters B

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).
Measurement of the cross section for the production of a $W$ boson in association with $b$-jets in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector

ATLAS Collaboration

1. Introduction

A measurement is presented of the cross section for the production of a $W$ boson with one or two $b$-jets in proton–proton collisions at a centre-of-mass energy of 7 TeV. Production via top decay is not included in the signal definition. This measurement provides an important test of quantum chromodynamics (QCD) as it is sensitive to heavy-flavour quarks in the initial state. The measurement is based on 35 pb$^{-1}$ of data collected with the ATLAS detector at the LHC. The $W + b$-jet cross section is defined for jets reconstructed with the anti-$k_t$ clustering algorithm with transverse momentum above 25 GeV and rapidity within ±2.1. The $b$-jets are identified by reconstructing secondary vertices. The fiducial cross section is measured both for the electron and muon decay channel of the $W$ boson and is found to be 10.2 ± 1.9(stat) ± 2.6(syst) pb for one lepton flavour. The results are compared with next-to-leading order QCD calculations, which predict a cross section smaller than, though consistent with, the measured value.

© 2011 CERN. Published by Elsevier B.V. All rights reserved.

Table 1
Definition of the phase space for the fiducial cross section. The $W$ transverse mass is defined as $m_T = \sqrt{2p_T^Wp_T^\ell(1 - \cos(\phi^\ell - \phi^W))}$.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton transverse momentum</td>
<td>$p_T^\ell &gt; 20$ GeV</td>
</tr>
<tr>
<td>Lepton pseudorapidity</td>
<td>$</td>
</tr>
<tr>
<td>Neutrino transverse momentum</td>
<td>$p_T^\nu &gt; 25$ GeV</td>
</tr>
<tr>
<td>$W$ transverse mass</td>
<td>$m_T &gt; 40$ GeV</td>
</tr>
<tr>
<td>Jet transverse momentum</td>
<td>$p_T^j &gt; 25$ GeV</td>
</tr>
<tr>
<td>Jet rapidity</td>
<td>$</td>
</tr>
<tr>
<td>Jet multiplicity</td>
<td>$n \leq 2$</td>
</tr>
<tr>
<td>$b$-jet multiplicity</td>
<td>$n_b = 1$ or $n_b = 2$</td>
</tr>
<tr>
<td>Jet-lepton separation</td>
<td>$\Delta R(j,\ell) &gt; 0.5$</td>
</tr>
</tbody>
</table>

The measurement is based on data corresponding to an integrated luminosity of 35 pb$^{-1}$ and is compared with QCD NLO predictions. A closely related measurement has been performed, using very similar techniques, in the $Z + b$-jet final state.

2. The ATLAS detector

The ATLAS detector consists of an inner detector tracking system (ID) surrounded by a superconducting solenoid providing a 2 T magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer (MS). The ID consists of pixel and silicon 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 $\eta = -\ln\tan(\theta/2)$. The distance $\Delta R$ in $\eta$–$\phi$ space is defined as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$.
microstrip detectors inside a transition radiation tracker. The electromagnetic calorimeter is a lead-liquid-argon (LAr) detector in the barrel ($|\eta| < 1.475$) and the endcap ($1.375 < |\eta| < 3.2$) regions. Hadron calorimetry is based on two different detector technologies. The barrel ($|\eta| < 0.8$) and extended barrel ($0.8 < |\eta| < 1.7$) calorimeters are composed of scintillator/steel, while the hadronic endcap calorimeters ($1.5 < |\eta| < 3.2$) are LAr/copper. The forward calorimeters ($3.1 < |\eta| < 4.9$) are instrumented with LAr/tungsten and LAr/copper, providing electromagnetic and hadronic energy measurements, respectively. The MS consists of three large superconducting toroids and a system of three stations of trigger chambers and precision tracking chambers.

3. Simulated event samples

Monte Carlo (MC) simulated event samples with full detector simulation [17], based on the GRANT4 program [18] corrected for all known detector effects, are used to model the W + b-jet signal and most of the backgrounds, as well as to unfold the measured W + b-jet yield to obtain the fiducial cross section.

The processes of W boson production in association with b-jets, c-jets and light-flavour jets are simulated separately using the Alpgen [19] generator, interfaced to HERWIG [20] for parton shower and fragmentation, and JIMMY [21] for the underlying event simulation. The MLM [22] matching scheme as implemented in Alpgen is used to remove overlaps between the n and n + 1 parton samples from the matrix element (ME) and the parton shower. In addition, overlap between heavy-flavour quarks that originate from ME production and those that originate from the parton shower is removed.

The Z + jets background is simulated with Alpgen interfaced to HERWIG using the same configuration as for W + jets. The di-boson (WW, WZ, ZZ) background is simulated with HERWIG. The t-channel and Wt-channel single top processes are simulated with AcerMC [23], while the s-channel process is simulated with MC@NLO [24]. The inclusive W + jets and Z + jets cross sections are normalized to NNLO predictions [25], and the cross sections of the other backgrounds are normalized to NLO predictions [26]. The tt background is simulated with MC@NLO interfaced to HERWIG. The tt normalisation is extracted from the data.

Multiple interactions per bunch crossing are accounted for by overlaying simulated minimum bias events. To match the observed instantaneous luminosity profile, the MC events are reweighted to yield the same distribution of the number of primary vertices as measured in the data.

4. Event selection

The analysis is based on the 2010 data set using 35 pb⁻¹ of integrated luminosity with an uncertainty of 3.4% [27,28]. The data are collected using a single electron or muon high p_T trigger. Trigger thresholds are low enough to ensure that leptons with p_T > 20 GeV lie in the efficiency plateau. All events are required to have a primary vertex that is reconstructed from at least three tracks with p_T > 150 MeV.

Final states are selected with exactly one isolated electron or muon. Electrons are required to satisfy ET > 20 GeV and |\eta| < 2.47. Electrons in the region between the barrel and the endcap electromagnetic calorimeters ($1.37 < |\eta| < 1.52$) are removed. In addition to the tight selection as defined in Ref. [29], a p_T- and \eta-dependent requirement on a combination of calorimeter and track isolation is designed to yield constant efficiency and to reduce the large background from multi-jet production. Muon candidates are constructed from matched ID and MS tracks and are required to satisfy p_T > 20 GeV and |\eta| < 2.4. Muons within a distance ΔR < 0.4 of a jet are rejected. In addition the calorimeter transverse energy and the sum of track transverse momenta within ΔR < 0.3 of the muon must both be less than 4 GeV.

Jets are reconstructed using the anti-k_T [30] algorithm with a radius parameter R = 0.4. To take into account the differences in calorimeter response to electrons and hadrons, a p_T- and \eta-dependent factor, derived from simulated events, is applied to each jet to provide an average energy scale correction [31] back to particle level. Events with one or two reconstructed jets are selected with jet p_T > 25 GeV and rapidity |y| < 2.1. All jets within ΔR < 0.5 of a selected electron are removed. Jets produced in additional interactions are removed by requiring that 75% of the sum of the transverse momenta of the tracks associated to each jet is consistent with originating from the primary vertex.

The reconstruction of the missing transverse momentum (E_T^{miss}) [29] is based on the energy deposits in calorimeter cells grouped into three-dimensional clusters. Corrections for electromagnetic to hadronic energy scale, dead material, out-of-cluster energy as well as muon momentum for the muon channel are applied. The W boson transverse mass (m_T) is calculated from the measured lepton momentum, the missing transverse momentum and the opening angle between the two according to the formula m_T = \sqrt{2p_T^e p_T^\nu(1 - \cos(\phi^e - \phi^\nu))}. For both lepton channels E_T^{miss} > 25 GeV and m_T > 40 GeV is required.

The algorithm used to tag b-jets, SV0 [32], is based on the decay length significance between the primary vertex and the displaced secondary vertex reconstructed in the jet. Jets with a decay length significance greater than 5.85 are considered to be b-jet candidates, referred to as b-tagged jets. This working point of the SV0 algorithm ensures about 35% efficiency for b-jets with a mistag rate of about 0.3%, and 8% for light- and c-jets, respectively. The b-tagging efficiency is measured in a sample enriched in b-jets by requiring that the jet contains a muon, which is expected to come predominantly from a semileptonic b hadron decay [33]. The muon momentum relative to the jet axis, referred to as p_T^\mu, is used to discriminate b-jets from c- and light-flavour jets. The ratio of the b-tagging efficiency measured in data and in the MC simulation is applied to the simulated samples in the form of a correction factor. This correction factor does not show any strong dependence on jet p_T or \eta and is consistent with unity. The total uncertainty on the correction factor ranges from 6% to 13%. These results are confirmed with independent b-tagging efficiency measurements in tt events and alternatively using partial reconstruction of b hadrons in jets in D^0 \mu final states [33].

The overall fraction of W + b-jet events with two b-tagged jets is negligible (2%). Most of the W + b-jet events with two true b-jets are reconstructed as events with one b-jet candidate. This is due to the requirement of central and high p_T jets and to the b-tagging efficiency of about 35%. In addition, events containing more than one b-jet candidate are predicted to be dominated by tt. Therefore events are selected with one and only one tagged jet despite the measurement also being sensitive to the production of W + b-jet with two true b-jets.

5. Background estimation and cross section extraction

Charm hadrons also have an appreciable lifetime which can result in reconstructed displaced secondary vertices. Light-flavour jets can also be misidentified as b-jets due to hadronic interactions and photon conversions in detector material, long-lived light-flavour hadrons like K^0 and wrongly reconstructed displaced vertices. The invariant mass of the secondary vertex, m_{3\ell}\nu, is correlated with the mass of the parent hadron and thus discriminates between b-, c- and light-flavour jets. The number of
**Fig. 1.** (Top) \(E_{\text{miss}}\) distribution in the electron channel in the combined 1- and 2-jet bin without applying the \(E_{\text{miss}}\) selection criterion. (Bottom) \(m_{T}\) distribution in the muon channel in the 1-jet bin without applying the \(m_{T}\) selection criterion after applying the \(b\)-tagging requirement. Non-multi-jet contributions are normalized to their MC predictions.

\(W + b\)-jet events is extracted from data by fitting the measured \(m_{SV}\) distribution with a linear combination of templates for \(b\)-, \(c\)- and light-flavour jets using a binned maximum likelihood fit, while the expected contributions from non-\(W +\) jets background processes are constrained in the fit using the estimated template shapes and normalisations. The \(m_{SV}\) is calculated from the tracks associated to the secondary vertex assuming they are pions. The fit procedure is validated with simulated pseudo-experiments with flavour compositions and background levels similar to the measured ones.

The non-\(W +\) jets background sources comprise top quark pair, single top, multi-jet and the other electroweak (EW) production processes, \(Z +\) jets and dibosons.

The \(t\bar{t}\) background is estimated from data by applying the same secondary vertex mass fit to a control region enriched in \(t\bar{t}\) using the same event selection except requiring four or more jets instead of one or two. Backgrounds to the \(t\bar{t}\) process are estimated in the same way as for the fit in the signal region. The \(W + b\)-jet contamination in the \(t\bar{t}\) control region is at the 5% level and is extrapolated from the measured yield in the signal region using \textsc{alpgen} and an uncertainty of \(\pm 100\%\). The measured \(t\bar{t}\) yield, \(n_{t\bar{t}}^{2+\text{jets}}\), is then projected into the signal region using MC simulation: \(n_{t\bar{t}}^{2+\text{jets}} = n_{t\bar{t}, \text{measured}}^{2+\text{jets}} \cdot \frac{1}{n_{t\bar{t}, \text{expected}}^{2+\text{jets}}}.\) This data-driven \(t\bar{t}\) yield estimate is in good agreement with MC@NLO prediction and has the advantage that it is almost completely independent of the \(b\)-tagging uncertainty. The \(ttm_{SV}\) template is modelled using MC simulation.

\(W + b\)-jet events is extracted from data by fitting the measured \(m_{SV}\) distribution with a linear combination of templates for \(b\)-, \(c\)- and light-flavour jets using a binned maximum likelihood fit, while the expected contributions from non-\(W +\) jets background processes are constrained in the fit using the estimated template shapes and normalisations. The \(m_{SV}\) is calculated from the tracks associated to the secondary vertex assuming they are pions. The fit procedure is validated with simulated pseudo-experiments with flavour compositions and background levels similar to the measured ones.

The non-\(W +\) jets background sources comprise top quark pair, single top, multi-jet and the other electroweak (EW) production processes, \(Z +\) jets and dibosons.

The \(t\bar{t}\) background is estimated from data by applying the same secondary vertex mass fit to a control region enriched in \(t\bar{t}\) using the same event selection except requiring four or more jets instead of one or two. Backgrounds to the \(t\bar{t}\) process are estimated in the same way as for the fit in the signal region. The \(W + b\)-jet contamination in the \(t\bar{t}\) control region is at the 5% level and is extrapolated from the measured yield in the signal region using \textsc{alpgen} and an uncertainty of \(\pm 100\%\). The measured \(t\bar{t}\) yield, \(n_{t\bar{t}}^{2+\text{jets}}\), is then projected into the signal region using MC simulation: \(n_{t\bar{t}}^{2+\text{jets}} = n_{t\bar{t}, \text{measured}}^{2+\text{jets}} \cdot \frac{1}{n_{t\bar{t}, \text{expected}}^{2+\text{jets}}}.\) This data-driven \(t\bar{t}\) yield estimate is in good agreement with MC@NLO prediction and has the advantage that it is almost completely independent of the \(b\)-tagging uncertainty. The \(ttm_{SV}\) template is modelled using MC simulation.

As the multi-jet background is difficult to model with simulation, data-driven techniques similar to those described in Ref. [11] are used to estimate this background in each jet multiplicity bin and lepton flavour. The multi-jet background in the electron channel arises mainly from non-prompt electrons and a small amount of fake electrons such as electrons from photon conversions and misidentified jets with high electromagnetic fractions. A binned likelihood template fit of the \(E_{\text{miss}}\) distribution is used to estimate the multi-jet background. The \(E_{\text{miss}}\) template for multi-jet events is modelled using a complementary data sample where the full event selection including the \(b\)-tagging requirement is satisfied but electrons are required to fail certain selection criteria and to satisfy a looser identification requirement. This selection is dominated by multi-jet events. The \(E_{\text{miss}}\) template for the other contributions is modelled using MC simulation. The results of the fit are shown in Fig. 1, where the fit region goes from 0–100 GeV. The \(m_{SV}\) template is modelled using the same control region used to model the multi-jet \(E_{\text{miss}}\) template.

The muon multi-jet background is dominated by non-prompt muons and extracted using the matrix method [11]. The method is based on the difference in efficiency for a ‘real’ (prompt) or a ‘fake’ (non-prompt) muon that satisfies a loose selection criterion, to also satisfy the standard selection criterion. Fig. 1 illustrates that the muon multi-jet background is well modelled with this method. The shape of the \(m_{SV}\) template is modelled using a control re-
and are converted into a region enriched in multi-jet events where the full event selection is the same as in the legend. 

The systematic uncertainties on the measured $W + b$-jet cross section are derived from the non-$W + b$-jet background estimate, the modelling of the $m_{SV}$ templates and the correction factor of the fitted $W + b$-jet event distributions to derive the cross section. All correlations between systematic uncertainties are accounted for.

The largest uncertainty is related to the calibration of the $b$-tagging efficiency, which impacts not only the $W + b$-jet acceptance and efficiency, but also the template shapes and the normalisation of the single top background. The uncertainty on the $b$-tagging efficiency is estimated to be between 6% for high jet $p_T$ to 13% at the low $p_T$ end of 25 GeV [33]. The uncertainty is driven by the $b$-decay modelling, the MC statistics, the modelling of the muon $p_T$ spectrum and the uncertainty on the jet energy scale. The impact of the $b$-tagging efficiency uncertainty on the $b$-jet background is strongly reduced since this background is extracted from data.

The systematic uncertainties on the $m_{SV}$ templates are evaluated from direct comparisons of the $m_{SV}$ shapes of the data and the simulation in three multi-jet control regions (an example of the agreement between data and simulation in such control regions can be seen in Fig. 19 of Ref. [33]). Two of these control regions are used to determine systematic uncertainties on the bottom and charm template shapes. Since both charm and bottom...

\[ \sigma_{W+b\text{-}jet} \times B(W \rightarrow \ell\nu) = \frac{n_{tag} \cdot f_{W+b\text{-}jet}}{L (dt \cdot U)}, \]

where the index $j = 1, 2$ indicated the jet multiplicity, $n_{tag}$ the number of selected events with exactly one $b$-tagged jet, $f_{W+b\text{-}jet}$ the fitted fraction of signal events, $f_{W+b\text{-}jet}$ the integrated luminosity, and $U$ the $W + b$-jet correction factor which includes the acceptance and efficiency effects.

The correction factor is calculated from the simulation as the ratio of $W + b$-jet events which satisfy the offline selection requirements to the $W + b$-jet events which satisfy the fiducial particle-level selection criteria, summarized in Table 1. At the particle level jets are reconstructed with the anti-$k_t$ algorithm using all stable particles ($\tau > 10$ ps), $b$-jets are defined by the presence of a $b$ hadron with $p_T > 5$ GeV associated to the jet requiring $\Delta R(jet, b$ hadron) $< 0.3$, and only weakly-decaying $b$ hadrons are considered. Leptons are defined by including the energy of all radiated photons within $\Delta R = 0.1$ around the lepton.

The small contributions to the measured $W + b$-jet yield from $W \rightarrow \tau\nu$ decays (less than 5%, where the $\tau$ decays to an electron or muon) are treated as background and corrected for. The final correction factors are 0.17 and 0.21 in the electron channel and 0.23 and 0.28 in the muon channel for the 1-jet and 2-jet bin, respectively. The correction factor is dominated by the $b$-tagging requirement which has an efficiency of about 35%. The correction factor in the electron channel is smaller than in the muon channel due to tighter electron selection in order to reduce the larger multi-jet background. Relative uncertainties on the correction factors vary between 12% and 14% and are dominated by the uncertainty on the $b$-tagging efficiency, as discussed below.

6. Systematic uncertainties

Systematic uncertainties on the measured $W + b$-jet cross section are derived from the non-$W + b$-jet background estimate, the modelling of the $m_{SV}$ templates and the correction factor of the fitted $W + b$-jet event distributions to derive the cross section. All correlations between systematic uncertainties are accounted for.

The largest uncertainty is related to the calibration of the $b$-tagging efficiency, which impacts not only the $W + b$-jet acceptance and efficiency, but also the template shapes and the normalisation of the single top background. The uncertainty on the $b$-tagging efficiency is estimated to be between 6% for high jet $p_T$ to 13% at the low $p_T$ end of 25 GeV [33]. The uncertainty is driven by the $b$-decay modelling, the MC statistics, the modelling of the muon $p_T$ spectrum and the uncertainty on the jet energy scale. The impact of the $b$-tagging efficiency uncertainty on the $b$-jet background is strongly reduced since this background is extracted from data.

The systematic uncertainties on the $m_{SV}$ templates are evaluated from direct comparisons of the $m_{SV}$ shapes of the data and the simulation in three multi-jet control regions (an example of the agreement between data and simulation in such control regions can be seen in Fig. 19 of Ref. [33]). Two of these control regions are used to determine systematic uncertainties on the bottom and charm template shapes. Since both charm and bottom...

Fig. 3. $m_{SV}$ distributions for the $b$-tagged jet in data and MC, where the $W + jets$ samples are normalized to the results of the maximum likelihood fit and non-$W + jets$ backgrounds are normalized to the estimates as given in the text.

Fig. 4. Invariant mass of the $W + b$-jet system in the electron channel. The neutrino $p_T$ is obtained by imposing the $W$ invariant mass and using the smallest in absolute value of the two solutions. The $W + jets$ samples are normalized to the results of the maximum likelihood fit and non-$W + jets$ backgrounds are normalized to the estimates as given in the text.
jet tags are caused by displaced tracks from real vertex decays, it is natural to determine their uncertainties together from control regions that enhance the heavy-flavour fractions. One of these control regions is taken from events in which two jets are $b$-tagged, increasing the probability that both of the selected jets are from heavy-flavour production. The other region is taken from $b$-tagged jets which are also required to contain muons, which is very rare for light-flavour jets. Both of these control regions are determined to have a light-flavour contamination of less than 10%. The bottom and charm $m_{SV}$ templates used in the $W$ + jets fit are then transformed simultaneously by multiplying by the ratio of the data to the simulation in the control region for each $m_{SV}$ bin. The shapes of the simulated heavy-flavour backgrounds (in particular the top backgrounds) are also transformed simultaneously. In each lepton channel, out of the two control regions, the transformation resulting in the larger variation is chosen to assess the systematic uncertainty.

Additional studies are performed to account for the possibility that the charm and the bottom templates may not transform in exactly the same manner. This is tested by transforming the charm and the bottom templates one at a time instead of together. It is observed that varying both the charm and the bottom templates together leads to the maximum systematic bias, with most of the effect coming from the distortion of the $b$-template shape, and only about a third of the effect coming from the distortion of the charm template shape. The reason that the charm shape plays such a small role in the fit results is that the template shapes below about $m_{SV} = 1.5$ GeV do not strongly influence the final fitted $b$-normalisation. The $b$-normalisation is mostly constrained by the high $m_{SV}$ tail where there is very little background, especially in the one jet fits. In fact, fitting the $m_{SV}$ distribution only for $m_{SV} > 1.5$ GeV does not considerably reduce the analysis sensitivity or bias the final results.

The uncertainty on the measured $t\bar{t}$ yield in the $\geq 4$-jet bin is dominated by the limited data statistics. The number of $t\bar{t}$ events is alternatively estimated using a tag-counting method [33]. The use of simulated $t\bar{t}$ samples for the projection from the $\geq 4$-jet bin gives rise to systematic uncertainties from the choice of generator, the amount of QCD initial and final state radiation (ISR/FSR) and uncertainties on the PDF. The uncertainty due to the choice of generator is evaluated by comparing the predictions of MC@NLO with those of POWHEG [34–36] interfaced to either HERWIG or PYTHIA [37]. The dominant uncertainty is represented ISR/FSR, and it is evaluated by studies using the AcerMC generator interfaced to PYTHIA, and by varying the parameters controlling ISR and FSR in a range consistent with experimental data [38]. The uncertainty in the PDFs used to generate $t\bar{t}$ events is evaluated using a range of current PDF sets with the procedure described in Ref. [38]. ISR/FSR and PDF uncertainties are evaluated in the same way for the single top background.

Both the $t\bar{t}$ and single top background are irreducible in the sense that both backgrounds contain a $W$ boson, at least one $b$-jet, and additional jets. While the $t\bar{t}$ background is extracted from the data, this is not possible for single top due to the limited statistics. Therefore, more details are given here on the single top background. The selection efficiency for single top is considerably larger than for the $W + b$-jet signal, mainly due to the different $p_T$ spectrum of the $b$-jet. The corresponding single top fiducial cross sections as defined in Table 1 for one lepton flavour are 1.4 pb and 1.8 pb in the 1-jet and 2-jet bin, respectively. The secondary vertex mass shapes for the single top background and $W + b$-jet signal are found to be in good agreement. The invariant mass distribution of $W + b$-jet in Fig. 4 illustrates good agreement between data and the fit results.

Uncertainties on the signal modelling are estimated by reweighting the spectra of both the $b$-jet $p_T$ and the opening angle between the $bb$ pairs to match either the HERWIG parton showering or the ALPGEN matrix element shapes. The parton shower model leads to softer $b$-jets and a narrower angle between the quarks in the $bb$ pairs. These modelling uncertainties affect both the acceptance and efficiency, and the fit templates. It should be noted, however, that even large changes in the bottom quark production model have very little effect on the fit template shapes. The fit template shape dependence on jet kinematics is weak. The shape also does not depend much on the mode of production for the heavy-flavour jets except in the rare cases when the two $b$ quarks are produced close to each other such that their fragments are not resolved in separate jets. Similarly, even large biases in the charm quark production kinematics (including varying the rate of $Wc$ production by $\pm 100\%$) have no significant effect on the fit template shapes.

The systematic uncertainty on the multi-jet background estimate in the electron channel is assessed by changing the requirements which define the control region to model the $E_{\text{T}}^{\text{miss}}$ template. The uncertainty on the $m_{SV}$ template shape is estimated in the same way. In addition the nominal $E_{\text{T}}^{\text{miss}}$ fit range (0–100 GeV) is reduced to both 10–100 and 0–60 GeV. Uncertainties on the EW and top contamination in the control region are found to be negligible. The uncertainty on the multi-jet background normalisation in the electron channel is estimated to be 50% and is limited by low statistics. The systematic uncertainty on the muon multi-jet background is dominated by the validity of the assumptions which go into the matrix method, which is assessed with closure tests in simulated samples. The uncertainty on the $m_{SV}$ template is

### Table 2

Fitted event yields for the $W + \text{jet}$ contributions including the statistical uncertainty from the binned likelihood fit, compared to the ALPGEN MC prediction normalized to the NNLO inclusive $W$ cross section [25], per lepton channel and jet multiplicity. The data-driven multi-jet and $t\bar{t}$ estimates and the other background estimates normalized to (NNLO cross sections are also shown. The statistical uncertainty on the MC prediction is negligible and is not shown. No attempt to compare the fitted and predicted event yields is made, therefore the systematic uncertainties on the predicted event yields have not been estimated. Uncertainties on the MC shapes that affect the $W + b$-jets fitted event yield are discussed in Section 6.

<table>
<thead>
<tr>
<th>$W \rightarrow \mu \nu$, 1-jet</th>
<th>$W \rightarrow \mu \nu$, 2-jet</th>
<th>$W \rightarrow e \nu$, 1-jet</th>
<th>$W \rightarrow e \nu$, 2-jet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred.</td>
<td>Fit result</td>
<td>Pred.</td>
<td>Fit result</td>
</tr>
<tr>
<td>$W + b$</td>
<td>25</td>
<td>28 $^{+13}_{-12}$</td>
<td>26</td>
</tr>
<tr>
<td>$W + c$</td>
<td>108</td>
<td>170 $^{+20}_{-20}$</td>
<td>45</td>
</tr>
<tr>
<td>$W + \text{light}$</td>
<td>38</td>
<td>21.2 $^{+9.9}_{-9.2}$</td>
<td>20</td>
</tr>
<tr>
<td>Multi-jets</td>
<td>8</td>
<td>–</td>
<td>10</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>11</td>
<td>–</td>
<td>44</td>
</tr>
<tr>
<td>Single top</td>
<td>17</td>
<td>–</td>
<td>23</td>
</tr>
<tr>
<td>Other backgrounds</td>
<td>3.9</td>
<td>–</td>
<td>2.5</td>
</tr>
<tr>
<td>Total Predicted</td>
<td>212</td>
<td>–</td>
<td>170</td>
</tr>
<tr>
<td>Data</td>
<td>261</td>
<td>–</td>
<td>217</td>
</tr>
</tbody>
</table>
7. Results and conclusions

The fiducial W + b-jet cross section in the phase space defined in Table 1 is measured in the 1- and 2-jet bin in the electron and muon channel. The results are combined across jet bins and lepton flavour by summing the corresponding measured cross sections as given in Eq. (1). This linear addition is also performed for each of the systematic variations considered, in order to properly take into account the correlations among the different jet bins and lepton channels due to common systematic uncertainties. The leading uncertainties are related to the $b$-tagging calibration and the $m_{SV}$ template shapes, the top quark background, both $t\bar{t}$ and single top, the modelling of the signal, the multi-jet background and the jet energy scale uncertainty. Most of these systematic uncertainties exhibit a strong correlation with each other between the jet bins and lepton channels and therefore the relative systematic uncertainties are only slightly reduced in the combination.

The results are presented in Table 3 and Fig. 5 and are compared with QCD NLO predictions [14] performed in the 5FS (5 flavour number scheme) described in Refs. [3,4,7]. This calculation requires the combination of two contributions. The first contribution has a $b\bar{b}$ pair in the final state, and the $b$ quarks are considered massive (4FS). The second one has a $b$ quark in the initial state and is treated in a scheme based on $b$ quark PDFs where the $b$ quark is assumed massless.

The 5FS prediction is obtained using $Q_{NLO}(m_Z) = 0.118$, $m_b = 4.7$ GeV and $V_{us} = V_{cd} = 0.227$. The NLO CTQED6.4 [40] PDF sets are used. The calculation is obtained with $\mu_R = \mu_F = \mu_0 = m_W + 2m_b$, where $\mu_R$ and $\mu_F$ are the renormalisation and factorization scale. The dependence of the result on the choice of $\mu_0$ is assessed by varying $\mu$ between $\mu_0/4$ and $4\mu_0$, as in Ref. [14]. These variations account for about a 25% uncertainty in the cross section. The PDF uncertainty, estimated to be at the most 7%, is obtained by comparing three different PDF sets: NNPDF2.1 [41], CT10 [42] and MSTW2008 [43].

This QCD NLO prediction is only available at the parton level with an uncovayed W boson. The implementation of the NLO 4FS in POWHEG [34-36] is used to calculate the W acceptance factor of $0.465 \pm 0.003$ (stat). To compare with data the non-perturbative effects of the hadronization and the underlying event have to be considered. The impact of these effects has been evaluated using the PYTHIA PERUGIA 2011 tune [44] on the POWHEG prediction by comparing the results with hadronization and underlying event model turned on and off. The non-perturbative correction to the cross section is $0.93 \pm 0.07$, dominated by particles from $b$ hadron decays landing outside the effective anti-$k_t$ jet cone. The systematic uncertainty accounts for the difference in the modelling of the non-perturbative physics in PYTHIA PERUGIA 2011, PYTHIA MC11,
Acknowledgements

We are grateful to Laura Reina and Doreen Wackeroth for helpful correspondence and discussions. 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.

References

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

Also at DSM/IRFU (Institut de Recherches sur les Lois Fondamentales de l'Univers), CEA Saclay (Commissariat à l'Energie Atomique), Gif-sur-Yvette, France.

Also at Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and CNRS/IN2P3, Paris, France.

Also at Department of Physics, Nanjing University, Jiangsu, China.

* Deceased.