Abstract

The production of a $W$ or $Z$ boson in conjunction with jets is an interesting process in its own right as well as a signal channel (and background) for many interesting standard model and beyond standard model physics signals. Final states with 2, 3, 4 or more jets accompanying a $W/Z$ boson will be observable at the LHC and serve as a crucial part of the ATLAS physics program. The variety of possible jet multiplicities allows for precision tests of jet reconstruction algorithms and techniques. In addition, the reconstruction of leptons and of missing transverse energy becomes more complex in the presence of a multi-jet final state. In this note, we will quantify the differences of lepton, missing transverse energy and jet reconstruction with respect to that observed in inclusive $W$ and $Z$ production.

The wide kinematic range for production of $W/Z +$ jets allows serves as a testing ground for perturbative QCD predictions, both fixed order alone and in conjunction with parton shower Monte Carlos. As an example, the possible large rapidity separations between pairs of jets allows for tests for the applicability and importance of BFKL-type logarithms.
## Contents

1 Introduction ..................................................... 3

2 Reference Cross Sections ...................................... 3
   2.1 Corrections from parton to hadron level .................. 3

3 Datasets .................................................................... 5

4 Definitions and cuts ............................................... 5
   4.1 Electrons .......................................................... 5
   4.2 Muons ............................................................... 5
   4.3 Missing Transverse Energy .................................. 5
   4.4 Jets .................................................................. 6

5 Trigger Paths .......................................................... 6

6 Tag and Probe Studies ............................................ 6
   6.1 Global results in comparison with the inclusive Pythia generated sample .................. 7
   6.2 Differential efficiencies ...................................... 7
   6.3 Background treatment ....................................... 7

7 Measurement of $Z + \text{Jet}$ Cross-Sections ................. 8
   7.1 Introduction ....................................................... 8
   7.2 $Z \rightarrow e^+ e^- + \text{jets} $ ............................. 8
      7.2.1 Signal and Background Distributions .............. 8
      7.2.2 Background Estimation ............................... 8
      7.2.3 Unfolding of Detector Effects ...................... 9
      7.2.4 Background Subtraction .............................. 12
      7.2.5 Comparison to Theory ............................... 13
   7.3 $Z \rightarrow \mu^+ \mu^- + \text{Jets} $ ............................ 15
      7.3.1 Signal and Background Distributions .............. 15
      7.3.2 Background Estimation ............................... 15
      7.3.3 Correction from parton to particle level .......... 17
      7.3.4 Comparison of Event generators and MCFM at the Hadron Level ............... 17
      7.3.5 Unfolding Detector Effects ......................... 17
      7.3.6 Systematic Errors ................................... 19
   7.4 $Z \rightarrow \mu^+ \mu^- + b \text{ Jets} $ ........................ 19
      7.4.1 Event Selection ......................................... 20

8 $W \rightarrow e\nu + \text{jets}$ ............................................. 23
   8.1 Datasets .......................................................... 23
   8.2 Signal-Background Analysis ............................... 23
      8.2.1 Data-driven Methods for the Background Extraction ......................... 26
   8.3 Jet Energy Scale Uncertainties ............................ 26
   8.4 PDF Reweighting Technique .............................. 27
   8.5 PDF Uncertainties ........................................... 28
Introduction

In this note, we will examine the triggering, reconstruction and analysis of events containing a $W$ or $Z$ boson plus jets in ATLAS. We will concentrate on channels with decays into electrons and muons, ignoring for the moment taus (except to account for backgrounds to the other decay channels). Much of the effort on the triggering and reconstruction of leptons and of missing transverse energy is in common with the inclusive $W/Z$ note [1]; thus, we will not reproduce all of the details from that note, but rather will comment on the impact of a multi-jet environment on these issues. Fully simulated signal and background event samples will be treated as pseudo-data. Comparisons of the reconstructed/corrected quantities, when possible, will be to truth-level hadron information (and in some cases to parton-level information with parton-to-hadron corrections applied). Our primary end-result will be hadron-level cross sections, similar to what we will have with the real ATLAS data. We will present expectations/yields/systematics scaled to $1 \text{ fb}^{-1}$, an integrated luminosity that may be accumulated in the first two years of running, but we will also comment on difficulties expected and analysis strategies to be adopted during the early running.

An important source of theoretical systematic uncertainties at the LHC is represented by the Parton Distribution Functions (PDFs). Thus, assessing the level of PDF uncertainties is crucial for studying both processes within and beyond the SM. Experimental systematic uncertainties are the main limitations to the possibility of improving our knowledge on PDFs. Among these, errors due to a miscalibration of the jet energy scale are expected to be the dominant systematics. A comparison of the impact of theoretical and experimental systematic uncertainties is therefore crucial. The impact of JES uncertainties of 1, 3 and 5% will be examined and compared to the level of pdf uncertainties. Experiences from the streaming tests will be reported.

Reference Cross Sections

Reference cross sections for the ATLAS CSC notes are collected in ATLAS note [3], but we briefly discuss here the cross sections relevant for this CSC note. NLO is the first order at which the $W/Z + \text{jets}$ cross sections have a realistic normalization (and realistic shapes for some kinematic distributions) [2]. The current state of the art for NLO calculations is for $W/Z + 2$ jets, although there is ongoing work for the calculation of 3 jet final states. Cross sections for $W/Z + 0, 1$ and 2 (3) final states can be conveniently calculated at NLO (LO) using the MCFM [3] program, and it is from this program that we determine our reference cross sections and uncertainties (both scale and pdf).

2.1 Corrections from parton to hadron level

Cross section measurements in data and LO/NLO predictions are to be compared at the hadron level (particle level). Hence, the data have to be unfolded with respect to the detector response and MCFM predictions have to be corrected with respect to the non-perturbative effects of fragmentation and underlying event (UE). For comparisons to NLO parton level predictions from MCFM, either the data needs...
to be corrected to the parton level or the theory corrected to the hadron level. We discuss the latter correction below for the specific case of $Z +$ jets (but which can also be applied without great error to the case of $W +$ jets). The fragmentation and underlying event corrections are extracted using Pythia Monte Carlo samples by comparing the hadron level results, with the current ATLAS underlying event tune, to corresponding results in samples in which fragmentation and multiple-parton-interactions have been switched off. The corrections are determined by dividing the hadron-level distribution of the observables from standard Pythia by the respective distributions from Pythia with the non-perturbative effects switched off. To the extent to which the 2 partons that can comprise a jet in MCFM mimic the effects of the parton shower in Pythia, the corrections derived from the procedure above can be applied to the MCFM output [2]. We study the hadronization corrections for both cone (R=0.4 and 0.7) and $k_T$ jet algorithms (D=0.4 and 0.6).

Figure 1 shows the correction from parton to hadron level from fragmentation alone (a) and from both fragmentation and underlying event (b) for Cone04 jets. The impact of fragmentation is to reduce the amount of energy in the jet cone. Thus, from fragmentation effects alone, jets at the hadron level tend to have lower $p_T$ than jets at the parton level. The impact of the underlying event is to add energy to the hadron level jet. In general, the underlying event tends to add more energy to the jet than lost by fragmentation, but the exact ratio depends on the radius of the jet.

Figure ?? shows the combined correction for fragmentation and UE for the other jet algorithms. Whereas the fragmentation corrections for Cone07 jets are smaller than for Cone04 jets, the UE corrections are larger due to the larger cone size. $K_T$4 shows the lowest combined corrections since fragmentation and UE effects cancel out. The performance of $K_T$6 jets is comparable to the one of Cone04 jets. Except for Cone07 jets, the non-perturbative effects are negligible for jets with $p_T > 40$ GeV in the current Pythia tune.

This study was carried out specifically for the case of $Z +$ jets but the same correction factors will be used for the $W +$ jets analyses. Differences are expected to be small. The MCFM predictions, corrected for the hadronization effects discussed above, will be compared to $Z +$ jets distributions (and the other channels if there is time).
3 Datasets

Most of the Monte Carlo data samples used in these studies were generated with ALPGEN interfaced with HERWIG using the leading order PDF set CTEQ6LL. The full datasets were obtained by merging samples of $W(Z) + n$ up to 5 partons, weighted according to the expected cross-sections, with a MLM matching cut at $p_T > 20$ GeV, and normalized to an integrated luminosity of 1 fb$^{-1}$. All datasets correspond to exclusive samples, with the exception of the highest multiplicity sample ($W(Z) + 5$ partons) which is inclusive. The samples correspond to the decays, $W \rightarrow e\nu$ $W \rightarrow \mu\nu$ $Z \rightarrow ee$ and $Z \rightarrow \mu^+\mu^- \hat{A}$ generator-level filter (VBF-loose) requires one truth Cone04 jet of $p_T > 20$ GeV and $|\eta| < 5.0$ and one(two for Z) electrons/muons with $p_T > 10$ GeV and $|\eta| < 2.7$ in the event. Pythia signal and background samples [7] were also produced, using the standard Atlas underlying event tune. Pythia $Z \rightarrow ee$ and $W \rightarrow e\nu$ events were generated with a generator-level filter requiring one truth electron with $p_T > 10$ GeV and $|\eta| < 2.7$. Pythia $Z \rightarrow \tau\tau$ events were generated with a filter requiring two electrons/muons with $p_T > 5$ GeV and $|\eta| < 2.8$. For $Z \rightarrow \tau\tau$ and $Z \rightarrow ee$ the dilepton mass is required to be larger than 60 GeV. A large QCD di-jet sample (3.3M events) was also produced, with a minimum hard-scattering transverse momentum of 14 GeV/c.

4 Definitions and cuts

We adopt as much as possible definitions and cuts in common with the other CSC notes, and in particular with the inclusive $W/Z$ note, with comparisons to alternate definitions/cuts where relevant. The reconstruction algorithms in this note are not necessarily optimized for the expected running conditions.

4.1 Electrons

Both the $W$ and $Z$ analyses impose common electron selection criteria. It is required that the transverse momentum of the electron candidate, $P_{Te} > 25$ GeV and that is lies in the range $|\eta| < 2.4$ excluding the crack region (1.37 < $|\eta|$ < 1.52). The electrons are required to fulfill the medium electron-ID [9], which consists of requirements on the calorimeter shower-shape and the matched track. In order to further suppress the contamination from fake electrons in the W selection an isolation requirement within a cone $\Delta R = 0.2$ of the candidate is imposed, requiring the number of tracks with $P_T > 1$ GeV to be less than 4 and the scalar sum of the $P_T$ to be less than 4 GeV. The Z selection requires two electron candidates with an invariant mass of 81 < $M_{ee}$ < 101 GeV and $\Delta R > 0.2$. No calorimeter isolation cuts are applied. There is an implicit isolation cut, however, present in the trigger [?].

4.2 Muons

As in the case of electrons, both the $W$ and $Z$ analyses impose common requirements to muon candidates. A muon candidate uses comprises the combined reconstruction of a inner detector track and a track in the muon spectrometer ?? . The muon is required to have a transverse momentum of $P_{T\mu} > 20$ GeV. In the Z analysis the muon is required to have $P_{T\mu} > 15$ GeV and $0.1 < |\eta| < 1.2$ or $1.3 < |\eta| < 2.4$ . An isolation is applied by requiring the energy deposition in the calorimeter to be the less than 15 GeV in a cone of $\Delta R = 0.2$ around the extrapolation of the muon track. The Z analysis selects two muon candidates with an invariant mass of 81 < $M_{\mu\mu}$ < 101 GeV.

4.3 Missing Transverse Energy

In the $W$ analysis a cut on the missing transverse transverse momentum (MET) is applied to suppress QCD backgrounds. The raw calorimeter-based MET is corrected for the presence of a muon or electron...
in the final state [?] (reference to MET note). Studies were performed to establish the optimal value of the cut on MET that reduces QCD backgrounds to acceptable levels, while not introducing significant biases to the cross-section measurement. An optimal cut of MET > 25 GeV was chosen.

4.4 Jets

Jet clustering algorithms can be divided into two main classes: cones and iterative recombination (e.g., the $k_T$ algorithm). Historically, cone algorithms have been used in hadron colliders, due to concerns about speed, especially at the trigger level, and of large systematic effects in busy multi-jet environments. Fast implementations of the $k_T$ algorithm [?], as well as detailed studies at the Tevatron performing precision measurements with the algorithm [?] call for a detailed comparison of the $k_T$ algorithm with cone-based ones at the LHC.

Many implementations of cone algorithms have been developed over the years. Many of the algorithms have been shown to suffer from infrared safety issues, i.e. the results of the algorithm can change if very soft particles are added or subtracted. Other algorithms, such as Midpoint [?] are stable under infrared correction for most (but still not all) cases. But, since they start clustering jets around energy depositions larger than a given value (seeds), the outcome, to some extent, will depend on the value of this threshold. The manner in which this will affect clustering under real experimental conditions is one of the questions we will address in this note. Finally, a seedless infrared-safe cone algorithm has recently emerged [?], providing most of the desirable features a cone algorithm needs from the theoretical point of view.

For the bulk of the analyses in this note we use jets clustered with a Seeded Cone algorithm with a radius of $\Delta R = 0.4$ [?] built from calorimeter towers and calibrated to the hadron level. In the $Z \to \mu^+\mu^-$ channel we also examine the impact of using clusters as a jet constituent for the same algorithm. The kinematic selection is common for the $Z$ and $W$ analyses. It is required that the jet transverse momentum be $P_T j > 20$ GeV in the range $|\eta| < 3.0$ excluding the crack region ($1.37 < |\eta| < 1.52$). The lepton and jet candidates must be separated by $\Delta R_{l j} > 0.4$. The cross section measurement itself is prepared for jets with $P_T j > 40$ GeV.

5 Trigger Paths

The trigger selection used in here is the same as that used in the inclusive analyses [1]. In the electron channel, $W \to e\nu +$ jet(s) events are required to pass the isolated single-electron trigger (e25i); $Z \to ee +$ jet(s) events are required to pass the isolated di-electron trigger (2e15i) or the isolated single-electron trigger. In the muon channel, $W \to \mu\nu +$ jet(s) events are required to pass the isolated single-muon trigger; $Z \to \mu^+\mu^- +$ jet(s) events are required to pass the isolated dimuon trigger.

The trigger efficiencies at the first, second and event filter levels were evaluated as a function of the jet multiplicity. The trigger efficiency was also studied as a function of the overall hadronic activity, the $P_T$ of the leading jet and the $Z,W$ transverse momentum. For this purpose, a MC truth and data-driven tag-and-probe method were implemented. A good agreement between the two methods was found.

The overall trigger efficiency, with respect to that for the off-line cuts, for the inclusive analysis has been compared to that obtained here. It was found that the trigger efficiency for the $Z,W+$jets analysis is 1.5 − 2% lower than that of the inclusive sample.

6 Tag and Probe Studies

The ‘tag and probe’ method is used in $Z \to e^+e^-+$jet events to determine the single object EM23I/e25i efficiency as well as the offline reconstruction and identification efficiencies from data [?]. As described
in detail in the tag and probe section of [?], a diagnostic sample of size $N_1$ is defined by requiring at least one electron candidate to satisfy tag requirements. A control sample of size $N_2$ is defined by additionally requiring that the second electron satisfies the probe requirements. $N_1$ and $N_2$ may be used to extract the single electron efficiency. Absolute reconstruction, identification and trigger efficiencies are calculated independently and may be multiplied together to give the overall efficiency for an electron. Results in this note are quoted at medium IsEM level and cuts used are those taken as standard in this note. Results are quoted at $1\text{fb}^{-1}$ integrated luminosity.

6.1 Global results in comparison with the inclusive Pythia generated sample

Comparisons have been made with an inclusive Pythia generated sample used in [1], which doesn’t have the generator level truth jet cuts that are present in this sample. The efficiencies measured from these events may be found in CSC notes [?] and [?]. The difference in global efficiencies between the two samples are summarised in table 1. It may be seen that the efficiency in the Alpgen sample is lower due to the higher hadronic activity in these events.

<table>
<thead>
<tr>
<th>Level</th>
<th>Efficiency (stat. error) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction</td>
<td>2.05 (0.17)</td>
</tr>
<tr>
<td>LVL1</td>
<td>0.74 (0.05)</td>
</tr>
<tr>
<td>LVL2</td>
<td>0.36 (0.05)</td>
</tr>
<tr>
<td>EF</td>
<td>0.42 (0.08)</td>
</tr>
<tr>
<td>Overall Trigger Efficiency</td>
<td>1.42 (0.09)</td>
</tr>
</tbody>
</table>

Table 1: Efficiency (Pythia) - Efficiency (Alpgen) (%)

6.2 Differential efficiencies

The dependence of efficiency on electron parameters ($\eta$, $\phi$, $p_T$) in the Pythia sample is well documented in [?] and [?] and the dependence observed is similar in the Alpgen sample. The effect of jet parameters on efficiencies are shown in figures 2. Some points noted:

- Lower efficiency seen in regions of higher jet activity (higher jet multiplicity, hadronic activity and $p_T$ of the most energetic jet. These events are more likely to fail isolation cuts.
- Drop in efficiency at low hadronic $p_T$ is due to the lower electron $p_T$ in this region.
- Drop at higher $\Delta R$ due to the spacepoint bug in the L2 tracking algorithm [?].

6.3 Background treatment

The method of fake rates as explained in [?] was used to estimate the background magnitude. It has been seen that the signal-background ratio is lower in the Alpgen sample. The measured efficiencies after background subtraction (as opposed to earlier results which are from signal analysis only) are tabulated in table 2. We see that the main discrepancy is seen at the reconstruction level, which is to be expected as this is where the background level is highest.
Table 2: Efficiency (Signal only)-Efficiency(after BG subtraction) (%)

<table>
<thead>
<tr>
<th>Level</th>
<th>ΔEfficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction</td>
<td>1.00</td>
</tr>
<tr>
<td>LVL1</td>
<td>-0.74</td>
</tr>
<tr>
<td>LVL2</td>
<td>0.04</td>
</tr>
<tr>
<td>EF</td>
<td>0.02</td>
</tr>
<tr>
<td>Whole trigger</td>
<td>-0.63</td>
</tr>
</tbody>
</table>

7 Measurement of \(Z + Jet\) Cross-Sections

7.1 Introduction

This section prepares the measurement of the inclusive \(Z(\rightarrow ll) +\) jets cross sections as a function of the jet transverse momentum and jet multiplicity. The measurement is compared with LO and NLO perturbative QCD predictions. It will be used in addition to validate event generators which are used to predict the \(W/Z +\) jets backgrounds for searches in ATLAS.

In order to prepare the measurement, we use full-simulated signal and background MC as pseudo data, with which we perform all steps of the analysis. We also compare the predictions of fully simulated signal MC sets from different event generators. The goal of the analysis simulation is to validate the lepton and jet reconstruction in high-multiplicity events, develop the necessary analysis techniques (unfolding, background extraction) and evaluate the statistic and systematic limitations in terms of probing the QCD predictions and discriminating between predictions of different Event generators.

Two separate studies were performed on the feasibility of the \(Z \rightarrow e^+e^-\) and the \(Z \rightarrow \mu^+\mu^-\) channels (see Sections 7.2 and 7.3).

7.2 \(Z \rightarrow e^+e^- + jets\)

7.2.1 Signal and Background Distributions

The presence of additional jets in the event has an impact on the kinematics of leptons and jets. Figure 3 shows the distribution of electron \(p_T\), jet \(p_T\), \(\Delta R\) between electrons and the minimum \(\Delta R\) between each electron and the jets for different jet multiplicities in fully-simulated \(Z \rightarrow ee\) samples. As expected, the electrons are more boosted (larger \(p_T\) and lower \(\Delta R\) between electrons) in events with jets and the distance between electrons and jets becomes smaller in high-multiplicity events. The average \(p_T\)_jet increases with the number of jets.

7.2.2 Background Estimation

The most important backgrounds to the \(Z \rightarrow ee + jets\) signal are processes with real electrons (\(t\bar{t}, W \rightarrow e\nu, Z \rightarrow \tau\tau\)) and QCD jet production. In lack of data all backgrounds are estimated from MC. \(Z \rightarrow \tau\tau\) and \(W \rightarrow e\nu\) are simulated with Pythia and \(t\bar{t}\) is simulated with MCatNLO. The QCD background is derived from a filtered Pythia Dijet sample.

The combined distribution of the invariant mass for signal and background events is shown in Fig. 4a-c for various jet multiplicities. Table 3 gives an overview of the accepted cross section expected from MC from signal and backgrounds. With increasing jet multiplicity \(t\bar{t}\) replaces QCD as the dominant background source.

Eventually the QCD background will be determined with data-driven methods. It can be separated by the \(t\bar{t}\) background by inverting electron-ID cuts and requiring low missing transverse energy. The
Figure 2: Tag and probe efficiencies (L1 × L2 × EF) compared to truth as a function of (a) Hadronic activity, (b) ΔR between the probe electron and the closest jet. Reconstruction efficiencies are the two top plots, and trigger efficiencies the bottom two.

simulation of the t\bar{t} background has to be validated separately with data, using control regions with large missing transverse energy.

Figure 4d-f shows the distribution of signal and backgrounds for the observables (jet multiplicity and $p_T^{\text{jet}}$).

7.2.3 Unfolding of Detector Effects

The reconstructed data has to be unfolded from the detector level to the hadron level, correcting for efficiency loss, resolution and non-linearities in electron and jet reconstruction. In this note, the individual unfolding corrections are assumed to factorize in leading approximation and the individual contributions are investigated separately. A more frequently used approach consists in unfolding the combined impact of jet reconstruction efficiency and jet resolution in a bin-by-bin unfolding procedure [?]. The corrections are derived with full-simulated MC. The simulation of jet calibration and resolution has to be validated in data (in-situ calibration [?]).

The event weight is corrected for the electron reconstruction efficiency, derived as a function of generated $|\eta|$ in four bins of generated $p_T$. It is also corrected for the global efficiency for the electron trigger with respect to the offline selection, determined as: $\text{eff}_{\text{trig}} = (99.63 \pm 0.11)\%$. The Errors on deriving the efficiency, stemming from the limited MC statistics are taken into account as systematic errors for the unfolding procedure.

The jet observables have to be corrected for shifts in the jet energy scale (mainly non-linearities at
Figure 3: Distribution of (a) electron $p_T$, (b) jet $p_T$, (c) $\Delta R$ between electrons and (d) the minimum $\Delta R$ between each electron and the jets for samples with a different minimum number of reconstructed jets with $p_T > 20\text{GeV}$. The events are generated with Pythia (inclusive sample) and Alpgen (Z+jets). Histograms are normalized to unity.

<table>
<thead>
<tr>
<th>Process</th>
<th>$Z \rightarrow ee^+ \geq 1\text{jet}$ xsec [fb]</th>
<th>fraction [%]</th>
<th>$Z \rightarrow ee^+ \geq 2\text{jets}$ xsec [fb]</th>
<th>fraction [%]</th>
<th>$Z \rightarrow ee^+ \geq 3\text{jets}$ xsec [fb]</th>
<th>fraction [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z \rightarrow ee$</td>
<td>23520±145</td>
<td>91.7±0.9</td>
<td>4894±45</td>
<td>86.5±1.6</td>
<td>900±15</td>
<td>78.5±3.5</td>
</tr>
<tr>
<td>QCD jets</td>
<td>1545±89</td>
<td>6.0±0.3</td>
<td>336±42</td>
<td>5.9±0.7</td>
<td>78±20</td>
<td>6.8±1.8</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>552±74</td>
<td>2.2±0.3</td>
<td>424±65</td>
<td>7.5±1.1</td>
<td>168±41</td>
<td>14.6±3.6</td>
</tr>
<tr>
<td>$Z \rightarrow \tau\tau$</td>
<td>3.2±1.2</td>
<td>0.01±0.005</td>
<td>(0.66±0.45)</td>
<td>(0.012±0.001)</td>
<td>(0.12±0.05)</td>
<td>(0.011±0.004)</td>
</tr>
<tr>
<td>$W \rightarrow e\nu$</td>
<td>(28±13)</td>
<td>(0.1±0.05)</td>
<td>(5.9±2.6)</td>
<td>(0.1±0.05)</td>
<td>(1.1±0.5)</td>
<td>(0.1±0.05)</td>
</tr>
</tbody>
</table>

Table 3: Accepted cross section [fb] from signal and background in the $Z \rightarrow ee^+ \text{jets}$ selection. Numbers in brackets are extrapolated from results obtained for a lower jet multiplicity.
Figure 4: Distribution of the di-electron mass for signal and background for (a) $Z \rightarrow ee + \geq 1$jet, (b) $Z \rightarrow ee + \geq 2$jets and (c) $Z \rightarrow ee + \geq 3$jets the number of events with $\geq N$ jets ($p_T > 40$ GeV) (d) and of the $p_T$ of the leading (e) and the next-to-leading (f) jet for $\int L dt = 1fb^{-1}$. 
low $p_T$), the jet energy scale resolution and the jet reconstruction efficiency. All corrections are derived for ten bins in $p_T$ with a comparable number of events in each bin in order to avoid large statistical fluctuations.

Figure 5 shows the reconstruction efficiency of truth jets and the shift in the population of the $p_T$ bins due the $p_T$ resolution, extracted from Alpgen $Z \to ee +$ jets MC. Jet $p_T$ scale and resolution are determined in a matching window of $\Delta R(\text{truth} - \text{reco}) < 0.2$ and $0.5 < p_T(\text{reco})/p_T(\text{truth}) < 1.5$. The jet reconstruction efficiency is determined as the fraction of truth jets which are matched to reconstructed jets applying the same requirements. As expected, the largest bias is observed for low values of $p_T$ jets. The impact of the resolution shown in Fig. 5b is derived by comparing the $p_T$ distribution of truth jets before and after a gaussian smearing with the resolution as determined above. Due to the large statistical uncertainties a global average correction of $98.4 \pm 0.5\%$ is extracted for jets with $p_T > 40$ GeV.

The reconstructed jet $p_T$ is corrected with the jet energy scale corrections and the event is weighted for each required jet with the correction for efficiency-loss in reconstruction and overpopulation due to the jet $p_T$ resolution.

(a) (b)

Figure 5: Jet reconstruction efficiency (a), and the impact of resolution on the jet $p_T$ spectrum (b).

The unfolding corrections are validated by comparing the distributions of truth jet variables with the ones for corrected reconstructed jets. Figure 6 compares examplarily the distribution of the $p_T$ of the leading and the next-to-leading jet in different unfolding stages with the $p_T$ distribution of truth jets. Within the statistic and systematic errors the $p_T$ distributions of truth jets and corrected reconstructed jets are in agreement.

7.2.4 Background Subtraction

The $Z \to \tau\tau$, $t\bar{t}$ and $W \to e\nu$ backgrounds are subtracted using the MC estimates. This will also be done for the collision data. In order to exercise the procedure, ideally the background estimate should be derived from a different sample than the sample used for the simulation of the data, which is not the case in this section. The QCD background is subtracted by weighting all events with a global factor, calculated by $1 - \text{QCD-fraction}$ (as of Table 3). In collisions data, the factor can be derived by a combined fit of signal and backgrounds of the invariant mass peak (sidebands), ideally for low missing transverse energy. This method assumes a comparable distribution of the jets in QCD events with two electron fakes and in $Z \to ee$ events. Alternatively, the background sample can be derived by inverting selected electron-ID cuts. The latter method is preferable but cannot be tested here for the lack of statistics of full simulated QCD MC events. For the final simulation of the measurement, signal and background samples are combined and unfolded to the hadron level as described above.
Figure 6: Comparison of the distribution of the $p_T$ of the leading jet (a) and the next-to-leading jet (b) in truth and corrected reconstructed MC

7.2.5 Comparison to Theory

The comparison of theory and measurement is prepared for theoretically well described quantities: the inclusive cross section of $Z \rightarrow \ell\ell + \geq 1\text{jet}$, $Z \rightarrow \ell\ell + \geq 2\text{jets}$ and $Z \rightarrow \ell\ell + \geq 3\text{jets}$ and the differential cross sections with respect to the $p_T$ of the leading and the next-to-leading jets. The MCFM predictions are corrected by the residual energy loss due to fragmentation for jets with $p_T > 40$ GeV, determined from the current Pythia tune of underlying event and fragmentation as $0.98 \pm 0.01$ for $Z \rightarrow ee+ \geq 1\text{jet}$, $0.98 \pm 0.02$ for $Z \rightarrow ee+ \geq 2\text{jets}$ and $0.95 \pm 0.06$ for $Z \rightarrow ee+ \geq 3\text{jets}$. PDF uncertainties on the MCFM predictions are calculated using the complete set of PDF error functions. The inclusive Pythia and Alpgen $Z \rightarrow e^+e^-$ samples are normalized to the NLO inclusive (DrellYan) MCFM $Z \rightarrow e^+e^-$ cross section at the generator level.

Figure 7 shows the comparison of the distribution of the observables (jet multiplicities and the $p_T$ of the leading and next-to-leading jets) at the hadron level for Alpgen and Pythia MC with LO and NLO MCFM calculations. The error bars are calculated only from intrinsic MC quantities as the quadratic sum of statistic errors from the MC sample size and systematic errors from the unfolding corrections derived by MC. The shape of the jet-$p_T$ distribution predicted by Alpgen agrees well with the shape predicted by MCFM. Due to the tuning of the leading soft radiation in the parton shower, Pythia predicts a larger inclusive cross section for $Z \rightarrow ee+ \geq 1\text{jet}$ but a clearly softer $p_T$ spectrum. The predictions for cross sections for $Z+\geq 1\text{jet}$ both by Alpgen and Pythia are lower than the theoretical cross sections.

As a next step towards real Atlas data, systematic errors are adapted to the values expected from the collision data and propagated to the measured cross section. Figure 7 shows the relative systematic uncertainty on the cross section (normalized to 1) expected for different uncertainties on the jet energy scale for the production of a Z with 1-4 jets ($p_T > 40$ GeV). Since the difference between LO and NLO cross section predictions are in the order to 30%, with a 3% uncertainty on the jet energy scale we are still able to differentiate between LO and NLO predictions whereas with an error of 10% on the jet energy scale this is not possible.

The uncertainty on the jet energy resolution and its impact on the unfolding procedure is an additional source of systematic errors on the cross section measurement. They are investigated for jet resolutions up to two times the ones currently expected. An uncertainty of 50% on the jet resolution propagates to an error of 2-4% on the $Z \rightarrow ee+ \geq 1\text{jet}$ cross section measurement. A wrong assumption of the jet-$p_T$ distribution in calculating the unfolding corrections from the jet energy resolution can also lead to a systematic shift in the cross section measurement. A comparison between the unfolding corrections derived from Pythia and from Alpgen yields a systematic uncertainty of up to 1.5%. 
Figure 7: Inclusive jet cross section (a), $p_T$ of the leading jet (b) and of the next-to-leading jet (c) in unfolded Pythia and Alpgen signal+background MC and predictions from MCFM for $\int L \, dt = 100 \text{pb}^{-1}$. Relative systematic uncertainty (d) on the cross section (normalized to 1) expected for different uncertainties on the jet energy scale.
### 7.3 $Z \rightarrow \mu^+\mu^- + \text{Jets}$

The impact of a high-multiplicity environment on the lepton reconstruction efficiency is also investigated in case of the muon. Muon reconstruction efficiencies for different isolation requirements were studied. Non isolated muons and cone 0.20 isolation present no bias for high jet multiplicity events. No bias in muon reconstruction efficiency for high $E_T$ jets was found.

#### 7.3.1 Signal and Background Distributions

Figure 8 shows jet multiplicities for the signal and backgrounds for Pythia (a) and Alpgen (b) MC datasamples. Figure 8 shows jet $p_T$ and $\eta$ distributions for signal and backgrounds (Pythia (a),(b) and Alpgen (c),(d)). As expected, most important backgrounds are $t\bar{t}$ and $b\bar{b}(\mu\mu)$ events with similar $p_T$ distributions.

#### 7.3.2 Background Estimation

The important backgrounds for $Z \rightarrow \mu^+\mu^- + \text{jets}$ analysis are processes with similar topologies ($t\bar{t},W \rightarrow \mu\nu,Z \rightarrow \tau^+\tau^-$) with real muons, and QCD jet production. All backgrounds are estimated using MC datasamples. For $t\bar{t}$, $W \rightarrow \mu\nu$, $Z \rightarrow \tau^+\tau^-$ Pythia and Alpgen datasamples are used. In the case of QCD background, datasamples generated with Pythia $b\bar{b}(\mu\mu)$ are used to increase muon statistics. It can be assumed that the dominating contribution of high energetic muons are from decays of $b\bar{b}$mesons. Figure 9 shows the distribution of the invariant mass for signal and background, using Pythia (a) and Alpgen (b).

Muons from $Z \rightarrow \mu\mu$ are high energetic and isolated muons. Therefore, muons coming from background can be rejected using isolation. In this analysis, different isolation cuts have been studied. Table 4 shows results for signal and background considering different isolation criteria, for Pythia and Alpgen. As the size of the muon isolation cone is increased, background is reduced, in particular $b\bar{b}$. For our analysis, a muon isolation cone $\Delta R = 0.20$ with $E_T < 15$ GeV is considered because $b\bar{b}$ background is reduced and muon reconstruction efficiency has no bias on high jet multiplicities. The invariant mass window considered is 81-101 GeV in order to increase the ratio signal/background, in particular for the $t\bar{t}$ background. Table 5 shows the accepted cross section and the fraction of signal and background for different jet multiplicities using Pythia and Alpgen. As jet multiplicity increases, $t\bar{t}$ background increases. For $Z \rightarrow \mu\mu + \geq 3\text{jets}$ events, $t\bar{t}$ background ration increases up to 12-20%. Top-quark pairs cross section is of the order of the signal, top-quark pairs decay mostly into a $W$boson and a $b$-quark,
Figure 8: Distribution of signal and backgrounds jet multiplicities, for Pythia (a) and Alpgen (b), with no invariant mass applied. Jet $p_T$ and $\eta$ distributions for signal and backgrounds using Pythia (c),(d) and Alpgen (e),(f) datasamples. $\eta$ distributions are for high-$E_T$ jets ($p_T > 40\text{GeV}$).
both of them decay further into muons, and have a significant large hadronic activity. $W \rightarrow \mu \nu$ and $b \bar{b}$ errors are extrapolated.

7.3.3 Correction from parton to particle level

Cross section measurements in data and LO/NLO predictions are to be compared at hadron level. MCFM predictions are corrected with respect to non-perturbative effects (fragmentation and underlying event). The global corrections from parton to hadron level from fragmentation and UE for Cone04 jets, applied to MCFM are: $0.98 \pm 0.01$ for $Z+1$jet, $0.98 \pm 0.02$ for $Z+2$jets and $0.95 \pm 0.06$ for $Z+3$jets, inclusive.

7.3.4 Comparison of Event generators and MCFM at the Hadron Level

Figure 10 (a) shows the comparison of the inclusive $Z \rightarrow \mu \mu + \geq N\text{jets}$, cross sections (number of events expected for 1fb$^{-1}$) as predicted by the event-generator Alpgen (using Atlfast reconstructed samples, where no jet filter cut is applied) and Pythia, and as predicted by MCFM. Alpgen, and also Pythia, predicts less events than MCFM in all multiplicity bins. Since it is not the purpose of this analysis to measure the inclusive $Z \rightarrow \mu \mu$ cross section, the event generator output can be normalized globally to the inclusive $Z \rightarrow \mu \mu$ cross section from data. From the lack of data, the cross sections of all Alpgen $Z \rightarrow \mu \mu$ samples and of the Pythia $Z \rightarrow \mu \mu$ sample are normalized to the NLO inclusive cross sections. The normalization factors derived are: Alpgen, $1.328 \pm 0.069$; Pythia, $1.049 \pm 0.056$.

Figure 10 (b) shows the comparison of the inclusive $Z \rightarrow \mu \mu + \geq N\text{jets}$ cross sections after the normalization.

Figure 10 shows the distribution of the $p_T$ of the leading jet (c) and next-to-leading jet (d), predicted by Alpgen, Pythia (at truth level of full-simulated samples, normalized to the inclusive cross section) and MCFM for an integrated luminosity of 1fb$^{-1}$. Pythia parton shower predicts a higher jet multiplicity in the low-$p_T$ region ($p_T < 100$ GeV), and Alpgen predicts more high-$p_T$ jets (a).

7.3.5 Unfolding Detector Effects

The reconstruction data is unfolded to the hadron level correcting for efficiency losses, resolution and non-linearities in muon and jet reconstruction. The unfolding procedure is derived (and validated) with
Table 5: Number of events expected from signal and background in the \( Z\rightarrow \mu\mu + \geq 1\text{jet} \) selection for \( \int Ldt = 1\text{fb}^{-1} \), for different jet multiplicities, using isolated cone 0.2 \( E_T \text{isol} \) 15 GeV muons, for an invariant mass window of 81-101 GeV.

& Process & \( Z\rightarrow \mu\mu + \geq 1\text{jet} \) & \( Z\rightarrow \mu\mu + \geq 2\text{jets} \) & \( Z\rightarrow \mu\mu + \geq 3\text{jets} \) \\
\hline
\text{pythia} & \begin{pmatrix} 72819 & 96.9 & 11660 & 89.4 \end{pmatrix} & \begin{pmatrix} 1974 & 87.7 \end{pmatrix} \\
\text{W} & \begin{pmatrix} 0 & 0.0 \end{pmatrix} & \begin{pmatrix} 0 & 0.0 \end{pmatrix} & \begin{pmatrix} 0 & 0.0 \end{pmatrix} \\
\text{QCD}(b\bar{b}) & \begin{pmatrix} 1225 & 1.6 \end{pmatrix} & \begin{pmatrix} 599 & 4.6 \end{pmatrix} & \begin{pmatrix} 277 & 12.3 \end{pmatrix} \\
t\bar{t} & \begin{pmatrix} 1141 & 1.5 \end{pmatrix} & \begin{pmatrix} 789 & 6.0 \end{pmatrix} & \begin{pmatrix} 277 & 12.3 \end{pmatrix} \\
total\ background & \begin{pmatrix} 2366 & 3.1 \end{pmatrix} & \begin{pmatrix} 1388 & 10.6 \end{pmatrix} & \begin{pmatrix} 277 & 12.3 \end{pmatrix} \\
\text{alpgen} & \begin{pmatrix} 59408 & 96.1 \end{pmatrix} & \begin{pmatrix} 12568 & 88.7 \end{pmatrix} & \begin{pmatrix} 2446 & 79.6 \end{pmatrix} \\
\text{W} & \begin{pmatrix} 22 & 0.0 \end{pmatrix} & \begin{pmatrix} 0 & 0.0 \end{pmatrix} & \begin{pmatrix} 0 & 0.0 \end{pmatrix} & \begin{pmatrix} 0.0 & 0.0 \end{pmatrix} \\
\text{QCD}(b\bar{b}) & \begin{pmatrix} 1225 & 1.6 \end{pmatrix} & \begin{pmatrix} 599 & 4.6 \end{pmatrix} & \begin{pmatrix} 277 & 12.3 \end{pmatrix} \\
t\bar{t} & \begin{pmatrix} 1163 & 1.9 \end{pmatrix} & \begin{pmatrix} 1008 & 7.1 \end{pmatrix} & \begin{pmatrix} 626 & 20.4 \end{pmatrix} \\
total\ background & \begin{pmatrix} 2310 & 3.9 \end{pmatrix} & \begin{pmatrix} 1607 & 11.3 \end{pmatrix} & \begin{pmatrix} 626 & 20.4 \end{pmatrix}

Figure 10: Number of events expected for 1 fb\(^{-1}\) at hadron level for Alpgen, Pythia and MCFM. On the top right, normalized to inclusive \( Z\rightarrow \mu\mu \) NLO MCFM cross section (b). \( p_T \) of the leading jet (c) and \( p_T \) of the next-to-leading jet (d) at hadron level from Alpgen, Pythia (MC truth normalized to \( Z\rightarrow \mu\mu \) inclusive cross section) and LO and NLO MCFM predictions.
full-simulated MC. Figure 11 (a) shows the $p_T$ distribution of the reconstructed leading jet unfolded (Alpgen and Pythia MC) compared to NLO and LO MCFM predictions at hadron level. From $p_T > 150$ GeV Pythia and Alpgen distributions can be distinguished. Figure 11 (b) shows the $p_T$ distribution of the reconstructed next-to-leading jet unfolded (Alpgen and Pythia MC) compared to NLO and LO MCFM predictions at hadron level. In this case, Pythia predicts a softer $p_T$ distribution than Alpgen and MCFM, including LO predictions.

(a) (b)

Figure 11: $p_T$ of the leading jet in unfolded Pythia and Alpgen signal MC and predictions from MCFM for $\int L dt = 1 fb^{-1}$.

7.3.6 Systematic Errors

Figure 12 shows the relative systematic error expected for different uncertainties on the Jet Energy Scale (1%, corresponding to the challenging Atlas goal, 3% expected for early data and 10% for the worst case). If a 3% error on JES is achieved in early data, the systematic error increases up to 10% for events $\geq 4$ jets. In the worst scenario, systematic errors increase up to 40% for $\geq 4$ jet and 25% for $\geq 3$ jet events. If the challenging JES uncertainty of 1% is achieved, systematic errors decrease to less than 0.5%.

7.4 $Z \rightarrow \mu^+ \mu^- + b$ Jets

The measurement of the $Z + b$ jet production cross section at LHC will provide an important test of perturbative quantum-chromodynamics (QCD). The cross section is sensitive to the $b$ quark density of the proton and its precise measurement will help in reducing the current uncertainty on the partonic content of the proton (PDF’s). Such uncertainty is presently affecting the potential for discovering new physics at LHC.

Table 6 gives the next-to-leading cross-sections for $Z +$ jets at the Tevatron and LHC [?]. As we can see from this table, at LHC the process $gb \rightarrow Zb$ clearly dominates w.r.t. $q\bar{q} \rightarrow Zb\bar{b}$. The selection of $Z+b$ jet events should be easier at LHC than at Tevatron for two reasons: we will have a cross section for $Z+b$ jet a factor 50 larger than at Tevatron and the relative production of $Z+c$ jet will be less important at LHC, thus reducing the probability of jet mis-tagging.

With the present analysis we provide a preliminary estimation of the selection efficiency for $Z \rightarrow \mu^+ \mu^- + b$ jets events and for the main backgroud channels. We also show the rejection power for $Z \rightarrow \mu^+ \mu^- + c$ and light jets, as a function of the $b$-tagging efficiency.
Figure 12: Relative systematic errors expected on the measured cross section from different uncertainties on the Jet Energy Scale.

<table>
<thead>
<tr>
<th>Cross Section(pb)</th>
<th>Tevatron</th>
<th>LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>ZQ inclusive</td>
<td></td>
</tr>
<tr>
<td>$gb \rightarrow Zb$</td>
<td>$13.4 \pm 0.9 \pm 0.8 \pm 0.8$</td>
<td>$1040^{+70+70+30}_{-60-100-50}$</td>
</tr>
<tr>
<td>$q\bar{q} \rightarrow Zb\bar{b}$</td>
<td>$6.83$</td>
<td>$49.2$</td>
</tr>
<tr>
<td>$gc \rightarrow Zc$</td>
<td>$20.3^{+1.8}<em>{-1.5} \pm 0.1^{+1.3}</em>{-1.2}$</td>
<td>$1390 \pm 100^{+60+40}_{-70-80}$</td>
</tr>
<tr>
<td>$q\bar{q} \rightarrow Zc\bar{c}$</td>
<td>$13.8$</td>
<td>$89.7$</td>
</tr>
</tbody>
</table>

Table 6: Next-to-leading-order inclusive cross section (pb) for $Z$ plus jets at LHC ($\sqrt{s} = 14$TeV $pp$) and at Tevatron ($\sqrt{s} = 1.96$TeV $p\bar{p}$). The calculations are limited to the case of a jet in a range $p_T > 15$GeV and $|\eta| < 2.5$ (LHC) or $|\eta| < 2.0$ (Tevatron).

7.4.1 Event Selection

We require two muons of opposite charge, with $p_T > 15$ GeV, $|\eta| < 2.4$ and $\Delta R > 0.4$ from the jet axis. $Z$ selection is done requiring $M_{\mu\mu}$ in the range $M_{Z} \pm 20$ GeV. We select Cone04 jets with $|\eta| < 2.5$ and require $N_{Jets} \geq 1$. The analysis has been performed for two jet energy cuts: $E_{T}^{jet} > 20$ GeV and $E_{T}^{jet} > 40$ GeV.

The b-tagging is performed cutting on the weight parameter, namely the combination of the secondary vertex (SV1) and impact parameter (IP3D) algorithms [13]. In this analysis we cut at weight $\geq 5$.

Figure 13 (a) shows the transverse momentum spectrum for all the muons found in the various data samples, before any cut is applied, while figure 13 (b) shows the invariant mass of the two muons, after all the muon cuts. The plots are normalized to the number of entries.

Figure 14 (a) shows the b-tagging efficiency as a function of the weight parameter: cutting at weight $\geq 5$, as we do in this analysis, ensures 60% efficiency for b jets. Figure 14 (b) shows the rejection factor for c and light quarks as a function of the b-tagging efficiency: 60% efficiency gives a rejection
Figure 13: Muon transverse momentum spectrum, before any cut (left). Invariant mass of the two muons, after all the muon cuts (right).

factor of 7 for c jets and 110 for light jets, in fairly good agreement with what is found for other physics channels [13].

Figure 14: B-tagging efficiency as a function of the weight parameter calculated for the sample Z+bJet (left). Rejection factor for c (black circles) and light (black stars) jets as a function of the b-tagging efficiency (right).

Figure 15 (a) shows the jet transverse energy spectrum for b, c and light jets in the $Z \rightarrow \mu \mu$ inclusive data sample, after all the cuts on the muons, including the cut on $M_{\mu \mu}$, and for $|\eta_{jet}| < 2.5$ and $E_{T}^{jet} > 20$ GeV. Figure 15 (b) shows the same events of figure (a) when also the cut on the b-tagging weight is added.

A summary of the efficiency for the various muons cuts, for all the samples is shown in Table 7.

The final efficiency after additional cuts on jets is shown in table 8 for the Z inclusive, Z+b jets and Z+c and light jets samples. Those efficiencies are calculated requiring transverse energy of the jet $E_{T}^{jet} > 20$ or $E_{T}^{jet} > 40$. 
March 5, 2008 – 05:45

Table 7: Efficiency after the cuts to identify the muons from Z decay

<table>
<thead>
<tr>
<th>Cuts</th>
<th>Z inclusive</th>
<th>Z+b jets</th>
<th>Z+c or light jets</th>
<th>t\bar{t}</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>at least 1 (\mu^+) and 1 (\mu^-)</td>
<td>56.9 ± 0.3%</td>
<td>91 ± 2%</td>
<td>93 ± 3%</td>
<td>11.4 ± 0.1%</td>
<td>2.51 ± 0.09%</td>
</tr>
<tr>
<td>(p_T^{\mu} &gt; 15) GeV</td>
<td>32.6 ± 0.2%</td>
<td>51 ± 1%</td>
<td>53 ± 2%</td>
<td>1.24 ± 0.04%</td>
<td>0.09 ± 0.01%</td>
</tr>
<tr>
<td>(</td>
<td>\eta^{\mu}</td>
<td>&lt; 2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta R_{\mu-jet} &gt; 0.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 GeV &lt; (M_Z) &lt; 110 GeV</td>
<td>30.9 ± 0.2%</td>
<td>49 ± 1%</td>
<td>50 ± 2%</td>
<td>0.29 ± 0.02%</td>
<td>0.004 ± 0.003%</td>
</tr>
</tbody>
</table>

Table 8: Final efficiency after all the cuts for the selection of \(Z \rightarrow \mu^+\mu^- + b\) jets.

Given the final efficiencies quoted in this table, for an integrated luminosity of 100\(pb^{-1}\) and a signal cross section of 41.25 \(pb\) we should be able to select 536 ± 23 events when cutting at \(E_T^{jet} > 20\) GeV and 289 ± 17 events when cutting at \(E_T^{jet} > 40\) GeV. The contamination from Z+c and light jets should be of the order of 30\% while the background from \(t\bar{t}\) and \(W \rightarrow \mu\nu\) is negligible.

Figure 15: Jet transverse energy spectrum for b, c and light jets in the \(Z \rightarrow \mu\mu\) inclusive data sample, before (left) and after (right) the cut on the b-tagging weight.
8 \( W \rightarrow e\nu + \text{jets} \)

The following analysis focuses on \( W(\rightarrow e\nu) + \text{jets} \) events, especially in regards to an evaluation of the Jet Energy Scale (JES) and Parton Distribution Function (PDF) uncertainties. Observable distributions will be shown for signal and backgrounds at detector level, after a standard cut-based selection. A data driven method to remove backgrounds will also be discussed. The impact of the jet energy scale on jet multiplicity distributions will be examined. The PDF reweighting technique will then applied to predict PDF uncertainties on electron and jet distributions and the PDF uncertainties on the measurement of the cross section, as a function of the cumulative jet multiplicity, will be compared to the experimental systematic uncertainties.

8.1 Datasets

The analysis is performed on \( W + \text{jets} \) events with the W boson decaying into the electron channel; the signal datasets have been described earlier in Section 3. The full dataset is obtained by merging samples of \( W + 0 \) up to 5 partons, weighted according to the expected cross sections, and normalized to an integrated luminosity of 1 fb\(^{-1}\).

Four kinds of background processes are considered: di-jets events (referred in the following as 'QCD background'), \( t\bar{t}, W \rightarrow \tau\nu, \) and \( Z \rightarrow ee \) events. The background datasets have also been described in Section 3.

8.2 Signal-Background Analysis

The number of signal events and the fraction of background events after this selection has been applied are given in Table 9, using the datasets mentioned in the previous section and assuming an integrated luminosity of 1 fb\(^{-1}\).

Table 9: Number of signal events selected for an integrated luminosity of 1 fb\(^{-1}\). For the background samples, the ratio \( \frac{N_B}{N_S} \) in percent is indicated. The quoted uncertainties are due to the Monte-Carlo statistics.

<table>
<thead>
<tr>
<th>Selection</th>
<th>( W \rightarrow e\nu ) (N)</th>
<th>QCD (f (%))</th>
<th>( W \rightarrow \tau\nu ) (f (%))</th>
<th>( Z \rightarrow ee ) (f (%))</th>
<th>( t\bar{t} ) (f (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger/offline e-id + ( E_{T}^{\text{miss}} &gt; 25 \text{ GeV} )</td>
<td>((1.053 \pm 0.001) \cdot 10^6)</td>
<td>10 ± 3</td>
<td>8.0 ± 0.1</td>
<td>1.57 ± 0.05</td>
<td>6.25 ± 0.04</td>
</tr>
</tbody>
</table>

The table indicates that the largest contribution to the \( W \rightarrow e\nu \) background comes from QCD di-jet production. It has to be noted that this QCD background is, among all background processes, the least understood both theoretically and experimentally. A rather large uncertainty (factor \( \approx \) of 2-3) on the level of this background is expected due to the large theoretical uncertainty. Furthermore, due to the large di-jet cross-section and to the high rejection power of the selection criteria, the available Monte Carlo statistics is not sufficient to evaluate the number of QCD background events by simply counting the events after all cuts to the Monte Carlo QCD background sample. To overcome this difficulty the QCD background contamination is obtained by using the product of the offline selection efficiency only, times the trigger selection efficiency. To properly take into account the correlation between these two quantities, the trigger selection efficiency is calculated as the ratio of the efficiency of the trigger and electron identification cuts and the electron identification efficiency.

Figure 16 shows the transverse mass distribution, \( M_T \), for signal and background events after the full selection.
Figure 16: W transverse mass distribution for signal and background normalized to 1 fb\(^{-1}\): \(W \rightarrow e\nu\) (white), QCD (yellow), \(W \rightarrow \tau\nu\) (red), \(Z \rightarrow ee\) (green), \(t\bar{t}\) (blue). Histograms are cumulative.

In Figure 17, we see the \(E_T\) spectrum for all jets in the events for the signal, \(W \rightarrow e\nu\), and the backgrounds. Despite the poor Monte-Carlo statistics for QCD events, the figure clearly indicates the increasing importance of the QCD and \(t\bar{t}\) backgrounds at higher jet \(E_T\) with respect to the steeply falling signal distribution.

Figure 17: Jet \(E_T\) distribution for signal and background normalized to 1 fb\(^{-1}\): \(W \rightarrow e\nu\) (white), QCD (yellow), \(W \rightarrow \tau\nu\) (red), \(Z \rightarrow ee\) (green), \(t\bar{t}\) (blue). Histograms are cumulative.

In Figure 18, we can see the expected number of events as a function of the jet multiplicity in any given event, for an integrated luminosity of 1 fb\(^{-1}\). Figure ?? shows the corresponding cumulative jet multiplicity. The term cumulative multiplicity refers to events with at least \(n\) jets in the final state, so that \(n = 0\) corresponds to the inclusive \(W \rightarrow e\nu\) sample. The distributions show that the \(t\bar{t}\) background dominates at large jet multiplicities, while QCD di-jet production is the largest background for small multiplicities.

However, it must be noted that in this analysis the QCD background has been estimated by using PYTHIA di-jet processes, which are known to underestimate the level of multi-jet production. The real amount of QCD background can only be reliably estimated from LHC data.
Figure 18: Jet multiplicity (upper left) and cumulative jet multiplicity (upper right) distributions for signal and backgrounds. Transverse mass distributions before (lower left) and after (lower right) the rejection of the $Z \rightarrow ee$ background (in green). Plots are normalised to 1 fb$^{-1}$ integrated luminosity.
8.2.1 Data-driven Methods for the Background Extraction

Whereas the $W \rightarrow \tau \nu$ and $Z \rightarrow ee$ backgrounds can be reliably estimated by Monte Carlo techniques, the $t\bar{t}$ and especially the QCD di-jet backgrounds can only be estimated by using LHC data. Thus, to minimize the uncertainty on the amount of background, an alternative, mostly data-driven, selection procedure is proposed. The $E_T^{\text{miss}}$ cut is efficient in rejecting the backgrounds, but results in a loss of information of the QCD distribution shape. The amount and shape of the background under the signal area will be hard to estimate. Therefore, a data-driven approach to reject the background has been developed (see the inclusive $W/Z$ CSC note for a detailed description of this technique).

In this approach, $Z \rightarrow ee$ events are efficiently removed by cutting on the invariant mass of an $e^+e^-$-pair, as shown in Figure 18.

The dominant QCD background can be parameterized and subsequently subtracted from the signal area by using a pure QCD di-jet sample. To accomplish this, photon candidates are selected which pass a photon trigger and the same calorimeter-based identification as electrons. Photon candidates are distinguished from electrons as no good quality track is found to be associated with the electro-magnetic cluster. Applying some additional requirements on track isolation, this sample is largely dominated by QCD jets faking a photon and can be parameterized as a function of the cell-based $E_T^{\text{miss}}$. As the electron and photon samples are kinematically very similar, the $E_T^{\text{miss}}$ spectrum for the QCD background has the same shape in both samples.

The background parameterization coming from the photon sample is then normalized to the electron sample. The normalization factor is obtained using, as a side-band, the low $E_T^{\text{miss}}$ region in the electron sample, where QCD is dominating. Only after having subtracted the QCD background, a cut on $E_T^{\text{miss}}$ is applied to remove the side-band.

In order to obtain a measurement of the $W \rightarrow e\nu$ cross section as function of the jet multiplicity, the above QCD-subtraction method can be applied for each given jet multiplicity. However, the limited Monte Carlo statistics of QCD events prevents us from applying this method in this analysis, but the principle has been tested in the inclusive cross section analysis.

8.3 Jet Energy Scale Uncertainties

The dominant experimental systematic uncertainty for $W + \text{jets}$ production at the Tevatron is given by the Jet Energy Scale (JES) uncertainty. Since the same is expected to be true at ATLAS, only this source of experimental systematics is considered in the following study.

The effect of jet energy scale uncertainties on the cross section as a function of cumulative jet multiplicity is investigated in the following section. A miscalibration of $\pm 1\%$, $\pm 3\%$ and $\pm 5\%$ is assumed on the jet energy and its effect on the jet transverse momentum is calculated. Because of time constraints, the jet energy is miscalibrated after having run the jet reconstruction algorithm. The transverse momentum is then recalculated from the mis-calibrated energy and the pseudorapidity. This simple method allows a quick estimate of the uncertainty on the cross section. It is sufficient for the scope of this study, where the aim is a comparison with the uncertainties due to PDFs. After the miscalibration, the event selection is applied and the jet multiplicity distribution is plotted for each value of the assumed miscalibration. From these distributions the error on the cross section is calculated.

Figures 19 show the cross section as a function of the leading jet transverse momentum and the cumulative jet multiplicity with the experimental uncertainty deriving from a $\pm 1\%$, $\pm 3\%$ and $\pm 5\%$ energy miscalibration.

Figure 19 indicates how the JES uncertainties increase significantly with the transverse momentum of the leading jet, rising above 30% for highly energetic jets.
Figure 19: JES Uncertainties on leading jet transverse momentum and on cumulative jet multiplicity assuming ±1%, ±3%, ±5% jet energy miscalibration. Upper left: Leading Jet Pt spectra with absolute uncertainties. Upper Right: Uncertainties expressed as percentage. Lower Left: Cumulative Multiplicity spectra with absolute uncertainties. Lower Right: Uncertainties expressed as percentage.

The main effect on the cross section as a function of the jet multiplicity results from the selection cut of 20 GeV applied on the jet transverse momentum. If the energy is miscalibrated by a positive correction, more jets pass the selection cut and therefore there will be more high jet multiplicity events and vice-versa when the energy is miscalibrated by a negative correction. The relative uncertainties on the measurement are shown in figure 19. When the jet energy is miscalibrated by 1%, the experimental systematic uncertainties on the cross section remain within 5%, but they increase significantly with higher miscalibration, rising above 20%. Uncertainties show a tendency to increase with the multiplicity. The first bin with zero jets is biased by the event filter cut applied at generator level and is therefore not representative.

8.4 PDF Reweighting Technique

The effect of PDF uncertainties on physical distributions is investigated applying PDF re-weighting. This is a powerful technique used in Monte Carlo (MC) simulations to evaluate PDF uncertainties on a physical quantity in a fast way. In fact, applying an event weight to the hard process, predictions for one PDF set can be derived from a sample generated with a different PDF set. The event weight is a function of momentum fraction $x$ of the two incoming partons and the scale $Q^2$ of the interaction:

$$ EW = \frac{f_{pdf_2}(x_1, Q^2)}{f_{pdf_1}(x_1, Q^2)} \cdot \frac{f_{pdf_2}(x_2, Q^2)}{f_{pdf_1}(x_2, Q^2)} $$

The scale $Q^2$ is calculated from the mass ($m_W$) and transverse momentum ($P_T$) of the $W$ boson, consistently with the convention used in ALPGEN:
\[ Q^2 = m_W^2 + P_T^2 \] (2)

The accuracy of PDF reweighting on events generated with Herwig was investigated in previous studies [?] and found to be better than 1%. A similar analysis was performed on ALPGEN events to test the accuracy of reweighting from a LO to a NLO PDF set. Two samples of 500,000 $W + 3$partons events were generated with two different PDF sets (CTEQ6LL and MRST2001NLO). Events generated with CTEQ6LL were then reweighted to MRST2001NLO and compared to those generated directly with MRST2001NLO. It was shown that the PDF reweighting technique is accurate to be better than 1% in the measurable kinematic regions.

8.5 PDF Uncertainties

The CTEQ group provides error sets for the Next to Leading Order PDF set CTEQ6M. Since the datasets used in this CSC note are generated with the PDF set CTEQ6LL, reweighting is applied from CTEQ6LL to the central value of CTEQ6M and the 40 error sets corresponding to the upper and lower uncertainty on each of the 20 eigenvectors. The upper around the NLO central value are calculated using the asymmetric error formula recommended in [?].

Figures ?? and ?? show the electron $P_T$ and $\eta$ distributions determined by the CTEQ6M central value with the upper and lower PDF uncertainties. The relative errors are shown in Figures ?? and ?? . PDF uncertainties are of the order of 5% and exhibit a dependence on both $P_T$ and $\eta$. They increase at very low and very high transverse momentum, where statistical fluctuations also become important. Uncertainties are also larger in the central $\eta$ region, but remain well below 10%.

Figure 20: PDF Uncertainties on the electron $P_T$ and $\eta$ at detector level, after a $P_T$ cut of 25 GeV. Upper left: Distribution before (blue) and after reweighting to the central CTEQ6M set (black) with upper (green) and lower (red) uncertainties. Upper Right: Relative uncertainties on the same distribution. Lower left: Upper (green) and lower (red) uncertainties on the $\eta$ spectra. Lower Right: Uncertainties expressed as percentage.
Figure 21: PDF uncertainties on the leading jet transverse momentum and $\eta$ at detector level. Upper left: $P_T$ spectra with upper (green) and lower (red) uncertainties. Upper right: Relative uncertainties on jet $P_T$. Lower left: Upper (green) and lower (red) uncertainties on the $\eta$ spectra. Lower right: Uncertainties expressed as percentage.

Similar results are obtained for jet variables. Figure 21 shows that the relative uncertainties on leading jet $P_T$ and jet $\eta$ are approximately 3-5% but can rise above 5% at very small and large $P_T$ and small $\eta$. An increase is also observed at the edges of the pseudorapidity spectra, outside the electron measurable $\eta$ region.

9 $W \rightarrow \mu \nu +$ jets

This section describes the $W \rightarrow \mu \nu +$ jets analysis with a brief look at the relative contributions from signal and backgrounds. This analysis will be expanded further in later revisions. The $W \rightarrow \mu \nu +$ jets signal is simulated with the ALPGEN generator followed by showering with the JIMMY program. Exclusive matching samples with up to 5 partons are combined in proportion to the relative cross sections to yield a combined signal sample. The calculated contributions are scaled to a $1 fb^{-1}$ dataset. We focus on the following four background processes: di-jet events (the so-called QCD background), $t\bar{t}$, $W \rightarrow \tau \nu \tau$ and $Z \rightarrow \mu \mu$. The details of the signal and background datasets are described in Section ???.

9.1 Event selection

Events consistent with $W \rightarrow \mu \nu +$ jet production are expected to pass the single mu20 trigger. Exactly one muon with $p_T > 20$ GeV and $|\eta| < 2.4$ must be isolated; specifically the energy in the $\Delta R = 0.2$ cone around the muon cannot exceed 15 GeV. True $W$ events should exhibit significant missing transverse energy from the neutrino; we require $E_T > 25$ GeV. Finally, we count jets with $E_T > 20$ GeV and $|\eta| < 5.0$, but we remove jets which are within $\Delta R = 0.4$ of the candidate muon.
9.2 Signal-Background Analysis

The number of signal and background events after the selection will be presented in Table 10.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>QCD</th>
<th>$W \rightarrow \mu \nu$</th>
<th>$W \rightarrow \tau \nu$</th>
<th>$Z \rightarrow \mu \mu$</th>
<th>$t \bar{t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Events</td>
<td>1.490$\cdot 10^6$</td>
<td>168.0</td>
<td>11.5</td>
<td>1.6</td>
<td>5.8</td>
</tr>
</tbody>
</table>

Table 10: Number of signal and background events selected for an integrated luminosity of 1$fb^{-1}$. For the background samples, the ratio of the number of background over the number of signal events ($N_B/N_S$) is given in percent.

Figure 22 shows both the expected total jet multiplicity and the jet $E_T$ distribution for selected signal and background events.

Figure 23 shows both the expected reconstructed $W$ transverse mass and the missing $E_T$ distribution for selected signal and background events.

10 Comparison of jet algorithms

The studies presented in the following section will involve comparisons between the $k_T$ algorithm (for the two different jet sizes of 0.4 and 0.6), the legacy Atlas cone and the seedless infrared cone algorithm (SISCone) (both with cone radius of 0.4). From previous experience, the Midpoint algorithm is expected to give very similar results to those from the SISCone algorithm. To examine the impact of algorithms not present in the ATLAS software, we use the SpartyJet framework \[7]. To perform our studies, we use the datasets produced in the context of this Atlas CSC note. In particular, we are interested in the behavior of jet algorithms in a multi-jet environment. The standard jet-lepton isolation cut ($\Delta R > 0.4$) has been used.

10.1 Jet Multiplicity

The first variable we examine is the jet multiplicity for events with a leptonically decaying $W$ and with a number of accompanying partons varying from 0 to 5.

The number of reconstructed jets with $p_T > 15$ GeV for the algorithms Kt04, Kt06, Atlas Cone04 and SISCone04 are shown respectively in black, red, green and blue in the left side of Figure 25, where
Figure 23: $W$ transverse mass (left) and missing $E_T$ (right) distributions for signal and background normalised to $1fb^{-1}$. The contribution from $W \rightarrow \mu \nu$ signal (blue) is shown with the contamination from jet events (green), $W \rightarrow \tau \nu$ (yellow), $Z \rightarrow \mu \mu$ (magenta) and $t\bar{t}$(red).

Figure 24: Number of reconstructed (left) and generated jets (right) for $W + n$ partons Monte Carlo samples, with the number of partons increasing (from 0 to 5) as the plot order.

each plot represents a different number of generated partons. The same quantity (with the same color code) is shown in the right side of the same figure, using as an input the four-vectors of generator level particles after fragmentation. As expected, the distribution of the reconstructed number of jets broadens with increasing number of partons, both at the reconstructed and generator level. Since only jets passing the 15 GeV $p_T$ cut are included, the multiplicity is higher for the Kt06 than for the Kt04 algorithm, since the Kt06 algorithm integrates over a larger area. We also see that SISCone tends to reconstruct a smaller number of jets (usually closer to the actual number of partons) than the other algorithms.

10.2 Matching efficiency

One of the most important characteristics of a jet algorithm is the ability to reconstruct jet 4-vectors at the detector level that are as close as possible to the generated ones. In all comparisons between generated and reconstructed quantities, generator level jets are reconstructed from stable particles at the hadron level, using the same algorithm as at the detector level. Matching efficiencies are defined as the number of hadron level jets in a given $p_T$ or $/eta$ bin that have a reconstructed jet within a given $\Delta R$ cut.

The $\Delta R$ distribution between the generated and the closest reconstructed jet is shown in Figure ?? for the four algorithms studied in the previous section, for a Monte Carlo dataset with $W + 2$ partons. We see that the Kt06 algorithm has the worst matching, probably due to fluctuations far from the core of the
jet. To better understand this difference and its effect on the matching efficiency we study its behaviour as a function of jet kinematics. Figure ?? shows the efficiency for various $p_T$ bins and a range of $\Delta R$ cuts for the algorithms considered in the previous session, on the Monte Carlo dataset of $W + 2$ partons. For all algorithms, an efficiency higher than 95% (in red) is reached at high jet momenta even for quite tight $\Delta R$ cuts, while differences among algorithms emerge at lower jet momenta. If we take slices of this 2d plot corresponding to $\Delta R < 0.1$ and $\Delta R < 0.2$, respectively, we obtain the results in Figures 26.

These plots are produced from a $W + 2$ partons dataset, but all other datasets exhibit a similar behaviour. The algorithms differ in efficiency especially in the low-$p_T$ region and for high-multiplicity events (see Figure ??). SISCone does a very good job under these difficult situations, and fears of the $k_T$ algorithm picking up too much underlying event seem justified only in the case of large jet size. Of course, this study was performed with no additional pileup events and the conclusion may change in the case of additional pileup.

We
References

[1] ATLAS Collaboration, W/Z inclusive cross section CSC note, this volume


[10] https://twiki.cern.ch/twiki/bin/view/Atlas/StreamingTestForUsers

