

# W and Z production in association with light and heavy-flavour jets at the LHC and the Tevatron

Lucia Di Ciaccio<sup>1,a</sup> on behalf of the ATLAS, CMS, CDF and D0 Collaborations

<sup>1</sup>LAPP, CNRS/IN2P3 and Université de Savoie, Annecy-le-Vieux, France

**Abstract.** Recent results on W and Z production in association with light and heavy-flavour jets at the Tevatron and the LHC are reviewed. The integrated luminosity which is used for physics analysis by each of the two Tevatron experiments (CDF and D0) corresponds to a maximum of  $\approx 10 \text{ fb}^{-1}$ , at an energy in the centre of mass of 1.96 TeV. The LHC analyses of ATLAS and CMS use the data sample collected in 2011 which corresponds to an integrated luminosity of  $\approx 5 \text{ fb}^{-1}$ , at an energy in the centre of mass of 7 TeV. Events with W and Z bosons accompanied by jets represent one of the main background samples in the analyses of rare Standard Model processes and in searches for New Physics. Furthermore these measurements provide a test of perturbative QCD calculations and probe the strange and heavy flavour content of the proton.

## 1 Introduction

The production of weak vector bosons in association with jets is an important part of the physics program at the high energy hadron colliders. A good understanding of these processes is of great relevance since it validates our current theory of the strong interactions and since events with weak vector boson plus jets represent one of the main background samples in the analyses of rare Standard Model processes (like the Higgs and the top quark production) and in searches for new particles and new interactions. To this extend, the results are compared with fixed-order QCD predictions and are used to test the performance of Monte Carlo (MC) event generators. Furthermore the study of weak vector bosons in association with c and b-jets puts constraints on the strange and heavy flavour content of the proton.

The Tevatron experiments, CDF[1] and D0[2], have been pioneers in the studies reported in these proceedings. It is of great interest to pursue these analyses at the LHC with the ATLAS[3] and CMS[4] detectors. At the LHC the contribution from process with heavy flavour in the initial state is enhanced and the larger data sets at higher energy allow to investigate regions of the phase space with higher jet multiplicity and higher jet transverse momentum. Larger data samples result in measurements with reduced experimental uncertainties, in particular for final states with heavy flavour jets. Higher order effects in QCD and QED perturbation theory can then be tested.

## 2 Main analysis issues

After selection and background subtraction, the data distributions are corrected and unfolded to account for detec-

<sup>a</sup>e-mail: lucia.di.ciaccio@cern.ch

tor effects like efficiency and resolution. In many analyses the comparison with the theory predictions is done at this level (particle level). When the measurements are compared with existing calculations carried out at parton level, the theory predictions are corrected for non perturbative effects like hadronization and underlying events. The size of this correction is of the order of 5 % in the low  $p_T^{\text{jet}}$  regime and converges with increasing  $p_T^{\text{jet}}$  towards values compatible with no correction.

The theoretical study of the production of weak vector bosons in association with jets is a very active field. Perturbative calculations at Next to Leading Order (NLO) in QCD exists for jet multiplicities up to five [5]. The analyses reviewed in these proceedings compare the experimental results with parton level predictions obtained using MCFM [6] and BlackHat interfaced with Sherpa [7]. MCFM performs fixed order perturbative QCD calculations at NLO of processes with Z or W plus one or two jets. Blackhat in conjunction with Sherpa is used to calculate NLO cross sections for events with a W or Z plus 1, 2, 3 and 4 jets. Efforts are on-going to improve the accuracy beyond the NLO QCD. A currently available theoretical calculation, including beyond NLO QCD corrections, is based on the Loopsim method[8] interfaced with MCFM predictions. This calculation provide an approximate NNLO prediction. The High Energy Jets (HEJ) resummation framework[9] is also compared with data (see section 3).

The recent analyses compare the experimental results also with several MC generators which provide theory predictions at particle level. Pythia[10], Herwig[11], Alpgen[12], Sherpa[13], Madgraph[14] interfaced with Pythia or Herwig, aMC@NLO[15] interfaced with Herwig, and Powheg[16, 17] interfaced with Pythia are used.

Pythia, Herwig are used to generate leading order (LO) matrix elements (ME) for the hard scattering, to evolve ISR and FSR parton showers, to simulate multiple interactions and to implement hadronization. Alpgen, Sherpa and Madgraph are based on LO matrix element calculations and include tree level multiparton emissions up to five partons. Alpgen and Madgraph use the MLM matching scheme in order to merge the LO matrix element results with the parton shower evolution, Sherpa uses the CKKW matching scheme. Another implementation of the NLO perturbative QCD (pQCD) calculations is provided by aMC@NLO and by the Powheg Monte Carlo. They are both interfaced with parton shower (PS) generators. Recent efforts have been made in combining the LO-ME plus PS approach based on the CKKW matching scheme with the Powheg NLO plus PS formalism, leading to the so-called MENLOPS approach as proposed in [18] and implemented in recent versions of the Sherpa MC.

The signature of events with weak vector bosons plus jets is quite distinctive. The analyses with Z bosons select events with two isolated leptons (muons or electrons) of high transverse momentum,  $p_T$ , typically  $p_T > 20$ -25 GeV. The measurements with W bosons require one high  $p_T$  isolated lepton (muon or electron) and significant missing transverse energy, typically  $E_T^{miss} > 25$  GeV. Jets, reconstructed mainly using the midpoint cone algorithm[19] in the Tevatron experiments and the anti- $k_T$  algorithm[20] in the LHC experiments, are required to have a minimum transverse momentum, typically above 25-30 GeV. The mean number of interactions occurring in the same bunch crossing,  $\bar{\mu}$ , is higher at LHC ( $\bar{\mu} \approx 9$  for the 2011 dataset) than at Tevatron ( $\bar{\mu} \approx 2$ ). The superposition (Pile Up) of soft interactions to the hard scattering process affects, in particular, the measurement of the lepton isolation and the reconstruction of  $E_T^{miss}$  and of jets. Robust algorithms and corrections are applied and the systematic uncertainty deriving from this source is negligible.

The measurements of weak vector bosons plus heavy flavour jets are more challenging since the cross sections of these processes are smaller and the background higher compared to the processes with weak vector bosons plus light jets. Jets originated from b and c quarks are identified by exploiting the semileptonic decays of the heavy flavour hadrons and/or their characteristic lifetime. Measurements of track impact parameters and decay lengths of secondary vertices are combined using statistical discriminants. In the W plus b-jet analysis[38], ATLAS uses a neural network to combine different b-tagging algorithms, CMS[39] combine b-tagging variables using a likelihood ratio technique. The chosen working point of these algorithms correspond typically to a b-tag efficiency of  $\approx 50\%$  and to a probability of misidentify c and light jets as b-jets of  $\approx 10\%$  and  $0.1\%$ , respectively. Due to their comparable lifetime, it is challenging to distinguish c and b-jet. The CDF collaboration has obtained the first measurement of the production cross-section of W bosons plus c-jets[28] by exploiting the correlation between the charge of the W boson and the charge of the lepton from the semi-leptonic decay of the charm hadron resulting from the hadroniza-

tion of the c-jet. These charges are of opposite-sign (effects due to slow-rate charmed hadron oscillations are neglected). In the W plus c-jet analyses of the ATLAS[30] and CMS[29] experiments, the charm quark is tagged by the presence of a  $D^\pm$  or a  $D^{*\pm}$  meson reconstructed exclusively and the signal yield is extracted using the charge correlation between the W and the hadron.

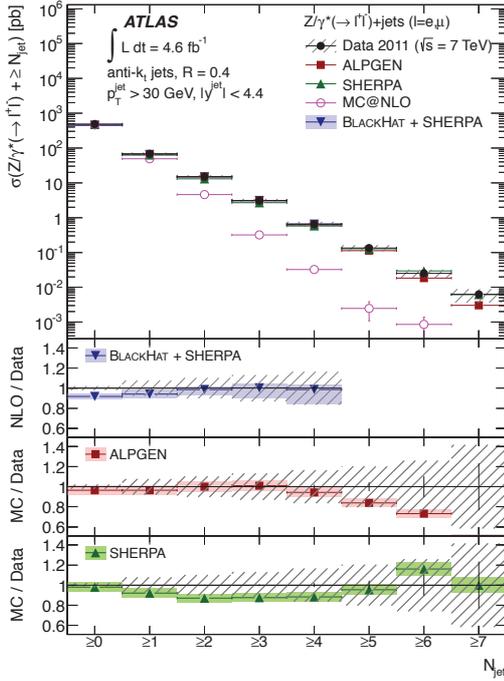
Multi-jet and  $t\bar{t}$  events represent the main sources of background processes. Typical values for the multi-jet background fraction range between 0.5% and few per cent and from few per cent up to 10% in the Z plus jet and W plus jet measurements, respectively. At high jet multiplicity and for high  $p_T^{jet}$ , the  $t\bar{t}$  background is particularly important. Typical values range between 0.5% and 30% and from 10% up to 30% in the Z plus jet and W plus jet measurements, respectively. The amount of the multi-jet and  $t\bar{t}$  background is estimated with data driven methods.

The main systematic uncertainty in all the measurements described in these proceedings derives from the uncertainty on the jet energy scale (JES) and, to a minor extend, from the jet energy resolution (JER). The impact of the JES uncertainty on the measurements is greatly reduced by taking the ratio of cross sections (as for example  $\sigma_{Z+j}/\sigma_{W+j}$ ). For the ratio measurements the uncertainty on the background evaluation often gives the main systematic contribution to the total systematic experimental error. In the measurements of weak vector bosons plus heavy flavour jets the b-tagging efficiency and the purity of the heavy flavour sample are also among the main sources of systematic uncertainty. The statistical uncertainty dominates the experimental error at high jet multiplicities and at high  $p_T^{jet}$ .

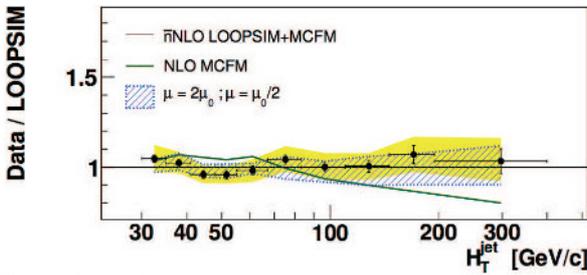
### 3 Results of Z or W plus light jets analyses

Cross sections for a Z boson produced in association with jets have been measured with the ATLAS detector using electron and muon decay modes of the Z boson[21]. The data are compared with predictions from the Alpgen, Sherpa and MC@NLO generators and with fixed-order calculations from BlackHat+Sherpa. Many distributions have been studied. Figure 1 presents the absolute cross sections as function of the inclusive jet multiplicity for multiplicities up to seven jets. The limit of current generators are reached: Alpgen and Sherpa employ ME calculations for multiplicities up to five partons, therefore higher multiplicities are generated by the parton showers. The Figure shows that data are consistent with calculations and with predictions of the generators and that as expected, the MC@NLO program underestimates the observed rate for additional jet emission. The ATLAS analysis concludes that in general the predictions of the matrix element plus parton shower generators and of the fixed-order calculations are consistent with the measured values over a large kinematic range and that effects of missing higher order in the fixed-order QCD or EWK calculations start to appear.

The recent study of the production of Z boson plus jets of the CDF Collaboration presents[22] a comprehensive analysis using the full Tevatron Run II dataset ( $\approx 10 \text{ fb}^{-1}$ ).



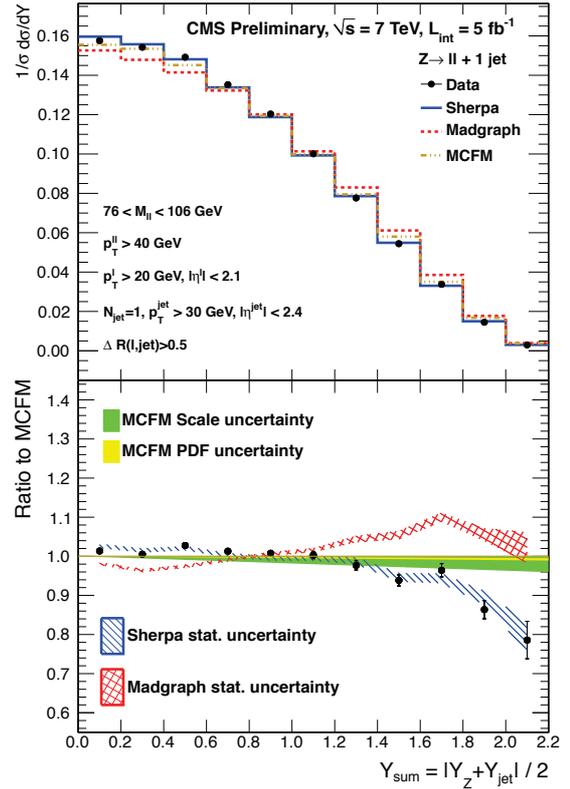
**Figure 1.** Fiducial cross section for Z+jets as a function of the inclusive jet multiplicity,  $N_{jet}$ . The lower panel show theory/data comparisons [21].



**Figure 2.** Ratio of the measured cross section to MCFM and Loopsim+MCFM predictions as a function of  $H_T^{jet}$  in events with a Z boson plus more than one jet[22]. This result has been obtained by the CDF Collaboration.

Many distributions are studied. One interesting example is the  $H_T^{jet}$  distribution, defined as  $H_T^{jet} = \sum_{jet} p_T^{jet}$ . Figure 2 shows the ratio of the measured cross section to the theoretical predictions as a function of  $H_T^{jet}$  for events with a Z plus more than one jet. The NLO MCFM prediction fails to describe the shape of the  $H_T^{jet}$  distribution, in particular it underestimates the measured cross section in the high  $H_T^{jet}$  region. The Alpgen interfaced with Pythia prediction is in good agreement with data, but suffers for a large LO scale uncertainty. Also the Powheg interfaced with Pythia is in good agreement with data, but is still affected by a large NLO scale uncertainty in the  $H_T^{jet}$  region. The approximate NNLO Loopsim interfaced with MCFM prediction properly model the data distribution, and shows a reduced scale uncertainty.

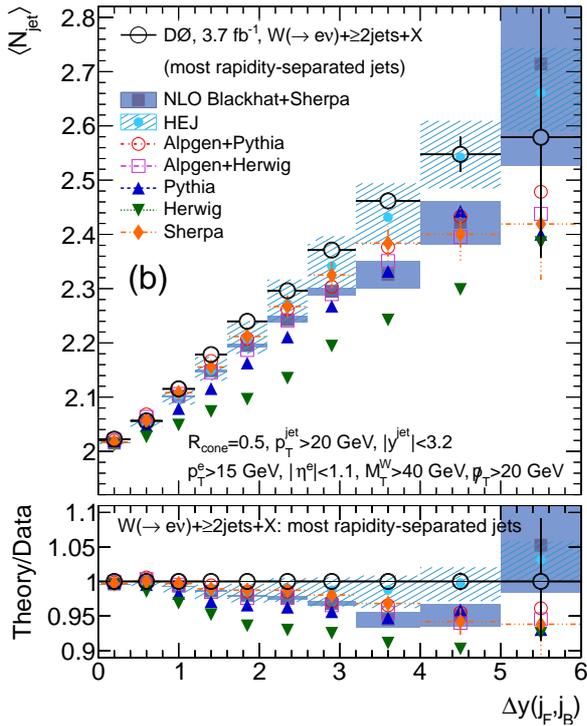
The CMS Collaboration has recently presented results on the azimuthal correlations between the Z boson and the jets[23] and on rapidity distributions for events with



**Figure 3.** Distribution of  $Y_{sum} = |Y_Z + Y_{jet}|/2$  (where  $Y_Z$  is the the Z boson rapidity and  $Y_{jet}$  is the jet rapidity) in events with a Z boson plus one jet. The lower panel show theory/data comparisons[24].

a Z boson in association with one jet[24]. In the analysis of the rapidity distributions, it is found that the Z and the jet rapidity ( $Y_Z$  and  $Y_{jet}$ ) are described to better than 5% by Sherpa, Madgraph and MCFM. In addition to these basic quantities, the variables  $Y_{dif} = |Y_Z - Y_{jet}|/2$  and of  $Y_{sum} = |Y_Z + Y_{jet}|/2$  are studied. The quantity  $Y_{sum}$  approximates the rapidity boost from the laboratory to the centre of mass of the Z-jet system and  $Y_{dif}$  is related to the polar scattering angle of the Z and the jet with respect to the incident proton direction in the two-object centre of momentum reference frame. These two variables are approximately uncorrelated quantities. Figure 3 shows that Sherpa describes better than Madgraph and MCFM the  $Y_{sum}$  distribution. Since both Madgraph and Sherpa use the same matrix elements and showering approaches, the difference is attributed to the matching algorithm, with the Sherpa CKKW scheme reproducing the data better.

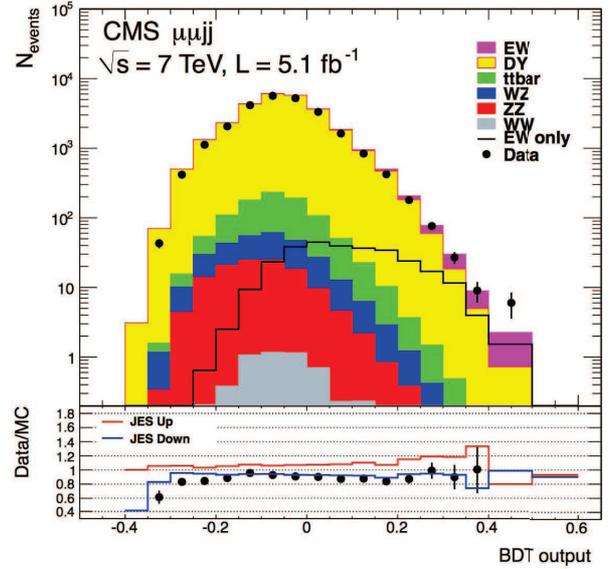
The D0 Collaboration has recently presented[25] a comprehensive analysis of final states with W bosons plus 1, 2, 3 and 4 jets using a  $3.7 \text{ fb}^{-1}$  dataset. Many differential cross sections, normalized to the total W boson cross section and fully corrected for detector effects, are studied and compared with theoretical expectations. Some distributions are found to be well described by NLO pQCD calculations. Nevertheless, in some cases differences are observed compared to the theoretical predictions especially to the Sherpa and Alpgen predictions which underestimate



**Figure 4.** Mean number of jets in inclusive W plus 2-jet events as a function of the dijet rapidity separation of the two most rapidly-separated jets. The lower panel show theory/data comparisons [25].

the emission probability of an additional jet in inclusive W+2-jet events for large dijet rapidity. For example, Figure 4 shows that Blackhat interfaced with Sherpa, Alpgen and Sherpa underestimate the data while the resummation predictions from HEJ are able to accurately describe the jet multiplicity dependence on the jet rapidity separation across the full interval studied, with high precision. These data can be used to improve the modeling of W plus  $n$ -jet production and the emission of QCD radiation.

The electroweak production of a Z boson associated with two forward-backward jets has been studied by the CMS Collaboration[26]. This process probes the triple gauge boson coupling and is an important benchmark in the searches for Higgs boson produced via the Vector Boson Scattering mechanism. The signal is very small with respect to the background which is mainly due to the Drell-Yan (DY) production of Z plus two jets. To distinguish the signal from the background a likelihood discriminator that separates the gluon-originated jets from the quark-originated jets has been used since the jets in the signal sample are quark-jets, while half of the jets in the DY sample originate from the gluons. The signal cross section has been extracted from the a fit to a Boosted Decision Tree(BDT) output distribution built by using jet and lepton kinematic variables (see Figure 5). The main systematic uncertainty derives from the JES and the JER uncertainty. The measurement, combining the muon and electron channels, gives the cross section  $\sigma_{\text{meas}} = 154 \pm 24_{\text{stat}} \pm 46_{\text{exp.syst}} \pm 27_{\text{th.syst}} \pm 3_{\text{lumi}} \text{ fb}$ , in agreement with the theoretical cross section of

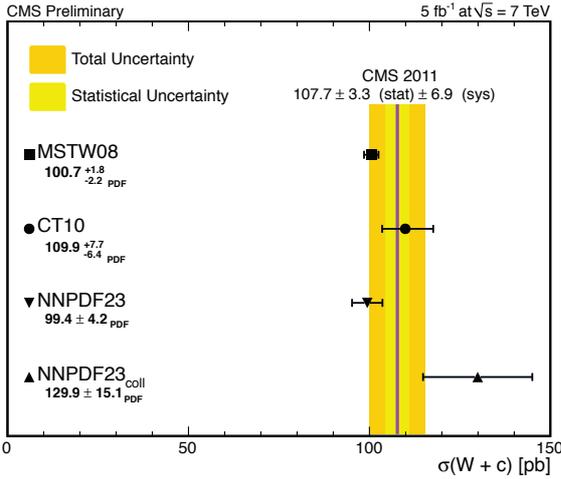


**Figure 5.** BDT output distribution in the study of the electroweak production cross section of a Z boson plus forward-backward jets. The bottom panel shows the ratio of data over the expected contribution of the signal plus backgrounds [26].

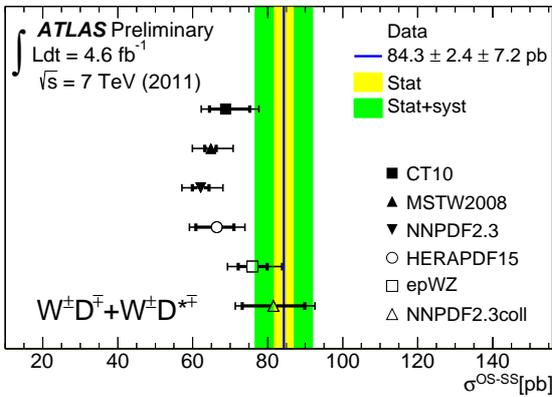
166 fb, calculated with the next-to-leading order QCD corrections[27].

## 4 Results of Z or W plus heavy flavour jet analyses

The measurement of the W boson production in association with a charm quark is sensitive to the gluon and s-quark parton density functions. The first observations of this final state has been reported by the CDF collaboration [28]. The production cross-section measurement, obtained using an integrated luminosity of  $\approx 4 \text{ fb}^{-1}$ , is found in agreement with NLO pQCD predictions of the MCFM calculation. The CMS experiment has also measured the production of W bosons plus charm jets and compared the fiducial cross section and several differential distributions with theoretical predictions obtained with NLO MCFM interfaced with four Parton Distribution Functions (PDF) differing in the s-quark content [29]. The results are consistent with the theory expectations within the uncertainties and a symmetric strange sea, as implemented in NNPDF2.3coll seems very slightly disfavored, as shown in Figure 6 where the comparison of the measured fiducial cross section with the theory predictions is presented. ATLAS has used aMC@NLO interfaced to six PDF sets for this study[30] and has measured the production cross section of  $W^{\pm}$  bosons in association with  $D^{*\mp}$  or  $D^{\mp}$  mesons as well as several differential distributions. The predictions of the shapes of the transverse momentum of  $D^{(*)}$  mesons distributions for the different PDF sets are found similar, but the predicted cross sections may differ by as much as 25%. Figure 7 shows the measurement of the fiducial cross section compared with the theoretical predictions. This result indicates that the NNPDF2.3coll PDF



**Figure 6.** Comparison of the measured cross section  $\sigma(W+c)$  with theoretical predictions computed with MCFM and four sets of PDFs[29].



**Figure 7.** Sum of measured cross sections  $\sigma(WD^* + WD^+)$  compared to theoretical predictions computed with aMC@NLO and six sets of PDFs. The inner error bars on the theoretical predictions show the 68% confidence level uncertainties obtained from the error sets provided with each PDF set, while the outer error bar represents the total theoretical uncertainty (quadratic sum of PDF uncertainty, scale uncertainty and fragmentation uncertainty) [30].

is favored. To reach a firm conclusion and to discriminate among different PDF sets larger samples are needed.

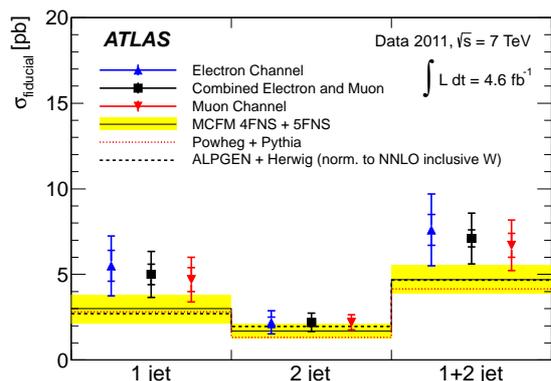
Recently published results of the D0 experiment [31] on the Z production in association with bottom quarks and the recent preliminary results of the CDF experiment [32] on this final state are in fair agreement with NLO MCFM, with Alpgen and Sherpa predictions. These analyses make use of the full data sample available at Tevatron and study several differential distributions. The ATLAS analysis [33] of the same final state also finds a good agreement of the measured cross section with NLO pQCD predictions from MCFM. This ATLAS study shows that the leading order generators are able to reproduce the measured average number of b-jets per Z event within the uncertainties of the measurement, although their predictions differ significantly from each other. The fiducial cross sections mea-

sured by CMS of a Z boson plus exactly one b jet and a Z boson plus at least 2 b jets events are found in good agreement with the MadGraph generator (interfaced with Pythia for the parton shower) in both the five Flavour Number Scheme(FNS) and the four FNS [34]. Comparisons of the kinematic properties with simulations show potential limitations of the MC event generators, which employ the ME plus PS approach at leading order with massless b quarks. Next-to-leading order simulations and/or simulations with massive quarks could possibly do better. Understanding the kinematics of these events is important for searches for yet undiscovered particles in similar topologies. The angular correlation between B hadrons produced in association with a Z boson is also studied using the CMS detector and  $4.6 \text{ fb}^{-1}$  of data [35]. In this analysis B hadrons are identified by displaced secondary vertices and therefore this analysis does not make use of jets. This allows for studying Z+bb production at small angular separations. The angular separation between B hadrons,  $\Delta R$ , is found to be in reasonable agreement with Madgraph in the four FNS and with aMC@NLO. Madgraph in the 5 FNS seems to undershoot the data for small value of  $\Delta R$ .

Using a data sample corresponding to a luminosity of  $1.9 \text{ fb}^{-1}$ , the CDF Collaboration had reported a  $\approx 2.8\sigma$  excess of events with W plus b-jet with respect to the NLO predictions of the MCFM calculations[36]. The new D0 cross section measurement [37] obtained analysing a data sample corresponding to a luminosity of  $6.1 \text{ fb}^{-1}$  is  $\sigma_{W+b} \cdot Br_{W \rightarrow lv} = 1.05 \pm 0.03 \pm 0.12 \text{ pb}$ . This is in better agreement with NLO prediction from MCFM ( $\sigma_{W+b} \cdot Br_{W \rightarrow lv} = 1.34 \pm 0.4 \pm 1 \text{ pb}$ ). At the LHC, W bosons plus b-jet events have been studied by the ATLAS and CMS experiments using  $\approx 5 \text{ fb}^{-1}$ . The recent measurements [38, 39] are complementary since ATLAS has selected events with W bosons and exactly one b-jet, while CMS at events with two jets tagged as b-jets. The fiducial cross section results from ATLAS are shown in Figure 8. All predictions describe reasonably the data, nevertheless the production cross section of a W boson plus one b-jet is  $\approx 1$  sigma above the calculations and the difference is found to increase with the transverse momentum of the b-jet. The measurement has been repeated including the single top background in the signal. The uncertainty on the result becomes smaller, and the tension between the data and the predictions increases. Larger data sample will help in understanding the comparison. The CMS measurement of the production cross section of events with a W boson plus two b-tagged jets is in agreement with NLO pQCD calculation (MCFN) and the kinematics is also well described by Madgraph interfaced with Pythia. Understanding the production of W plus two jets tagged as b-jets is crucial as these events constitute a major background in the search for the standard model Higgs boson decaying to  $b\bar{b}$  when produced in association with a weak vector boson.

## 5 Conclusion

The study of weak gauge boson plus jets is a very active field: many analyses have been performed. They extend



**Figure 8.** Fiducial cross-sections of the production of W+jet. The yellow bands represent the total uncertainty on the predictions[38].

the previously probed phase space, test and challenge the predictions based on very recent calculations. The Tevatron analyses have paved the way, the higher statistic at LHC allows more precise measurements in particular of W or Z plus heavy flavour jets. In general NLO predictions describe fairly well the data, nevertheless in some distributions some tension between calculations and measurements start to appear. This is attributed to missing higher order (QCD or EWK) or to the matching schemes and needs to be understood. Efforts have started and discussions are going on among experiments and with theorists. Large data samples will help. The LHC data recorded in 2012 are still in the process to be analysed: a reduction of the statistical uncertainty (in high multiplicity, high  $p_T$  bins, and in heavy flavour channels) are expected as well as of some systematic uncertainties depending on the size of the data sample, (like the jet energy scale). Understanding the W or Z plus light or heavy jet production is a key for finding New Physics.

## References

- [1] CDF Collaboration, Phys. Rev. D **71**, 052003 (2005).
- [2] D0 Collaboration, Nucl. Instrum. Methods Phys. Res. A **565**, 463 (2006).
- [3] ATLAS Collaboration, JINST **3**, S08003 (2008).
- [4] CMS Collaboration, JINST **3**, S08004 (2008).
- [5] L.Dixon, CERN Academic Training 2013. Matteo Cacciari, Stefano A. Pozzorini, Leif Lonnblad, this proceedings.
- [6] John Campbell, R. K. Ellis., Phys. Rev. D **65**, 113007 (2002).
- [7] C. F. Berger et al., Phys. Rev. D **82**, 074002 (2010). H. Ita et al., arXiv:1108.2229
- [8] Mathieu Rubin et al, J. High Energy Phys. **1009** 084 (2010).
- [9] Jeppe R. Andersen et al, J. High Energy Phys. **1001**, 039 (2010).
- [10] T. Sjöstrand et al., J. High Energy Phys. **0605**, 026 (2006).
- [11] G. Corcella et al., J. High Energy Phys. **0101**, 010 (2001).
- [12] M.L. Mangano et al., J. High Energy Phys. **0307** 001 (2003).
- [13] T. Gleisberg et al., J. High Energy Phys. **0902** 007 (2009).
- [14] J. Alwall et al., J. High Energy Phys. **6** 128 (2011).
- [15] S. Frixione and B. R. Webber, J. High Energy Phys. **0206**, 029 (2002).
- [16] P. Nason, J. High Energy Phys. **0411**, 040 (2004).
- [17] S. Frixione et al., J. High Energy Phys. **0711**, 070 (2007).
- [18] Stefan Hoeche et al, arXiv:1009.1127
- [19] G. C. Blazey et al., FERMILAB-CONF-00-092-E (2000)
- [20] M. Cacciari et al., J. High Energy Phys. **0804**, 063 (2008).
- [21] ATLAS Collaboration, arXiv:1304.7098
- [22] S. Camarda, FERMILAB-THESIS-2012-12
- [23] CMS Collaboration, Phys. Lett. B **722** (2013) 238
- [24] CMS Collaboration, CMS-PAS-SMP-12-004
- [25] D0 Collaboration, arXiv:1302.6508
- [26] CMS Collaboration, arXiv:1305.7389
- [27] K. Arnold et al., arXiv:1107.4038
- [28] CDF Collaboration, Phys.Rev.Lett. **110**, 071801 (2013).
- [29] CMS Collaboration, CMS-PAS-SMP-12-002
- [30] ATLAS Collaboration, ATLAS-CONF-2013-045
- [31] D0 Collaboration, Phys. Rev. D **87**, 092010 (2013).
- [32] CDF Collaboration, CDF Note 10594.
- [33] ATLAS Collaboration, Phys. Lett. B **706**, 295 (2012).
- [34] CMS Collaboration, CMS-PAS-SMP-13-004
- [35] CMS Collaboration, CMS-PAS-EWK-11-015
- [36] CDF Collaboration, Phys. Rev. Lett. **104**, 131801 (2010).
- [37] D0 Collaboration, Phys.Lett. B **718**, 1314-1320 (2013).
- [38] ATLAS Collaboration, arXiv:1302.2929
- [39] CMS Collaboration, CMS-PAS-SMP-12-026