W/Z+jets with the CMS detector

Massimo Nespolo
on behalf of the CMS collaboration
Università di Padova and INFN, Via Marzolo 8, 35131 Padova, Italy
E-mail: massimo.nespolo@gmail.com

Abstract. We present a study of jet production in association with W and Z bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV using the full 2010 data set collected by the CMS experiment corresponding to an integrated luminosity of $36\pm4$ pb$^{-1}$. The production of vector bosons with jets provides a stringent and important test of perturbative QCD, and is an important background to the study of top quark properties and in searches for new physics. We report the inclusive rates of jets produced (normalized to the total cross section) as well as the ratios $\sigma(V+(n+1)\text{-jets})/\sigma(V+n\text{-jets})$ for a jet threshold of 30 GeV. We also present a measurement of $Z \rightarrow ll + b$ and the first test of the Berends-Giele scaling at $\sqrt{s} = 7$ TeV.

1. Motivation

The production of vector bosons with jets provides a stringent test of perturbative QCD, and is an important background to the study of top quark properties and in searches for new physics. Events with jets produced in association with W and Z bosons are usually simulated with Matrix Element + Particle Showers generators (MadGraph$^1$+PYTHIA$^2$). This means that the description of this process is at the tree level, so the overall normalization is uncertain, and a procedure to combine partons and fragmentation is needed. This situation has improved due to next-to-leading order (NLO) calculations, which were recently performed for up to $W+4$ jets$^3$, and up to $Z+3$ jets$^4$. However, also in these cases the precision varies from 10% to 30%, due to uncertainties on parton distribution function, and due to the choice of renormalization and factorization scale.

It is therefore essential to use events tagged by the presence of $W/Z$ bosons at the hadron colliders to probe QCD calculations in an environment which is relatively clean. In this way, it is possible to extract background normalizations for lepton(s)+MET+jets final states, which are important not only for Standard Model (SM) measurements (top, Higgs), but also for searches beyond the Standard Model (BSM), such as SUSY.

Moreover, we can observe that this analysis in itself may also be potentially sensitive to new physics, since heavy particles predicted by several BSM theories can show their presence through an excess of jets at high $p_T$, or through a deviation from the expected scaling.

2. CMS Detector

The central feature of the CMS detector is a 6 m diameter superconducting solenoid which provides a 3.8 T axial field. Located within the field volume are the silicon pixel and strip tracker, the crystal electromagnetic calorimeter and the brass/scintillator hadron calorimeter.
Muons are measured in gas-ionisation detectors embedded in the steel return yoke. A detailed description of the detector and its performance can be found in [5].

3. Selection and First Results

For the muon channel, we only use events which, according to the High Level Trigger (HLT) system of the CMS detector, have at least one muon with $p_T > (9-15)$ GeV (the threshold was raised following the increasing instantaneous luminosity of the LHC). In offline analysis, we require presence of one muon well fitted from hits in the tracker and in the muon chambers and consistent with ID designed to reject punch-through and decay in flight. This muon must have $p_T > 20$ GeV, $|\eta| < 2.1$, and must be isolated: $(\sum p_T(tk) + \sum E_T(had + em))/p_T \mu < 15\%$. If there is a second muon with $p_T > 10$ GeV, $\eta < 2.4$ and $50$ GeV $< M_{\mu\mu} < 120$ GeV, the event is put in the $Z$ sample; otherwise, if $M_T(\mu\nu) > 20$ GeV, it is put in the $W$ sample.

Similarly, in the electron channel we select only events with one electron with $p_T > (10-17)$ GeV at the HLT. For the final analysis, we require presence of one electron satisfying tight (80% efficient) isolation, electron identification, and conversion rejection criteria. This electron must have $p_T > 20$ GeV, $|\eta| < 2.5$ ($1.44 < |\eta| < 1.57$ excluded), and must match the trigger primitive. Events were assigned to the $W$ or to the $Z$ sample using the same criteria described for the muon channel, with the only difference that the second electron was required to pass a loose (95% efficient) ID, and that the $\eta$ cut was identical for the two electrons.

Jets are reconstructed using the anti-kt clustering algorithm [6] with $\Delta R = 0.5$ (the default in CMS), applied to a list of “particles” identified by the Particle Flow framework [7] (PF-jets). These jets are accepted only if they have $E_T > 30$ GeV and $|\eta| < 2.4$ (which is the tracker acceptance). We also apply a loose ID to remove calorimetric noise [8], apply data-driven energy calibration, and pile-up energy offset removal procedure based on the event energy density (see [9] and [10]). Leptons from Vector Boson decay should not be counted as jets, so isolated muons are removed from the list of particles before PF jets are clustered, and jets which have $\Delta R < 0.3$ from the selected electron are also not considered. Additional details are given in [11].

The transverse momentum distribution of the leading jet for $W+1$ jet and $Z+1$ jet is shown in figures 1-4, while the distribution for exclusive jet multiplicity in $W/Z+n$ jets ($n=0$ to 6) is shown in figures 5-8. To remove further background, an additional cut of $M_T > 50$ GeV is applied to the $W$ events. The data is in very good agreement with the MadGraph predictions normalized to the NNLO cross sections.

4. Signal Extraction and Corrections

The MC simulations show that the procedure described in the previous section to identify the $Z$ boson produces a very clean resonant peak with extremely low background for all jet multiplicity; the corresponding signal is therefore extracted with an Unbinned Maximum-Likelihood (UML) fit to the di-lepton invariant mass spectra. In this case, a functional form is assumed (Cruijff for signal, exponential accounting for all backgrounds) with floating background and the signal parameters also are left floating, but are kept equal for all jet multiplicity.

The situation is more complicated for $W$ events, because its experimental signature (a single isolated high-$p_T$ lepton with missing transverse energy) can be mimicked by many processes for physical reason (i.e. presence of a real $W$) or for detector effects (mainly acceptance and fake leptons). The residual background in the transverse mass spectra can be divided into two categories: not peaking at $M_W$ (mainly QCD), and peaking at $M_W$ (top, which is important in $W+3,4$ jets). To cope with this, an UML 2-dimensional fit is performed on the $M_T$ distribution vs the number of $b$-jets $n^{b\text{-jets}}_{\text{tagged}}$, the latter being identified with the “Track Counting High Purity” $b$-tagging algorithm for the muon channel and with the “Track Counting High Efficiency” algorithm for the electron channel [12]. Also in this case a functional form is assumed, and the fit is constrained in the range $20 < M_T < 150$ GeV. Two parameters of the fit (b-tagging
efficiency and mis-tag probability) are fixed to the corresponding values extracted from MC, and are cross-checked with a di-leptonic top-enriched data sample.

Many corrections had to be applied to the fitted yields: first of all, we took into account the lepton efficiency measured with Tag & Probe method on Z events, and factorized into reconstruction, selection, and trigger efficiency. For the W candidates, we corrected (using MC and data) for the fact that the range $0 < M_T < 20$ GeV was not included in the fit. Finally, to remove the effects of imperfect energy resolution and reconstruction efficiency, we used MadGraph simulation to extract the migration matrix $R(n_{\text{RECO}}, n_{\text{Gen}})$, and applied a Singular Value Decomposition (SVD, see [13]) to “unsmear” the $n_{\text{jet}}$ distribution.
5. Ratios, Berends-Giele Scaling and Systematics

Finally we measured two types of ratios: the production of ≥ n jets over the total cross-section, and the ratio of \((V+ ≥ n\text{-jets})/(V+ ≥ (n-1)\text{-jets})\). The usage of ratios greatly reduces the systematics. We obtain in fact a full cancellation of the contribution due to luminosity (which had a big uncertainty, approximately 10% at the time this measurement was done) and a partial cancellation of the effects produced by the jet energy scale, so only lepton efficiency versus the number of jets matters. The main source of the remaining systematic uncertainty is the inaccuracy in the determination of the jet energy which moves jets above and below certain threshold, changing the measured jet multiplicity. This effect is produced by the uncertainties on

---

**Figure 5.** Exclusive reconstructed jet multiplicity with \(W \rightarrow e\nu\), not corrected for efficiency and detector effects.

**Figure 6.** Exclusive reconstructed jet multiplicity with \(W \rightarrow \mu\nu\), not corrected for efficiency and detector effects.

**Figure 7.** Exclusive reconstructed jet multiplicity with \(Z \rightarrow e^+e^-\), not corrected for efficiency and detector effects.

**Figure 8.** Exclusive reconstructed jet multiplicity with \(Z \rightarrow \mu^+\mu^-\), not corrected for efficiency and detector effects.
data-driven JEC, by the JEC flavour dependence (estimated from MC), and by pile-up removal (500 MeV in excess for each jet in MC). Moreover, we also considered jet energy resolution, the pile-up residual effect after pile-up subtraction, and for the $W$, the effect of mismeasurement on MET (from fitting $M_T$ on data). Finally, the effect of reconstruction/selection efficiency (close to the previous one) and signal extraction (important only at high jet multiplicity) were evaluated.

Figure 9. Ratios for $W \rightarrow \mu\nu$, compared with expectation from MadGraph and PYTHIA.

Figure 10. Ratios for $W \rightarrow e\nu$, compared with expectation from MadGraph and PYTHIA.

Figure 11. Ratios for $Z \rightarrow e^+e^-$, compared with expectation from MadGraph and PYTHIA.

Figure 12. Ratios for $Z \rightarrow \mu^+\mu^-$, compared with expectation from MadGraph and PYTHIA.
The results are shown in figures 9-10 for the $W$+jets (muon and electron channel respectively), and in figures 11-12 for the $Z$+jets (again, for the muon and electron channel respectively). The measured ratios are in excellent agreement with MC predictions from ME+PS (MadGraph), while PYTHIA alone start to fail for $n_{\text{jets}} \geq 2$ in $W$+jet events. In $Z$+jet events both calculations are compatible with data because of limited statistics and hence larger errors.

For what concerns the Berends-Giele scaling [14], the LO calculation would predict a constant...
value for the ratio $C_n = \sigma_n/\sigma_{n+1}$, but the NLO corrections and/or phase space effects could violate this simple proportionality. To check possible deviations, we assumed a linear dependence from $n$ ($C_n = \alpha + \beta n$), and we performed a 2-parameter fit on the exclusive jet multiplicity spectra.

The results obtained (excluding events with no jet recoiling against the Vector Boson, and taking into account the bin-to-bin migration) for $W$ and $Z$ are shown in figure 13 and 14 respectively. In the first case, $\beta$ is compatible with 0 within 1$\sigma$, and the difference between the two decay channels are due to the $\Delta R$ cut (which is not applied to the muons); the $Z$ plots show that $\beta$ is even more compatible with 0 (within 0.5$\sigma$), and that there is a very good agreement with the ME+PS prediction.

6. $Z + b$ jets
The production of $Z$ decaying into leptons in association with $b$-quarks is a benchmark channel for the identification of Higgs (especially MSSM Higgs) at the LHC, and is also a background for SM and BSM searches involving leptons + $b$-jet final states. However, theoretical calculations of the corresponding cross section cannot be precise enough, since there is a 30% difference ([15], p.53) between the numbers obtained with the fixed-flavour scheme (where $b$-quarks are produced explicitly in pairs from gluon splitting), and those predicted using the variable-flavour scheme (which allows the $b$-quarks to participate directly in the hard scattering, by integrating the gluon splitting process into the PDF).

For this measurement (see [16]), events were selected requiring one $Z$ (decaying either into muons or electrons) and at least one PF jet with $E_T > 25$ GeV, $|\eta| < 2.1$ (which implied that the jet was contained completely inside the tracker), $\Delta R(\text{jet, lepton}) > 0.5$, and one or more secondary vertices reconstructed inside the jet itself. Two versions of b-tagging called Simple Secondary Vertex (SSV, see [12]) with high efficiency and high purity were used, and top candidates were rejected by applying the cuts MET > 40 GeV, and 60 GeV < $M_{ll}$ < 120 GeV.

![Figure 15.](image1.png)  
**Figure 15.** Invariant mass of opposite-sign lepton pairs for the high-purity selection, compared with MC simulation.

![Figure 16.](image2.png)  
**Figure 16.** Binned likelihood fit of the mass of the secondary vertex in data events, for the high-efficiency selection, compared with MC simulation.

The resulting invariant mass spectrum for the High-Purity SSV b-tagging algorithm is shown
in figure 15, and has been compared with the MC expectation for $Z +$ different quark flavour. The purity of this sample was extracted with a binned likelihood fit of the mass of the secondary vertex (see figure 16 for the High-Efficiency SSV), using templates from MC. Both the measured purity and the ratio $(Z + b)/(Z + j)$ are in good agreement with MadGraph simulation within theoretical and experimental uncertainties.

7. Conclusions

We have studied the production of jets in association with a $W$ or a $Z$ boson, and the production of $b$-jets in association with a $Z$, where the Vector Boson decays into electrons or muons with the full dataset of $36 \pm 4 \text{ pb}^{-1}$ collected by the CMS detector in 2010.

We measured the leading jet $p_T$ spectra, the ratios $\sigma(V+n\text{-jets})/\sigma(V)$ and $\sigma(V+n\text{-jets})/\sigma(V+(n-1)\text{-jets})$, the Berends-Giele scaling, and the ratio $(Z + b)/(Z + j)$. In all cases, the agreement with MC calculation (ME+PS, in particular) is very good, and the precision achieved shows that CMS is ready for searches beyond the Standard Model.

References

[1] F. Maltoni and T. Stelzer, “MadEvent: Automatic event generation with MadGraph”, JHEP 02 (2003) 027, arXiv:hep-ph/0208156.
[2] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, JHEP 05 (2006) 026, arXiv:hep-ph/0603175.
[3] C. F. Berger, Z. Bern, L. J. Dixon et al., “Precise Predictions for $W + 4$-Jet Production at the Large Hadron Collider”, Phys. Rev. Lett. 106 (Mar, 2011) 092001.
[4] C. F. Berger, Z. Bern, L. J. Dixon et al., “Next-to-leading order QCD predictions for $Z, \gamma^* + 3$-jet distributions at the Tevatron”, Phys. Rev. D 82 (Oct, 2010) 074002.
[5] CMS Collaboration, “The CMS experiment at the CERN LHC”, JINST 3 (2008) S08004.
[6] M. Cacciari, G.P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, JHEP 0804 (2008) 063.
[7] CMS Collaboration, “Commissioning of the Particle-Flow Event Reconstruction using high-energy LHC Collisions in the CMS Detector”, CMS PAS PFT 1-002 (2010).
[8] CMS Collaboration, “Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter”, JINST 5 (2010) T03014, arXiv:0911.4881.
[9] M. Cacciari and G.P. Salam, “Pileup subtraction using jet areas (hep-ph/0707.1378)”, Physics Letters B 659 (2008) 119126.
[10] M. Cacciari, G.P. Salam and G. Soyez, “The catchment area of jets (hep-ph/0708.1188)”, JHEP 0804 (2008) 005.
[11] The CMS Collaboration, “Rates of Jets Produced in Association with $W$ and $Z$ Bosons”, CMS PAS EWK 10-012 (2011).
[12] CMS Collaboration, “Commissioning of b-jet identification with pp collisions at $\sqrt{s} = 7 \text{ TeV}$”, CMS PAS BTV 10-001 (2010).
[13] A. Hocker and V. Kartvelishvili, “SVD Approach to Data Unfolding”, Nucl. Instrum. Meth. A 372 (1996) 469481, arXiv:hep-ph/9509307.
[14] F. A. Berends, W. T. Giele, H. Kuijf et al., “Multi-jet production in $W, Z$ events at pp colliders, Phys. Lett. B 224 (1989) 237.
[15] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, arXiv:hep-ph/1101.0593.
[16] The CMS Collaboration, “Observation of $Z + b, Z \to e\mu, \mu\mu$ with CMS at $\sqrt{s} = 7 \text{ TeV}$”, CMS PAS EWK 10-015 (2011).