Jet production in association with vector bosons at CMS

Matteo MARONE[∗]**, on behalf of the CMS collaboration**

University of Trieste and INFN

E-mail: marone@cern.ch

The associated production of jets and vector bosons allows for stringent tests of perturbative QCD calculations and is sensitive to the possible presence of new physics beyond the Standard Model. Measurements of jet production rates in association with W, Z or photons, in proton-proton collisions at a 7 TeV center-of-mass energy are presented, using data collected with the CMS detector. In particular, we compare data to the theory predictions on jet rates, dijet invariant mass, angular correlations and event shapes distributions.

[∗]Speaker.

In this report an update of the most recent studies on the production of vector boson associated with jets is presented. Common aim of the analyses described in this document is to provide comparisons between experimental data and the predictions available from several Monte Carlo (MC) generators accessible at the moment (MADGRAPH, SHERPA, PYTHIA, POWHEG, aMC@NLO).

This summary is divided into three sections, each one is devoted to a different measurement. Analyses profit from the use of different datasets, corresponding to different integrated luminosity: 2010 data consists of a luminosity of 36 pb^{-1}, while the full 2011 dataset collected by CMS[1] is 5.0 fb^{-1}.

Z and W bosons are detected through their decay in electrons or muons, relying on the excellent CMS detector performance and reconstruction algorithms. The event selection is done in several steps: an hardware-based trigger system selects leptons with a minimum associated transverse energy (depending on instantaneous luminosity). Then, the contamination from fake leptons and hadronic decays is reduced using several isolation variables, considering the detector activity in a cone of $\Delta R \leq 0.3$ around the leptons. Further cuts are applied in order to remove misidentified electrons from converted photons and to eliminate the contribution of muons from cosmics.

Jets are reconstructed from the particle collection created with the Particle-Flow (PF) algorithm and formed with the anti-kt clustering algorithm with a size parameter of R = 0.5. Jet energy corrections (JEC) are applied to account for the jet energy response as a function of η and E_T.

Efficiencies are determined in a data driven way, using the well known technique called "Tag and Probe" (TAP). An unfolding procedure is applied to remove the effect of the detector, and to allow a direct comparison with the theoretical models.

Common sources of systematic uncertainties are essentially the jet energy scale, the effect of the pile-up and its subtraction, the estimation of the efficiencies and the unfolding method used to allow generator level comparison.

1. Rates of Jets Produced in Association with W and Z Bosons in pp Collisions at $\sqrt{s}$ = 7 TeV

In this section, the study of jet production in association with W and Z bosons (V+jets) in proton-proton collisions at $\sqrt{s}$ = 7 TeV, using the full 2010 dataset collected by CMS is presented [2].

This process is an important test of perturbative QCD calculation, given the availability of "Next-to-leading order" (NLO) predictions up to four (three) jets for the Z (W) boson. The precision of the NLO predictions is limited by the uncertainties on parton density functions (PDF): it can vary from 10% up to 30%. V+jets production is also an important background for the searches of new physics, such as the Higgs boson or SUSY.

Signal selection starts with the identification of the leptons within the CMS fiducial region and requiring leptons with at least 20 GeV. After the identification of the first lepton, the energy threshold is lowered to 10 GeV in order to find possible second leading leptons. If the second leading lepton is found, a Z boson candidate is acquired: the invariant mass between the two leptons is then calculated and it must lie within the window 60-120 GeV. If such lepton is not found, the event is

assigned to the W+jets sample.

For W+jets samples, the missing transverse energy is calculated in order to obtain the event transverse mass M_T, and thus selecting events with $M_T > 20$ GeV.

The signal yield is estimated by means of an extended maximum-likelihood fit to the leptons invariant mass for the Z boson candidates, while a fit to the transverse mass M_T distribution is done for the W (Figure 1). For the Z+jets sample, backgrounds arise essentially from $t\bar{t}$ and W+jets: their
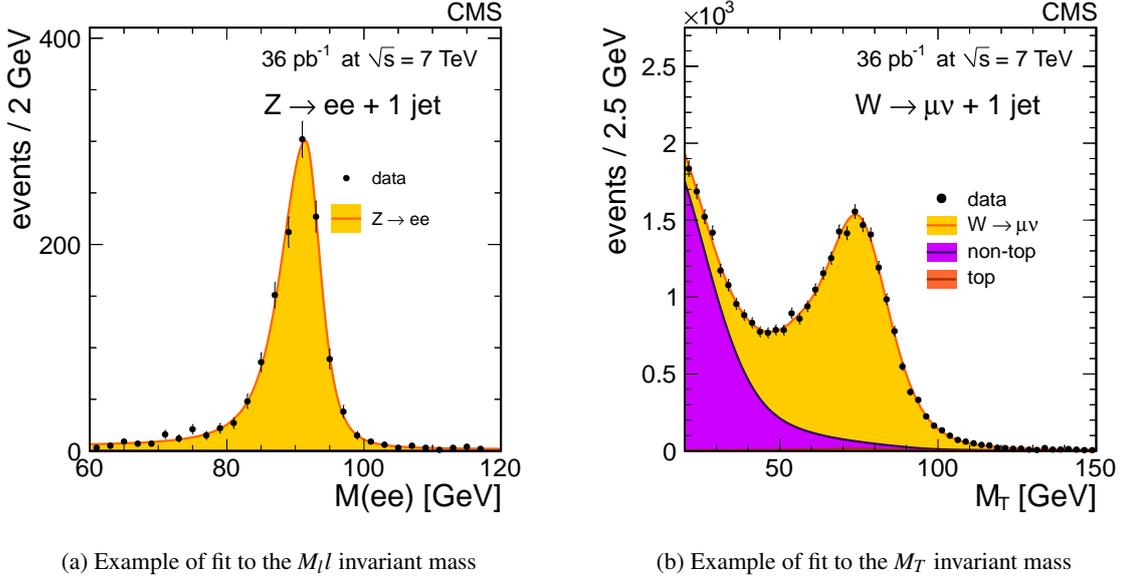
(a) Example of fit to the $M_l l$ invariant mass   (b) Example of fit to the M_T invariant mass

Figure 1: (left) Fit results for the Z + 1 jet sample. Background contribution is too small to be visible. (right) Fit results for the W + 1 jet sample. Fit results are shown by the colour-filled areas.

contribution, comparing the number of signal and background events under the peak, is found to be very small (see Figure 1). For W+jets events, background can be classified in two components. One component shows a peaking structure in M_T (mainly from $t\bar{t}$) and another one without it (in this case, W+jets is the dominant contribution).

V+jets events are finally counted when at least one jet has transverse energy above 30 GeV and when it falls within the acceptance $|\eta| \leq 2.4$. For further jets, same cuts are applied as for the first one.
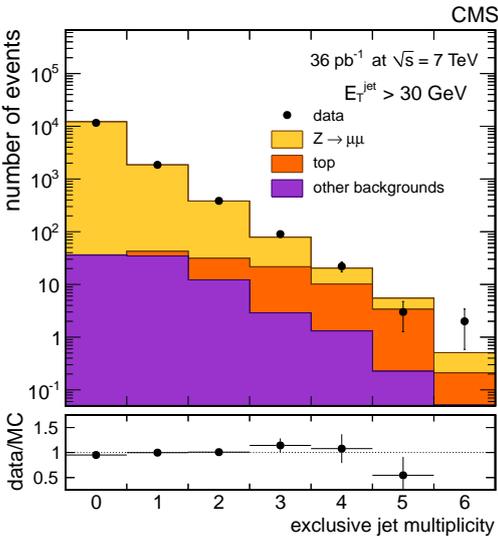
1.1 Results

Figure 2 shows the final distribution of selected events, divided in bins of exclusive jet multiplicity, together with the various backgrounds included in this analysis.
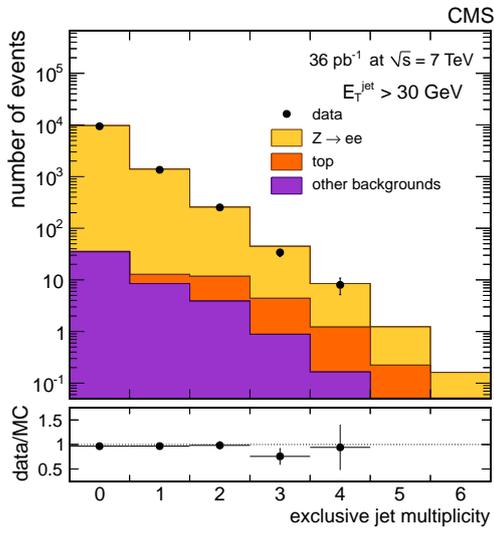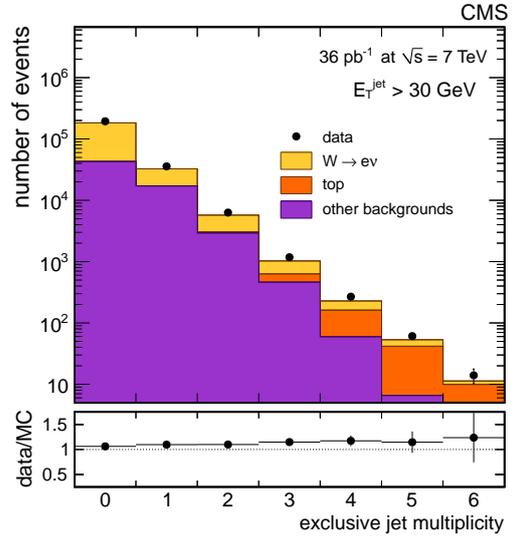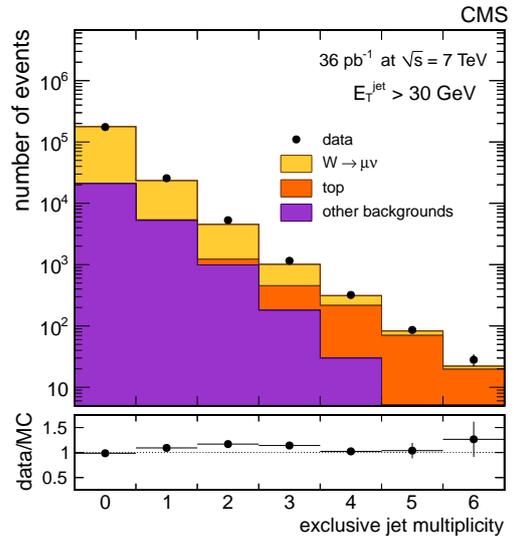
Figure 2: Exclusive numbers of reconstructed jets in events with W or Z bosons decaying in electrons or muons. MC predictions and backgrounds are superimposed.

From the unfolded inclusive jet multiplicity distributions, obtained from Figure 1 applying an unfolding technique (Bayesian method), two sets of ratios are then obtained. The first one (see Figure 3) is $\frac{\sigma(V+njets)}{\sigma(V)}$, where $\sigma(V)$ is the inclusive cross section. The second set, $\frac{\sigma(V+njets)}{\sigma(V+(n-1)jets)}$, is also shown in Figure 3. In these ratios, contributions from the two main sources of systematics (jet

(a) Z→ee sample

(b) W→ νe sample

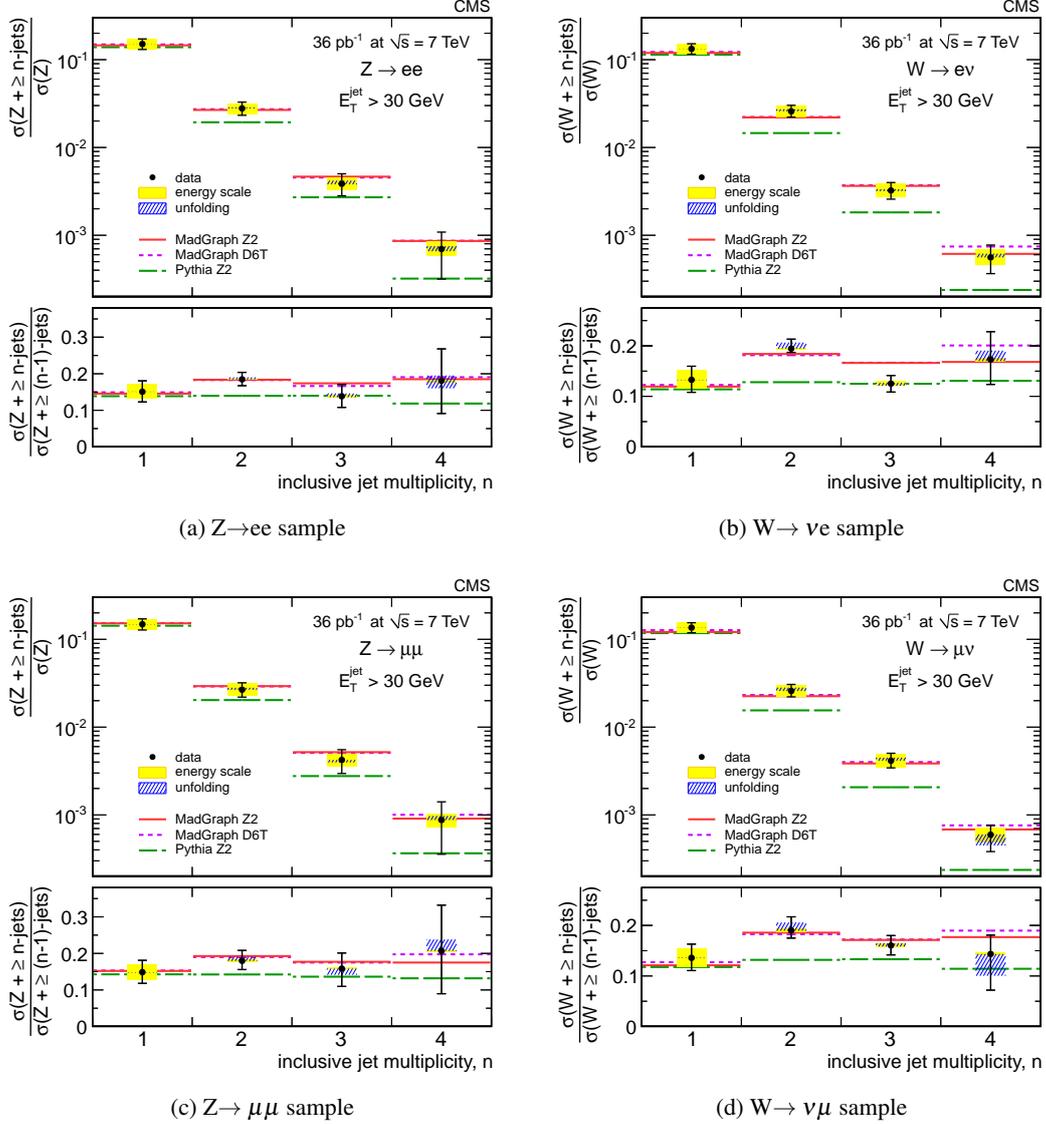(c) Z→ μμ sample

(d) W→ νμ sample

Figure 3: Ratio $\frac{\sigma(V+njets)}{\sigma(V)}$ and $\frac{\sigma(V+njets)}{\sigma(V+(n-1)jets)}$, with $\sigma(V)$ the inclusive cross section. The uncertainties due to the energy scale and unfolding procedure are superimposed. Error bars represent the total uncertainty.

energy scale and unfolding) are superimposed, as error bands.

When n ≥ 2 jets, PYTHIA pure parton shower simulation fails to describe the experimental spectrum, while MADGRAPH presents an overall good agreement. Sensitivity to the tuning of the underlying events (D6T and Z2 tunes have been studied) is negligible, as shown in Figure 3, given the high jet energy threshold (30 GeV).

The ratio of W+jets and Z+jets cross section as a function of the jet multiplicity is shown in Figure 4, together with the corresponding MC predictions (MADGRAPH and PYTHIA, Z2 tune only). An overall agreement is again found between data and Monte Carlo predictions.

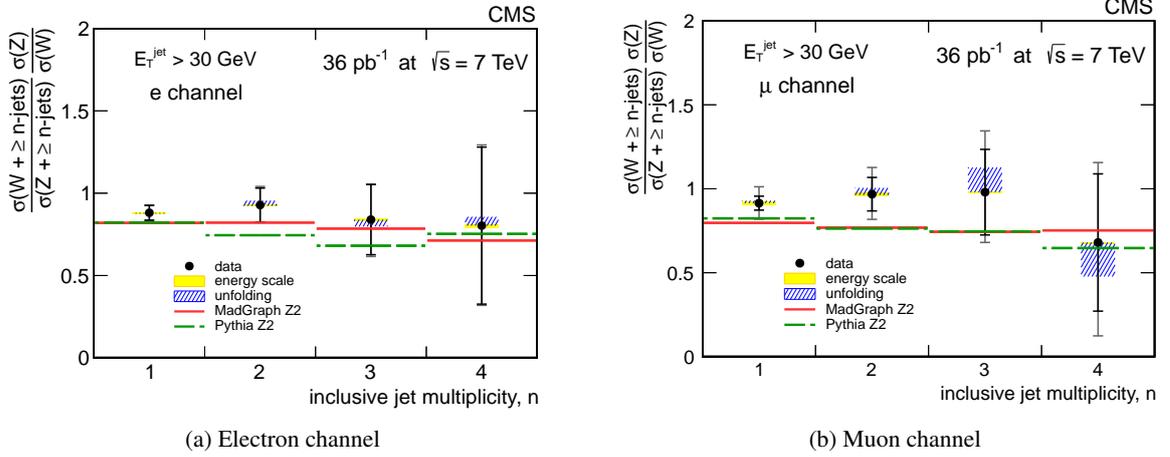(a) Electron channel   (b) Muon channel

Figure 4: Ratio of W+jets and Z+jets cross section as a function of the jet multiplicity. The uncertainties due to the energy scale and unfolding procedure are superimposed. Error bars represent the total uncertainty.

The W charge asymmetry, defined as A=$\frac{W^+ - W^-}{W^+ + W^-}$ is presented in Figure 5. Charge asymmetry

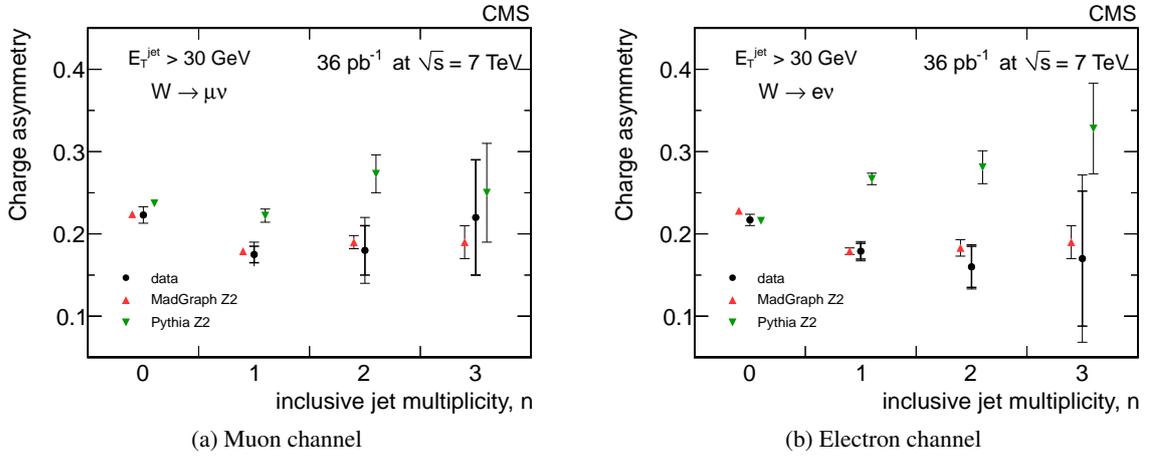(a) Muon channel   (b) Electron channel

Figure 5: W charge asymmetry A=$\frac{W^+ - W^-}{W^+ + W^-}$. Error bars on the data points show the statistical and total uncertainties. Error bars on the simulation points correspond to the statistical uncertainty only.

depends on the number of associated jets because the fraction of u(d) quarks contributing to the process is different in each case. Good agreement with the predictions based on MADGRAPH is found. PYTHIA predictions do not describe properly the W charge asymmetry.

At the end, the results of the fit for α and β parameter in the "Berends-Giele" scaling, for both W and Z plus jets is presented in Figure 6. The exclusive jet multiplicity bins is fitted with the
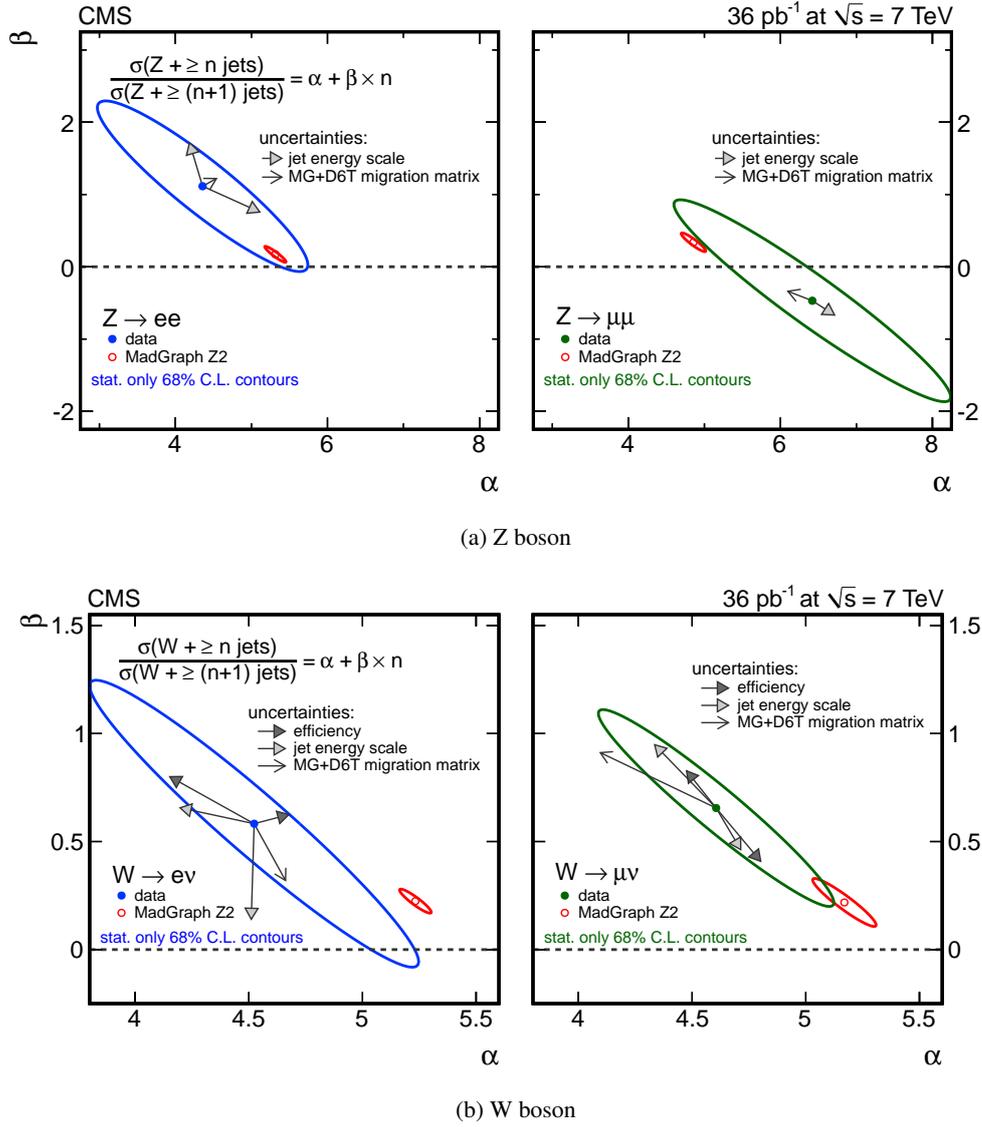
(a) Z boson

(b) W boson

Figure 6: Berends-Giele scaling, for both W and Z plus jets. The ellipses corresponds to % 68 confidence level contours considering the statistical uncertainty only, for both data and simulation.

assumption that they conform to the scaling function $C_n = \frac{\sigma_n}{\sigma_{n+1}} = \alpha$, with σ_n is $\sigma(V + n \text{ jets})$ and α is a constant. A second parameter, β, is introduced to measure any possible deviation from the constant scaling law ($C_n = \alpha + \beta n$). The data are in agreement with the predictions from MAD-GRAPH + PYTHIA with Z2 tune, within one or two standard deviation, and thus no discrepancy from the scaling function C_n is found.

2. Azimuthal correlations and event shapes in Z + jets production in pp collisions at $\sqrt{s}$ = 7 TeV

In this section, event shapes and azimuthal correlations for the production of the Z bosons in association with jets are reviewed. The analysis takes advantage from the 5.0 fb^{-1} of the 2011 dataset. Studies are performed as a function of the jet multiplicity for the Z boson inclusive production and for a boosted system in which the Z boson has transverse momentum greater than 150 GeV.[3]

D0 and Tevatron collaborations have already performed previous studies of angular correlations between the Z boson and the two jets with highest p_T in the events. This analysis aims to extend the measurement at higher jet multiplicity, and to compare the models currently available for highly boosted Z bosons p_T distributions. This part of the phase space is very important for searches that are based on large imbalance in the total transverse momentum in the event.

The analysis procedure is structured in the following steps: Z bosons decaying in leptons candidates are selected requiring the selections described previously (lepton $p_T \geq 20$ GeV). Jets candidates are chosen with an increased transverse energy threshold of 50 GeV and within the acceptance $|\eta| \leq$ 2.5. In Figure 7, the distribution of Z$\rightarrow \mu\mu$ candidate events as a function of the jet multiplicity and as a function of the p_T of the muon pair is presented. A nice agreement between data and MC predictions is shown. In addition, an angular separation of $\Delta R(l,J) \geq 0.4$ between the two leading leptons and any other jets is required.

Main backgrounds (foremost $t\bar{t}$ contribution, but also dibosons background (WW, WZ, ZZ) and W+jets) are subtracted and the resulting distributions are unfolded (i.e. correcting for efficiency and detector effects) to the particle level. Then, the two channels are combined together in order to obtain the global Z$\rightarrow$ll channel.

Three observables are introduced in this analysis:

- $\Delta\phi(\mathbf{Z,}J_i)$, azimuthal angle between Z boson and the p_T vector of the ith leading jet in the event

- $\Delta\phi(J_i,J_j)$, azimuthal angles among the three jets of leading p_T (i,j=1,2,3)

- $\tau_T = 1 - max\frac{\sum_i |\vec{p}_{T,i}\cdot\vec{n}_\tau|}{\sum_i \vec{p}_{T,i}}$, transverse thrust. In a limit of a perfectly balanced, pencil-like Z + 1-jet events, this variable tends to zero. Additional jets cause an increase of the variable up to the largest value that is reached in the limit of a spherical events. In this analysis, $\ln\tau_T$ is quoted rather than τ_T, in order to make the results more visible.

2.1 Results

MADGRAPH, SHERPA, POWHEG Z + 1-jet (NLO) and PYTHIA generators are used to compare the normalized differential cross sections of the observables presented in the previous section. Figure 8 shows the $\Delta\phi(Z,J_1)$, as a function of the jet multiplicity, with $N_{jets} \geq$1,2,3. Predictions from MADGRAPH are in good agreement with the experimental data, while an undershoot (overshoot) of about 10% is observed for PYTHIA (SHERPA).

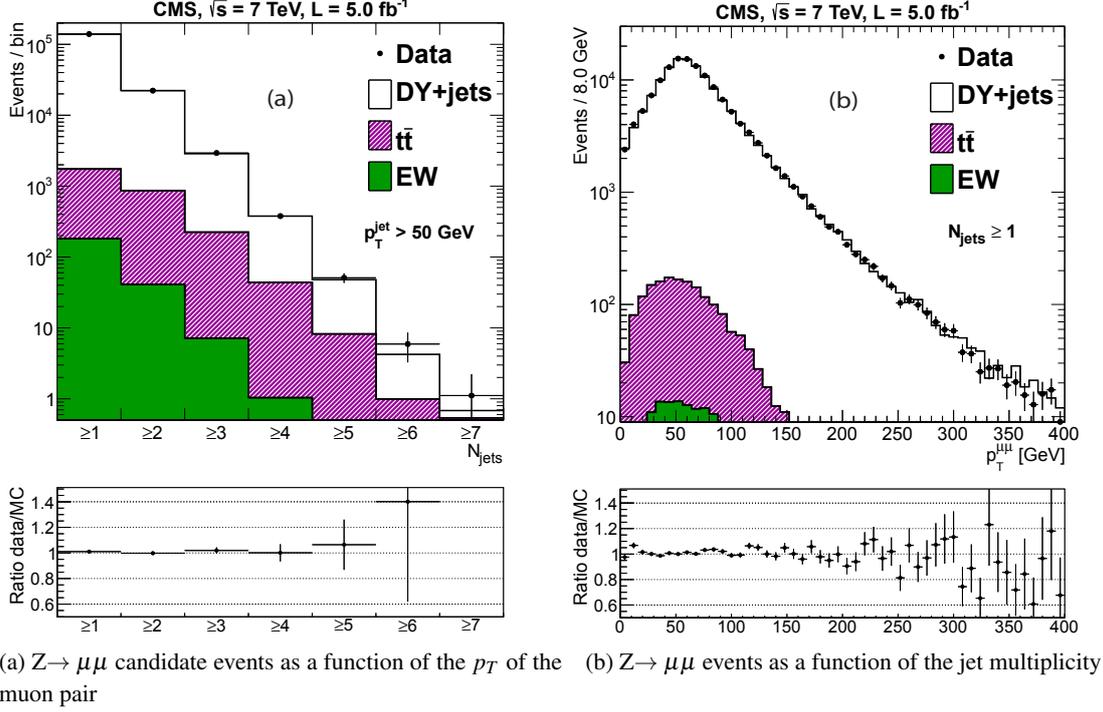Figure 9 show the $\Delta\phi(J_i,J_j)$, where i,j span jet indices in order of decreasing p_T, while Figure 10

(a) Z→ μμ candidate events as a function of the p_T of the muon pair

(b) Z→ μμ events as a function of the jet multiplicity

Figure 7: Distribution of Z→ μμ candidate events as a function of the jet multiplicity (right) and as a function of the p_T of the muon pair (left). The dibosons and W+jets backgrounds are denoted as EWK in the legends. In the ratio plots, error bars represent the statistical uncertainties.

presents the $\ln\tau_T$ distribution. For the boosted regime, an excess at $\ln\tau_T$ = -2 is shown. MAD-GRAPH and POWHEG generators are almost everywhere consistent with the data, except at the large negative values of thrust. SHERPA and PYTHIA again show a discrepancy around 20%, if compared to data distributions.

To summarize, the stand-alone PYTHIA predicts the event topologies in a proper way when the space phase available for parton emission increases. MC models combining multi parton QCD LO matrix element with parton-shower (MADGRAPH, SHERPA) evolution provide an overall good agreement with the experimental data. POWHEG (Z+1 Jet NLO) shows agreement with data at large jet multiplicity, despite the fact that, beyond the leading jet, additional radiation comes exclusively from parton shower.
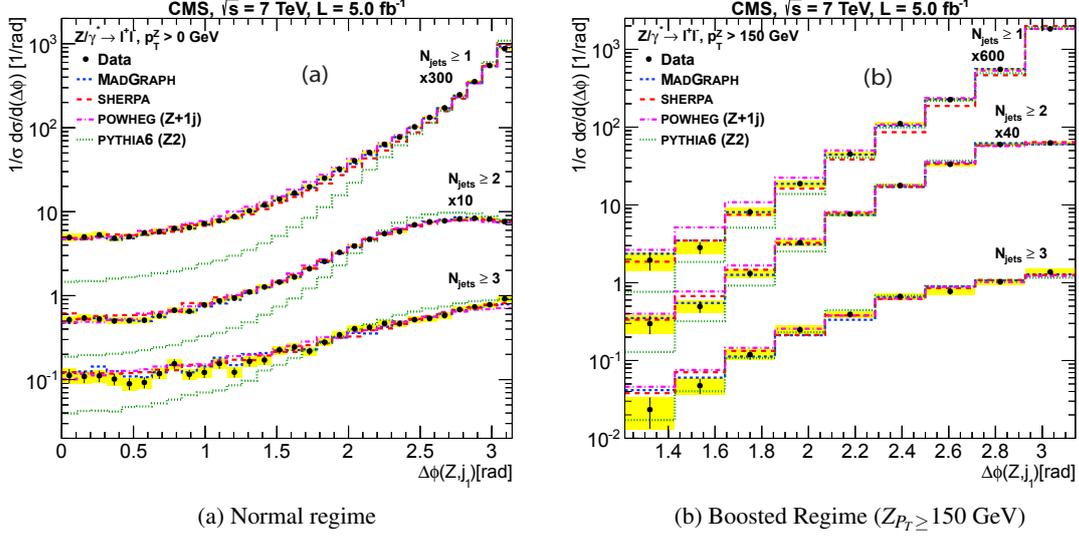
9

(a) Normal regime

(b) Boosted Regime ($Z_{P_T} \geq 150$ GeV)

Figure 8: Normalized $\Delta\phi(Z,J_1)$, as a function of the jet multiplicity, with $N_{jets} \geq 1,2,3$. The error bars on the data points represent their statistical uncertainties, while the solid shaded yellow bands represents the sum of the statistical and systematic uncertainties.

(a) Normal regime

(b) Boosted Regime ($Z_{P_T} \geq 150$ GeV)
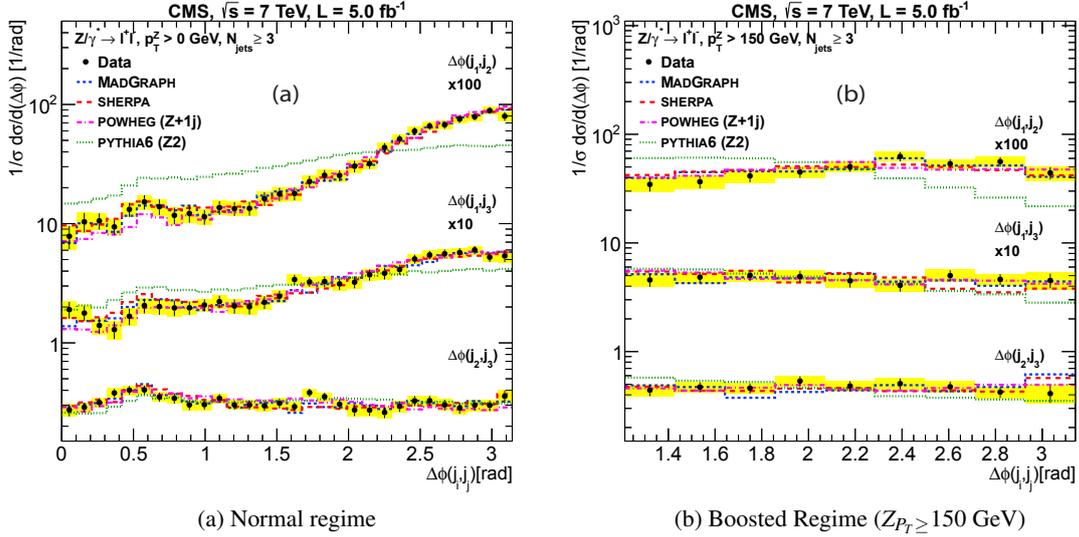
Figure 9: Normalized $\Delta\phi(J_i,J_j)$, as a function of the jet multiplicity, with $N_{jets} \geq 1,2,3$. The error bars on the data points represent their statistical uncertainties, while the solid shaded yellow bands represents the sum of the statistical and systematic uncertainties.

(a) Normal regime   (b) Boosted Regime ($Z_{P_T} \geq 150$ GeV)

Figure 10: Normalized $\ln \tau_T$, as a function of the jet multiplicity, with $N_{jets} \geq 1,2,3$. The error bars on the data points represent their statistical uncertainties, while the solid shaded yellow bands represents the sum of the statistical and systematic uncertainties.
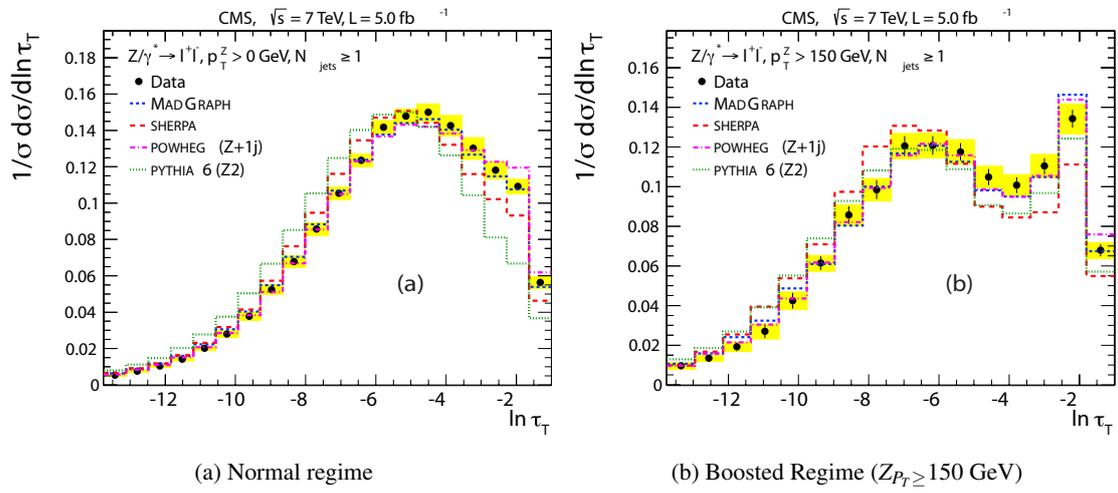
3. Measurement of the Z/γ* + b-jet cross section in pp collisions at $\sqrt{s}$ = 7 TeV

The production of at least one b jet in association with a Z boson is reviewed in this section. The analysis has been carried on profiting of an integrated luminosity of 2.2 fb^{-1} (first part of the total 2011 dataset). [4]

This measurement is very relevant, being this process a standard model background for searches of new physics states, in particular for the Higgs Boson production in association with b quarks.

Event candidates are selected from Z bosons decaying in leptons, requiring 25 (20) GeV transverse energy threshold and $|\eta| \leq 2.5$ (2.1) in order to select electrons (muons).

Main backgrounds arise from the production of Z/γ^* with jets of other flavors (misidentified) and from $t\bar{t}$ events. Negligible contributions come from dibosons, QCD multijets, single top and W + jets.

Jets are selected when their $p_T \geq 25$ GeV, requiring an additional separation from each of the other leading leptons by at least $\Delta R = 0.5$. A further cut on the jet geometric acceptance ($|\eta| \leq 2.1$) is applied.

3.1 B Tagging

Jets originating from b quarks are tagged by taking advantage of the long b-hadron lifetime, using the "Simple Secondary Vertex" (SSV) algorithm discriminant. Values of the discriminant greater than one indicate the presence of a secondary vertex. To improve the purity, only secondary vertices built from at least three tracks are considered (high purity, HP). The distribution of the SSV High purity vertices and the number of three-track secondary vertices are shown in Figure 11. The discriminant value to define b jets is selected to be 2.0. In this way the rate of tagging a light jets (mis-tagging rate) is kept below 0.1%. The MC b tagging efficiency is found to be higher than data, as visible in Figure 11. A weight is applied to the MC events to reproduce the data b-tagging efficiency.

The p_T spectrum for the b-tagged jet with the largest momentum (leading jet) is shown in Figure 12, after the dilepton + b-jet selection, together with the distribution of $\Delta\phi$(Z, b jet) between the leading b jet and the Z boson.

3.2 Results

The cross section for the production of a Z boson in association with at least one b jet can be derived in the following way:

$$\sigma_{hadron}(Z/\gamma^* + b, Z/\gamma^* \to ll) = \frac{N(ll+b) \cdot (P - f_{t\bar{t}})}{A_l \cdot C_{hadron} \cdot \varepsilon_l \cdot \varepsilon_b \cdot L} \tag{3.1}$$

where P is the purity of the b-jet, $f_{t\bar{t}}$ the fraction of genuine $t\bar{t}$ events, ε_b is the b-tagging efficiency (evaluated using the MC Z+b signal, corrected to match b-tagging efficiency), ε_l the lepton selection efficiency, A_l the detector acceptance (evaluated using MCFM, SHERPA and aMC@NLO + HERWIG) and finally C_{hadron} a correction factor for reconstruction and detector effects (obtained from comparison between detector level and generator level event yields, using MADGRAPH +

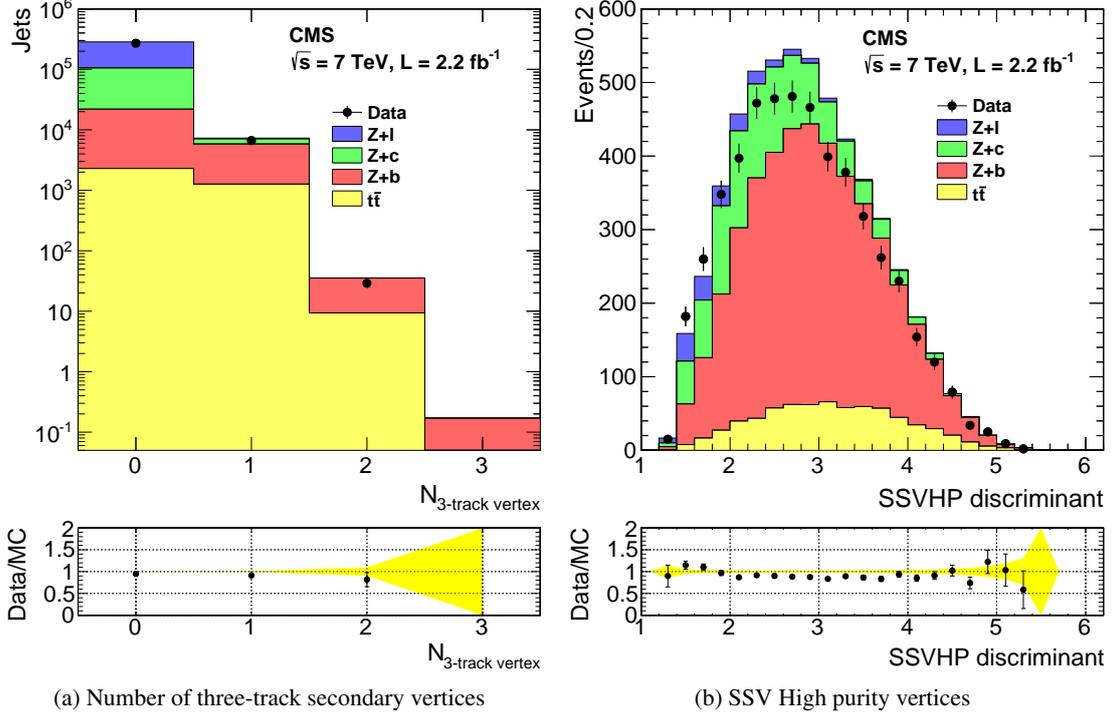(a) Number of three-track secondary vertices

(b) SSV High purity vertices

Figure 11: Distribution of the SSV High purity vertices (right) and the number of three-track secondary vertices (left). Yellow bands in the ratio plot represents the MC statistical uncertainties.

Variable	ee+b	$\mu\mu + b$
Selected Events	1990	3362
P(%)	83.4 $\pm$3.6	81.5 $\pm$2.9
$f_{t\bar{t}}$	18.7 $\pm$2.2	18.4 $\pm$2.3
ε_b	35.3 $\pm$3.5	34.9 $\pm$3.5
ε_l	63.2 $\pm$2.6	84.4 $\pm$1.7
C_{hadron}	$84.2^{5.8}_{-0.6}$	$95.0^{6.6}_{-0.5}$
A_l	$55.0^{3.6}_{-2.1}$	$57.2^{3.7}_{-2.4}$
$\sigma_{hadron}(Z/\gamma^* + b, Z/\gamma^* \to ll)$	$5.61 \pm 0.13 \pm 0.73^{0.24}_{-0.53}$	$5.97 \pm 0.10 \pm 0.73^{0.25}_{-0.57}$

Table 1: Extraction of the cross section $\sigma_{hadron}(Z/\gamma^* + b, Z/\gamma^* \to ll)$. Estimation of the parameters defined in (3.1) are also reported

PYTHIA). Purity P is extracted from a fit to the distribution of secondary-vertex mass of the leading b jet in the event. An estimation of the parameters defined above, together with the calculation of the final cross sections can be seen in Table 1.

Combining the two lepton channels, the final cross section is measured as:

$$\sigma_{hadron}(Z/\gamma^* + b, Z/\gamma^* \to ll) = 5.84 \pm 0.08(stat) \pm 0.72(syst)^{0.25}_{-0.55} \text{ pb}. \qquad (3.2)$$

The results are then compared to the NLO calculation obtained with MCFM tool, using the same

(a) P_T spectrum for the b-tagged jet

(b) $\Delta\phi(Z, b\ jet)$ between the leading b jet and the Z boson
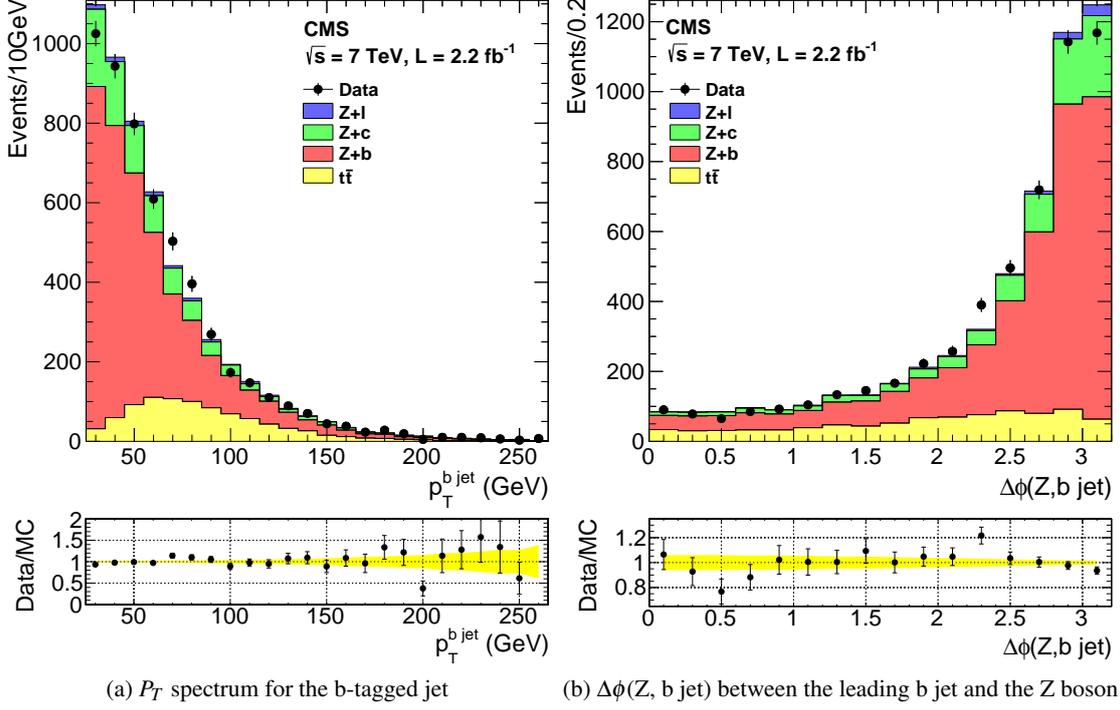
Figure 12: P_T spectrum for the b-tagged jet with the largest momentum (leading jet), together with the distribution of $\Delta\phi(Z, b\ jet)$ between the leading b jet and the Z boson. Yellow bands in the ratio plot represents the MC statistical uncertainties.

acceptances for leptons and parton jets:

$$\sigma_{parton}^{MCFM} = 4.73 \pm 0.54 \text{ pb}. \tag{3.3}$$

To conclude, a general fair agreement (see Figure 12) is found between data distributions and MADGRAPH. The residual discrepancy may be a consequence of the higher order terms absent in the MADGRAPH tree-level simulation.

References

[1] S. Chatrchyan *et al.* [CMS Collaboration], "The CMS experiment at the CERN LHC", JINST 3:S08004, 2008.

[2] S. Chatrchyan *et al.* [CMS Collaboration], "Jet Production Rates in Association with *W* and *Z* Bosons in *pp* Collisions at $\sqrt{s} = 7$ TeV," JHEP 01 (2012) 010 [arXiv:1110.3226 [hep-ex]].

[3] S. Chatrchyan *et al.* [CMS Collaboration], "Event shapes and azimuthal correlations in *Z* + jets events in *pp* collisions at $\sqrt{s} = 7$ TeV," ISSN 0370-2693, 10.1016/j.physletb.2013.04.025.

[4] S. Chatrchyan *et al.* [CMS Collaboration], "Measurement of the Z/gamma*+b-jet cross section in pp collisions at 7 TeV," JHEP 06 (2012) 126 [arXiv:1204.1643 [hep-ex]].