Associated production of W and Z bosons with jets from light and heavy quarks at CMS experiment

Roberto Castello on behalf of CMS collaboration
Centre for Cosmology, Particle Physics and Phenomenology
Université Catholique de Louvain
B-1348 Louvain la Neuve, BELGIUM

1 Motivations

The measurement of the vector boson production in association with jets at the CMS experiment is of fundamental importance for testing the Standard Model prediction and providing better knowledge of the leading background in the related physics searches. For example the measurement of the production rate of vector boson V in association with $n-$jets will shed light on the current perturbative QCD calculation: NLO predictions are available up to n=4 for V=Z or W only with a 10-30 % precision [1] [2]. At the same time, the measurement of the production of a Z boson in association with jets originating from b-quarks is of fundamental importance both as a benchmark channel to the production of the Higgs boson in association with b quarks, as well as a Standard Model background to Higgs and new physics searches in final states with leptons and b-jets.

2 Vector boson plus jets production rate

This measurement exploits the data sample recorded by the CMS experiment [3] during 2010, amounting to an integrated luminosity of 36 ± 4 pb^{-1}. Events containing leptonic decays of Z and W bosons are selected according to the following categories: lepton pairs (tight and loose requirements) in the invariant mass range $60 < M_{ll} < 120$ GeV are classified as Z-sample. Those falling out, satisfying an additional requirement on the $M_T > 20$ GeV, are classified as W-sample. Tracks and deposits in a jet are clustered with the PF anti-KT algorithm, required to be inside the Tracker acceptance and being well separated by electrons ($\Delta R > 0.3$).
The signal yield is estimated using an extended likelihood fit to M_{ll} for the Z + jets sample and to M_T for the W +jets sample, as shown in Figure 1. The overall normalization of the true distribution is allowed to float within a Poissonian constraint on the number of observed events. For the Z event samples, the contamination from the

main background processes, dominated by $t\bar{t}$ and W + jets, is small and does not produce a peak in the invariant mass distribution: M_{ll} is fitted with two components, one for the signal and one that accounts for all background processes. For the W sample, background contributions can be divided into two components, one which exhibits a peaking structure in M_T, dominated by $t\bar{t}$ and another which does not, dominated by QCD multi-jet events. Therefore a two-dimensional fit to the M_T distribution and to the b-jets multiplicity in the event is performed. In order to estimate the scaling rule

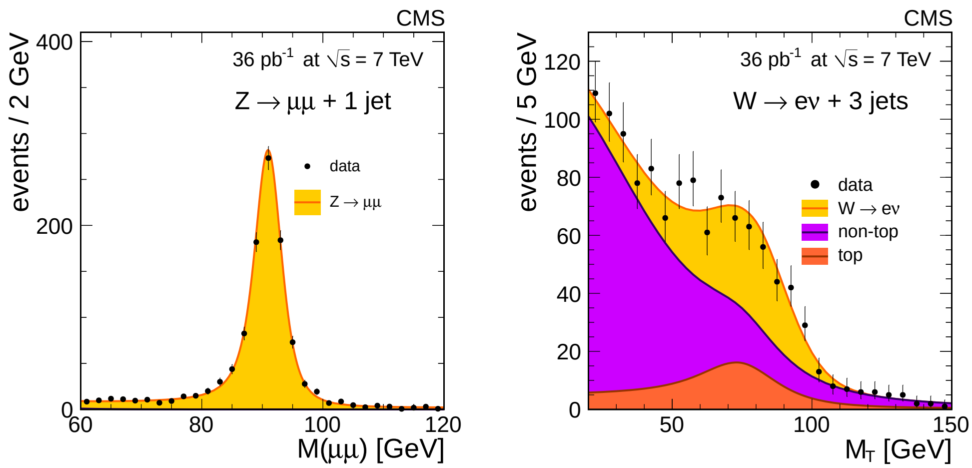

Figure 1: Left: di-lepton mass fit for the Z +1 jet sample, in the muon channel. Right: fit results for the $W(e\nu)$ + n-jet sample with n=3 in the M_T projection

of the jets at the *particle* level, an unfolding procedure that removes the effects of imperfect jet energy resolution and reconstruction efficiency is applied. Furthermore, in order to provide model-independent estimate, the results are not corrected for the acceptance, but rather quoted within acceptance, as defined by the lepton and jet fiducial and kinematic cuts. The efficiencies for lepton reconstruction, identification, isolation and trigger are obtained from data by means of a tag-and-probe method performed on Z/γ +jets data samples, evaluated in different jet multiplicity bins.
In order to reduce the systematic uncertainties associated with the integrated luminosity measurement, the jet energy scale and the lepton reconstruction and trigger efficiencies, the yields are normalized according to the inclusive W and Z cross sections. The cross section ratios is also measured $\sigma(V+n-jets)/\sigma(V+(n-1)-jets)$ where n stands for the inclusive number of jets in the event. An example of the results is shown in left plot of Figure 2 together with the most relevant uncertainties for each measured value. Within the uncertainties the measured ratios are found to be in agreement with `MADGRAPH` theoretical predictions. From the former normalization, the staircase scaling is tested as well and is found to be in reasonable agreement

with the theoretical expectations with deviations that are within one or two standard deviations depending on the channel [4].
With a similar lepton selection and ad-hoc requirements on jet topology [5], the di-jet invariant mass has been investigated in the W+jets channel and compared to the observation from the CDF experiment [6], exploiting the full 2011 statistics. Using an unbinned likelihood technique to fit the background in the sidebands and a parametrized fit shape for the W+jets signal component, no enhancement has been observed in the mass region $120 < M_{jj} < 160$ GeV, as visible in Figure 2 (right).

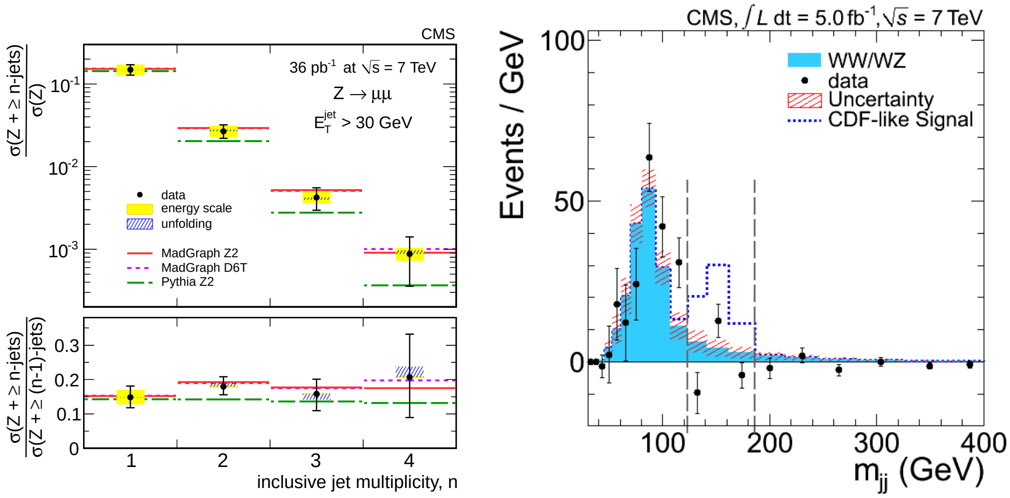

Figure 2: Left: Ratio $\sigma(Z + n - jets)/\sigma(Z)$ in the muon channel compared to expectations from MADGRAPH and PYTHIA. Right: Distribution of the invariant mass spectrum of the leading two jets after subtraction of all SM components except the electroweak diboson WW/WZ. Error bars correspond to the statistical uncertainty. The band represents the systematic uncertainty in the sum of the SM components.

3 Z boson production in association with b-jets

Calculations of the theoretical cross section for this process, driven by perturbative QCD, are currently derived in two schemes: fixed-flavour [7] and variable flavour [8]. The main experimental backgrounds arise from the production of Z with jets of other flavours misidentified as b jets and from $t\bar{t}$ + jets events: is therefore essential to reduce them as much as possible and finally quantify the remaining contribution in a precise and reliable way. The production of b jets in association with a Z boson has been studied in 2.1 fb^{-1} of proton-proton collision data at a centre-of-mass energy of 7 TeV and recorded by the CMS detector. Jets and leptons are reconstructed according to standard criteria described in [9]. Opposite charges for the leptons are required when forming pairs, and the lepton invariant mass M_{ll} is required to lie between 60 and 120 GeV. Separation between leptons and jets is also applied ($\Delta R > 0.5$). Jets originating from b quarks are tagged by taking advantage of the long b-hadron lifetime. The Simple Secondary Vertex (SSV) algorithm, requiring secondary vertices built from at least three tracks in order to improve the purity of the selection, is exploited. The cross section for the production of a Z boson in association with at least one hadron level b jet is then extracted from the selected numbers of dilepton+b-jet events ($N_{\ell\ell+b}$), taking into account the b-jet purity $\mathcal{P}$, the fraction fitted of $t\bar{t}$ events ($f_{t\bar{t}}$), the b-tagging efficiency ϵ_b, the lepton efficiency ϵ_ℓ, the correction factor C_{hadron} for detector and reconstruction effects, and the lepton acceptance A_l, using the following equation:

$$\sigma_{hadron}(Z+b, Z \to \ell\ell) = \frac{N_{\ell\ell+b} \times (\mathcal{P} - f_{t\bar{t}})}{\mathcal{A}_\ell \times \mathcal{C}_{hadron} \times \epsilon_\ell \times \epsilon_b \times \mathcal{L}}$$

The extraction of the purity $\mathcal{P}$ is based on a (data-driven) template fit of the mass

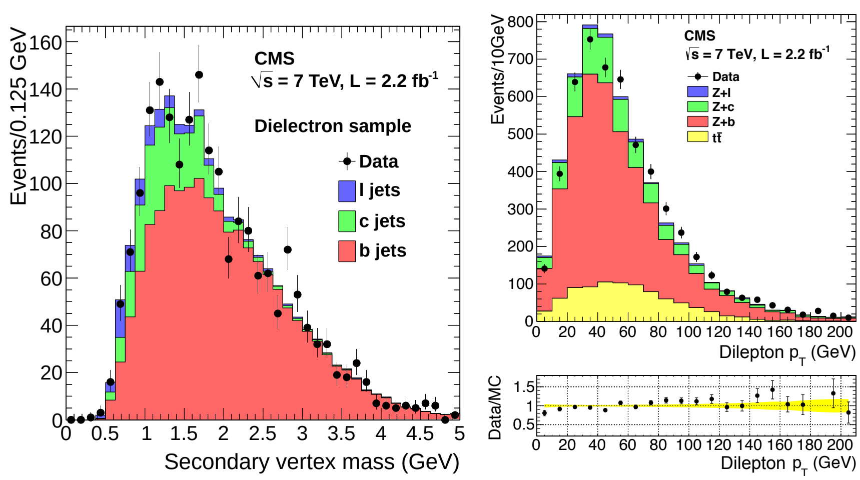

Figure 3: Left: Secondary-vertex mass of the leading-p_T b jet, with MC distributions constructed from inclusive jet samples, for the dielectron channel, after the b-purity maximum-likelihood fit. Right: p_T of the lepton pairs after the di-lepton+b-jet selection. The yellow bands in the lower plots represent the statistical uncertainty on the MC yield.

of the secondary vertex of the leading-p_T b-jet, as shown in Figure 3, resulting in a fraction of $83.4 \pm 3.6\%$ ($81.5 \pm 2.9\%$) for the di-electron (di-muon) channel. The $t\bar{t}$ contribution is extracted from extrapolation of upper sideband of M_{ll} under the signal region [60-120] GeV,and it is found to be of the order of $18.7 \pm 2.2\%$ ($18.4 \pm 2.3\%$).

The MADGRAPH simulation interfaced with PYTHIA is used to derive the correction from the reconstructed level to the hadron level and to evaluate the other efficiencies.
The Z+b-jets cross section, for events with Z$\to \ell\ell$ where $\ell\ell = ee$ or $\mu\mu$, lepton pair invariant mass $60 < M_{\ell\ell} < 120\,\mathrm{GeV}$, and at least one b jet at the hadron level with $p_T > 25\,\mathrm{GeV}$ and $|\eta| < 2.1$, with a separation between leptons and jets of $\Delta R > 0.5$, is found to be $5.84 \pm 0.08(stat.) \pm 0.72(syst.)^{+0.25}_{-0.55}(theory)\,\mathrm{pb}$. The shapes of the kinematic variables are found to be in fair agreement with the predictions made by the MADGRAPH event generator, and normalised to the integrated luminosity in data using the cross-section value which includes the NNLO corrections to the inclusive Z production. The residual discrepancy, for example noticeable in the dilepton $\mathrm{p_T}$ spectrum of Figure 3 (right), may be a consequence of the higher order terms absent in the MADGRAPH tree-level simulation in the variable-flavour scheme with massless b quarks and it has been noticed also in subsequent measurements [10].

References

[1] C. F. Berger *et al.* , arXiv:1009.2338 [hep-ph]

[2] H. Ita *et al.* , arXiv:1108.2229 [hep-ph]

[3] S. Chatrchyan *et al.* [CMS Collaboration], JINST **3**, S08004 (2008).

[4] [CMS Collaboration], J. High Energy Phys. 01 (2012) 010

[5] S. Chatrchyan *et al.* [CMS Collaboration], arXiv:1208.3477 [hep-ex].

[6] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **106**, 171801 (2011) [arXiv:1104.0699 [hep-ex]].

[7] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau and P. Torrielli, JHEP **1109**, 061 (2011) [arXiv:1106.6019 [hep-ph]].

[8] J. M. Campbell, R. K. Ellis, F. Maltoni and S. Willenbrock, Phys. Rev. D **69** (2004) 074021 [hep-ph/0312024].

[9] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1206**, 126 (2012) [arXiv:1204.1643 [hep-ex]].

[10] [CMS Collaboration], CMS-PAS-SMP-12-003