Study of $Z$ production in association with jets in proton-proton collisions at $\sqrt{s} = 10$ TeV with the CMS detector at the CERN LHC

The CMS Collaboration

Abstract

We present a simulation study of $Z$ boson production in association with jets in $pp$ collisions at $\sqrt{s} = 10$ TeV and demonstrate the feasibility of significant measurements up to four inclusive jets with $\mathcal{O}(100)$ pb$^{-1}$ of early CMS data at the LHC. QCD predicts a constant ratio of $Z+n$ jets over $Z+(n+1)$ jets yields while several new physics models are expected to produce an excess of events at high jet multiplicity. We present the measurement of this ratio in the dielectron+jets and dimuon+jets final states using tracker-based, calorimetry-based, and Particle Flow jet definitions. We discuss the $Z$+jets sample as a ‘candle’ for both physics and detector commissioning.
1 Introduction

Important standard model (SM) and new physics (NP) processes at the LHC are expected to produce final states with a vector boson (VB=W, Z) and multiple jets. The VB+jets associated production has been used at the Tevatron both as a stringent test of perturbative QCD predictions and as a handle on the accurate description of backgrounds to NP [1–3].

We present a data-driven strategy to study Z+jets production in final states with dielectrons and dimuons. We focus on the LHC startup and assume O(100) pb\(^{-1}\) of data collected with the CMS detector [4] at a center-of-mass energy \(\sqrt{s}=10\) TeV. We use two independent jet definitions: one based on calorimetry deposits (calo-jets) and one based on tracks (track-jets) to test the jet counting with different detector effects and to allow sampling of different parts of the phase space. We further validate the results with corrected calo-jets and Particle Flow (PF) jets (PF-jets) [5].

Within the SM the Z+n jets cross section is \(O(\alpha^n_s)\). The Z+n jets over Z + (n+1) jets yield ratio is then nearly constant as a function of n for \(p\bar{p} \sqrt{s} = 630\) GeV and \(p\bar{p} \sqrt{s} = 1.8\) TeV both at the parton level and in data ([6–9], [1–3] and references therein; this is also referred to as ‘Berends-Giele’ scaling). The purpose of this analysis is: i) to measure the ratio at different jet multiplicities at \(p\bar{p} \sqrt{s} = 10\) TeV and investigate to what extent the ratio is in fact independent of jet multiplicity, ii) to develop an analysis strategy that increases the available statistics of the signal using track-jet counting (and, by extension, PF-jet counting), iii) to select a pure Z+jets sample that can be used for detector and physics commissioning at the LHC startup and iv) to investigate the Z+n jets over Z + (n+1) jets ratio as a probe of new physics processes with multijets and real Z bosons in the final state (e.g. [10]). In the absence of any excess the comparison of the measured and computed power-law coefficients for the ratio will provide a benchmark for validation and tuning of QCD phenomenological models and provide a reference for comparisons of the data with higher order calculations [6–9] when these become available.

2 The CMS detector

A detailed description of the Compact Muon Solenoid (CMS) experiment can be found elsewhere [11]. The central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter. Within the field volume are the silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass-scintillator hadronic calorimeter (HCAL). Muons are measured in gas chambers embedded in the iron return yoke. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry.

CMS has a two-level trigger system. The Level-1 trigger, based on custom hardware, is designed to reduce the collision rate of 40 MHz to approximately 100 kHz. The High Level Trigger (HLT) employs a set of sophisticated software algorithms that analyze the complete event information and further reduce the accepted event rate for permanent storage and analysis.

3 Signal and background samples

The Z(\(\rightarrow\ell\ell\)) + n-jets events (\(\ell = e, \mu\)) in pp collisions at \(\sqrt{s} = 10\) are produced with the MadGraph [12] Monte Carlo (MC) event generator and processed through the full detector simulation and reconstruction chain of the CMS experiment. The MadGraph event generation is based on a leading-order calculation of the matrix element (ME) for final states with at most four primary partons with transverse momentum (\(p_T\)) larger than 10 GeV/c. PYTHIA [13] is used for
the parton shower, hadronization and the underlying event description. Parton shower matching is applied to avoid double counting of emissions in overlapping phase space regions. The MLM [14] matching algorithm with $k_T$ clustering is used with matching threshold 15 GeV/$c$. The lepton pair invariant mass is required to be $m(\ell\ell) > 50$ GeV/$c^2$ at the generator level. The CTEQ6L1 [15] parton distribution functions are used.

The largest background component for this analysis comes from multi-jet production. This is studied using a sample of Monte Carlo events generated with PYTHIA, fully simulated and reconstructed. Using a filter that selects electron and muon enriched QCD samples, the generation includes $b\bar{b}$, $c\bar{c}$, decays of long-lived light mesons as sources of muons, and loosely isolated hadrons or jets with an increased electromagnetic fraction as a source of electrons. The filter also requires an outgoing parton with $p_T > 20$ GeV/$c$. The $Z(\rightarrow \tau^+\tau^-)+$ jets events contribute to the background and are generated as part of the full $Z(\rightarrow \ell\ell)+$ jets samples. The $W(\rightarrow \ell\nu)+ n$-jets (with $\ell = e, \mu, \tau$) background processes are generated with MadGraph and PYTHIA and the same phase space requirements and parton shower matching settings as the signal. The $t\bar{t}$+jets are generated with MadGraph interfaced with PYTHIA with the associated $p_T > 20$ GeV/$c$ and matching threshold 30 GeV/$c$. Other potential backgrounds such as single top and diboson production are not considered since they are found to be negligible (cf also [16]).

4 Event reconstruction and selection

4.1 Trigger selection

The events are selected by the CMS Level-1 (L1, as emulated in the simulation) and High-Level (HLT) single electron and muon triggers with no requirement on the lepton isolation. The trigger $p_T$ thresholds are those determined in CMS for low luminosity running ($L = 10^{32}$ cm$^{-2}$ s$^{-1}$). The trigger paths used for both the W and Z selection are the HLT single ‘non-isolated’ muon and electron with thresholds 15 GeV/$c$ and L1 thresholds 12 GeV/$c$ and 10 GeV/$c$ for electrons and muons respectively.

4.2 Lepton reconstruction and selection

Muons are reconstructed using the algorithm combining the information from muon chambers and the silicon tracker [17], and very loose muon isolation is imposed by considering a cone around the muon defined as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \leq R_{cone} = 0.5$ and requiring that the sum of the $p_T$ of the tracks in this cone, excluding the muon track, is less than 30% of the muon transverse momentum. Electrons are reconstructed as single tracks matched to electromagnetic energy deposits in the ECAL. Electron identification is based on a standard set of criteria including various track-matching and shower shape variables in the electromagnetic calorimeter barrel and end-cap regions. In addition, a loose electron isolation criterion based on tracking requires that the scalar sum of transverse momenta of tracks with $p_T > 1.5$ GeV/$c$, consistent with coming from a vertex common to the electron, and within a cone of size $R_{cone} = 0.4$ around the electron direction, be smaller than 15% of the electron candidate transverse momentum.

The lepton selection is driven by the general requirements of i) retaining high efficiency for the $Z$+jets signal, ii) avoiding trigger and other threshold effects and iii) establishing a robust procedure to extract the signal. The $p_T$ requirement of the leading lepton is 20 GeV/$c$. The muons are selected with $|\eta| < 2.1$ and the electrons with $|\eta| < 2.5$. 
4.3 Z boson reconstruction and selection

The event reconstruction and selection is based on forming the Z boson candidates using all combinations of muon or electron pairs in the event. The candidates are selected with $60 < m_{\ell\ell} < 110 \text{ GeV}/c^2$. After the Z selection is applied, the fraction of events with multiple Z candidates is found to be very small. In the presence of multiple Z candidates, the combination with the highest $p_T$ leptons is found almost always to match the true candidate. The reconstructed vertex closest to the best Z candidate is taken as the primary vertex of the event and it is found to be the highest $\sum p_T$ vertex. The primary vertex is used to project the calorimeter deposits and to select the tracks when jets are formed.

4.4 Jet reconstruction and selection

The event selection is based on the leptons of the Z-boson and the counting of the associated jets. The expectation (validated by the results presented here) is that any jet definition can be used to construct the $Z+n$ jets over $Z+(n+1)$ jets ratio without altering the analysis strategy; the exception would be jets that are so inclusive that the first few jet clusterings use up all the available phase space, as discussed in [18].

We consider two scenarios based on the expected understanding of detector effects on the jet clustering and jet counting: i) At LHC start-up we consider the calorimetric response as not yet fully understood. In this scenario we use ‘raw’ calo-jets and track-jets [19] reconstructed from calorimeter deposits and tracks respectively using the Seedless Infrared Safe Cone (SISCone) jet algorithm [20] with a cone size $R_{\text{cone}} = 0.5$ in the $(\eta \times \phi)$ space. The two sets of jets allow for probing different parts of the phase space and are independent in terms of detector effects. ii) The second scenario assumes enough understanding of the detector to allow fully corrected calo-jets and PF-jets; this scenario would allow a quantitative direct comparison with parton-level QCD predictions (as they become available).

Events are selected with one or more calo-jets (track-jets) within $|\eta|<3.0$ ($|\eta|<2.4$) and $p_T>30 \text{ GeV}/c$ ($p_T>15 \text{ GeV}/c$). Track-jets are reconstructed from tracks with $|\eta|<2.4$ consistent with the event primary vertex. PF-jets are clustered within $|\eta|<3$ with best performance within $|\eta|<2.4$. The detailed description of Particle Flow jet reconstruction at CMS can be found elsewhere [5]. The leptons from the best Z candidate in the event are not considered as jets.

In the $Z+n$ jets over $Z+(n+1)$ jets ratio, systematic errors due to the mapping from partons to jets, the parton distribution functions, and other corrections substantially cancel [21]. Given the CMS high precision silicon tracker that offers a very good momentum resolution, track-jets and PF-jets can probe a part of the phase space where the calorimeter response is low and provide higher statistics $Z$+jets samples despite their limited $\eta$ acceptance compared to that of calo-jets. In all cases the jet counting is inclusive ($\geq n$) resulting in a statistical correlation between the successive bins.

4.5 Maximum likelihood fit

To determine the number of $Z$+jets events for each jet multiplicity bin ($n \geq 1$) we perform a one dimensional unbinned extended maximum likelihood (ML) fit based on the dilepton invariant mass ($m(\mu\mu)$ or $m(ee)$)

$$L = \frac{e^{-(N_S+N_B)}}{(N_S+N_B)} \prod_i \{N_S \cdot P_S(m(\ell\ell)_i) + N_B \cdot P_B(m(\ell\ell)_i) \}$$  \hspace{1cm} (1)

where $N_S$ ($N_B$) is the number of signal (background) events in the selected samples and $P_S(m(\ell\ell)_i)$ ($P_B(m(\ell\ell)_i)$) is the signal (background) probability density function (PDF) for the variable
Tables 1 and 2 show the signal event yields in the $Z(\rightarrow \mu\mu) + \text{jets}$ and $Z(\rightarrow ee) + \text{jets}$ selection for calo-jet and track-jet counting for 100 pb$^{-1}$ of integrated luminosity. The quoted errors are statistical only, related to the size of the available datasets. Tables 4 and 3 show the signal event yields in the $Z\rightarrow \mu\mu$ and $Z\rightarrow ee$ channels despite the muon acceptance being lower than the electron one, thanks to the possibility of lowering the jet $p_T$ threshold.

The $m(\ell\ell)$ and the event $i$.

### Table 1: Expected signal and backgrounds yields in the $Z(\rightarrow \mu\mu) + \text{jets}$ at $\sqrt{s} = 10$ TeV with $100 \text{ pb}^{-1}$ as a function of the number of jets. Shown are the calo-jet counting yields and in parentheses the track-jet counting ones.

<table>
<thead>
<tr>
<th>$\geq$ 1 jets</th>
<th>$\geq$ 2 jets</th>
<th>$\geq$ 3 jets</th>
<th>$\geq$ 4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z+jets</td>
<td>Z other</td>
<td>W+jets</td>
<td>t$\bar{t}$+jets</td>
</tr>
<tr>
<td>4007 ± 37 (9305 ± 56)</td>
<td>11 ± 2 (14 ± 2)</td>
<td>44 ± 4 (34 ± 3)</td>
<td>109 ± 2 (119 ± 2)</td>
</tr>
<tr>
<td>555 ± 14 (1741 ± 24)</td>
<td>2 ± 1 (2 ± 1)</td>
<td>17 ± 2 (18 ± 2)</td>
<td>66 ± 2 (90 ± 2)</td>
</tr>
<tr>
<td>72 ± 5 (338 ± 11)</td>
<td>1 ± 1 (1 ± 1)</td>
<td>7 ± 2 (16 ± 2)</td>
<td>25 ± 1 (43 ± 1)</td>
</tr>
<tr>
<td>11 ± 2 (66 ± 5)</td>
<td>(−)</td>
<td>4 ± 1 (8 ± 2)</td>
<td>17 ± 10 (35 ± 15)</td>
</tr>
</tbody>
</table>

### Table 2: Expected signal and backgrounds yields in the $Z(\rightarrow ee) + \text{jets}$ at $\sqrt{s} = 10$ TeV with $100 \text{ pb}^{-1}$ as a function of the number of jets. Shown are the calo-jet counting yields and in parentheses the track-jet counting ones.

<table>
<thead>
<tr>
<th>$\geq$ 1 jets</th>
<th>$\geq$ 2 jets</th>
<th>$\geq$ 3 jets</th>
<th>$\geq$ 4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z+jets</td>
<td>Z other</td>
<td>W+jets</td>
<td>t$\bar{t}$+jets</td>
</tr>
<tr>
<td>3135 ± 32 (7106 ± 49)</td>
<td>9 ± 2 (18 ± 2)</td>
<td>15 ± 2 (26 ± 3)</td>
<td>70 ± 1 (78 ± 1)</td>
</tr>
<tr>
<td>411 ± 11 (1334 ± 21)</td>
<td>1 ± 1 (3 ± 1)</td>
<td>4 ± 1 (9 ± 2)</td>
<td>34 ± 1 (35 ± 1)</td>
</tr>
<tr>
<td>58 ± 4 (268 ± 9)</td>
<td>(−)</td>
<td>1 ± 1 (3 ± 1)</td>
<td>9 ± 1 (23 ± 1)</td>
</tr>
<tr>
<td>7 ± 2 (52 ± 4)</td>
<td>(−)</td>
<td>(−)</td>
<td>(2 ± 1)</td>
</tr>
</tbody>
</table>

### Table 3: Expected signal event yields in the $Z(\rightarrow \mu\mu) + \text{jets}$ at $\sqrt{s} = 10$ TeV with $100 \text{ pb}^{-1}$ as a function of PF-(15 GeV/c) and corrected calo-(58 GeV/c) jet multiplicity.

<table>
<thead>
<tr>
<th>PF-jets (15 GeV/c)</th>
<th>track-jets (15 GeV/c)</th>
<th>calo-jets (30 GeV/c)</th>
<th>calo-jets$^{\text{corr}}$ (58 GeV/c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq$ 1 jets</td>
<td>24409 ± 90</td>
<td>9305 ± 56</td>
<td>4007 ± 37</td>
</tr>
<tr>
<td>$\geq$ 2 jets</td>
<td>8725 ± 54</td>
<td>1741 ± 24</td>
<td>555 ± 14</td>
</tr>
<tr>
<td>$\geq$ 3 jets</td>
<td>2889 ± 31</td>
<td>338 ± 11</td>
<td>72 ± 5</td>
</tr>
<tr>
<td>$\geq$ 4 jets</td>
<td>885 ± 17</td>
<td>66 ± 5</td>
<td>11 ± 2</td>
</tr>
<tr>
<td>$\geq$ 5 jets</td>
<td>243 ± 9</td>
<td>11 ± 8</td>
<td></td>
</tr>
</tbody>
</table>

### 4.5.1 Signal and background parameterization

The $m(\ell\ell)$ signal distributions are parameterized by a Gaussian-like function with asymmetric widths and non-Gaussian tails:

$$f(x; m, \sigma_L, \sigma_R, a_L, a_R) = N_s \cdot e^{-\frac{(x-m)^2}{2\sigma^2} + a_L(x-m) + a_R(x-m)^2}$$

(2)
4.5 Maximum likelihood fit

<table>
<thead>
<tr>
<th></th>
<th>PF-jets (15 GeV/c)</th>
<th>track-jets (15 GeV/c)</th>
<th>calo-jets (30 GeV/c)</th>
<th>calo-jets$^{corr}$ (58 GeV/c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 1 jets</td>
<td>16750 ± 23</td>
<td>7106 ± 49</td>
<td>3135 ± 32</td>
<td>3098 ± 32</td>
</tr>
<tr>
<td>≥ 2 jets</td>
<td>5865 ± 44</td>
<td>1334 ± 21</td>
<td>411 ± 11</td>
<td>411 ± 12</td>
</tr>
<tr>
<td>≥ 3 jets</td>
<td>1928 ± 25</td>
<td>268 ± 9</td>
<td>58 ± 4</td>
<td>57 ± 4</td>
</tr>
<tr>
<td>≥ 4 jets</td>
<td>581 ± 14</td>
<td>52 ± 4</td>
<td>7 ± 2</td>
<td>6 ± 1</td>
</tr>
<tr>
<td>≥ 5 jets</td>
<td>151 ± 71</td>
<td>7 ± 6</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 4: Expected signal event yields in the $Z(\rightarrow ee)$ + jets selection at $\sqrt{s} = 10$ TeV with 100 pb$^{-1}$ as a function of PF-(15 GeV/c), track-(15 GeV/c), calo-(uncorrected 30 GeV/c) and corrected calo-(58 GeV/c) jet multiplicity.

where $\sigma=\sigma_L$ and $\alpha=\alpha_L$ ($\sigma=\sigma_R$ and $\alpha=\alpha_R$) for $x < m(x > m)$. Within the precision of the targeted luminosity we find that the fit parameters are independent of the jet multiplicity. The background in both electron and muon final states is dominated by the multijets component. The shape of the background is studied in the ‘anti-lepton’ sample, obtained by inverting the tracking lepton isolation requirement in the lepton enriched QCD samples. The $m(\ell\ell)$ distributions for Monte Carlo events are shown in Figure 1 for the events selected by the $Z+ \geq 1$ track-jet analysis and those falling in the anti-lepton sample. In the case of electrons the selection is looser than the nominal in order to obtain adequate statistics to extract the shape and perform the comparison. The yields are normalized to the ones shown in tables 1 and 2. The anti-lepton samples provide the control data samples for the validation of the analytic function describing the multijet background in the fit. The presence of the other background is accounted for in the fit by floating the shape parameters. Using the PYTHIA [13] and GEANT4 [22] modeling of the multijets background in the CMS detector, we expect the $m(\ell\ell)$ distributions to be well-described by either an exponential or a second-order polynomial as shown in Figure 1 for track-jets. Similar distributions are obtained for the other jet definitions.

Figure 1: $m(\mu\mu)$ (left) and $m(ee)$ (right) for track-jet counting in the lepton enriched multijets sample (points) and the control anti-lepton sample (histogram). Similar results are obtained using other jet definitions.

4.5.2 Fit results and tests

By performing a set of pseudo-experiments for each jet multiplicity, we estimate the expected statistical error on the signal yield for 100 pb$^{-1}$. The fits are performed on Monte Carlo samples generated from the distributions obtained from the full simulation. This allows to perform the fit with unweighted events and to properly compute the statistical error of the fit result. We obtain for both dielectron and dimuon channels a precision of $\sim2\%$ for $Z+ \geq 1$ calo-jets ($\sim16\%$ for $Z \geq 3$ calo-jets). The corresponding statistical errors for track-jets are smaller due
to the larger statistics. These Monte Carlo tests demonstrate that the ML fit is unbiased and that the 68% confidence interval computed using the likelihood ratio correctly covers the true number of events. In Figures 2 and 3 the result of the fit is shown for the dimuon+jets and dielectron+jets final states with track-jets counting.

Figure 2: Projection of the likelihood at maximum on $m_{\mu\mu}$ for $Z(\rightarrow \mu\mu)\geq 1$ track-jets (top-left), $Z(\rightarrow \mu\mu)\geq 2$ track-jets (top-right), $Z(\rightarrow \mu\mu)\geq 3$ track-jets (bottom-left), and $Z(\rightarrow \mu\mu)\geq 4$ track-jets (bottom-right). The 'data' sample corresponding to 100 pb$^{-1}$ statistics is overlaid. The error bars correspond to the expected precision.

5 The $Z+\geq n$ jets over $Z+\geq (n+1)$ jets yield ratio

With the ML fit to the four different jet multiplicity samples we measure the yields of $Z+n$ jets as a function of jet multiplicity. Defining $C$ as the $Z+n$ jets over $Z+(n+1)$ jets yield ratio, we expect $C$ to be independent of $n$, within errors. Under the assumption that $C$ is a constant, the ratio of inclusive $Z+n$ jets ($\geq n$ jets) over inclusive $Z+(n+1)$ jets ($\geq n+1$) is identical to the ratio of exclusive $Z+n$ jets ($=n$) over exclusive $Z+(n+1)$ jets ($=n+1$). Thus physically $C$ represents the cost of adding an extra jet to $Z+n$ jet production at some fixed order in $\alpha_s$. The extracted value of $C$ depends on the jet definition: e.g. increasing the jet $p_T$ threshold for a fixed cone size increases $C$, while decreasing the cone size for a fixed jet $p_T$ threshold also increases $C$. Indeed, the difference in the $C$ values extracted from the calo-jet counting versus track-jet counting is largely due to the fact that track-jets probe a lower $p_T$ region of the phase space. By using both track-jets and calo-jets counting, the prediction of a constant $C$ can be verified in different regions of the phase space and using independent detector elements. Additionally, by using corrected calo-jets or PF-jets a detailed quantitative comparison with the parton-level QCD predictions could eventually be made. The loose selection used in this analysis allows us to confirm the expected behavior already with a data sample of 100 pb$^{-1}$, directly accessing $Z+n$ jets events up to the four jets inclusive bin. The fit of the measured yields to an exponential, shown in Figure 4, confirms the validity of the constant ratio assumption, returning fit probabilities between 75% and 94%. Here the errors reflect the expected statistical precision on data, as estimated from pseudo-experiment MC simulations without taking into
account the statistical correlation between the successive bins. We expect a similar picture to emerge from the first LHC data.

The overall selection efficiency $\epsilon_S$ within each jet multiplicity bin, as estimated from Monte Carlo simulation, is constant as a function of the number of jets. We obtain $\epsilon_S = (41.4 \pm 0.5)\%$ ($\epsilon_S = (42 \pm 11)\%$) for $Z(\rightarrow ee)+ \geq 1$ ($Z(\rightarrow ee)+ \geq 2$) calo-jets and $\epsilon_S = (47.5 \pm 0.5)\%$ ($\epsilon_S = (48 \pm 4)\%$) for $Z(\rightarrow \mu\mu)+ \geq 1$ ($Z(\rightarrow \mu\mu)+ \geq 4$). The stability of $\epsilon_S$ is a benefit of the loose $Z$ selection. We verify that the efficiency correction of the yields has a small impact on the results, inducing a shift in the slope $C$ smaller than the expected precision in 100 pb$^{-1}$. The use of different jet definitions demonstrates the robustness of the results. The output of the fit results shown in Figure 4 is i) $C_{(tj)}^{\mu\mu}=7.3 \pm 0.3$ for the $Z(\rightarrow \mu\mu)$ + calo-jets and $C_{(tj)}^{\mu\mu}=5.3 \pm 0.1$ for the $Z(\rightarrow \mu\mu)$ + track-jets and, ii) $C_{(tj)}^{ee}=7.6 \pm 0.4$ for the $Z(\rightarrow ee)$ + calo-jets and $C_{(tj)}^{ee}=5.3 \pm 0.2$ for the $Z(\rightarrow ee)$ + track-jets. The results are consistent with lepton universality ($C_{(tj)}^{\mu\mu}/C_{(tj)}^{ee}$ is consistent with 1).

The value of $C_{(tj)}^{\mu\mu}$ for calo-jets corresponds to the value obtained for generator-level jets in the same rapidity range for $p_T$ threshold 58 GeV, in agreement with the expected calorimeter response. With understood data, the slope, as extracted from corrected calo-jets and PF-jets, could be directly compared to QCD predictions, represented here by the generator-level jets from leading order QCD Monte Carlo with jet-parton matching. We validate that this is the case taking a 58 GeV/$c$ $p_T$ threshold for both corrected calo-jets and PF-jets. We obtain for corrected calo-jets $C_{cor-cj}^{ee(\mu\mu)} = 7.5 (7.6) \pm 0.5$ and for PF-jets $C_{PF-j}^{ee(\mu\mu)} = 7.5 (7.6) \pm 0.5$ as expected. The scaling for PF-jets counting is included in Figure 4. For track-jets, the $C_{(tj)}^{\mu\mu}$ value that corresponds to the value obtained with generator-level jets, within the rapidity region $|\eta| <2.4$, is obtained for a $p_T$ threshold of 30 GeV/$c$; this is compatible with the expectation of the jet charged fraction. With PF-jets the measurement can be performed with a threshold as low or
Figure 4: The \((dN/dn_{\text{jet}})\) distributions and exponential fit for \(Z(\rightarrow \mu\mu) + \geq 1 \text{ jets}\) (left) and \(Z(\rightarrow e\mu) + \geq 1 \text{ jets}\) (right) for different jet definitions. The resulting constant ratio of \(Z + n \text{ jets}\) to \(Z + (n + 1) \text{ jets}\) for each case is also shown with the uncertainty band from the fit.
lower than 15 GeV/$c$, producing a candle sample of optimal statistics. There are known QCD effects that can cause deviations from a constant slope. The first is that the $n$th jet has associated a factor of $\alpha_s(Q_n)$ whose physical scale $Q_n$ can be substantially lower than the scale $Q$ of the original hard subprocess, thus reducing the cost of this jet by $\alpha_s(Q)/\alpha_s(Q_n)$. The second is the Sudakov suppression of the extra hard branching needed to produce an extra jet. These two effects, which contribute in opposite directions, are included in the MadGraph matching, and the combined effect on the slope is small. Another effect is from higher order virtual contributions not included in the MadGraph matrix elements; an estimate of this effect awaits a full NLO calculation of $Z+3$ jets production. Given the results we present here using the high statistics $Z+1$ jets and $Z+2$ jets multiplicity bins (using either calo-jet or track-jet counting), we can estimate the $Z+3$ jets rate to within less than $\sim 10\%$. This is about equal to the precision expected from NLO calculations in the coming years. Further studies of $Z$+jets as a function of the boson or jet $p_T$ would be required to test the predictions in the high $p_T$ part of the phase space, where also larger integrated luminosity is required.

6 The $Z$+ jets sample as a candle

Some examples that appear in the literature of the $Z$+jets boson used as a candle include i) $Z$+jets as a normalization reference for the estimate of the $Z$ invisible decays after tuning the MC simulation to data [23] ii) $Z$+jets as a handle for extracting jet energy corrections or jet reconstruction efficiency (e.g. $Z$-jet balancing [24, 25]) and iii) $Z$ as a reference for characterizing the MET [26].

6.1 $Z$+ jets derived MET corrections for $W$+jets

An example is shown here of how the $Z(\rightarrow \mu\mu)$+jets selected data sample can be used as a calibration reference process for the missing transverse energy in $W$+jets events.

Due to the similarities between $Z$+jets and $W$+jets topologies the $Z(\rightarrow \mu\mu)$+jets sample can be used to calibrate the MET in $W(\rightarrow \mu\nu)$+jets events. The MET is decomposed into two orthogonal components, denoted $U_\parallel$ and $U_\perp$, that correspond respectively to the MET components parallel and perpendicular to the muon associated with the $W$ boson candidate. A $W$-like view of $Z(\mu\mu)$+jets events can be considered by treating one of the muons from the $Z$ decay as an escaping neutrino. The selected $Z(\rightarrow \mu\mu)$ + jets sample is used to compute corrections to $U_\parallel$ and $U_\perp$ by comparing the values from the calorimetric MET with the values obtained from the dimuon kinematics. Each event of the candle sample enters with a $m_{\mu\mu}$-based weight following the sPlot technique [27] which provides an optimal background subtraction. The $W$ transverse mass is shown before and after the corrections in $W(\rightarrow \mu\nu)$+ $\geq 1$ jets in Figure 5. After the corrections the characteristic Jacobian edge of the $W$ is mostly recovered.

6.2 $Z$+ jets and new physics

$Z$ bosons and jets produced through a new mechanism at the LHC could induce a large deviation from a constant slope in jet counting. This is the case for example in SUSY models with real $Z$ production and high jet multiplicity in the final states [28]. This kind of production mechanism could induce an excess of events at high jet multiplicity and a discrepancy between the observed yield and the predicted one, as obtained from the $Z+ \geq 1$ jets and $Z+ \geq 2$ jets yields. The presence of NP events could also bias the prediction, since the jet counting is inclusive and the NP events will be ‘contaminating’ all the jet multiplicity bins.

To demonstrate the sensitivity of the analysis on the breaking of the Berends-Giele scaling, we
consider an mSUGRA benchmark SUSY point that includes production of \( Z \) bosons in decays of the neutralinos (LM4 [28]). For a given number of NP events in the \( Z+ \geq 1 \) jets sample, we generate a set of MC pseudo-experiments according to the SM and background probability densities obtained in Section 4.5.

The SUSY sample contains events with real \( Z \) bosons as well as fake \( Z \) candidates from leptons produced in the decay chains of SUSY particles. The fit can distinguish between fake and true \( Z \) candidates, but it cannot separate NP events from SM ones. This results in a discrepancy between the observed yields at high jet multiplicities and the predicted values using the Berends-Giele scaling at lower jet multiplicities.

The result is shown in Figure 6 as a function of the total number of NP events (including those with fake \( Z \) candidates) added to a 100 pb\(^{-1}\) of SM events.

A simultaneous departure from the prediction in calo-jet and track-jet counting could not easily be attributed to systematic effects. If such a discrepancy is seen in data, beyond what the QCD effects discussed in Sec 5 could induce, one could use the sPlots to characterize the excess events, by studying effects of stable weakly interacting dark matter candidates in the MET distribution.

The \( Z+ \) jets candle analysis can also provide a measurement of the \( Z(\rightarrow\nu\nu)+\) jets irreducible background to MET-based new physics searches in hadronic final states. This is shown in Fig. 7 where we compare the expected distribution of MET for \( Z(\rightarrow\nu\nu)+\) jets with the sPlots of \( Z(\rightarrow\mu\mu)+\) jets events. The distributions are normalized to 100 pb\(^{-1}\).

### 7 Conclusions

We have demonstrated the possibility of measuring the \( Z+n \) jets over \( Z+(n+1) \) jets ratio, and verifying its constancy in \( n \) as predicted by QCD within the expected uncertainties of 100 pb\(^{-1}\) of data collected at \( pp \sqrt{s} = 10 \) TeV. In measuring this ratio systematic uncertainties related to the jet definition and counting are mostly suppressed. The cancellation of systematic uncertainties is predominantly due to the correlation in the jet counting uncertainties (and jet}

![Figure 5: The \( W \) transverse mass in \( W(\rightarrow\mu\nu)^+ \geq 1 \) jets before and after correcting the MET using the \( Z+\) jets candle derived corrections.](image-url)
Reconstructed SUSY Events [LM4]

Figure 6: Comparison of the $Z(\rightarrow \mu\mu)^+ \geq 4$ track-jets measured yields (filled dots) and the prediction using the $\geq 1$ and $\geq 2$ jet multiplicity yields (empty dots) as a function of the number of new physics events in the $Z^+ \geq 1$ jets sample. The LM4 point [28] is taken as a benchmark for the jet multiplicity in NP events. To quantify the breaking of the Berends-Giele scaling the average pull is also shown. distribution of $Z(\rightarrow \nu\nu)^+$

Figure 7: MET distribution of $Z(\rightarrow \nu\nu)^+ \geq 1$ calo-jets events (histogram) for events with MET $> 50$ GeV. The $Z(\rightarrow \nu\nu)^+ \geq 1$ calo-jets events are compared to the sPlots distribution of $Z(\rightarrow \mu\mu)^+ \geq 1$ calo-jets. The filled (empty) points correspond to the measured sPlot distribution (the distribution after scaling for the selection efficiency and the cross-section ratio).
energy corrections and uncertainties) between the numerator and denominator. The analysis with track-jet counting requires less integrated luminosity at startup, while the optimally largest statistics candle sample can be obtained using the Particle Flow jet counting with appropriately low jet threshold. We have shown how the validated $Z$+jets samples can serve as a reference for the initial commissioning of the MET and as a probe for new physics if an excess from SM predictions is observed at large jet multiplicities.

References


