

---

# DRAFT

## CMS Physics Analysis Summary

*The content of this note is intended for CMS internal use and distribution only*

---

2009/06/18  
Archive Id:  
Archive Date:

### 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 discuss a feasibility study of Z boson production in association with jets with the first LHC data. 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. The study targets the first  $\mathcal{O}(100)$   $\text{pb}^{-1}$  collected with the CMS detector at  $\sqrt{s} = 10$  TeV.

PDFAuthor:	M. Bona, E. Di Marco, J. Lykken, P. Meridiani, M. Pierini, C. Rogan, C. Rovelli, I. Segoni, M. Spiropulu, T. Tomei, M. Zanetti
PDFTitle:	Z+j candle@CMS LHC “(s’)=10 TeV
PDFSubject:	Z+j candle@CMS LHC “(s’)=10 TeV
PDFKeywords:	LHC,CMS, physics, proton-proton collisions, W,Z plus jets associated production, backgrounds SUSY



## 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  $\mathcal{O}(100)$  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  $\mathcal{O}(\alpha_s^n)$ . 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]. The purpose of this analysis is: i) to measure the ratio at different jet multiplicities at  $pp \sqrt{s} = 10$  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 than 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. Besides the barrel and endcap detectors, CMS has extensive forward calorimetry.

## 3 Signal and background samples

The  $Z(\rightarrow\ell\ell) + n$ -jets events ( $\ell = e, \mu$ ) are studied with Monte Carlo simulation, using the MADGRAPH [12] event generator, based on a leading-order calculation of the matrix element (ME). ME calculation is performed for final states with at most four primary partons, requiring that the parton  $p_T$  exceeds 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. 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 parton  $p_T > 20$  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) 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 the low luminosity running ( $L=10^{32}\text{cm}^{-2}\text{s}^{-1}$ ). The trigger paths used here 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 the 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 tracking electron isolation criterion is applied by requiring the sum  $p_T$  of the tracks compatible with the electron vertex and  $p_T > 1.5$  GeV/ $c$  around the electron candidate track in a cone of size  $R_{cone} = 0.4$  to be less than 15% of the electron candidate 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  $60 < m_{\ell\ell} < 110$  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 towers 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 towers and tracks respectively using the Seedless Infrared Safe Cone (SISCone) jet algorithm [20] with a cone size  $R_{cone} = 0.5$  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  $p_T > 30$  GeV/ $c$  ( $p_T > 15$  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 the limited  $\eta$  acceptance compared to calo-jets.

#### 4.5 Maximum likelihood fit

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

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

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

In Tables 1 and 2 the event yields are shown for signal and background events in the  $Z(\rightarrow\mu\mu)$ +jets and  $Z(\rightarrow ee)$ +jets selection for calo-jet and track-jet counting for  $100 \text{ pb}^{-1}$  of integrated luminosity. The quoted errors are statistical only, related to the size of the available datasets. Tables 3 and 4 show the signal event yields in the  $Z(\rightarrow ee)$  and  $Z(\rightarrow\mu\mu) +$  jets selection at  $\sqrt{s} = 10$  TeV with  $100 \text{ pb}^{-1}$  as a function of PF-jet multiplicity and with PF-jets of  $|\eta| < 3$  and  $p_T$  threshold 15 GeV/ $c$  compared to the yields obtained with the track-jet and calo-jet selection (both uncorrected and corrected as per discussion in Section 5). With PF-jets the candle analysis produces a candle sample with more than double the statistics compared to using the track-jet selection, due to the lower  $p_T$  threshold for jets. The dimuon channel is always statistically more powerful than the dielectron channel despite the muon acceptance being lower than the electron one,

	Z+jets	Z other	W+jets	$t\bar{t}$ +jets	QCD
$\geq 1$ jets	$4007 \pm 37$ ( $9305 \pm 56$ )	$11 \pm 2$ ( $14 \pm 2$ )	$44 \pm 4$ ( $34 \pm 3$ )	$109 \pm 2$ ( $119 \pm 2$ )	$900 \pm 70$ ( $1000 \pm 80$ )
$\geq 2$ jets	$555 \pm 14$ ( $1741 \pm 24$ )	$2 \pm 1$ ( $2 \pm 1$ )	$17 \pm 2$ ( $18 \pm 2$ )	$66 \pm 2$ ( $90 \pm 2$ )	$450 \pm 50$ ( $450 \pm 50$ )
$\geq 3$ jets	$72 \pm 5$ ( $338 \pm 11$ )	$1 \pm 1$ ( $1 \pm 1$ )	$7 \pm 2$ ( $16 \pm 2$ )	$25 \pm 1$ ( $43 \pm 1$ )	$90 \pm 25$ ( $160 \pm 30$ )
$\geq 4$ jets	$11 \pm 2$ ( $66 \pm 5$ )	-(-)	$-(1 \pm 1)$	$4 \pm 1$ ( $8 \pm 2$ )	$17 \pm 10$ ( $35 \pm 15$ )

Table 1: Expected signal and backgrounds yields in the  $Z(\rightarrow \mu\mu) + \text{jets}$  selection 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.

	Z+jets	Z other	W+jets	$t\bar{t}$ +jets	QCD
$\geq 1$ jets	$3135 \pm 32$ ( $7106 \pm 49$ )	$9 \pm 2$ ( $18 \pm 2$ )	$15 \pm 2$ ( $26 \pm 3$ )	$70 \pm 1$ ( $78 \pm 1$ )	$1500 \pm 500$ ( $2000 \pm 900$ )
$\geq 2$ jets	$411 \pm 11$ ( $1334 \pm 21$ )	$1 \pm 1$ ( $3 \pm 1$ )	$4 \pm 1$ ( $9 \pm 2$ )	$34 \pm 1$ ( $55 \pm 1$ )	$200 \pm 100$ ( $900 \pm 600$ )
$\geq 3$ jets	$58 \pm 4$ ( $268 \pm 9$ )	$-$ ( $2 \pm 1$ )	$1 \pm 1$ ( $3 \pm 1$ )	$9 \pm 1$ ( $23 \pm 1$ )	$-$ ( $200 \pm 200$ )
$\geq 4$ jets	$7 \pm 2$ ( $52 \pm 4$ )	$-$ ( $-$ )	$-(1 \pm 1)$	$2 \pm 1$ ( $7 \pm 1$ )	$-$ ( $100 \pm 100$ )

Table 2: Expected signal and backgrounds yields in the  $Z(\rightarrow ee) + \text{jets}$  selection 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 parenthesis the track-jet counting ones.

due the higher muon reconstruction efficiency.

	PF-jets (15 GeV/c)	track-jets (15 GeV/c)	calo-jets (30 GeV/c)	calo-jets <sup>corr</sup> (58 GeV/c)
$\geq 1$ jets	$16750 \pm 23$	$7106 \pm 49$	$3135 \pm 32$	$3098 \pm 32$
$\geq 2$ jets	$5865 \pm 44$	$1334 \pm 21$	$411 \pm 11$	$411 \pm 12$
$\geq 3$ jets	$1928 \pm 25$	$268 \pm 9$	$58 \pm 4$	$57 \pm 4$
$\geq 4$ jets	$581 \pm 14$	$52 \pm 4$	$7 \pm 2$	$6 \pm 1$
$\geq 5$ jets	$151 \pm 71$	$7 \pm 6$	-	-

Table 3: Expected signal event yields in the  $Z(\rightarrow ee) + \text{jets}$  selection at  $\sqrt{s} = 10$  TeV with  $100 \text{ pb}^{-1}$  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.

126

#### 127 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:

$$f(x; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N_s \cdot e^{-\frac{(x-m)^2}{2\sigma^2 + \alpha(x-m)^2}} \quad (2)$$

128 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 tar-  
 129 geted luminosity we find that the fit parameters are independent of the jet multiplicity. The  
 130 background in both electron and muon final states is dominated by the QCD component. The  
 131 shape of the background is studied in the ‘‘anti-lepton’’ sample, obtained by inverting the track-  
 132 ing lepton isolation requirement in the lepton enriched QCD sample. The  $m_{\ell\ell}$  distributions for  
 133 Monte Carlo events are shown in Figure 1 for the events selected by the  $Z + \geq 1$  track-jet anal-  
 134 ysis and those falling in the anti-lepton sample. In the case of electrons the selection is looser  
 135 than the nominal in order to obtain adequate statistics to extract the shape and perform the  
 136 comparison. The yields are normalized to the ones shown in tables 2 and 1. The anti-lepton  
 137 sample provides the control data sample for the validation of the analytical function describing  
 138 the multijet background in the fit. The presence of the other background is accounted for in the

	PF-jets (15 GeV/c)	track-jets (15 GeV/c)	calo-jets (30 GeV/c)	calo-jets <sup>corr</sup> (58 GeV/c)
$\geq 1$ jets	$24409 \pm 90$	$9305 \pm 56$	$4007 \pm 37$	$3693 \pm 35$
$\geq 2$ jets	$8725 \pm 54$	$1741 \pm 24$	$555 \pm 14$	$493 \pm 12$
$\geq 3$ jets	$2889 \pm 31$	$338 \pm 11$	$72 \pm 5$	$58 \pm 4$
$\geq 4$ jets	$885 \pm 17$	$66 \pm 5$	$11 \pm 2$	$11 \pm 2$
$\geq 5$ jets	$243 \pm 9$	$11 \pm 8$	–	–

Table 4: Expected signal event yields in the  $Z(\rightarrow \mu\mu) + \text{jets}$  selection at  $\sqrt{s} = 10$  TeV with  $100 \text{ pb}^{-1}$  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.

139 fit by floating the shape parameters. Using the PYTHIA [13] and GEANT4 [22] modeling of the  
 140 QCD background in the CMS detector, we expect the  $m_{ll}$  distributions to be well-described by  
 141 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.

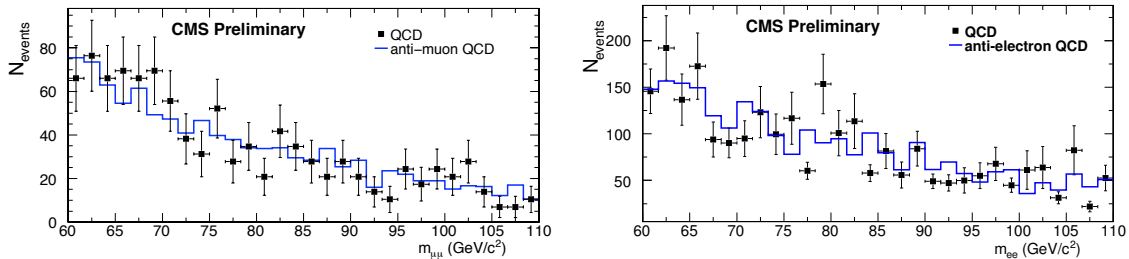


Figure 1:  $m(\mu\mu)$  (left) and  $m(ee)$  (right) for track-jet counting in the lepton enriched QCD sample (points) and the control “anti-lepton” QCD sample (histogram). Similar results are obtained using other jet definitions.

142

#### 143 4.5.2 Fit results and tests

144 By performing a set of pseudo-experiments for each jet multiplicity, we estimate the expected  
 145 statistical error on the signal yield for  $100 \text{ pb}^{-1}$ . The fits are performed on Monte Carlo samples  
 146 generated from the distributions obtained from the full simulation. This allows to perform the  
 147 fit with un-weighted events and to properly compute the statistical error of the fit result. We  
 148 obtain for both dielectron and dimuon channels a precision of  $\sim 2\%$  for  $Z + \geq 1$  calo-jets ( $\sim 16\%$   
 149 for  $Z \geq 3$  calo-jets). The corresponding values for track-jets are smaller due to the larger  
 150 statistics. These Monte Carlo tests demonstrate that the ML fit is unbiased and that the 68%  
 151 confidence interval computed using the likelihood ratio correctly covers the true number of  
 152 events. In Figures 2 and 3 the result of the fit is shown for the dimuon+jets and dielectron+jets  
 153 final states with track-jets.

## 154 5 The $Z+n$ jets over $Z + (n+1)$ jets yield ratio

155 With the ML fit to the four different jet multiplicity samples we measure the yields of  $Z+n$  jets  
 156 as a function of jet multiplicity. Defining  $C$  as the  $Z+n$  jets over  $Z + (n+1)$  jets yield ratio, we  
 157 expect  $C$  to be independent of  $n$ , within errors. Under the assumption that  $C$  is a constant, the  
 158 ratio of inclusive  $Z+n$  jets ( $\geq n$  jets) over inclusive  $Z + (n+1)$  jets ( $\geq n+1$ ) is identical to  
 159 the ratio of exclusive  $Z+n$  jets ( $= n$ ) over exclusive  $Z + (n+1)$  jets ( $= n+1$ ). Thus physically  $C$   
 160 represents the cost of adding an extra jet to  $Z+n$  jet production at some fixed order in  $\alpha_s$ . The

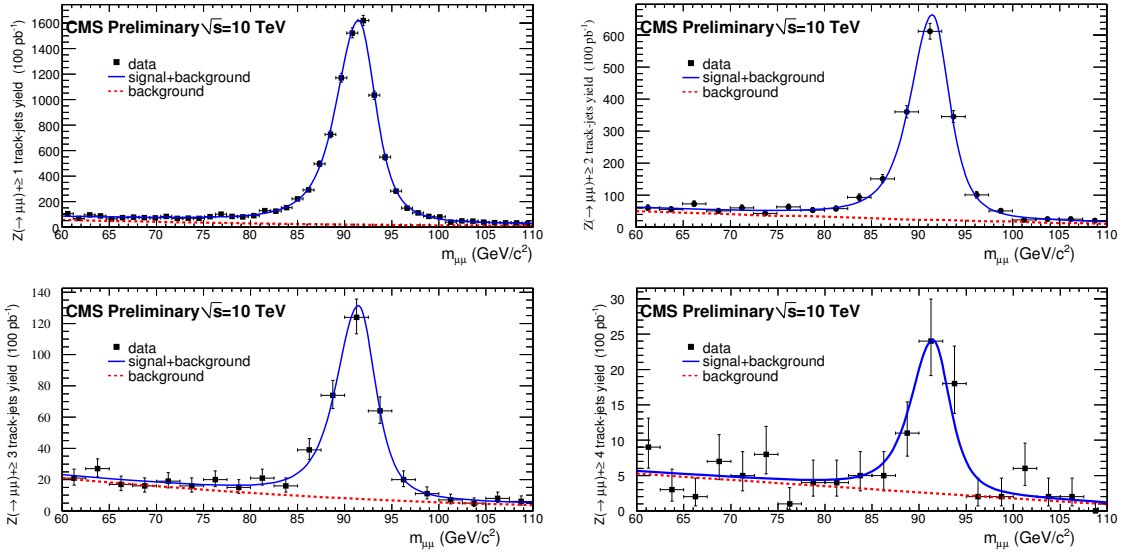


Figure 2: Projection of the likelihood at maximum on  $m_{\mu\mu}$  for  $Z(\rightarrow \mu\mu) + \geq 1$  track-jets (top-left),  $Z(\rightarrow \mu\mu) + \geq 2$  track-jets (top-right),  $Z(\rightarrow \mu\mu) + \geq 3$  track-jets (bottom-left), and  $Z(\rightarrow \mu\mu) + \geq 4$  track-jets (bottom-right).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

161 extracted value of  $C$  depends on the jet definition: e.g. increasing the jet  $p_T$  threshold for a fixed  
 162 cone size increases  $C$ , while decreasing the cone size for a fixed jet  $p_T$  threshold also increases  
 163  $C$ . Indeed, the difference in the  $C$  values extracted from the calo-jet counting versus track-jet  
 164 counting is largely due to the fact that track-jets probe a lower  $p_T$  region of the phase space.  
 165 By using both track-jets and calo-jets counting, the prediction of a constant  $C$  can be verified  
 166 in different regions of the phase space and using independent detector elements. Additionally,  
 167 by using corrected calo-jets or PF-jets a detailed quantitative comparison with the parton-level  
 168 QCD predictions could eventually be made. The loose selection used in this analysis allows us  
 169 to confirm the expected behavior already with a data sample  $100 \text{ pb}^{-1}$ , directly accessing  $Z$ -jets  
 170 events up to the four jets inclusive bin. The fit of the measured yields to an exponential, shown  
 171 in Figure 4, confirms the validity of the constant ratio assumption, returning fit probabilities  
 172 between 75% and 94%. Here the errors reflect the expected statistical precision on data, as  
 173 estimated from toy Monte Carlo tests. We expect a similar picture to emerge from the first LHC  
 174 data.

175 The overall selection efficiency  $\epsilon_S$  within each jet multiplicity bin, as estimated from Monte  
 176 Carlo simulation, is constant as a function of the number of jets. We obtain  $\epsilon_S = (41.4 \pm 0.5)\%$   
 177 ( $\epsilon_S = (42 \pm 11)\%$ ) for  $Z(\rightarrow ee) + \geq 1$  ( $Z(\rightarrow ee) + \geq 4$ ) calo-jets and  $\epsilon_S = (47.5 \pm 0.5)\%$   
 178 ( $\epsilon_S = (48 \pm 4)\%$ ) for  $Z(\rightarrow \mu\mu) + \geq 1$  ( $Z(\rightarrow \mu\mu) + \geq 4$ ). The stability of  $\epsilon_S$  is a consequence of  
 179 the loose  $Z$  selection. We verify that the efficiency correction of the yields has a small impact  
 180 on the results, inducing a shift in the slope  $C$  smaller than the expected precision in  $100 \text{ pb}^{-1}$ .  
 181 The use of different jet definitions demonstrates the robustness of the results. The output of the  
 182 fit results shown in Figure 4 is i)  $C_{c_j}^{\mu\mu} = 7.3 \pm 0.3$  for the  $Z(\rightarrow \mu\mu) +$  calo-jets and  $C_{t_j}^{\mu\mu} = 5.2 \pm 0.1$   
 183 for the  $Z(\rightarrow \mu\mu) +$  track-jets and ii)  $C_{c_j}^{ee} = 7.6 \pm 0.4$  for the  $Z(\rightarrow ee) +$  calo-jets and  $C_{t_j}^{ee} = 5.3 \pm 0.2$   
 184 for the  $Z(\rightarrow ee) +$  track-jets. The results are consistent with lepton universality ( $C_{c(t)_j}^{\mu\mu} / C_{c(t)_j}^{ee}$  is  
 185 compatible with 1).

186 The value of  $C_{c_j}^{\ell\ell}$  for calo-jets corresponds to the value obtained for generator-level jets in the

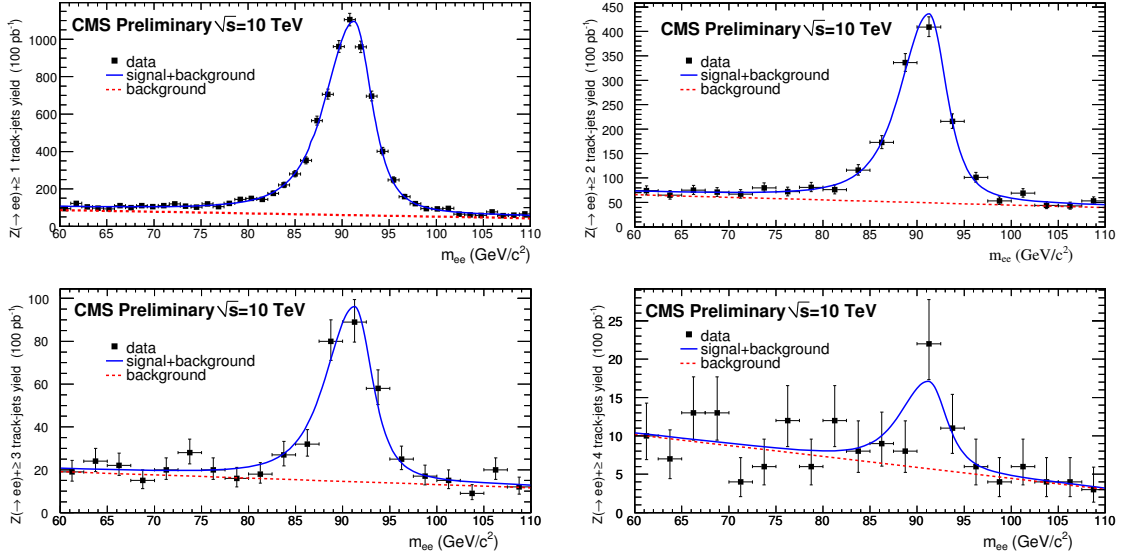


Figure 3: Projection of the likelihood at maximum on  $m_{ee}$  for  $Z(\rightarrow ee) + \geq 1$  track-jets (top-left),  $Z(\rightarrow ee) + \geq 2$  track-jets (top-right),  $Z(\rightarrow ee) + \geq 3$  track-jets (bottom-left), and  $Z(\rightarrow ee) + \geq 4$  track-jets (bottom-right). The “data”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

187 same rapidity range for  $p_T$  threshold 58 GeV, in agreement with the expected calorimeter re-  
 188 sponse. With understood data, the slope, as extracted from corrected calo-jets and PF-jets,  
 189 could be directly compared to QCD predictions, represented here by the generator-level jets  
 190 from leading order QCD Monte Carlo with jet-parton matching. We validate that this is the  
 191 case taking a 58 GeV/c  $p_T$  threshold for both corrected calo-jets and PF-jets. We obtain for corrected  
 192 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. For  
 193 track-jets, the  $C_{tj}^{ee}$  value that corresponds to the value obtained with generator-level jets, within  
 194 the rapidity region  $|\eta| < 2.4$ , is obtained for a  $p_T$  threshold of 30 GeV; this is compatible with the  
 195 expectation of the jet charged fraction. With PF-jets the measurement can be performed with a  
 196 threshold as low or lower than 15 GeV/c, producing a optimally larger statistics candle sample.  
 197 The slope linearity test with PF-jets of 15 GeV/c is shown in Figure 5. There are known QCD  
 198 effects that can cause deviations from a constant slope. The first is that the  $n$ th jet has associated  
 199 a factor of  $\alpha_s(Q_n)$  whose physical scale  $Q_n$  can be substantially lower than the scale  $Q$  of the  
 200 original hard subprocess, thus reducing the cost of this jet by  $\alpha_s(Q)/\alpha_s(Q_n)$ . The second is the  
 201 Sudakov suppression of the extra hard branching needed to produce an extra jet. These two  
 202 effects, which work in opposite directions, are included in the MADGRAPH matching, and the  
 203 net effect on the slope is small. Another effect is from higher order virtual contributions not  
 204 included in the MADGRAPH matrix elements; an estimate of this effect awaits a full NLO calcu-  
 205 lation of  $Z+3$  jets production. Given the results we present here using the high statistics  $Z+1$   
 206 jets and  $Z+2$  jets multiplicity bins (using either calo-jet or track-jet counting), we can estimate  
 207 the  $Z+3$  jets rate to within less than  $\sim 10\%$ . This is about equal to the precision expected from  
 208 NLO calculations in the coming years. Further studies of  $Z$ +jets as a function of the boson or  
 209 jet  $p_T$  would be required to test the predictions in the high  $p_T$  part of the phase space, where  
 210 also larger integrated luminosity is required.

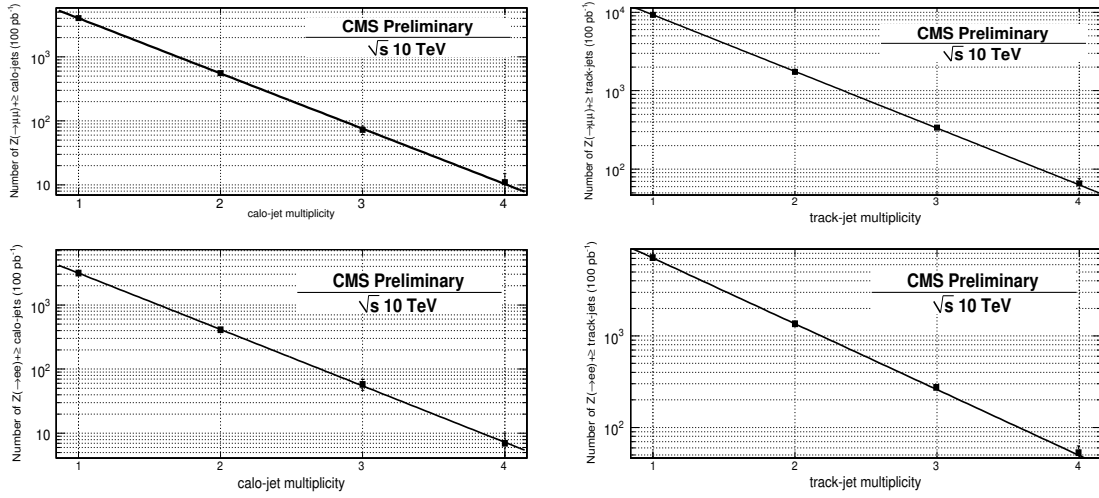


Figure 4: The  $(dN/dn_{\text{jet}})$  distributions and exponential fit for  $Z(\rightarrow \mu\mu) + \geq 1$  calo-jets (top left) and track-jets (top right) and  $Z(\rightarrow ee) + \geq 1$  calo-jets (bottom left) and track-jets (bottom right). The resulting constant ratio  $Z + n$  jets over  $Z + (n + 1)$  jets values for all cases is given in the text.

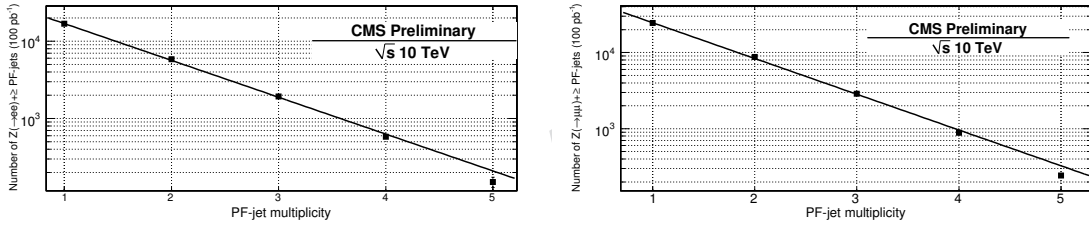


Figure 5: The  $(dN/dn_{\text{jet}})$  distribution and exponential fit for  $Z(\rightarrow ee) + \geq 1$  (left) and  $Z(\rightarrow \mu\mu) + \geq 1$  (right) PF-jets.

## 6 The Z boson as a “candle”

211

212 Some examples that appear in the literature of the Z boson used as a “candle” include i) Z+jets  
 213 as a normalization reference for the estimate of the Z invisible decays after tuning the MC to  
 214 data [23] ii) Z as a handle for measuring trigger efficiencies (e.g. “tag-and-probe”) or extracting  
 215 jet energy corrections (e.g. Z-jet balancing [24]) and iii) Z as a reference for characterizing the  
 216 MET [25].

### 6.1 Candle derived MET corrections for W+jets and MET characterization

217

218 An example is shown here of how the  $Z(\rightarrow \mu\mu)$ +jets selected data sample can be used as a  
 219 calibration reference process for the missing transverse energy in W+jets events that profits  
 220 from the effective background subtraction technique offered by the sPlot formulation [26].

221 Due to the similarities between Z+jets and W+jets topologies the  $Z(\rightarrow \mu\mu)$ +jets sample can be  
 222 used to calibrate the MET in  $W(\rightarrow \mu\nu)$ +jets events. The MET is decomposed into two orthog-  
 223 onal components, denoted  $U_{\parallel}$  and  $U_{\perp}$ , that correspond respectively to the MET components  
 224 perpendicular and parallel to the muon associated with the W boson candidate. Analogously, a  
 225 “W-like” view of  $Z(\mu\mu)$ +jets events can be considered by treating one of the muons from the Z  
 226 decay as an escaping neutrino. This is a reasonable approximation given the small magnitude

227 of expected calorimetric depositions from the muons. Using this approach the candle is the  
 228 selected  $Z(\rightarrow\mu\mu)+\text{jets}$  sample after the effective background subtraction using the  $m_{\ell\ell}$ -based  
 229 sPlots technique. The  $U_{\parallel}$  and  $U_{\perp}$  are calculated in this sample using both the fully measured Z  
 230 boson kinematics (the “real” dimuon specific values) and independently using the calorimetric  
 231 MET (the “measured” values). The former are then used to calculate corrections for the latter.  
 232 The corrections for  $U_{\parallel}/U_{\perp}$  are derived in bins of  $p_T^{\mu}$  and  $U_{\parallel}/U_{\perp}$ . The W transverse mass is  
 233 shown before and after the corrections in  $W(\rightarrow\mu\nu)+\geq 1$  jets in Figure 6. After the corrections  
 the characteristic Jacobian edge of the W is recovered.

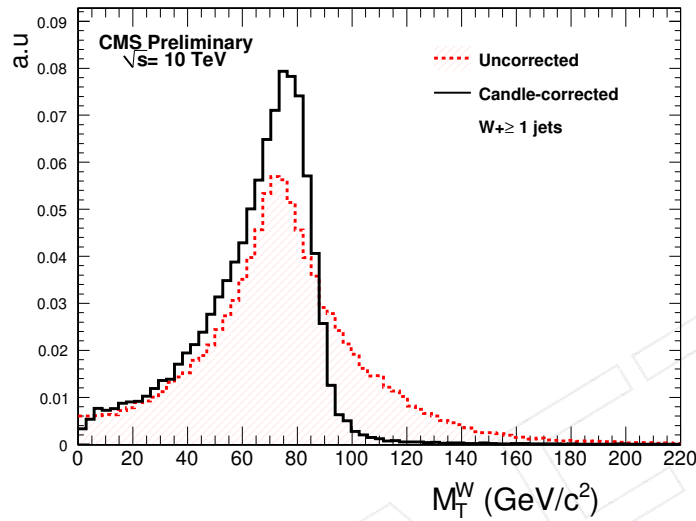


Figure 6: The W transverse mass in  $W(\rightarrow\mu\nu)+\geq 1$  jets before and after the Z candle derived MET corrections.

234

## 235 6.2 The Z+jets candle and new physics

236 Z bosons and jets produced through a new mechanism at the LHC could induce a large devi-  
 237 ation from a constant slope in jet counting. This is the case for example (and without loss of  
 238 generality) in SUSY models with real Z production and high jet multiplicity in the final states  
 239 [27]. This kind of production mechanism could induce an excess of events at high jet multiplicity  
 240 and a discrepancy between the observed yield and the predicted one, obtained from  $Z+\geq 1$   
 241 jets and  $Z+\geq 2$  jets yields. The presence of NP events could also bias the prediction, since the  
 242 jet counting is inclusive and the NP events will be “contaminating” all the jet multiplicity bins.

To show the sensitivity of the analysis to a breaking in linearity, we consider an mSUGRA benchmark SUSY point that includes production of Z bosons in decays of the neutralinos (LM4 [27]). For a given number of NP events in the  $Z+\geq 1$  jets sample, we perform a set of toy Monte Carlo experiments using the SM signal and backgrounds as described in Section 4.5, while NP events are generated separately. We consider both the events with real Z bosons and events with fake Z candidates from leptons produced from the decay chains of SUSY particles. The ML 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 scaling at lower jet multiplicities.

$$N_{\geq 3j}^{\text{PREDICTED}} = \frac{(N_{\geq 2j}^{\text{OBSERVED}})^2}{N_{\geq 1j}^{\text{OBSERVED}}} ; \quad N_{\geq 4j}^{\text{PREDICTED}} = \frac{(N_{\geq 2j}^{\text{OBSERVED}})^3}{(N_{\geq 1j}^{\text{OBSERVED}})^2} . \quad (3)$$

243 This is shown in Figure 7 as a function of the total number of NP events (including both events  
 244 with real and fake  $Z$  candidates) added to a  $100 \text{ pb}^{-1}$  of SM events. A simultaneous departure  
 245 from the prediction in calo-jet and track-jet counting could not easily be attributed to systematic  
 246 effects. If a similar discrepancy is seen in data, beyond what the QCD effects discussed in Sec 5  
 247 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.

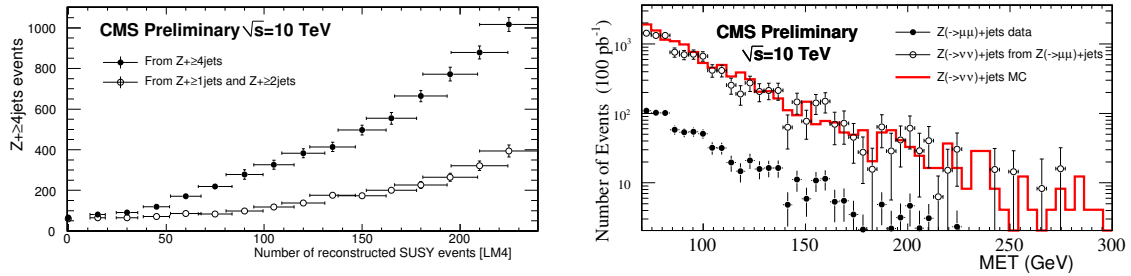


Figure 7: (Left) Comparison of the  $Z(\rightarrow \mu\mu) + \geq 4$  track-jets measured yields (filled dots) and the prediction from Eq. 3 (empty dots) as a function of the number of new physics events in the  $Z + \geq 1$  jets sample. The LM4 point [27] is taken as a benchmark for the jet multiplicity in NP events. (Right) MET Distribution of  $Z(\rightarrow \nu\nu) + \geq 1$  calo-jets events (histogram) for events with  $MET > 50 \text{ GeV}$ . 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).

248

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

## 253 7 Conclusions

254 We have demonstrated the possibility of measuring the  $Z+n$  jets over  $Z + (n+1)$  jets ratio,  
 255 and verifying its constancy in  $n$  as predicted by QCD within the expected uncertainties of  $100$   
 256  $\text{pb}^{-1}$  of data collected at  $pp \sqrt{s} = 10 \text{ TeV}$ . In measuring this ratio systematic uncertainties  
 257 related to the jet definition and counting are suppressed. The cancellation of systematic  
 258 uncertainties is predominantly due to the correlation in the jet counting uncertainties (and jet  
 259 energy corrections and uncertainties) between the numerator and denominator. The analysis  
 260 with track-jet counting requires less integrated luminosity at startup, while the optimally  
 261 largest statistics candle sample can be obtained using the Particle Flow jet counting. We have  
 262 shown how the validated  $Z$ +jets samples can serve as a reference for the initial commissioning  
 263 of the MET and as a probe for New Physics if an excess from SM predictions is observed at  
 264 large jet multiplicities.

## 265 References

- 266 [1] CDF Collaboration, T. Aaltonen et al., "Measurement of the cross section for  $W$ -boson  
 267 production in association with jets in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96 \text{ TeV}$ ," *Phys. Rev. D* **77**  
 268 (2008) 011108, arXiv:0711.4044. doi:10.1103/PhysRevD.77.011108.

- 269 [2] **CDF - Run II** Collaboration, T. Aaltonen et al., "Measurement of Inclusive Jet Cross  
270 Sections in  $Z/\gamma^*(\rightarrow ee)+$  jets Production in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96$  TeV," *Phys. Rev.*  
271 *Lett.* **100** (2008) 102001, arXiv:0711.3717.
- 272 [3] **D0** Collaboration, V. M. Abazov et al., "Measurement of the shape of the boson  
273 transverse momentum distribution in  $p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee+X$  events produced at  
274  $\sqrt{s}=1.96$  TeV," *Phys. Rev. Lett.* **100** (2008) 102002, arXiv:0712.0803.  
275 doi:10.1103/PhysRevLett.100.102002.
- 276 [4] **CMS** Collaboration, R. Adolphi et al., "The CMS experiment at the CERN LHC," *JINST*  
277 **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.
- 278 [5] **CMS** Collaboration, G. L. Bayatian et al., "Particle Flow Event Reconstruction in CMS  
279 and Performance for Jets, Taus and MET,". PF-09-001 (2009).
- 280 [6] J. M. Campbell and R. K. Ellis, "Next-to-leading order corrections to  $W + 2$ jet and  $Z + 2$ jet  
281 production at hadron colliders," *Phys. Rev.* **D65** (2002) 113007,  
282 arXiv:hep-ph/0202176. doi:10.1103/PhysRevD.65.113007.
- 283 [7] J. M. Campbell, R. K. Ellis, and D. L. Rainwater, "Next-to-leading order QCD predictions  
284 for  $W + 2$ jet and  $Z + 2$ jet production at the CERN LHC," *Phys. Rev.* **D68** (2003) 094021,  
285 arXiv:hep-ph/0308195. doi:10.1103/PhysRevD.68.094021.
- 286 [8] R. K. Ellis, K. Melnikov, and G. Zanderighi, "Generalized unitarity at work: first NLO  
287 QCD results for hadronic  $W^+ 3$ jet production," arXiv:0901.4101.
- 288 [9] C. F. Berger et al., "Precise Predictions for  $W + 3$  Jet Production at Hadron Colliders,"  
289 arXiv:0902.2760.
- 290 [10] H. Baer, V. Barger, and G. Shaughnessy, "SUSY backgrounds to Standard Model  
291 calibration processes at the LHC," arXiv:0806.3745.
- 292 [11] **CMS** Collaboration, R. Adolphi et al., "The CMS experiment at the CERN LHC," *JINST*  
293 **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.
- 294 [12] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph,"  
295 *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.
- 296 [13] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05**  
297 (2006) 026, arXiv:hep-ph/0603175.
- 298 [14] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and  
299 shower evolution for top- quark production in hadronic collisions," *JHEP* **01** (2007) 013,  
300 arXiv:hep-ph/0611129.
- 301 [15] S. Kretzer, H. L. Lai, F. I. Olness, and W. K. Tung, "CTEQ6 parton distributions with  
302 heavy quark mass effects," *Phys. Rev.* **D69** (2004) 114005, arXiv:hep-ph/0307022.  
303 doi:10.1103/PhysRevD.69.114005.
- 304 [16] **CMS** Collaboration, G. L. Bayatian et al., "Measurement of the  $Zbb, Z \rightarrow \ell\ell$  cross section  
305 with  $100 \text{ pb}^{-1}$  of Early CMS data at the LHC,". CMS-EWK-08-001 (2008).
- 306 [17] **CMS** Collaboration, G. L. Bayatian et al., "CMS physics: Technical design report,"  
307 CERN-LHCC-2006-001.

- 308 [18] F. A. Berends, H. Kuijf, B. Tausk, and W. T. Giele, "On the production of a  $W$  and jets at  
309 hadron colliders," *Nucl. Phys.* **B357** (1991) 32–64.  
310 doi:10.1016/0550-3213(91)90458-A.
- 311 [19] CMS Collaboration, G. L. Bayatian et al., "Jet Reconstruction with charged tracks only in  
312 CMS," CMS-EWK-08-001 (2008).
- 313 [20] G. P. Salam and G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm," *JHEP*  
314 **05** (2007) 086, arXiv:0704.0292.
- 315 [21] E. Abouzaid and H. J. Frisch, "The Ratio of  $W + N$  jets to  $Z^0/\gamma^* + N$  jets versus  $N$  as a  
316 test precision test of the standard model," *Phys. Rev.* **D68** (2003) 033014,  
317 arXiv:hep-ph/0303088. doi:10.1103/PhysRevD.68.033014.
- 318 [22] GEANT4 Collaboration, S. Agostinelli et al., "GEANT4: A simulation toolkit," *Nucl.*  
319 *Instrum. Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.
- 320 [23] CDF Collaboration, A. A. Affolder et al., "Search for gluinos and scalar quarks in  $p\bar{p}$   
321 collisions at  $\sqrt{s} = 1.8$  TeV using the missing energy plus multijets signature," *Phys. Rev.*  
322 *Lett.* **88** (2002) 041801, arXiv:hep-ex/0106001.  
323 doi:10.1103/PhysRevLett.88.041801.
- 324 [24] CMS Collaboration, G. L. Bayatian et al., "Jet Energy correction using  $Z(\text{mumu})+\text{jet}$  PT  
325 balance," JME-EWK-09-009 (2009).
- 326 [25] D0 Collaboration, V. Abazov et al., "Measurement of the  $W$  Boson Mass with  $1 \text{ fb}^{-1}$  of D0  
327 Run II Data," D0 Note 5893-CONF (2009).
- 328 [26] M. Pivk and F. R. Le Diberder, "sPlot: a statistical tool to unfold data distributions," *Nucl.*  
329 *Instrum. Meth.* **A555** (2005) 356–369, arXiv:physics/0402083.  
330 doi:10.1016/j.nima.2005.08.106.
- 331 [27] CMS Collaboration, G. L. Bayatian et al., "CMS technical design report, volume II:  
332 Physics performance," *J. Phys.* **G34** (2007) 995–1579.  
333 doi:10.1088/0954-3899/34/6/S01.