
CMS Physics Analysis Summary

Contact: cms-pag-conveners-ewk@cern.ch

2011/03/18

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

The CMS Collaboration

Abstract

We present a study of jet production in association with W and Z bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV using the full 2010 data set collected by CMS corresponding to an integrated luminosity of 36 ± 4 pb⁻¹. We report the normalized inclusive rates of jets produced as well as the ratios $\sigma(V + \geq n \text{ jets}) / \sigma(V + \geq (n - 1) \text{ jets})$ with a jet threshold of 30 GeV. We present the first test of the Berends-Giele scaling at $\sqrt{s} = 7$ TeV.

1 Introduction

The study of jets produced in association with W and Z bosons provides a stringent and important test of perturbative QCD calculations. Next-to-leading order (NLO) predictions are available for $V + n$ jets, with n up to four for the W and three for the Z [1–4], but are only known with a precision varying from 10% up to 30% due to uncertainties on parton distribution functions (PDF) and on the perturbative nature of the calculations, which makes them dependent on the choice of renormalization and factorization scales.

The production of vector bosons with jets constitutes an important background in searches for new physics and for studies of the top quark, so a precise measurement of the cross section and an understanding of the jet and lepton kinematics is essential.

This note presents the results obtained with the full data sample recorded with the CMS experiment during 2010, amounting to an integrated luminosity of $36 \pm 4 \text{ pb}^{-1}$. In order to reduce the systematic uncertainties associated with the integrated luminosity measurement, the jet energy scale (JES) and the lepton reconstruction and trigger efficiencies, we normalize the yields according to the inclusive W and Z cross sections. We also measure the cross section ratios $\sigma(V + n \text{ jets})/\sigma(V + (n - 1) \text{ jets})$ where n stands for the inclusive number of jets. From the former we are able to test Berends-Giele scaling [5] in a quantitative manner. The CMS measurements of the inclusive W and Z cross sections are reported in [6]. The ATLAS Collaboration reported measurements of $V + n$ jets production in [7, 8].

2 The CMS Experiment

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the field volume are a silicon pixel and strip tracker, an electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL). Muons are detected in gas-ionization detectors embedded in the steel return yoke. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry.

The inner tracker measures charged particle trajectories in the pseudorapidity range $|\eta| \leq 2.5$. It consists of 1440 silicon pixel and 15148 silicon strip detector modules. It provides an impact parameter resolution of $15 \mu\text{m}$ and a transverse momentum (p_T) resolution of about 1% for charged particles with p_T around 40 GeV. The electromagnetic calorimeter consists of nearly 76000 lead tungstate crystals which provide coverage in pseudorapidity $|\eta| \leq 1.479$ in a cylindrical barrel region (EB) and $1.479 \leq |\eta| \leq 3.0$ in two endcap regions (EE). A preshower detector consisting of two planes of silicon sensors interleaved with a total of $3 X_0$ of lead is located in front of the EE. The ECAL has an ultimate energy resolution of better than 0.5% for unconverted photons with transverse energies above 100 GeV. The energy resolution is 3% or better for the range of electron energies relevant for this analysis. The hadronic calorimeter is a sampling device with brass as passive material and scintillator as active material. The combined calorimeter cells are grouped in projective towers of granularity $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ at central rapidities and 0.175×0.175 at forward rapidities. Muons are detected in the pseudorapidity window $|\eta| \leq 2.4$, with detection planes based on three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. A high- p_T muon originating from the interaction point produces track segments in typically three or four muon stations. Matching these segments to tracks measured in the inner tracker results in a p_T resolution between 1% and 2% for p_T values up to 100 GeV. The first level (L1) of the CMS trigger system, composed of custom hardware processors, is designed to select the most interesting events in less than 1 ms using information from the calorimeters and muon detectors. The High Level Trigger (HLT)

processor farm further decreases the event rate to a few hundred hertz, before data storage. A more detailed description of CMS can be found elsewhere [9].

3 Data and Simulation Samples

A hardware-based trigger system selects electrons with an associated energy deposit in the ECAL of at least 5 GeV or 8 GeV, depending on the luminosity conditions, and muons with a transverse momentum exceeding 4 GeV. The events are then filtered in the online cluster with algorithms that have evolved following the rapid rise of the LHC luminosity. For each data taking period, we retained the electrons that passed the trigger with the lowest p_T threshold that was not prescaled. The largest sample of electrons were collected with an online requirement $p_T \geq 17$ GeV. We retained muons based on the logical “OR” of those unprescaled triggers with the lowest p_T threshold. The threshold of $p_T \geq 15$ GeV pertains to most of the data collected.

Simulated data samples are needed as a representation of the theoretical predictions, and also as the basis for unfolding the jet multiplicity distributions. Backgrounds are estimated with data-driven techniques, as explained below.

Samples of events with a W or a Z boson are generated with MADGRAPH [10] interfaced with PYTHIA [11]. MADGRAPH produces parton-level events with a vector boson and up to four jets on the basis of a matrix-element calculation. This sample serves as the base line for comparisons with data. In addition, a sample is generated with PYTHIA for which the parton shower is modified in such a way that the hardest emission is modeled using the exact matrix element calculation for one additional real emission. MADGRAPH is expected to be more accurate than PYTHIA for large jet multiplicities. Events with a $t\bar{t}$ quark pair are generated with MADGRAPH.

The full list of simulated samples is given in Table 1. Whenever available, the NNLO or NLO cross section [12–14] is used to normalize the simulation samples compared to the data distributions.

Generated events are processed through the full detector simulation based on GEANT4 [15, 16], followed by a detailed trigger emulation and the CMS event reconstruction. Several minimum-bias events are superimposed to the hard interactions to simulate event pile-up according to the distribution of multiple proton-proton collisions observed during the 2010 data taking period. A signal sample without pile-up is used for purposes of comparison. The PYTHIA parameters for the underlying event have been set according to the “Z2” tune, a modification of the “Z1” tune described in [17]. Comparisons are also made to the “D6T” tune [18].

4 Signal Selection and Jet Rates

4.1 Signal Selection in the Electron Channel

Electrons are selected following the standard established by the measurement of the inclusive W and Z cross sections [6].

A fiducial region in the ECAL is defined which excludes electrons close to the barrel/endcap transition, and electrons in the first endcap trigger tower which lies in the shadow of cables and services exiting between the barrel and endcap. A fiducial region cut is applied on the position of the ECAL supercluster which is required to have $|\eta| < 2.5$, with $1.4442 < |\eta| < 1.566$ excluded.

Electron candidates are further identified based on the spatial matching between the ECAL

Table 1: Summary of simulated datasets for the various signal and background processes used in this analysis.

Generator	Process	Kinematic cuts (in GeV, $c = 1$)	σ (pb)
MADGRAPH	$W \rightarrow \ell\nu$	no cuts	3.1×10^4 (NNLO)
MADGRAPH	$Z \rightarrow \ell^+\ell^-$	$m_{\ell\ell} > 50$	3.0×10^1 (NNLO)
MADGRAPH	$t\bar{t}$	no cuts	1.6×10^2 (NLO)
MADGRAPH	single top tW channel	no cuts	1.1×10^1 (LO)
MADGRAPH	single top s and t channels	no cuts	3.5 (NLO)
PYTHIA	$W \rightarrow e\nu$	$ \eta_e < 2.7$	8.2×10^3 (NNLO)
PYTHIA	$W \rightarrow \mu\nu$	$ \eta_\mu < 2.5$	7.7×10^3 (NNLO)
PYTHIA	$W \rightarrow \tau\nu$	no cuts	1.0×10^4 (NNLO)
PYTHIA	$Z \rightarrow \ell^+\ell^-$	$m_{\ell\ell} > 20$	5.0×10^3 (NNLO)
PYTHIA	Inclusive μ QCD	$p_T > 20, p_T^\mu > 10, \eta_\mu < 2.5$	3.4×10^5 (LO)
PYTHIA	EM-enriched QCD	$20 < p_T < 170$	5.4×10^6 (LO)
PYTHIA	$b/c \rightarrow e$	$20 < p_T < 170$	2.6×10^5 (LO)
PYTHIA	γ +jet	no cuts	8.5×10^7 (LO)

supercluster and the track in the η and ϕ coordinates, the supercluster shower spread along the η direction and the energy leakage in the HCAL detector.

In order to reduce the contamination from fake electrons and hadronic decays, we build three isolation variables based, respectively, on the sums of tracks p_T in the tracker, of individual channel transverse energies in the ECAL, and in the HCAL. The sums are computed in regions of $\Delta R < 0.3$ around either the supercluster position or the track direction at the vertex with an inner exclusion region which removes the electron “footprint”, resulting from showering in the tracker and ECAL. In the case of the HCAL excludes the region summed for the HCAL leaking variable which is calculated using the HCAL energy found within $\Delta R < 0.15$ of the ECAL seed cluster, divided by the seed cluster energy.

The sums are divided by the electron E_T and a cut is applied on these ratios.

A large fraction of misidentified electrons stem from converted photons. Three discriminants have been applied to improve the purity of the electron sample: the absence of track hits in tracker layers between the vertex and the first measured hit of the track, the presence of a conversion partner, and the measurement of a significant transverse impact parameter.

V + jets signal selection begins with the identification of a charged lepton within the fiducial region and with $p_T > 20$ GeV. This so-called “leading lepton” must match the object which triggered the event readout, and it must also pass tight quality requirements which correspond to a lepton efficiency of roughly 80% as evaluated with a MADGRAPH +PYTHIA simulated sample (as described in Section 5, the efficiency values subsequently used for the cross sections determination are measured on data). We look for an additional charged lepton, called the “second leading lepton” which is required to have $p_T > 10$ GeV and to pass looser quality requirements, *i.e.* which correspond to a lepton efficiency of roughly 95% as evaluated with a MADGRAPH +PYTHIA simulated sample.

If such a second leading lepton is found, and its invariant mass with the first leading lepton is between 60 GeV and 120 GeV, then the event is assigned to the Z + jets sample. If the such leading lepton is not found the event is assigned to the W + jets sample, thereby ensuring that there is no overlap between the two samples.

For the W + jets sample we compute the missing transverse energy \cancel{E}_T using the particle flow

(PF) algorithm [19]. Although we do not place any direct requirement on \cancel{E}_T , we do use it to calculate the transverse mass, $M_T = \sqrt{2p_T \cancel{E}_T (1 - \cos \Delta\phi)}$, where $\Delta\phi$ is the angle in the xy -plane between the lepton p_T and the \cancel{E}_T direction and we select events where $M_T > 20$ GeV to avoid a problematic region at low M_T containing essentially no signal.

4.2 Signal Selection in the Muon Channel

The muon reconstruction and identification are identical to that used for the measurement of the W and Z cross sections [6, 20].

We define a relative isolation variable, $I_{\text{iso}} = \sum(p_T^{\text{track}} + p_T^{\text{HCAL}} + p_T^{\text{ECAL}}) / p_T^\mu$, which includes the p_T for tracks, ECAL and HCAL towers in a cone $\Delta R < 0.3$ around the muon direction. The muon and its energy deposits are excluded from the sum. A muon is isolated if $I_{\text{iso}} < 0.15$. The $V + \text{jets}$ event selection starts from requiring the presence of a high p_T isolated muon in the region $|\eta| < 2.1$ with $p_T > 20$ GeV. It must be a high quality muon with an impact parameter cut, $|d_{xy}| < 2$ mm to suppress cosmic ray muon background. As for the electron channel we then search for a second leading muon ($p_T > 10$ GeV) accepted in the range $|\eta| < 2.5$ such that the di-muon invariant mass lies within the region 60 GeV to 120 GeV. If such second leading muon is (is not) found the event is assigned to the $Z + \text{jets}$ ($W + \text{jets}$) sample. For the $W + \text{jets}$ sample selection we require $M_T > 20$ GeV.

4.3 Jet Rates

Jets are reconstructed from the particle collection created with the particle flow algorithm and are formed with the anti- k_T clustering algorithm [21] with a size parameter of $R = 0.5$. Jet energy corrections (JEC) are applied to improve the accuracy of the jet p_T measurement and to flatten the jet energy response as a function of η and p_T [22].

We require $|\eta| < 2.4$ so that the jets fall within the tracker acceptance. Jets are required to pass identification criteria which eliminate jets originating or being seeded by noisy channels in the calorimeter[23].

Pile-up, *i.e.* overlapping minimum-bias events coming from the same bunch crossing and the underlying event have an effect on jet counting by contributing additional energy to the measured jet energy and therefore “promoting” jets from below to above the p_T threshold for counting. This effect is taken into account by evaluating an event energy density not related to the hard interaction activity, as $\rho = \text{median}(p_t^{\text{jet}} / \text{Area}(\text{jet}))$, assumed to be a constant in the event. This amount is subtracted from each jet [24, 25]. The contamination from real di-jet events is found to be negligible for the jet threshold used in this measurement and it is not accounted for.

Electrons can be reconstructed as jets, or they can overlap a hadronic jet. Therefore, jets which fall within $\Delta R < 0.3$ from an electron produced in a W or Z decay are not included in the jet count. Muons can also overlap with a jet and we rely on the PF algorithm to exclude muons from jet clustering.

One of the most important backgrounds in the W sample comes from $t\bar{t}$ events that contain two b -quark jets. We count the number of b -tagged jets, $n_{b\text{-jet}}$, using the “Track Counting High Efficiency” b -tagging algorithm for the electron channel and the “Track Counting High Purity” algorithm for the muon channel [26]. The $n_{\text{jet}}^{b\text{-tagged}}$ variable is then used in the fitting method to extract the number of signal events in data, as described in Section 6.

The observed, uncorrected transverse momentum distributions for the leading jet are shown

in Figs. 1 and 2. The data is in good agreement with the MADGRAPH predictions normalized to the NNLO cross sections and the luminosity. For the W sample, we have required $M_T > 50$ GeV in order to reduce backgrounds.

After passing the selection, we assign events to exclusive bins of jet multiplicity by counting the number of jets in the events with $p_T > 30$ GeV. The observed, uncorrected distributions of the exclusive numbers of reconstructed jets in the W and Z samples are shown in Figs. 3 and 4, respectively. The distributions from simulation, including background processes, are also shown. Overall, a good agreement is found up to $n = 6$ jets.

5 Acceptance and Efficiency

In order to provide model-independent results, we do not correct for the acceptance, but rather quote the results within acceptance as defined by the lepton and jet fiducial and kinematic cuts given above.

The efficiencies for lepton reconstruction, identification, isolation and trigger are obtained from data by means of a tag-and-probe method performed on $Z/\gamma^* + \text{jets}$ data samples. The tag-and-probe sample for the measurement of a given efficiency contains events selected with two lepton candidates. One lepton candidate, called the “tag”, satisfies all selection requirements including being matched to a trigger object. The other lepton candidate, called the “probe” is selected with criteria that depend on the efficiency being measured. The invariant mass of the tag and probe lepton candidates must fall in the range [60–120] GeV. The signal yields are obtained for two exclusive subsamples of events in which the probe lepton passes or fails the selection criteria considered. Fits are performed to the invariant-mass distributions of the pass and fail subsamples, including a term that accounts for the background. The measured efficiency is deduced from the relative level of signal in the pass and fail subsamples. A systematic error is obtained from varying the line shape to describe the Z signal shape. The lepton selection efficiency is the product of three components: 1) the reconstruction efficiency, 2) the identification and isolation efficiency, and 3) the trigger efficiency. Each of these efficiencies are calculated as a function of the jet multiplicity in the event.

For electrons we find that the efficiencies are roughly 70% (60%) for the $W + \text{jets}$ ($Z/\gamma^* + \text{jets}$) signal events with variations of a few percent across different jet multiplicity bins.

For muons, the efficiencies are measured as a function of p_T and η in the highest statistics bins ($n = 0$ and $n = 1$). Due to the isolation requirement, the efficiencies also exhibit a significant dependence on the observed jet multiplicity. To account for this properly, since the statistical precision in the bins with $n > 1$ is insufficient, the efficiencies for these bins are extrapolated by using the p_T and η shape of the $n = 1$ bin with a scale determined by the average value, for each bin of the trigger efficiency, which is found to have the largest dependence on the jet multiplicity. We find an average efficiency close to 82% for the leading p_T muon and of above 90% for the second leading muon.

6 Signal Extraction

The signal yield is estimated using an extended likelihood fit to $M_{\ell+\ell-}$ for the $Z + \text{jets}$ sample and to M_T for the $W + \text{jets}$ sample; 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

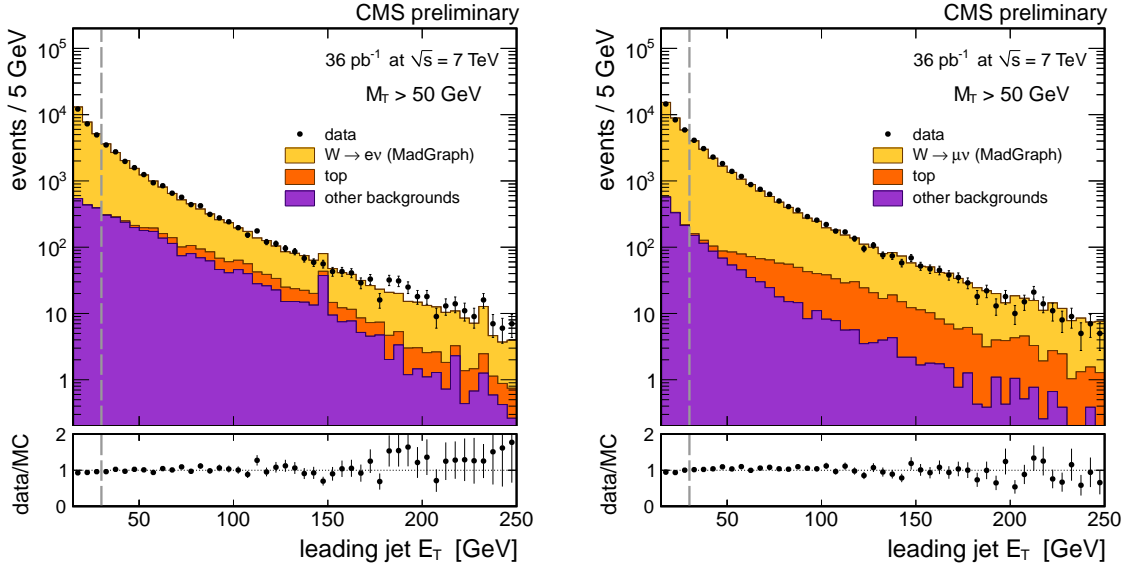


Figure 1: Distributions of the uncorrected p_T for the leading jet in the $W + 1$ jet sample for the electron channel (left) and for the muon channel (right). The ratio between the data and the simulation is also shown. The line at $p_T = 30$ GeV corresponds to the threshold imposed for counting jets.

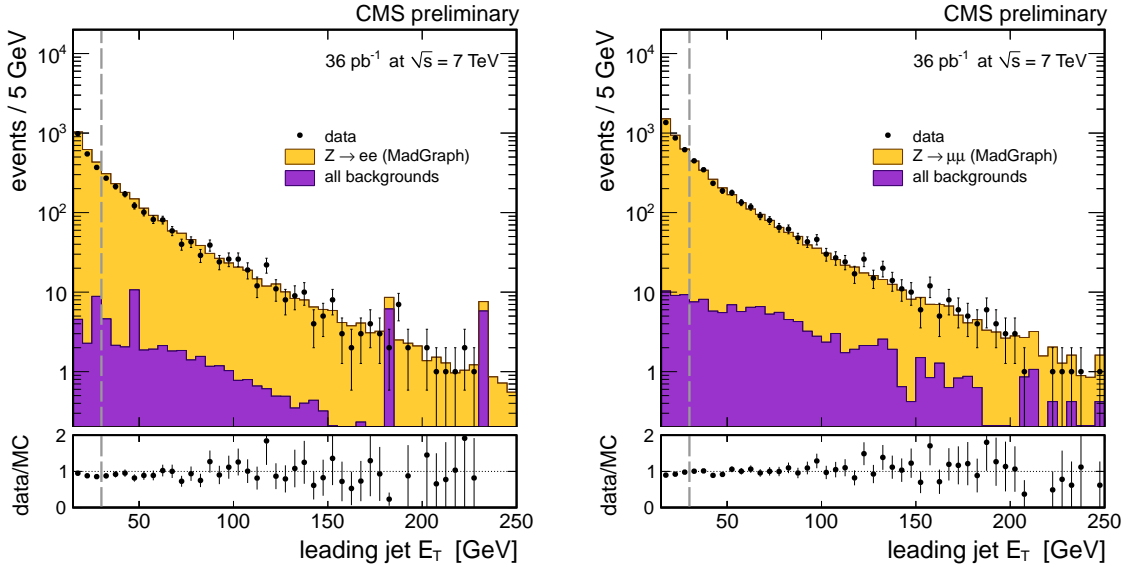


Figure 2: Distributions of the uncorrected p_T for the leading jet in the $Z + 1$ jet samples, similar to Fig. 1.

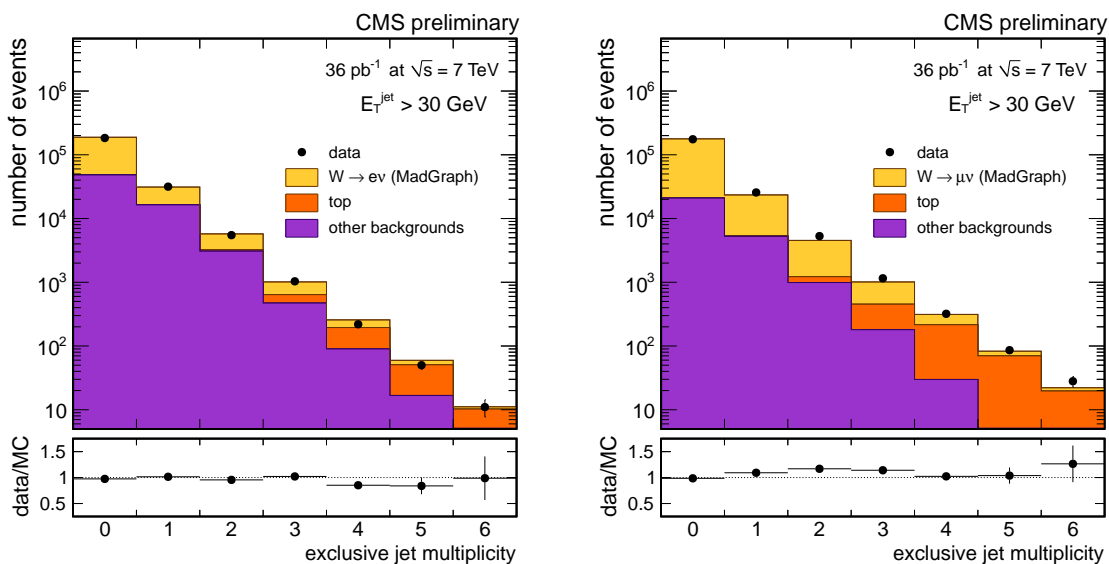


Figure 3: Exclusive number of reconstructed jets in events with $W \rightarrow e\nu$ (left) and $W \rightarrow \mu\nu$ (right). The histograms represent the expectations based on simulated events. These distributions have not been corrected for detector effects or selection efficiency.

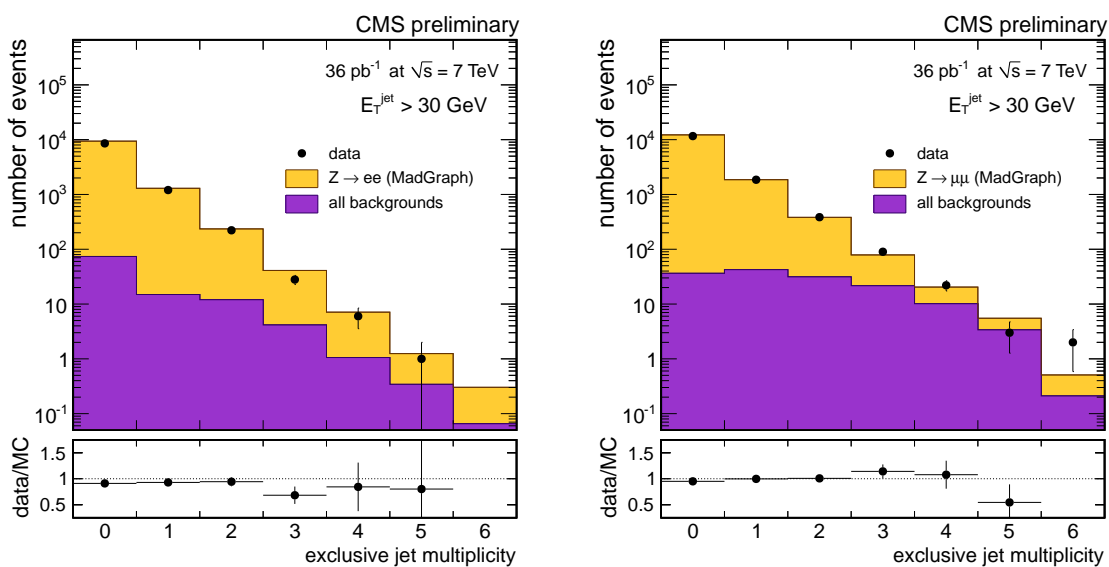


Figure 4: Exclusive number of reconstructed jets in events with $Z \rightarrow e^+e^-$ (left) and $Z \rightarrow \mu^+\mu^-$ (right). The histograms represent the expectations based on simulated events. These distributions have not been corrected for detector effects or selection efficiency.

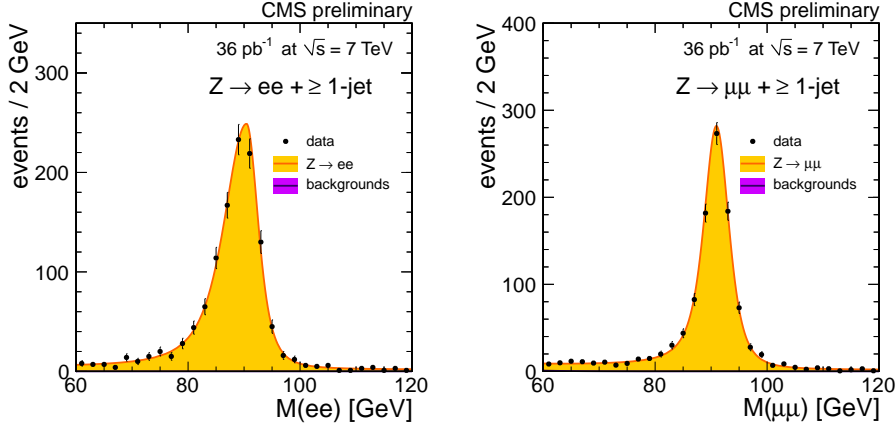


Figure 5: Di-lepton mass fit for the $Z + 1$ jet samples, in the electron channel (left) and the muon channel (right). The background is very low, therefore it is hardly visible in the figure.

by $t\bar{t}$ and $W + \text{jets}$, is small and does not produce a peak in the $M_{\ell^+\ell^-}$ distribution, so we fit the $M_{\ell^+\ell^-}$ distribution to 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. We perform a two-dimensional fit to the M_T distribution and the number of b -jets, $n_{b\text{-jet}}$. The M_T distribution allows the statistical separation of the signal from the non-peaking backgrounds, while $n_{b\text{-jet}}$ distinguishes the signal and the background from $t\bar{t}$.

The fits are done in exclusive jet multiplicity bins for $n \leq 3$; jet counting is instead done inclusively for the last bin of jet multiplicity, *i.e.* $n \geq 4$. Examples of fits for $Z + 1$ jet are shown in Figure 5. Figures 6 and 7 show fits in M_T and $n_{b\text{-jet}}$ projections for $W + n$ jets ($n=1$ and $n=3$) channel. The presence of the top background is evident comparing the $n = 1$ and $n = 3$ exclusive multiplicity bins.

In the electron channel, observed exclusive $V + \text{jets}$ rates are corrected for electron efficiencies as discussed in Section 5. In the muon channel, efficiencies depend on the lepton p_T and η and on the jet multiplicity. In order to account for these variations, every event is assigned a weight and the fit is performed to a weighted distribution.

A second fit is performed in order to test Berends-Giele scaling and measure the parameters that characterize it. Events are assigned to exclusive jet multiplicity bins (inclusively for $n \geq 4$) and the yields are fit with the assumption that they conform to a scaling function:

$$C_n = \frac{\sigma_n}{\sigma_{n+1}} \quad (1)$$

where $\sigma_n = \sigma(V + n \text{ jets})$. Leading-order calculations would predict $C_n = \alpha$, where the constant α is proportional to the inverse of the strong coupling constant, α_s^{-1} . Next-to-Leading-Order effects will violate this simple proportionality, as will phase space effects for high n . We introduce a second parameter, β , to allow for a deviation from a simple constant scaling law:

$$C_n = \alpha + \beta n. \quad (2)$$

Due to the different production kinematics of the $n = 0$ sample, where no reconstructed jets are recoiling against the vector boson, the scaling expressed in Eq. (1) is not expected to hold, so we do not include the $n = 0$ sample in the fit.

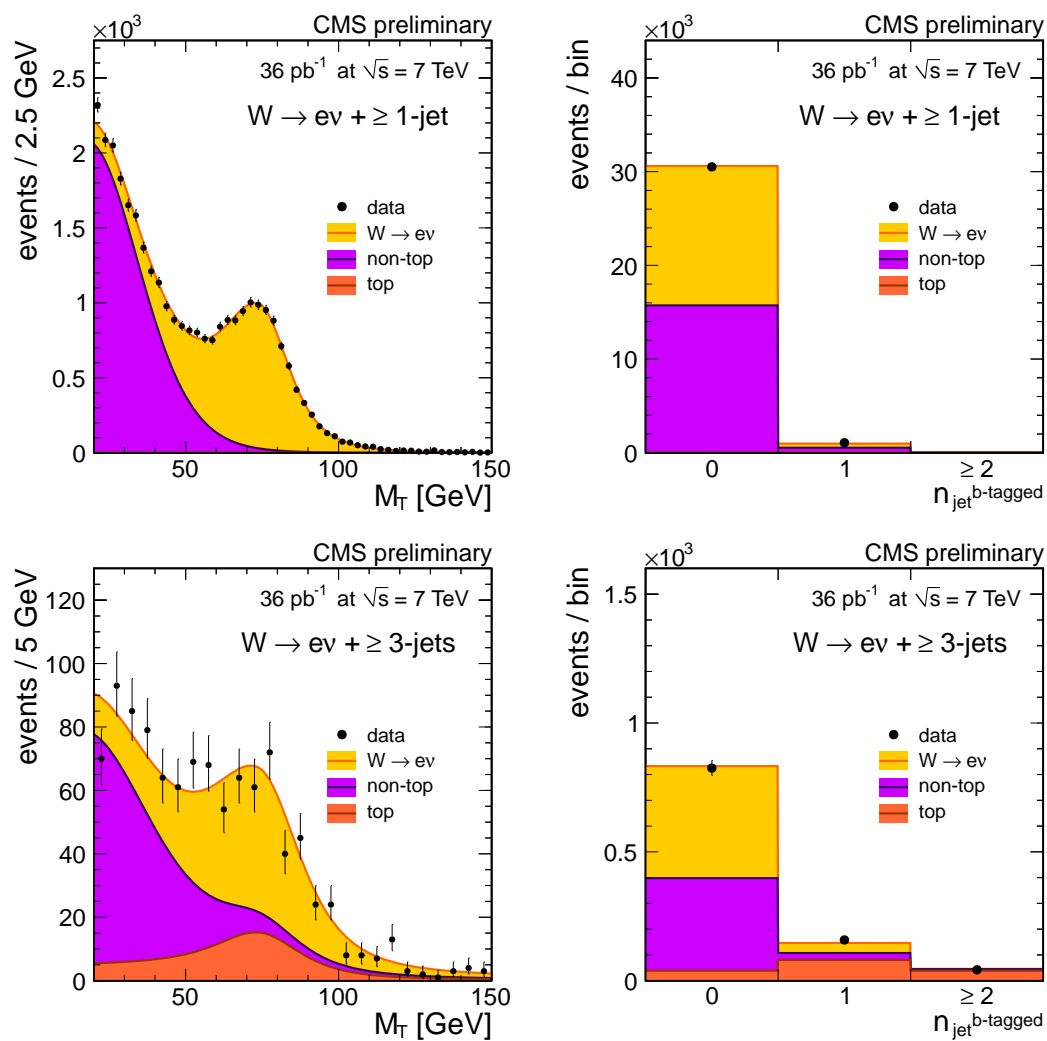


Figure 6: Fit results for the $W(ev) + n$ jet sample with $n = 1$ (upper row) and $n = 3$ (lower row). On the left we show the M_T projection, and on the right $n_{b\text{-jet}}$.

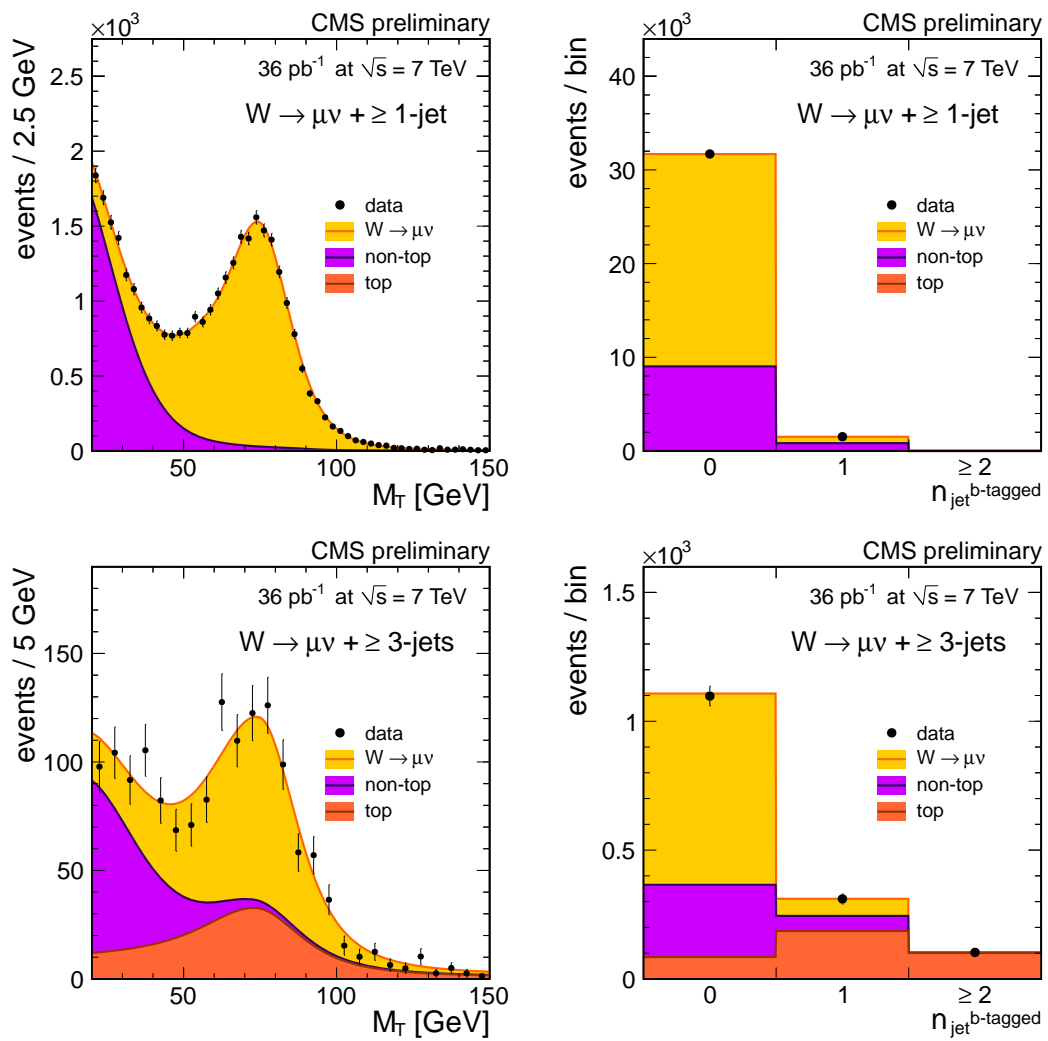


Figure 7: Same as Fig. 6, for the muon channel.

7 Unfolding

In order to estimate the scaling rule of the jets at the *particle level*, we apply an unfolding procedure that removes the effects of imperfect jet energy resolution and reconstruction efficiency. The migration matrix, which relates a number n' of produced jets at particle level to an observed number n of reconstructed jets, is derived from simulated samples of $Z + \text{jets}$ and $W + \text{jets}$. The unfolding procedure takes into account the statistical uncertainties in the measurement.

We employ two well-known unfolding methods. Our base line method is the “singular value decomposition” (SVD) method [27]. As a cross check, we apply the iterative or “Bayesian” method [28]. Both algorithms require a regularization parameter to prevent the statistical fluctuations in the data from appearing as structure in the unfolded distribution. The regularization parameter determines the relative weight placed on the data, compared to the training sample truth. For the SVD method, the regularization parameter is chosen to be $k_{SVD} = 5$, corresponding to the number of bins. This case corresponds to an inversion of the response matrix, and studies show this gives the most accurate uncertainty estimate. For the iterative algorithm, the regularization parameter $k_{Bayes} = 4$ is used, as suggested by the RooUnfold authors. This parameter specifies the number of iterations, starting with the training sample truth, *i.e.*, iteration zero.

The response matrices are derived from simulated events in which the leptons and jets pass the selection. The simulation is based on the MADGRAPH sample with the Z2 tune, and there is no pile-up.

8 Systematic Uncertainties

One of the dominant sources of systematic uncertainties in the $W/Z + \text{jets}$ measurements is the determination of the jet energy, which affects the jet counting. We consider three sources of uncertainty related to the jet-counting:

- corrections applied to the measured jet energy to account for the detector response and inhomogeneities, jet energy scale. Corrections are derived with data driven methods and are available as a function of p_T and η [22].
- the difference between the flavor composition of jets in $W/Z + \text{jets}$ events and the flavor composition of the jet sample used to extract the corrections, is accounted for by an additional 2% uncertainty on the jet energy.
- the pile-up subtraction method used is found to systematically remove 500 MeV to jets in events without pile-up. The same amount has been added to the systematic uncertainty on the jet energy.

All above uncertainties on the jet energy have been added in quadrature and their effect is evaluated on the jet multiplicity distribution using simulation. Compatible results have been found in all the channels, for both W and Z events, and they are shown in Table 2. In addition for W events, the effect of the mismeasurement on \cancel{E}_T and therefore on M_T due to mis-measured jet energy has been evaluated in the fitting procedure and the results are also shown in Table 2.

Uncertainty in jet energy resolution also affects jet multiplicity. The resolution is known to be underestimated in the simulation by about 10%, with a similar uncertainty [29]. The effect of this uncertainty on the jet multiplicity has been studied in simulated $W + \text{jets}$ events and found to be below 2%.

The pile-up subtraction was also tested comparing the jet multiplicity in two simulated signal samples, one without pile-up and one with pile-up and pile-up subtraction applied. The difference, due to residual effects and to contamination from di-jet events, is found to be below 5%.

In Table 2 all the sources of systematic uncertainty in the jet counting are summarized.

In Tables 3 and 4 the relative systematic uncertainties on jet rates are shown, including those due to the signal extraction procedure and the selection efficiency. While the systematic uncertainty in the jet counting is correlated among the different jet multiplicities, all other uncertainties are uncorrelated between different jet multiplicities. The relative statistical uncertainty is also shown for comparison.

All statistical and systematic uncertainties are propagated in the unfolding procedure. Finally, to estimate uncertainties in the unfolding procedure itself, we calculated the difference in unfolding using the Bayes algorithm versus the SVD algorithm and using two different simulations, MADGRAPH and PYTHIA, for the unfolding matrix, and two different tunes, Z2 and D6T for the unfolding matrix. The resulting uncertainties are shown together with final results in the next section.

9 Results

From the unfolded exclusive jet multiplicity distributions we derive inclusive jet multiplicities and calculate two sets of ratios. The first set of ratios is $\sigma(V + n \text{ jets})/\sigma(V)$, where $\sigma(V)$ is the inclusive cross section, see Tables 5 and 6 and Figs. 8–11. The systematic uncertainties associated with the JES and the unfolding are shown as error bands. Due to the jet threshold $p_T > 30 \text{ GeV}$, the sensitivity to the underlying event is rather small. For a large number of jets, the PYTHIA pure parton shower simulation fails to describe the data, while the MADGRAPH simulation agrees well, as expected.

The second set of ratios is $\sigma(V + n \text{ jets})/\sigma(V + (n - 1) \text{ jets})$, reported in Tables 7 and 8. In the table the statistical uncertainty quoted includes only the statistical errors on the fit results. The systematic uncertainty that is combined with the statistical comes from the systematics on the fit and the statistical and systematic uncertainty on the efficiency. The JES uncertainty was calculated by scaling the numbers from the fits either higher or lower. Those numbers were then unfolded and the difference in the output from the actual fit value is quoted as the JES

Table 2: Systematic uncertainties on the jet counting in W and Z events. In each row the uncertainties are fully correlated or anti-correlated, depending on the sign.

Systematic uncertainty on jet counting [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet Energy Scale	∓ 1	± 6	$+9$ -8	$+12$ -11	$+14$ -13
E_T (W only)	$+0.6$ -0.7	$+3.5$ -3.1	$+4.5$ -3.9	$+5.2$ -4.5	$+6$ -5
Jet Energy Resolution		$+0.6$ -0.5	$+0.8$ -0.7	$+1.0$ -0.9	$+1.1$ -1.0
Pile-up	∓ 5	± 5	± 5	± 5	± 5
Total in W events	∓ 5	± 8	$+11$ -10	$+14$ -12	$+16$ -15
Total in Z events	∓ 5	± 8	± 10	$+13$ -12	$+15$ -14

uncertainty. The uncertainty from the different unfolding algorithms and from changing the tune or generator used in the response matrix are also included in the tables.

Finally, we show the results of the fit for α and β in our treatment of Berends-Giele scaling, for both cases $W + \text{jets}$ and $Z + \text{jets}$ in Fig. 12. The results are given in the (α, β) plane and are compared to the results obtained from the MADGRAPH sample with the Z2 tune in the same lepton and jet acceptance of the selection. The electron and muon expected values differ mostly because of the $\Delta R > 0.3$ cut between the jets and the leptons, which is applied only in the electron channel. When the same cut is applied also in the muon channel, the two values

Table 3: Relative systematic and statistical uncertainties on the measured jet multiplicity in W events.

Uncertainties on jet rate in $W \rightarrow e\nu$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Lepton efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Signal extraction		± 0.1	± 0.4	± 2.9	± 8.5
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 0.3	± 1.0	± 2.4	± 7.5	± 22
Uncertainties on jet rate in $W \rightarrow \mu\nu$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Lepton efficiency	± 3	± 6	± 4	± 10	± 17
Signal extraction		± 0.1	± 0.4	± 2.9	± 8.5
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+19}_{-17}$	± 26
Statistical uncertainty	± 0.2	± 0.8	± 2.3	± 6.5	± 27

Table 4: Relative systematic and statistical uncertainties on the measured jet multiplicity in Z events.

Uncertainties on jet rate in $Z \rightarrow e^+e^-$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 1.0	± 3.0	± 8.0	± 20	± 47
Uncertainties on jet rate in $Z \rightarrow \mu^+\mu^-$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 1.1	± 2.7	± 5.2	± 18	± 35

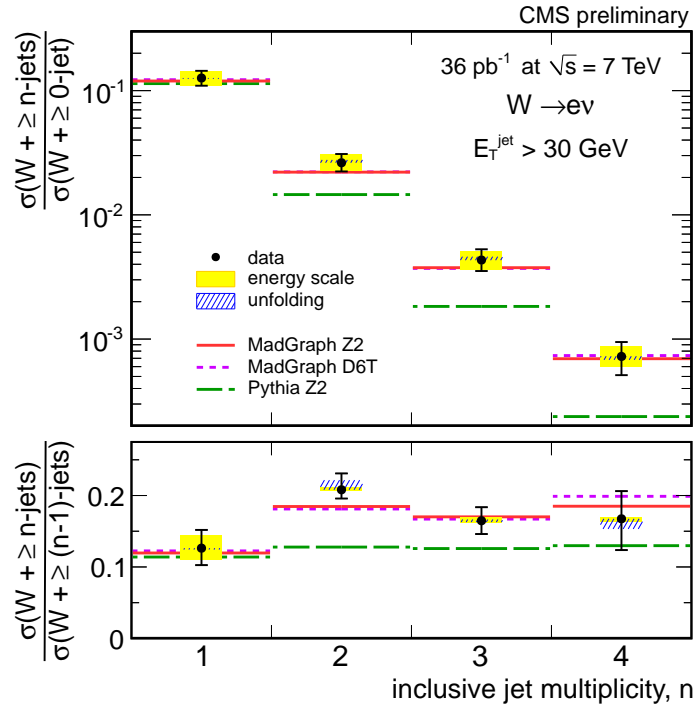


Figure 8: The ratio $\sigma(W + n \text{ jets})/\sigma(W)$ in the electron channel compared to expectations from MADGRAPH and PYTHIA.

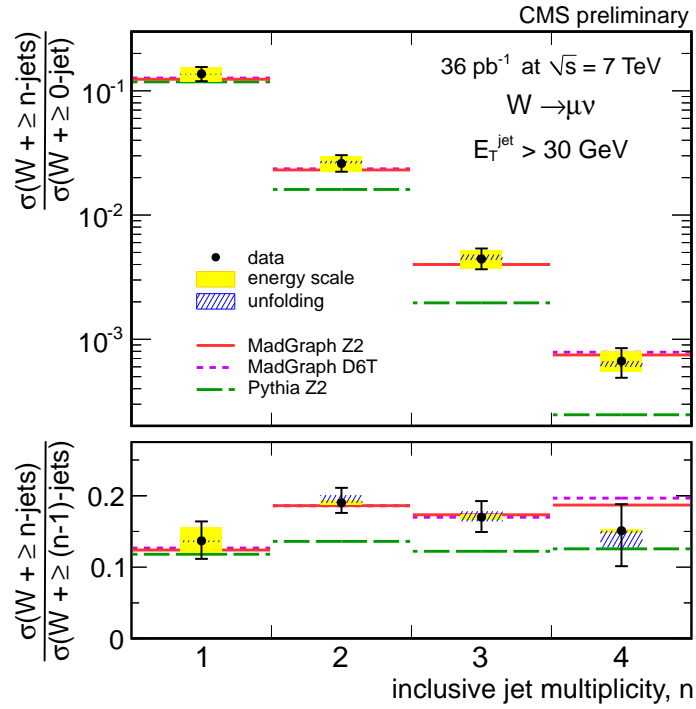


Figure 9: The ratio $\sigma(W + n \text{ jets})/\sigma(W)$ in the muon channel compared to expectations from MADGRAPH and PYTHIA.

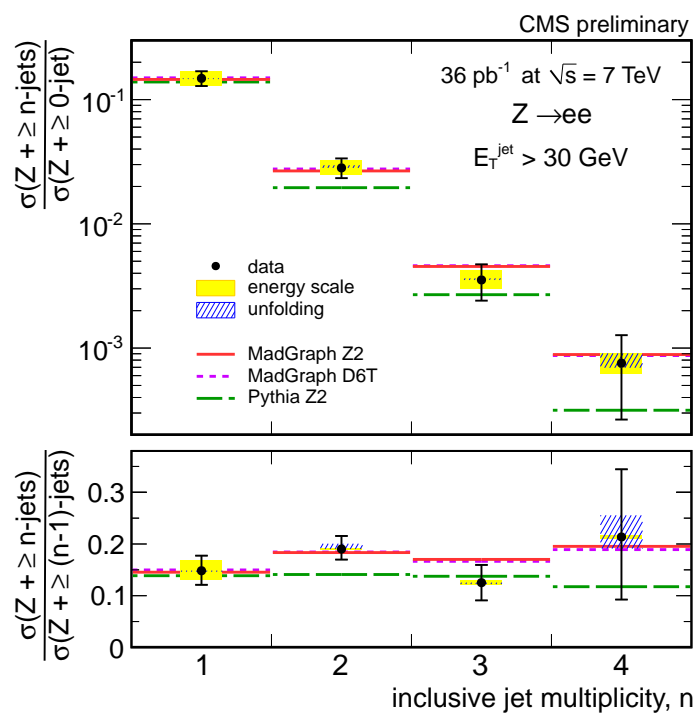


Figure 10: The ratio $\sigma(Z + n \text{ jets})/\sigma(Z)$ in the electron channel compared to expectations from MADGRAPH and PYTHIA.

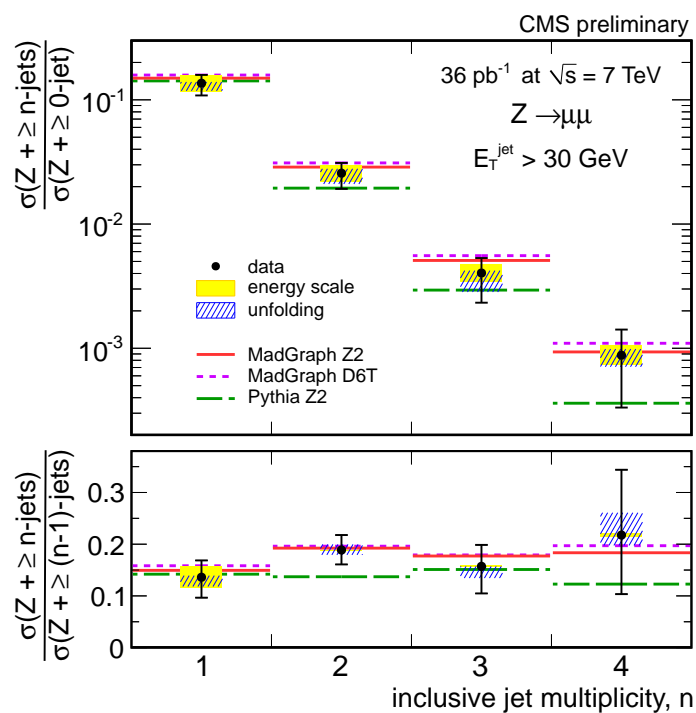


Figure 11: The ratio $\sigma(Z + n \text{ jets})/\sigma(Z)$ in the muon channel compared to expectations from MADGRAPH and PYTHIA.

are much closer. If the value of the ΔR cut is further increased to 0.5 they become compatible with each other, which may be related to some difference in the final state radiation. The ellipses correspond to 68% C.L. if statistical uncertainty only is considered: the arrows show the effect on the central value due to the most important sources of systematic uncertainty. Our measurements agree well in the $Z + \text{jets}$ channels, and fairly well in the $W + \text{jets}$ channel. The β parameter is within one standard deviation from zero for the $W + \text{jets}$ case and within 0.5 standard deviation for the $Z + \text{jets}$. The values for $W + \text{jets}$ and $Z + \text{jets}$ agree with one another, as expected in the standard model. The results and associated systematic uncertainties for the Berends-Giele scaling parameters are also reported in Tables 9 and 10. The data 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.

Table 5: $\sigma(W + \geq n \text{ jets})/\sigma(W)$, the jet multiplicities normalized to the inclusive cross section.

num jets	σ ratio	stat	stat + fit and efficiency	JES	unfolding
electron channel					
$\geq 1 / \geq 0$ jets	0.126	0.001	0.004	+0.018 -0.016	+0.000 -0.002
$\geq 2 / \geq 0$ jets	0.026	0.000	0.002	+0.004 -0.004	+0.001 -0.000
$\geq 3 / \geq 0$ jets	0.0043	0.0002	0.0005	+0.0008 -0.0007	+0.0003 -0.0000
$\geq 4 / \geq 0$ jets	0.0007	0.0000	0.0002	+0.0002 -0.0001	+0.0000 -0.0000
muon channel					
$\geq 1 / \geq 0$ jets	0.137	0.001	0.007	+0.019 -0.017	+0.000 -0.002
$\geq 2 / \geq 0$ jets	0.026	0.000	0.001	+0.004 -0.004	+0.001 -0.000
$\geq 3 / \geq 0$ jets	0.0044	0.0001	0.0005	+0.0008 -0.0007	+0.0004 -0.0001
$\geq 4 / \geq 0$ jets	0.0007	0.0000	0.0002	+0.0001 -0.0001	+0.0000 -0.0001

Table 6: $\sigma(Z + \geq n \text{ jets})/\sigma(Z)$, the jet multiplicities normalized to the inclusive cross section.

num jets	σ ratio	stat	stat + fit and efficiency	JES	unfolding
electron channel					
$\geq 1 / \geq 0$ jets	0.148	0.003	0.007	+0.020 -0.019	+0.000 -0.002
$\geq 2 / \geq 0$ jets	0.028	0.001	0.003	+0.004 -0.004	+0.001 -0.000
$\geq 3 / \geq 0$ jets	0.0035	0.0005	0.0010	+0.0007 -0.0005	+0.0001 -0.0000
$\geq 4 / \geq 0$ jets	0.0008	0.0000	0.0005	+0.0002 -0.0001	+0.0002 -0.0001
muon channel					
$\geq 1 / \geq 0$ jets	0.136	0.003	0.009	+0.022 -0.020	+0.003 -0.018
$\geq 2 / \geq 0$ jets	0.026	0.001	0.003	+0.004 -0.004	+0.002 -0.005
$\geq 3 / \geq 0$ jets	0.0040	0.0005	0.0011	+0.0007 -0.0006	+0.0002 -0.0012
$\geq 4 / \geq 0$ jets	0.0009	0.0000	0.0005	+0.0002 -0.0001	+0.0001 -0.0002

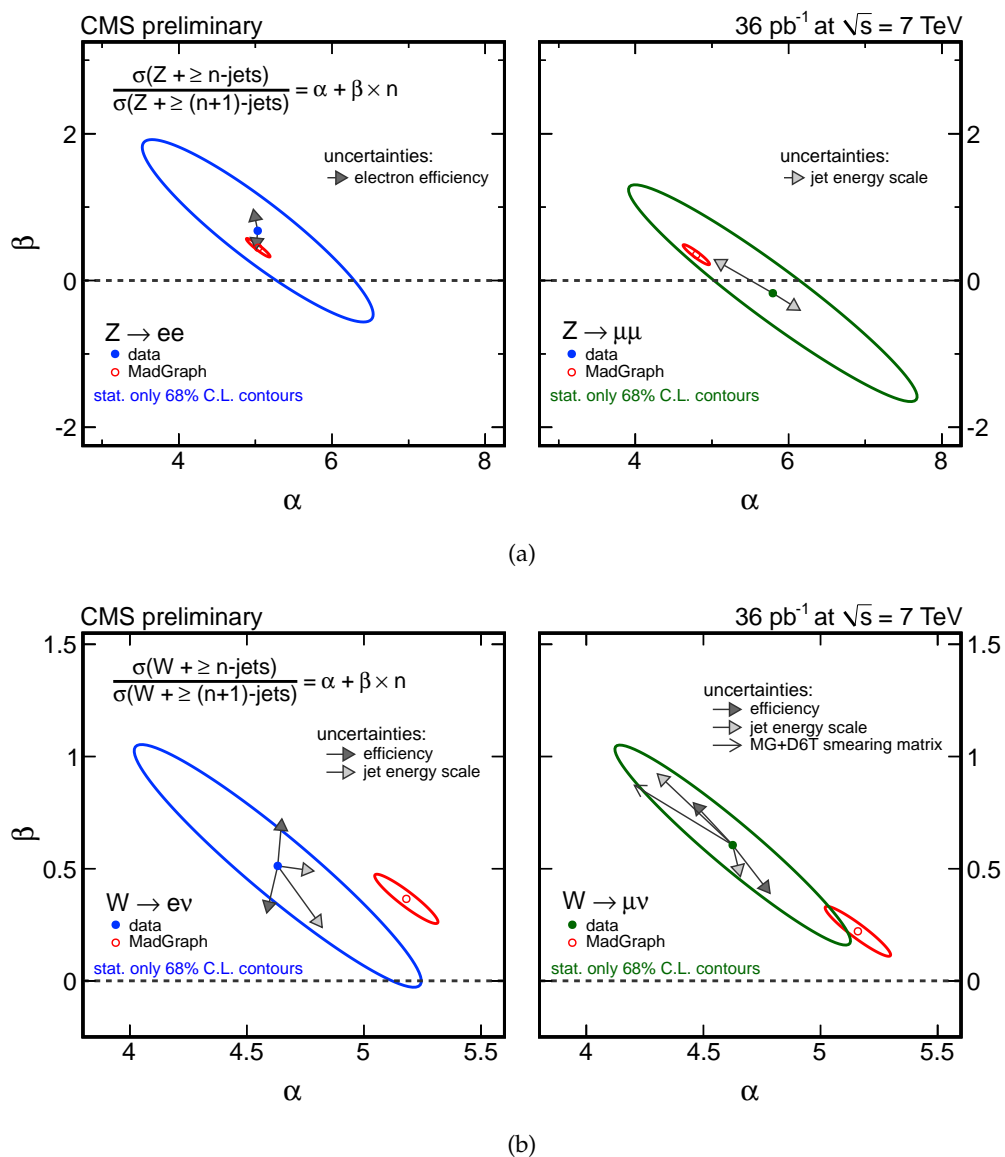


Figure 12: Fit results on the Berends-Giele scaling parameters α and β after pile up subtraction, efficiency corrections, and unfolding for detector resolution effects. The data are compared to MADGRAPH with the Z2 tune. a) shows $Z/\gamma^* + \text{jets}$, b) shows $W + \text{jets}$.

10 Conclusions

We measured the rate of jet production in association with a W or Z vector boson using pp collision data at $\sqrt{s} = 7$ TeV. The data were collected by the CMS Collaboration in 2010 and correspond to an integrated luminosity of 36 ± 4 pb $^{-1}$. The W + jets and Z + jets samples were reconstructed in the electron and muon decay channels. We used particle flow jets with $p_T > 30$ GeV clustered with the anti- k_T algorithm with a size parameter $k = 0.5$.

The leading jet p_T spectrum agrees well with simulations based on MADGRAPH + PYTHIA and the Z2 tune.

We unfolded the exclusive jet multiplicity distributions and measured the ratios of cross sections $\sigma(V + n \text{ jets})/\sigma(V)$ and $\sigma(V + n \text{ jets})/\sigma(V + (n - 1) \text{ jets})$ where n is the inclusive number

Table 7: $\sigma(W + \geq n \text{ jets})/\sigma(W + \geq (n - 1) \text{ jets})$, the ratio of jet multiplicities.

num jets	σ ratio	stat	stat + fit and efficiency	JES	unfolding
electron channel					
$\geq 1 / \geq 0$ jets	0.126	0.002	0.004	+0.018 -0.016	+0.002 -0.000
$\geq 2 / \geq 1$ jets	0.208	0.009	0.012	+0.003 -0.002	+0.000 -0.013
$\geq 3 / \geq 2$ jets	0.165	0.015	0.018	+0.004 -0.004	+0.002 -0.002
$\geq 4 / \geq 3$ jets	0.167	0.035	0.039	+0.002 -0.003	+0.014 -0.000
muon channel					
$\geq 1 / \geq 0$ jets	0.137	0.001	0.007	+0.019 -0.017	+0.002 -0.000
$\geq 2 / \geq 1$ jets	0.190	0.005	0.013	+0.004 -0.003	+0.000 -0.011
$\geq 3 / \geq 2$ jets	0.170	0.011	0.018	+0.004 -0.003	+0.006 -0.008
$\geq 4 / \geq 3$ jets	0.151	0.025	0.037	+0.003 -0.002	+0.023 -0.000

Table 8: $\sigma(Z + \geq n \text{ jets})/\sigma(Z + \geq (n - 1) \text{ jets})$, the ratio of jet multiplicities.

num jets	σ ratio	stat	stat + fit and efficiency	JES	unfolding
electron channel					
$\geq 1 / \geq 0$ jets	0.148	0.006	0.007	+0.020 -0.019	+0.002 -0.000
$\geq 2 / \geq 1$ jets	0.190	0.020	0.020	+0.002 -0.001	+0.000 -0.010
$\geq 3 / \geq 2$ jets	0.125	0.034	0.034	+0.004 -0.004	+0.003 -0.000
$\geq 4 / \geq 3$ jets	0.214	0.117	0.117	+0.003 -0.004	+0.022 -0.042
muon channel					
$\geq 1 / \geq 0$ jets	0.136	0.005	0.009	+0.022 -0.020	+0.018 -0.003
$\geq 2 / \geq 1$ jets	0.189	0.017	0.025	+0.001 -0.001	+0.009 -0.011
$\geq 3 / \geq 2$ jets	0.157	0.038	0.041	+0.002 -0.001	+0.023 -0.000
$\geq 4 / \geq 3$ jets	0.218	0.109	0.110	+0.002 -0.004	+0.020 -0.043

Table 9: Results for the Berends-Giele parameters in the electron channel compared with expectations from MadGraph Z2 at particle level.

		data	stat	JES	$\epsilon(\ell)$	Theory
Z	α	5.0	± 1.0	$+0.1$ -0.0	$+0.00$ -0.06	5.04 ± 0.10
	β	0.7	± 0.8	$+0.08$ -0.04	$+0.3$ -0.6	0.45 ± 0.08
W	α	4.6	± 0.4	$+0.2$ -0.0	-0.05 $+0.02$	5.18 ± 0.09
	β	0.5	± 0.4	$+0.0$ -0.3	± 0.2	0.36 ± 0.07

Table 10: Results for the Berends-Giele parameters in the muon channel compared with expectations from MadGraph Z2 at particle level.

		data	stat	JES MC	$\epsilon(\ell)$	D6T tune	Theory
Z	α	5.8	± 1.2	± 0.6	± 0.1	+0.3	4.8 ± 0.1
	β	-0.2	± 1.0	± 0.3	± 0.1	-0.0	0.35 ± 0.09
W	α	4.3	± 0.3	± 0.2	± 0.2	-0.4	5.16 ± 0.09
	β	0.7	± 0.3	± 0.2	± 0.3	+0.3	0.22 ± 0.06

of jets.

Finally, we made a quantitative test of Berends-Giele scaling parametrized as a function of two parameters determined by a fit. The results show good agreement between $W + \text{jets}$ and $Z + \text{jets}$ and fair agreement with the simulation.

Acknowledgements

We wish to congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC machine. We thank the technical and administrative staff at CERN and other CMS institutes. This work was supported by the Austrian Federal Ministry of Science and Research; the Belgium Fonds de la Recherche Scientifique, and Fonds voor Wetenschappelijk Onderzoek; the Brazilian Funding Agencies (CNPq, CAPES, FAPERJ, and FAPESP); the Bulgarian Ministry of Education and Science; CERN; the Chinese Academy of Sciences, Ministry of Science and Technology, and National Natural Science Foundation of China; the Colombian Funding Agency (COLCIENCIAS); the Croatian Ministry of Science, Education and Sport; the Research Promotion Foundation, Cyprus; the Estonian Academy of Sciences and NICPB; the Academy of Finland, Finnish Ministry of Education, and Helsinki Institute of Physics; the Institut National de Physique Nucléaire et de Physique des Particules / CNRS, and Commissariat à l'Énergie Atomique, France; the Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft, and Helmholtz-Gemeinschaft Deutscher Forschungszentren, Germany; the General Secretariat for Research and Technology, Greece; the National Scientific Research Foundation, and National Office for Research and Technology, Hungary; the Department of Atomic Energy, and Department of Science and Technology, India; the Institute for Studies in Theoretical Physics and Mathematics, Iran; the Science Foundation, Ireland; the Istituto Nazionale di Fisica Nucleare, Italy; the Korean Ministry of Education, Science and Technology and the World Class University program of NRF, Korea; the Lithuanian Academy of Sciences; the Mexican Funding Agencies (CINVESTAV, CONACYT, SEP, and UASLP-FAI); the Pakistan Atomic Energy Commission; the State Commission for Scientific Research, Poland; the Fundação para a Ciência e a Tecnologia, Portugal; JINR (Armenia, Belarus, Georgia, Ukraine, Uzbekistan); the Ministry of Science and Technologies of the Russian Federation, and Russian

Ministry of Atomic Energy; the Ministry of Science and Technological Development of Serbia; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Swiss Funding Agencies (ETH Board, ETH Zurich, PSI, SNF, UniZH, Canton Zurich, and SER); the National Science Council, Taipei; the Scientific and Technical Research Council of Turkey, and Turkish Atomic Energy Authority; the Science and Technology Facilities Council, UK; the US Department of Energy, and the US National Science Foundation.

Individuals have received support from the Marie-Curie programme and the European Research Council (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Associazione per lo Sviluppo Scientifico e Tecnologico del Piemonte (Italy); the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRRIA-Belgium); and the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium).

References

- [1] C. F. Berger, Z. Bern, L. J. Dixon et al., "Precise Predictions for $W + 4$ -Jet Production at the Large Hadron Collider", *Phys. Rev. Lett.* **106** (Mar, 2011) 092001.
doi:10.1103/PhysRevLett.106.092001.
- [2] C. F. Berger, Z. Bern, L. J. Dixon et al., "Next-to-leading order QCD predictions for $W + 3$ -jet distributions at hadron colliders", *Phys. Rev. D* **80** (Oct, 2009) 074036.
doi:10.1103/PhysRevD.80.074036.
- [3] R. K. Ellis, K. Melnikov, and G. Zanderighi, " $W + 3$ jet production at the Tevatron", *Phys. Rev. D* **80** (Nov, 2009) 094002. doi:10.1103/PhysRevD.80.094002.
- [4] C. F. Berger, Z. Bern, L. J. Dixon et al., "Next-to-leading order QCD predictions for $Z, \gamma^* + 3$ -jet distributions at the Tevatron", *Phys. Rev. D* **82** (Oct, 2010) 074002.
doi:10.1103/PhysRevD.82.074002.
- [5] F. A. Berends, W. T. Giele, H. Kuijff et al., "Multi-jet production in W, Z events at p anti- p colliders", *Phys. Lett.* **B224** (1989) 237. doi:10.1016/0370-2693(89)91081-2.
- [6] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s}=7$ TeV", arXiv:1012.2466.
- [7] ATLAS Collaboration, "Measurement of the production cross section for W -bosons in association with jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", arXiv:1012.5382.
- [8] ATLAS Collaboration, "Measurement of the production cross section for Z/γ in association with jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector", Technical Report ATLAS-CONF-2011-001, CERN, Geneva, Feb, 2011.
- [9] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004.
doi:10.1088/1748-0221/3/08/S08004.
- [10] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.
- [11] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

- [12] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $O(\alpha(s)^2)$ ”, *Phys. Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:10.1103/PhysRevD.74.114017.
- [13] J. M. Campbell, R. Frederix, F. Maltoni et al., “Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders”, *Phys. Rev. Lett.* **102** (2009) 182003, arXiv:0903.0005. doi:10.1103/PhysRevLett.102.182003.
- [14] R. Kleiss and W. J. Stirling, “Top quark production at hadron colliders: some useful formulae”, *Z. Phys.* **C40** (1988) 419–423. doi:10.1007/BF01548856.
- [15] GEANT4 Collaboration, “GEANT4: A simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.
- [16] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270. doi:10.1109/TNS.2006.869826.
- [17] R. Field, “Early LHC Underlying Event Data - Findings and Surprises”, arXiv:1010.3558.
- [18] P. Z. Skands, “Tuning Monte Carlo Generators: The Perugia Tunes”, *Phys. Rev.* **D82** (2010) 074018, arXiv:1005.3457. doi:10.1103/PhysRevD.82.074018.
- [19] CMS Collaboration, “Commissioning of the Particle-Flow Event Reconstruction using high-energy LHC Collisions in the CMS Detector”, *CMS PAS PFT* **1-002** (2010).
- [20] CMS Collaboration, “Performance of CMS muon identification in pp collisions at $\sqrt{s} = 7$ TeV”, *CMS PAS MUO-2010-002* (2010).
- [21] M. Cacciari, G.P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, *JHEP* **0804** (2008) 063.
- [22] CMS Collaboration, “Jet Energy Corrections determination at $\sqrt{s} = 7$ TeV”, *CMS PAS JME* **10-010** (2010).
- [23] CMS Collaboration, “Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter”, *JINST* **5** (2010) T03014, arXiv:0911.4881. doi:10.1088/1748-0221/5/03/T03014.
- [24] M. Cacciari and G.P. Salam, “Pileup subtraction using jet areas (hep-ph/0707.1378)”, *Physics Letters* **B 659** (2008) 119–126. doi:10.1016/j.physletb.2007.09.077.
- [25] M. Cacciari, G.P. Salam and G. Soyez, “The catchment area of jets (hep-ph/0802.1188)”, *JHEP* **0804** (2008) 005. doi:10.1088/1126-6708/2008/04/005.
- [26] CMS Collaboration, “Commissioning of b-jet identification with pp collisions at $\sqrt{s} = 7$ TeV”, *CMS PAS BTV* **10-001** (2010).
- [27] A. Hocker and V. Kartvelishvili, “SVD Approach to Data Unfolding”, *Nucl. Instrum. Meth.* **A372** (1996) 469–481, arXiv:hep-ph/9509307. doi:10.1016/0168-9002(95)01478-0.
- [28] G. D’Agostini, “A Multidimensional unfolding method based on Bayes’ theorem”, *Nucl. Instrum. Meth.* **A362** (1995) 487–498. doi:10.1016/0168-9002(95)00274-X.
- [29] CMS Collaboration, “Jet Resolution Determination at $\sqrt{s} = 7$ TeV”, *CMS PAS JME* **10-014** (2010).