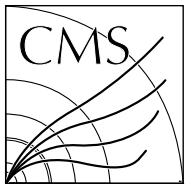
Available on CMS information server

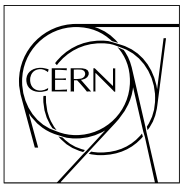
CMS AN -2010/342



The Compact Muon Solenoid Experiment

Analysis Note

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



01 November 2010

$\mathbf{Z} \longrightarrow \mathbf{ee}$ production in association with jets in proton-proton collisions at $\sqrt{\mathbf{s}} = \mathbf{7}$ TeV

M. Mozer *et. al.*

Abstract

We present results on the ratio of $Z{+}n$ jets over $Z + (n{+}1)$ jets yields using the first 3.1 pb^{-1} LHC data. The analysis path and methodology have been presented in AN-2009/92 and published in EWK-08-006. An improved fitting method has been developed that takes into account the correlations in the inclusive jet counting. Systematic uncertainties studies are also included. The results are currently being updated with the full integrated luminosity recorded by CMS.

The Compact Muon Solenoid Experiment

CMS Draft Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

2010/10/28
Head Id: 9438
Archive Id: 15960:19645M
Archive Date: 2010/06/16
Archive Tag: trunk

Study of the $Z \to ee$ production in associtation with jets in proton-proton collisions at $\sqrt{s} = 7$ TeV

Eamanuele[2], Chiara[2], Maurizio[2], Maria[2], Will[2], Sara[2], Yi[1], Chris[1], Paolo[1], Lukas[1], Ilaria[2], and XXX who else XXX[1]

[1] XXX where? XXX
[2] CMG

Abstract

This is an update of AN-2009/92. In this note we discuss a robust selection of a pure $Z \to ee$+jets data sample and its application on both physics and detector calibrations. The $Z \to ee+ \geq 1$ jets data sample and the measured $d\mathrm{N}_{\mathrm{events}}/\mathrm{dN}_{\mathrm{jets}}$ slope can be used to absolutely normalize the Z+jets invisible decays in searches for new physics using the missing transverse energy plus multijet signatures. In addition the $Z \to ee+ \geq N$ jets samples provide a reference for the missing transverse energy performance and calibration as a function of the jet multiplicity. The analysis is indented to be commissioned, in all its physics and detector candle aspects, with 10 pb^{-1} of LHC collision data.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: the vecbos people
PDFTitle: CMS Paper Template 2006 LaTeX/PdfLaTeX version
PDFSubject: CMS
PDFKeywords: CMS, physics, Z+jets

Please also verify that the abstract does not use any user defined symbols

Contents

1 Introduction 2
2 Event Samples 2
2.1 Monte Carlo Samples 2
2.2 Data Samples 3
3 Event Reconstruction and Selection 3
3.1 Electron Reconstruction and Selection 3
3.2 Jet Reconstruction 5
3.3 Selection Efficiencies 6
4 Signal extraction 6
4.1 Expected Yields 6
4.2 The Maximum Likelihood Fit 10
4.3 Ratios 17
5 Systematic Uncertainties 20
5.1 Jet Energy Scale 20
5.2 Background Shape Uncertainty 20
5.3 Signal Shape Uncertainty 21
5.4 Selection Efficiencies 22
6 Results from Early Data 23
7 Conclusion 28

1 Introduction

The analysis presented here is designed to tag and isolate a pure sample of a Standard Model process in a robust way that can be used as a calibration *candle*. The $Z \rightarrow ee$+jets process is ideally suited for this, as the clear di-electron signature can be selected with low backgrounds with a simple, cut based analysis. Such a sample has several complimentary applications.

The ratio of $N/N+1$ jet cross sections is expected to be approximately independent of N in QCD calculations (XXX scaling [**?**]). Previous results [1, 2] show this to be approximately true, but improvements are expected from measurements at the LHC, where the higher center of mass energy compared to the Tevatron considerably increases the available phase space for the jets.

The candle data sample may also be used to evaluate trigger and reconstruction efficiencies, lepton or jet energy scale and corrections, global event missing transverse energy corrections as well as absolute physics calibration (normalization) of the Monte Carlo predictions of challenging processes via the candle well measured and calculated Standard Model ones.

Since the electron p_T is precisely measured, the boson p_T measured from the di-electron system (boson recoil) can serve as a high level calibration of the calorimetric jets and the calorimetric missing transverse energy (MET). The study of $Z \rightarrow ee$+ jets events offers a study-ground for the MET from $Z(\rightarrow \nu\bar{\nu})$ + jets events. Note that the invisible Z decays have not been directly measured but only inferred so far. They constitute an interesting (the source of the missing energy is indeed neutrinos from the decay of a massive object and not mis-measurements) and irreducible background to searches for Dark Matter (DM) production at the LHC. We note that in combination with the $Z \rightarrow \mu\mu$ + jets [**?**] and W+jets in muon [**?**] and electron [**?**] decay modes we develop an extended V+jets Standard Model candle program that uses the lepton universality related ratios and the W/Z+jets double ratios.

The analysis proceeds by the selection of events based on the CMS trigger decision and the presence of a reconstructed primary vertex. The reconstructed electrons in the event are grouped into two (overlapping) categories, tight and loose, with selection criteria described in more detail in section 3.1.3. Z candidates are constructed from combinations of these two collections. Jets are selected according to the criteria discussed in section 3.2 and events classified according to their jet multiplicities. Unbinned maximum likelihood fits (ML fits) are used to extract signal and background yields for the different jet multiplicities and the results are used to measure the $N/N+1$ jet multiplicity ratio.

2 Event Samples

2.1 Monte Carlo Samples

All Monte Carlo samples used in this study are subjected to the full CMS detector simulation and standard CMSSW reconstruction chain. The MADGRAPH generator is used to generate the signal ($Z \rightarrow ee$) sample, using leading order matrix elements with up to four partons in the final state, matched to a parton shower simulation.MADGRAPH is also used to generate samples for electroweak ($W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$) and top backgrounds.

Backgrounds from QCD processes are produced using the leading order PYTHIA Monte Carlo generator. In addition to a dedicated γ+jet sample, where the photon is included in the matrix element, two samples are produced by enriching basic QCD processes in either EM particles or heavy quarks at the generator level.

The samples are summarized in table 1. The effective integrated luminosities of the samples are computed using the NLO cross section where available, i.e. the Z, W and top samples. Backgrounds from single top and multiples vector boson production have been shown to be small [3] and are neglected in this note.

sample	cross section		integrated luminosity (pb^{-1})
ZJets-madgraph	2.4 (2.95)	nb	350
WJets-madgraph	24.2 (31.0)	nb	315
ttbarJets-madgraph	95 (162)	pb	$7.9 \cdot 10^3$
QCD-BCtoE-Pt20to30	0.108	μb	20.9
QCD-BCtoE-Pt30to80	0.138	μb	6.34
QCD-BCtoE-Pt80to170	9.46	nb	0.92
QCD-EMEnriched-Pt20to30	1.72	μb	17.6
QCD-EMEnriched-Pt30to80	3.48	μb	9.80
QCD-EMEnriched-Pt80to170	0.135	μb	37.3
PhotonJet-Pt0to15	$8.446 \cdot 10^7$	pb	$1.36 \cdot 10^{-3}$
PhotonJet-Pt15to20	$1.147 \cdot 10^5$	pb	0.944
PhotonJet-Pt20to30	$5.718 \cdot 10^4$	pb	1.05
PhotonJet-Pt30to50	$1.652 \cdot 10^4$	pb	6.66
PhotonJet-Pt50to80	$2.723 \cdot 10^3$	pb	40.6
PhotonJet-Pt80to120	$4.462 \cdot 10^2$	pb	151
PhotonJet-Pt120to170	84.43	pb	$1.46 \cdot 10^3$
PhotonJet-Pt170to300	22.55	pb	$5.44 \cdot 10^3$
PhotonJet-Pt300to500	1.545	pb	$69.4 \cdot 10^3$
PhotonJet-Pt500toInf	$9.230 \cdot 10^{-2}$	pb	$0.618 \cdot 10^6$

Table 1: Cross sections and integrated luminosities for the Monte Carlo samples used in this analysis. Cross sections are quoted at leading order, NLO cross sections are given in brackets where available. The integrated luminosity of each sample is computed using the NLO cross section where applicable. For samples that are enriched in certain final states (QCD-BCtoE and QCD-EMEnriched), the cross section has been corrected for the enrichment factor.

2.2 Data Samples

The data sample consists of the events collected by the CMS experiment at 7 TeV center-of-mass energy during the Run2010A data-taking period, starting from March 2010, corresponding to an integrated luminosity of 2.8 pb^{-}1. Events collected in the `EG` dataset are considered.

3 Event Reconstruction and Selection

A baseline selection is applied to reduce the number of background events while keeping the signal efficiency as large as possible, starting with a HLT requirement. In addition, we exploit the possibility of using isolation and impact-parameter requirements (optimized using toy Monte Carlo studies).

3.1 Electron Reconstruction and Selection

3.1.1 Trigger

All events are required to pass an appropriate HLT requirement, but no matching between HLT objects and offline reconstructed electrons is required. All events are required to pass at least one triggers with the maximal requirement of pixel-matched supercluster of $E_T > 17$ GeV with

requirements on cluster shape and hadronic energy behind the supercluster, both looser than the offline selection. In data where no such trigger is unprescaled, the selection is supplemented by a trigger that requires a second supercluster of $E_T > 8$ GeV, the other requirements being identical for both. As none of the triggers contains any isolation requirements, it is not expected that trigger efficiency depends on the jet multiplicity, such that the ratio of Z+(N/N+1) jet yields is expected to be unaffected by possible inefficiencies in the trigger.

3.1.2 Electron Reconstruction

The gsf-electrons which are the result of the standard CMSSW reconstruction chain are used in this analysis. Their reconstruction is described in detail in [4].

3.1.3 Electron Selection

Reconstructed electrons are required to lie well within the tracker acceptance of $|\eta| < 2.5$ and are required to have a transverse energy of E_{T} >20(10) GeV for the leading (second) leg, respectively. The following variables are used in the selection of electrons:

$\Delta\eta$ The difference in pseudorapidity between the center of the electrons supercluster and the extrapolation of the electrons track at the vertex. This variable is degraded in the endcap due to misalignment and a cut is only applied in the barrel.

$\Delta\phi$ The difference in ϕ between the center of the electrons supercluster and the extrapolation of the electrons track at the vertex. This variable has a noticeably worse resolution than $\Delta\eta$ as radiative effects affect the shape of the supercluster and track estimate in the ϕ direction but not in the pseudorapidity.

$\sigma_{i\eta i\eta}$ The log(energy) weighted η-width of the supercluster in units of crystals, taking into account the $5x5$ crystals surrounding the superclusters seed crystal.

H/E The ratio of the energy in the HCAL behind the electron to the electrons super cluster energy. The HCAL energy is collected in a cone of radius 0.15 around the electron.

trackiso The E_{T} sum of selected tracks surrounding the electron in a radius of $\Delta R = 0.3$ divided by the E_{T} of the electron. Tracks are required to have at least 0.7 GeV p_{T} and originate no further than 0.2 mm from the vertex associated to the electron. Tracks within 0.15 in η of the electron are excluded to avoid tracks from converting bremsstrahlung.

ecaliso The E_{T} sum of ECAL Rechits surrounding the electron in a radius of $\Delta R = 0.3$ divided by the E_{T} of the electron. Rechits in the barrel (endcap) must exceed an energy (E_{T}) threshold of 80 MeV (100 MeV) to reduce noise. To avoid contributions from bremsstrahlung and energy leakage from the cluster, rechits are disqualified if they lie within 1.5 crystal widths in η or 3 crystal widths in ΔR of the electron, respectively.

hcaliso The E_{T} sum of all HCAL towers surrounding the electron in a radius of $\Delta R = 0.3$ divided by the E_{T} of the electron. Towers within $\Delta R < 0.15$ to overlaps with the H/E variable described above.

D_{xy} The transverse distance between the electrons track and the primary vertex. Cuts on this variable are designed to remove backgrounds involving heavy quarks, which are expected to produce secondary electrons after measurable decay lengths.

D_z The longitudinal distance between the electrons track and the primary vertex at the point of closest approach. This variable is used to reject backgrounds from pileup, where the jets (defining the primary vertex) and electrons originate from separate collisions.

missing hits The number of layers of the tracking detector passed by the electron track without producing a hit, counting from the interaction point. This criterion removes electrons produced by the conversion of photons in the tracker material.

The selection criteria applied in this analysis follow the "WP95" selection discussed in detail in [5, 6]. The cuts are summarized in table 2. As the isolation cuts are expected to induce a dependence of the efficiency on the jet number, we study the nominal WP95 cuts as well as the WP95 cuts without the isolation requirements.

	variable	Barrel			Endcpap		
acc.	E_T	20/10 GeV	$<$		20/10 GeV	$<$	
	η	-2.5	-	2.5	-2.5	-	2.5
Id	$\sigma_{i\eta i\eta}$		$<$	0.01		$<$	0.03
	Missing hits		$\leq$	1		$\leq$	1
	$\Delta\eta$		$<$	0.007		no cut	
	$\Delta\phi$		$<$	0.8		$<$	0.7
	H/E		$<$	0.15		$<$	0.07
vtx.	D_z	-0.6 mm	-	0.6 mm	-0.6 mm	-	0.6 mm
	D_{xy}	-0.35 mm	-	0.35 mm	-0.35 mm	-	0.35 mm
isol.	Track iso		$<$	0.15		$<$	0.08
	ECAL iso		$<$	2.0		$<$	0.06
	HCAL iso		$<$	0.12		$<$	0.05

Table 2: Selection criteria for electrons. The cut on D_{xy} and D_z implies the requirement on the presence of a primary vertex.

3.1.4 Z reconstruction

The four-momenta of all selected electron candidates are combined pairwise to Z-candidates. Note that no requirement on the electrons charge is made to avoid uncertainties due to charge-misidentification. Only Z-candidates in the mass range $60 < m_{ee} < 110$ GeV are considered for further analysis. In the case that more than one candidate exists, the one consisting of the highest E_T electrons is chosen.

3.2 Jet Reconstruction

Jets are clustered using the *anti-Kt* [7] algorithm with cone size $\Delta R = 0.5$. Two different jet constituents are considered: Calorimeter towers ("calo jets") and particle flow objects produced from a combination of calorimeter and track information [**?**] ("PFjets"). Appropriate correction factors are applied to correct for the η dependence of the calorimeter response as well as the overall energy scale[1].

Due to the inclusion of track information, PFjets can be reconstructed reliably down to lower energies than this is possible for calojets, where the corrections for energy lost in the material and low energy hadrons bent in the magnetic field become too large. PFjets are considered above $E_T > 15$ GeV, while calo jets are required to surpass $E_T > 30$ GeV. PFjets are required to

[1] While the corrections used in this study are Monte Carlo based, we expect data driven corrections to exist by the time the integrated luminosity assumed here is reached.

have $|\eta| < 2.4$, to match the acceptance of the tracker, while calorimeter jets are confined to the calorimeters, $|\eta| < 3.0$. As the electrons from the Z decay deposit significant amounts of energy in the ecal, they are generally reconstructed as jets as well as electrons. To avoid counting these electron jets, all jets are required to have a distance of $\Delta R > 0.3$ from both electrons of the reconstructed Z decay. The jet selection criteria are summarized in tables 3 and 4.

variable	Range		
η	-3.0	$-$	3.0
E_{T}	30 GeV	$<$	
ΔR		$<$	0.3

Table 3: Selection applied to Calo jets.

variable	Range		
η	-2.4	$-$	2.4
E_{T}	15 GeV	$<$	
ΔR		$<$	0.3

Table 4: Selection applied to PF jets.

3.3 Selection Efficiencies

Tables 5 shows the signal efficiency for each selection step for the various jet multiplicities. For this table the ΔR cut has been applied between reconstructed jets and generator electron, to allow proper jet counting even in the absence of a reconstructed Z candidate. These results are compatible with previous results [3] if hte slightly differing selection criteria are taken into account. The final signal yield is dominated by acceptance effects. The product of efficiency and acceptance is stable as function of jet multiplicity, an important factor in the analysis of the $N/N+1$ jet ratios. However, the acceptance alone increases $\sim 10\%$ as a function jet multiplicity as a higher number of jets is correlated with a higher Z transverse momentum. Conversely, additional jets cause the efficiency to drop by a similar amount due to the additional activity spoiling electron Id and isolation variables. While the cancellation allows us to draw conclusions directly about the cross section, it complicates the comparison to external calculations which will usually account for acceptance effects but not efficiencies.

In addition to the baseline selection, we also investigate the same criteria without the isolation requirements. While this removes a criterion which noticeably depends on the jet multiplicity it also increases the expected background contributions. Efficiencies for this selection are given in table 6.

4 Signal extraction

4.1 Expected Yields

The expected signal and background yields with the baseline selection are given in table 7, assuming an integrated luminosity of 10 pb^{-1}. The yields for the baseline selection without the isolation requirement are shown in table 8. In the non-isolated selection background from misidentified QCD processes becomes so large, that this selection will not be considered for the extraction of the signal cross section. However, such a sample is still useful for studies of systematic uncertainties, particularly of the isolation and as a control sample for other analyses.

Selection	**PFjets** > 15 **GeV**							
	$Z+ \geq$ 1jets		$Z+ \geq$ 2jets		$Z+ \geq$ 3jets		$Z+ \geq$ 4jets	
	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff
HLT	0.941 ± 0.000695	0.941 ± 0.000695	0.958 ± 0.00107	0.958 ± 0.00107	0.971 ± 0.00171	0.971 ± 0.00171	0.981 ± 0.00275	0.981 ± 0.00275
$e^{\pm}$ reconstruction	0.69 ± 0.00141	0.649 ± 0.00141	0.748 ± 0.00237	0.717 ± 0.00241	0.79 ± 0.00421	0.768 ± 0.00431	0.827 ± 0.0076	0.811 ± 0.0078
acceptance	0.835 ± 0.00136	0.542 ± 0.00147	0.821 ± 0.00242	0.589 ± 0.00263	0.808 ± 0.00458	0.62 ± 0.00495	0.809 ± 0.00868	0.656 ± 0.00946
electron ID	0.881 ± 0.0013	0.478 ± 0.00148	0.863 ± 0.00239	0.508 ± 0.00267	0.845 ± 0.00469	0.524 ± 0.00509	0.831 ± 0.00921	0.545 ± 0.00991
isolation	0.946 ± 0.000963	0.452 ± 0.00147	0.937 ± 0.00182	0.476 ± 0.00267	0.926 ± 0.00368	0.485 ± 0.0051	0.91 ± 0.00772	0.496 ± 0.00995
conv. rej.	0.989 ± 0.000465	0.447 ± 0.00147	0.988 ± 0.000834	0.47 ± 0.00266	0.988 ± 0.00161	0.479 ± 0.00509	0.987 ± 0.00317	0.49 ± 0.00995
existence of PV	1 ± 0	0.447 ± 0.00147	1 ± 0	0.47 ± 0.00266	1 ± 0	0.479 ± 0.00509	1 ± 0	0.49 ± 0.00995
$\lvert z_e - z_{PV} \rvert \leq 600\mu m$	$1 \pm 3.39e\text{-}05$	0.447 ± 0.00147	1 ± 0	0.47 ± 0.00266	1 ± 0	0.479 ± 0.00509	1 ± 0	0.49 ± 0.00995
$\lvert \delta_{xy} \rvert \leq 350\mu m$	1 ± 0	0.447 ± 0.00147	1 ± 0	0.47 ± 0.00266	1 ± 0	0.479 ± 0.00509	1 ± 0	0.49 ± 0.00995
muon veto	0.995 ± 0.000305	0.445 ± 0.00147	0.991 ± 0.000729	0.466 ± 0.00266	0.987 ± 0.00168	0.473 ± 0.00509	0.981 ± 0.00392	0.48 ± 0.00995
$60 < m(e^+e^-) < 110$ GeV/c^2	0.955 ± 0.000915	0.425 ± 0.00146	0.945 ± 0.00179	0.44 ± 0.00265	0.944 ± 0.00342	0.446 ± 0.00507	0.949 ± 0.00633	0.456 ± 0.00992
Selection	**calo jets** > 30 **GeV**							
	$Z+ \geq$ 1jets		$Z+ \geq$ 2jets		$Z+ \geq$ 3jets		$Z+ \geq$ 4jets	
	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff
HLT	0.957 ± 0.000831	0.957 ± 0.000831	0.979 ± 0.00132	0.979 ± 0.00132	0.989 ± 0.00232	0.989 ± 0.00232	0.991 ± 0.00443	0.991 ± 0.00443
$e^{\pm}$ reconstruction	0.722 ± 0.00187	0.691 ± 0.00189	0.798 ± 0.00371	0.781 ± 0.00379	0.854 ± 0.00775	0.844 ± 0.00792	0.888 ± 0.015	0.88 ± 0.0154
acceptance	0.835 ± 0.00182	0.577 ± 0.00202	0.826 ± 0.00393	0.645 ± 0.00438	0.828 ± 0.00896	0.699 ± 0.01	0.835 ± 0.0187	0.735 ± 0.0208
electron ID	0.858 ± 0.00188	0.495 ± 0.00204	0.834 ± 0.00424	0.538 ± 0.00457	0.791 ± 0.0106	0.553 ± 0.0109	0.767 ± 0.0233	0.563 ± 0.0234
isolation	0.943 ± 0.00134	0.467 ± 0.00204	0.931 ± 0.00316	0.501 ± 0.00458	0.924 ± 0.00777	0.511 ± 0.0109	0.905 ± 0.0184	0.51 ± 0.0236
conv. rej.	0.988 ± 0.000639	0.462 ± 0.00204	0.987 ± 0.00145	0.495 ± 0.00458	0.986 ± 0.00359	0.504 ± 0.0109	0.987 ± 0.00751	0.503 ± 0.0236
existence of PV	1 ± 0	0.462 ± 0.00204	1 ± 0	0.495 ± 0.00458	1 ± 0	0.504 ± 0.0109	1 ± 0	0.503 ± 0.0236
$\lvert z_e - z_{PV} \rvert \leq 600\mu m$	$1 \pm 6.26e\text{-}05$	0.462 ± 0.00204	1 ± 0	0.495 ± 0.00458	1 ± 0	0.504 ± 0.0109	1 ± 0	0.503 ± 0.0236
$\lvert \delta_{xy} \rvert \leq 350\mu m$	1 ± 0	0.462 ± 0.00204	1 ± 0	0.495 ± 0.00458	1 ± 0	0.504 ± 0.0109	1 ± 0	0.503 ± 0.0236
muon veto	0.993 ± 0.000516	0.458 ± 0.00203	0.985 ± 0.0016	0.487 ± 0.00458	0.969 ± 0.00535	0.489 ± 0.0109	0.956 ± 0.0137	0.481 ± 0.0236
$60 < m(e^+e^-) < 110$ GeV/c^2	0.948 ± 0.00133	0.435 ± 0.00202	0.916 ± 0.00364	0.446 ± 0.00455	0.92 ± 0.00848	0.449 ± 0.0109	0.926 ± 0.0178	0.445 ± 0.0235

Table 5: Reconstruction efficiency for Z+jets signal samples as a function of jet multiplicity, estimated from Monte Carlo. The number of events with at least one jet is taken as the normalization reference. The jet counting for the normalization is done using the generator-level electrons to apply the $\Delta R_{e,jet}$ requirement.

Selection	**PFjets** $>$ 15 **GeV**							
	$Z+\geq$ 1jets		$Z+\geq$ 2jets		$Z+\geq$ 3jets		$Z+\geq$ 4jets	
	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff
HLT	0.941 ± 0.000695	0.941 ± 0.000695	0.958 ± 0.00107	0.958 ± 0.00107	0.971 ± 0.00171	0.971 ± 0.00171	0.981 ± 0.00275	0.981 ± 0.00275
$e^{\pm}$ reconstruction	0.69 ± 0.00141	0.649 ± 0.00141	0.748 ± 0.00237	0.717 ± 0.00241	0.79 ± 0.00421	0.768 ± 0.00431	0.827 ± 0.0076	0.811 ± 0.0078
acceptance	0.835 ± 0.00136	0.542 ± 0.00147	0.821 ± 0.00242	0.589 ± 0.00263	0.808 ± 0.00458	0.62 ± 0.00495	0.809 ± 0.00868	0.656 ± 0.00946
electron ID	0.881 ± 0.0013	0.478 ± 0.00148	0.863 ± 0.00239	0.508 ± 0.00267	0.845 ± 0.00469	0.524 ± 0.00509	0.831 ± 0.00921	0.545 ± 0.00991
isolation	1 ± 0	0.478 ± 0.00148	1 ± 0	0.508 ± 0.00267	1 ± 0	0.524 ± 0.00509	1 ± 0	0.545 ± 0.00991
conv. rej.	0.988 ± 0.000462	0.472 ± 0.00148	0.987 ± 0.000833	0.501 ± 0.00267	0.987 ± 0.00158	0.517 ± 0.0051	0.988 ± 0.00298	0.539 ± 0.00992
existence of PV	1 ± 0	0.472 ± 0.00148	1 ± 0	0.501 ± 0.00267	1 ± 0	0.517 ± 0.0051	1 ± 0	0.539 ± 0.00992
$\lvert z_e - z_{PV}\rvert \leq 600\mu m$	$1 \pm 3.21e\text{-}05$	0.472 ± 0.00148	1 ± 0	0.501 ± 0.00267	1 ± 0	0.517 ± 0.0051	1 ± 0	0.539 ± 0.00992
$\lvert\delta_{xy}\rvert \leq 350\mu m$	1 ± 0	0.472 ± 0.00148	1 ± 0	0.501 ± 0.00267	1 ± 0	0.517 ± 0.0051	1 ± 0	0.539 ± 0.00992
muon veto	0.995 ± 0.000295	0.47 ± 0.00148	0.991 ± 0.000702	0.497 ± 0.00267	0.987 ± 0.0016	0.511 ± 0.0051	0.981 ± 0.00372	0.528 ± 0.00994
$60 < m(e^+e^-) < 110$ GeV/c^2	0.951 ± 0.000928	0.447 ± 0.00147	0.94 ± 0.0018	0.467 ± 0.00266	0.936 ± 0.0035	0.478 ± 0.00509	0.94 ± 0.00651	0.497 ± 0.00995
Selection	**calo jets** $>$ 30 **GeV**							
	$Z+\geq$ 1jets		$Z+\geq$ 2jets		$Z+\geq$ 3jets		$Z+\geq$ 4jets	
	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff	Relative Eff.	Cumulative Eff
HLT	0.957 ± 0.000831	0.957 ± 0.000831	0.979 ± 0.00132	0.979 ± 0.00132	0.989 ± 0.00232	0.989 ± 0.00232	0.991 ± 0.00443	0.991 ± 0.00443
$e^{\pm}$ reconstruction	0.722 ± 0.00187	0.691 ± 0.00189	0.798 ± 0.00371	0.781 ± 0.00379	0.854 ± 0.00775	0.844 ± 0.00792	0.888 ± 0.015	0.88 ± 0.0154
acceptance	0.835 ± 0.00182	0.577 ± 0.00202	0.826 ± 0.00393	0.645 ± 0.00438	0.828 ± 0.00896	0.699 ± 0.01	0.835 ± 0.0187	0.735 ± 0.0208
electron ID	0.858 ± 0.00188	0.495 ± 0.00204	0.834 ± 0.00424	0.538 ± 0.00457	0.791 ± 0.0106	0.553 ± 0.0109	0.767 ± 0.0233	0.563 ± 0.0234
isolation	1 ± 0	0.495 ± 0.00204	1 ± 0	0.538 ± 0.00457	1 ± 0	0.553 ± 0.0109	1 ± 0	0.563 ± 0.0234
conv. rej.	0.988 ± 0.000642	0.489 ± 0.00204	0.986 ± 0.00147	0.531 ± 0.00457	0.985 ± 0.00353	0.545 ± 0.0109	0.988 ± 0.00681	0.557 ± 0.0234
existence of PV	1 ± 0	0.489 ± 0.00204	1 ± 0	0.531 ± 0.00457	1 ± 0	0.545 ± 0.0109	1 ± 0	0.557 ± 0.0234
$\lvert z_e - z_{PV}\rvert \leq 600\mu m$	$1 \pm 5.91e\text{-}05$	0.489 ± 0.00204	1 ± 0	0.531 ± 0.00457	1 ± 0	0.545 ± 0.0109	1 ± 0	0.557 ± 0.0234
$\lvert\delta_{xy}\rvert \leq 350\mu m$	1 ± 0	0.489 ± 0.00204	1 ± 0	0.531 ± 0.00457	1 ± 0	0.545 ± 0.0109	1 ± 0	0.557 ± 0.0234
muon veto	0.993 ± 0.000497	0.485 ± 0.00204	0.985 ± 0.00152	0.523 ± 0.00457	0.971 ± 0.00495	0.53 ± 0.0109	0.96 ± 0.0124	0.535 ± 0.0235
$60 < m(e^+e^-) < 110$ GeV/c^2	0.942 ± 0.00136	0.458 ± 0.00203	0.908 ± 0.00367	0.475 ± 0.00457	0.909 ± 0.00863	0.481 ± 0.0109	0.904 ± 0.019	0.483 ± 0.0236

Table 6: Reconstruction efficiency for Z+jets signal samples as a function of jet multiplicity, estimated from Monte Carlo. The number of events with at least one jet is taken as the normalization reference. The jet counting for the normalization is done using the generator-level electrons to apply the $\Delta R_{e,jet}$ requirement.

jets		Z+jets	Z (other)	W+jets	$t\bar{t}$	QCD (uds)	QCD (bc)	γ+jets
PFjets > 15 **GeV**	≥ 0	3.8e+03 $\pm$ 9.7	5.2 $\pm$ 0.39	6.4 $\pm$ 0.45	5.7 $\pm$ 0.084	58 $\pm$ 7.4	15 $\pm$ 4.2	3.9 $\pm$ 2.2
	≥ 1	1.4e+03 $\pm$ 6.1	1.9 $\pm$ 0.23	3.1 $\pm$ 0.32	5.6 $\pm$ 0.084	28 $\pm$ 5.1	9.8 $\pm$ 3.7	3.7 $\pm$ 2.2
	≥ 2	4.3e+02 $\pm$ 3.5	0.63 $\pm$ 0.13	1 $\pm$ 0.18	5.2 $\pm$ 0.081	11 $\pm$ 3.3	6.8 $\pm$ 3.2	3.1 $\pm$ 2.1
	≥ 3	1.2e+02 $\pm$ 1.8	0.11 $\pm$ 0.057	0.32 $\pm$ 0.1	3.3 $\pm$ 0.064	2.2 $\pm$ 1.5	3.6 $\pm$ 2.3	0 $\pm$ 0
	≥ 4	32 $\pm$ 0.95	0.029 $\pm$ 0.029	0.095 $\pm$ 0.055	1.6 $\pm$ 0.045	1.1 $\pm$ 1.1	0 $\pm$ 0	0 $\pm$ 0
	≥ 5	8 $\pm$ 0.48	0 $\pm$ 0	0.032 $\pm$ 0.032	0.66 $\pm$ 0.029	1.1 $\pm$ 1.1	0 $\pm$ 0	0 $\pm$ 0
calo jets > 30 **GeV**	≥ 0	3.8e+03 $\pm$ 9.7	5.2 $\pm$ 0.39	6.4 $\pm$ 0.45	5.7 $\pm$ 0.084	58 $\pm$ 7.4	15 $\pm$ 4.2	3.9 $\pm$ 2.2
	≥ 1	7.3e+02 $\pm$ 4.5	1 $\pm$ 0.17	2.1 $\pm$ 0.26	5.6 $\pm$ 0.084	19 $\pm$ 4.2	8.9 $\pm$ 3.6	3.4 $\pm$ 2.1
	≥ 2	1.4e+02 $\pm$ 2	0.2 $\pm$ 0.075	0.48 $\pm$ 0.12	4.7 $\pm$ 0.077	2.7 $\pm$ 1.6	2.1 $\pm$ 1.7	0.049 $\pm$ 0.041
	≥ 3	25 $\pm$ 0.85	0.029 $\pm$ 0.029	0.095 $\pm$ 0.055	2.2 $\pm$ 0.053	0.27 $\pm$ 0.27	0 $\pm$ 0	0 $\pm$ 0
	≥ 4	5.3 $\pm$ 0.39	0.029 $\pm$ 0.029	0 $\pm$ 0	0.77 $\pm$ 0.031	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
	≥ 5	1 $\pm$ 0.17	0 $\pm$ 0	0 $\pm$ 0	0.24 $\pm$ 0.018	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0

Table 7: Breakdown of events in for the different jet multiplicities of signal and background after the full Z+jets selection is applied. The "Z(other)" column denotes the contribution from $Z \to \mu\mu$ and $Z \to \tau\tau$ decays. Events are normalized to 10 pb^{-1}

jets		Z+jets	Z (other)	W+jets	$t\bar{t}$	QCD (uds)	QCD (bc)	γ+jets
PFjets > 15 **GeV**	≥ 0	3.9e+03 $\pm$ 9.9	6.9 $\pm$ 0.44	35 $\pm$ 1.1	12 $\pm$ 0.12	1.2e+03 $\pm$ 33	6e+02 $\pm$ 40	32 $\pm$ 6
	≥ 1	1.4e+03 $\pm$ 6.3	2.6 $\pm$ 0.27	19 $\pm$ 0.77	12 $\pm$ 0.12	6.5e+02 $\pm$ 25	3.5e+02 $\pm$ 34	16 $\pm$ 4.3
	≥ 2	4.6e+02 $\pm$ 3.6	0.86 $\pm$ 0.16	7.6 $\pm$ 0.49	11 $\pm$ 0.12	2.6e+02 $\pm$ 15	1.5e+02 $\pm$ 25	8.2 $\pm$ 3.1
	≥ 3	1.3e+02 $\pm$ 1.9	0.23 $\pm$ 0.081	2.3 $\pm$ 0.27	7.9 $\pm$ 0.1	87 $\pm$ 8.8	60 $\pm$ 17	0.8 $\pm$ 0.37
	≥ 4	35 $\pm$ 0.99	0.086 $\pm$ 0.049	0.89 $\pm$ 0.17	4.5 $\pm$ 0.075	25 $\pm$ 4.7	16 $\pm$ 11	0.34 $\pm$ 0.25
	≥ 5	8.8 $\pm$ 0.5	0.029 $\pm$ 0.029	0.29 $\pm$ 0.095	2.1 $\pm$ 0.051	8.1 $\pm$ 2.6	0 $\pm$ 0	0.25 $\pm$ 0.25
calo jets > 30 **GeV**	≥ 0	3.9e+03 $\pm$ 9.9	6.9 $\pm$ 0.44	35 $\pm$ 1.1	12 $\pm$ 0.12	1.2e+03 $\pm$ 33	6e+02 $\pm$ 40	32 $\pm$ 6
	≥ 1	7.7e+02 $\pm$ 4.6	1.5 $\pm$ 0.21	13 $\pm$ 0.64	12 $\pm$ 0.12	4.4e+02 $\pm$ 20	2.5e+02 $\pm$ 30	12 $\pm$ 3.7
	≥ 2	1.5e+02 $\pm$ 2.1	0.37 $\pm$ 0.1	3 $\pm$ 0.31	10 $\pm$ 0.11	1.1e+02 $\pm$ 9.7	60 $\pm$ 20	2.4 $\pm$ 1.5
	≥ 3	27 $\pm$ 0.88	0.086 $\pm$ 0.049	1 $\pm$ 0.18	5.9 $\pm$ 0.086	20 $\pm$ 3.7	23 $\pm$ 15	0.34 $\pm$ 0.25
	≥ 4	5.7 $\pm$ 0.4	0.029 $\pm$ 0.029	0.32 $\pm$ 0.1	2.5 $\pm$ 0.056	4 $\pm$ 1.7	1.6 $\pm$ 1.6	0.0018 $\pm$ 0.0018
	≥ 5	1.1 $\pm$ 0.17	0 $\pm$ 0	0.13 $\pm$ 0.063	0.87 $\pm$ 0.033	2.1 $\pm$ 1.2	0 $\pm$ 0	1.6e-05 $\pm$ 1.6e-05

Table 8: Breakdown of events in for the different jet multiplicities of signal and background events after the Z+jets selection is applied without isolation. Events are normalized to 10 pb^{-1}

4.2 The Maximum Likelihood Fit

The signal yield is extracted performing an extended and unbinned maximum likelihood fit of the m_{ee} and N_{jet} distributions on the events selected as described in the previous sections.

The final signal extraction step uses a two-dimensional (in m_{ee} and N_{jet}) fit-function. However, to ascertain the validity of the shapes used to describe signal and backgrounds over the different jet bins, the parameterizations are first studied separately in each jet bin.

4.2.1 Validity of the Parametrization

For this preliminary study, the fit likelihood is written in terms of two signal species: i) the Z+jets signal ii) the background, dominated by QCD and top events.

We write the likelihood as:

$$\mathcal{L} = \frac{e^{(N_Z+N_B)}}{(N_Z+N_B)!}\prod_i^N [N_Z \cdot P_Z(m_{ee_i}) + N_B \cdot P_B(m_{ee_i})].$$

where N_Z (N_B) is the number of signal (background) events for a given jet multiplicity.

4.2.2 m_{ee} Parametrization

The m_{ee} signal distribution is parametrized using a Cruijff functions (see Eq. 1). The parameters obtained from the signal Monte Carlo samples as function of the jet multiplicity are given in Table 9. The fit results are shown in Figure 1. These results indicate that the shape of the signal does not depend significantly on the number of jets, with the exception of the $N_{jet} = 0$ bin[2].

$$P_Z(x) = \begin{cases} e^{-\frac{(x-m)^2}{\sigma_R^2+\alpha_R(x-m)^2}}, & \text{if } x > m \\ e^{-\frac{(x-m)^2}{\sigma_L^2+\alpha_L(x-m)^2}}, & \text{if } x < m \end{cases} \quad (1)$$

Two different parametrizations are investigated for the background: A 2nd order polynominal

$$P_B(x) = 1 + a * x + b * x^2 \quad (2)$$

and a modified exponential of the form:

$$P_B(x) = e^{\frac{(x-x_0)}{a(1+b(x-x_0))}}, \quad (3)$$

where x_0 is fixed to the lower bound of the mass range in question. The parameters obtained from the Monte Carlo samples for the background as a function of the jet multiplicity are given in Table 10. The shapes of the background fluctuate noticeably between jet bins, mostly due to the influence of few high weight events from the low statistics simualtion samples.

4.2.3 Tests of The Fit

The ML fit setup is tested using toy Monte Carlo experiments: we generate an ensemble of toy datasets according to the m_{ee} distribution for signal and background and we repeat the fit on each of this sample, maximizing the likelihood as a function of the two yields N_Z and N_B

[2]The difference is caused by the different kinematic in this bin: Due to the recoiling jet the Z in the higher N_{jet}-bins all have non-negligible p_t, while the $N_{jet} = 0$ bin contains a large fraction of low p_t Z's.

Table 9: Results of a ML fit to a sample of signal Monte Carlo events.

jets		Parameters Eqn 1				
		m	σ_L	σ_R	α_L	α_R
PFjets > 15 **GeV**	≥ 0	91.026±0.028	3.398±0.012	2.0976±0.00036	0.23036±0.000111	0.20155±0.0000461
	≥ 1	91.038±0.077	3.257±0.070	2.0902±0.061	0.23971±0.00365	0.20178±0.00498
	≥ 2	91.069±0.138	3.238±0.125	2.0344±0.107	0.24265±0.00650	0.20478±0.00878
	≥ 3	91.069±0.245	3.063±0.227	1.9671±0.186	0.24358±0.0119	0.21623±0.0155
	≥ 4	91.561±0.666	3.674±0.584	1.7476±0.460	0.24095±0.0268	0.25019±0.0338
calo jets > 30 **GeV**	≥ 0	91.026±0.028	3.398±0.012	2.0976±0.00036	0.23036±0.000111	0.20155±0.0000461
	≥ 1	91.058±0.107	3.184±0.097	2.0903±0.085	0.25422±0.00502	0.20122±0.00690
	≥ 2	91.096±0.251	3.291±0.233	1.9567±0.188	0.24993±0.0118	0.21320±0.0152
	≥ 3	90.978±0.584	2.989±0.519	1.8278±0.424	0.25744±0.0261	0.24122±0.0332
	≥ 4	90.414±1.273	2.840±1.008	2.2477±1.013	0.25636±0.0548	0.24376±0.0772

Table 10: Results of a ML fit to background Monte Carlo events.

jets		Parameters Eqn 2		Parameters Eqn 3	
		a	b	a	b
PFjets > 15 **GeV**	≥ 0	$-0.0158\pm0.774\cdot10^{-3}$	$69.0\cdot10^{-6}\pm6.72\cdot10^{-6}$	−6.98±2.99	0.0857±0.0396
	≥ 1	$-0.0114\pm1.35\cdot10^{-3}$	$28.5\cdot10^{-6}\pm12.0\cdot10^{-6}$	−23.2±17.5	0.0107±0.0238
	≥ 2	$-0.0114\pm0.764\cdot10^{-3}$	$22.4\cdot10^{-6}\pm6.98\cdot10^{-6}$	−56.3±16.5	−0.0140±0.00251
	≥ 3	$-0.0170\pm1.82\cdot10^{-3}$	$73.1\cdot10^{-6}\pm16.8\cdot10^{-6}$	−37.7±15.2	−0.0111±0.00466
	≥ 4	4.80±3.36	−0.0426±0.0297	−96.8±91.2	−0.0153±0.00579
calo jets > 30 **GeV**	≥ 0	$-0.0158\pm0.774\cdot10^{-3}$	$69.0\cdot10^{-6}\pm6.72\cdot10^{-6}$	−6.98±2.995	0.0857±0.0396
	≥ 1	$-0.0169\pm1.89\cdot10^{-3}$	$76.7\cdot10^{-6}\pm16.8\cdot10^{-6}$	−14.0±7.04	0.0280±0.0231
	≥ 2	6.11±4.39	−0.04939±0.0354	−1040±9210	800±530
	≥ 3	0.527±1.44	$-4.04\cdot10^{-3}\pm11.0\cdot10^{-3}$	−517±1280	−0.0178±0.00767
	≥ 4	$-0.0169\pm7.66\cdot10^{-3}$	$81.0\cdot10^{-6}\pm64.7\cdot10^{-6}$	−21.4±147	0.0261±0.279

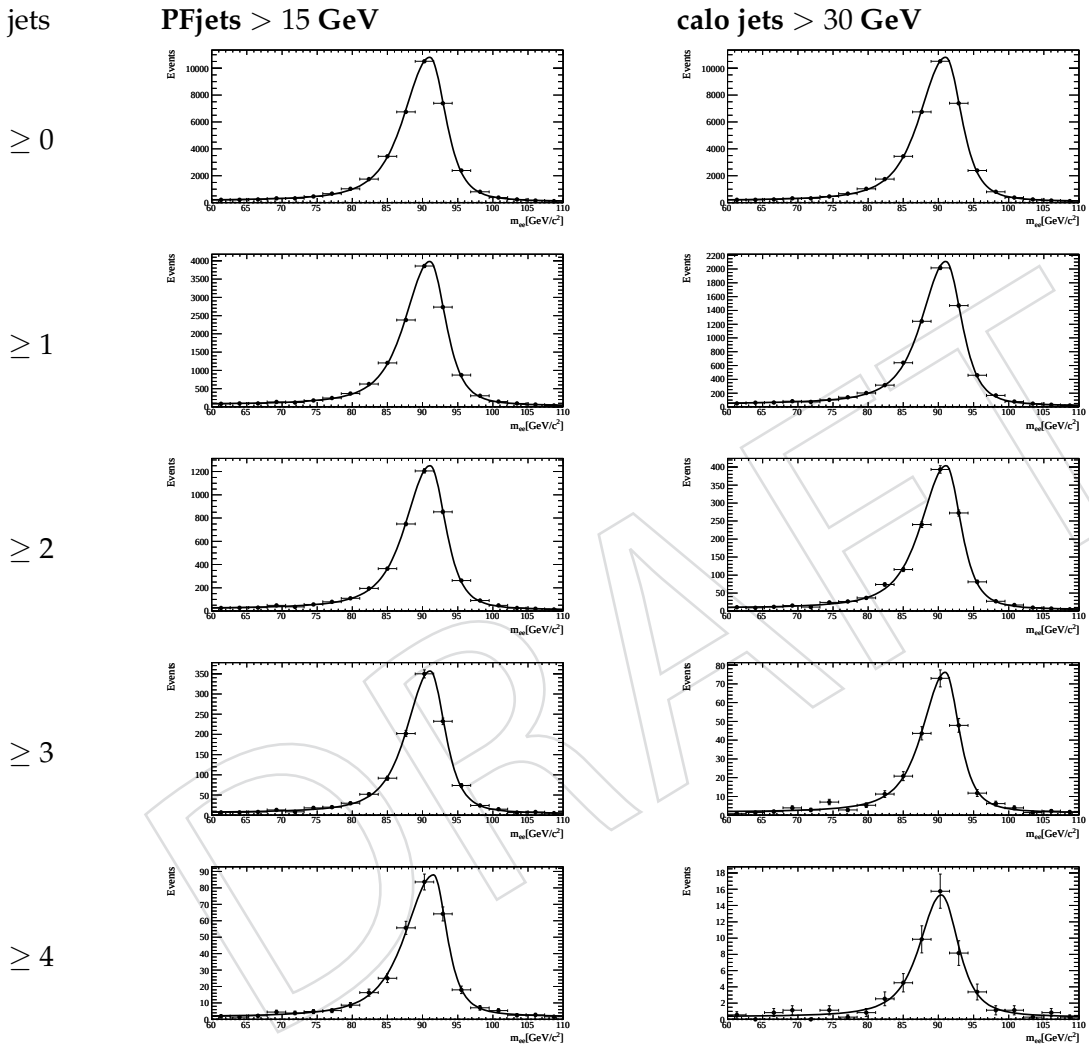


Figure 1: Signal MC distributions for different jet multiplicities and the corresponding fit result.

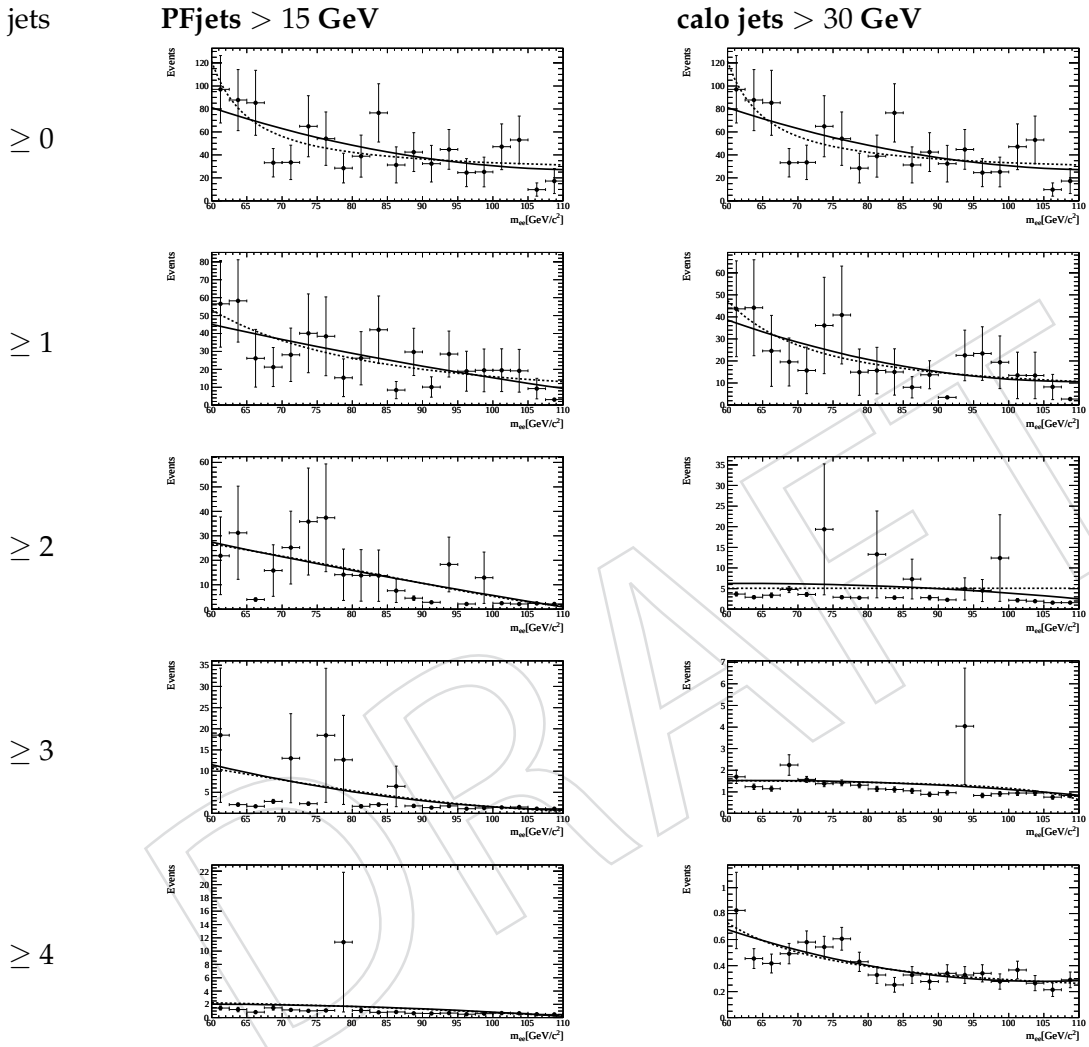


Figure 2: Background MC distributions for different jet multiplicities and the corresponding fit results. Solid lines represent the polynomial fit (2), while dashed lines represent the modified exponential fit (3).

and, shape parameters of the background distribution and the peak position and peak widths (σ_L and σ_R) of the signal distribution. The values of α_L and α_R are kept fixed at the values obtained from the simulation to reduce signal background correlations and associated biases. The actual number of events in each toy is generated according to a Poisson distribution around the expected yield. The test is performed independently for each fit multiplicity.

Table 11: m, σ, and expected error as a function of the jet definition and multiplicity, as obtained from a set of toy Monte Carlo experiments using the polynomial background model.

jets		$m(pull)$	$\sigma(pull)$	$< \sigma(N_i) >$	$< \sigma(N_i)/N_i >$	pull[events]
PFjets > 15 **GeV**	≥ 1	-0.121	1.01	42.1	0.031	5.1
	≥ 2	-0.182	0.989	22.4	0.056	4.1
	≥ 3	-0.236	0.989	10.3	0.11	2.4
	≥ 4	-0.296	0.971	5.00	0.22	1.5
calo jets > 30 **GeV**	≥ 1	-0.111	1.01	31.2	0.043	3.5
	≥ 2	-0.250	0.987	13.9	0.10	3.5
	≥ 3	-0.306	0.951	5.74	0.25	1.8
	≥ 4	-0.0791	0.912	2.33	0.46	0.18

We quote in Tab.11 and 12 the values of m and σ for a Gaussian fit of the pull distribution, together with the expected value for the statistical error on the signal yield. The pull for the experiment i is defined as:

$$pull_i = \frac{N_i - N_{GEN}}{\sigma(N_i)} \tag{4}$$

where $N_i \pm \sigma(N_i)$ is the result of the fit for the variable N and N_{GEN} is the value used in generation. The distribution of the pull variable for the different jet multiplicities is given in Figs. 3 for background parametrization 2 and Fig. 4 for background parametrization 3.

The bias in the extracted signal yields is larger for the polynomial backgrounds and grows with decreasing statistics in the test sample. This suggests that the background parameterization, in particular the plynomial, may be overly flexible, picking up signal fluctuations. However, as the background shape is not well known the bias in the fit need to be weighted against the additional systematic uncertainty that arises if the background parameterization is too reestrictive. The modified exponential background shape most likely provides a reasonable compromise, providing adequate flexibility (as discussed in section 5.2), while retainig overall low levels of bias. In the high statistics bins ($N_{jet} = 1, 2$) the bias remains below 10% of the statistical uncertainty, while remaining at less than one event in most of the low statistics bins.

4.2.4 Global Fit for Yields

Extracting the N_i from direct fits to the m_{ee} distributions of each jet bin separately introduces unwanted correlations in the N_i as the same events can enter multiple fits. While the correlations cannot be avoided they can at least be properly determined by a simultaneous fit to

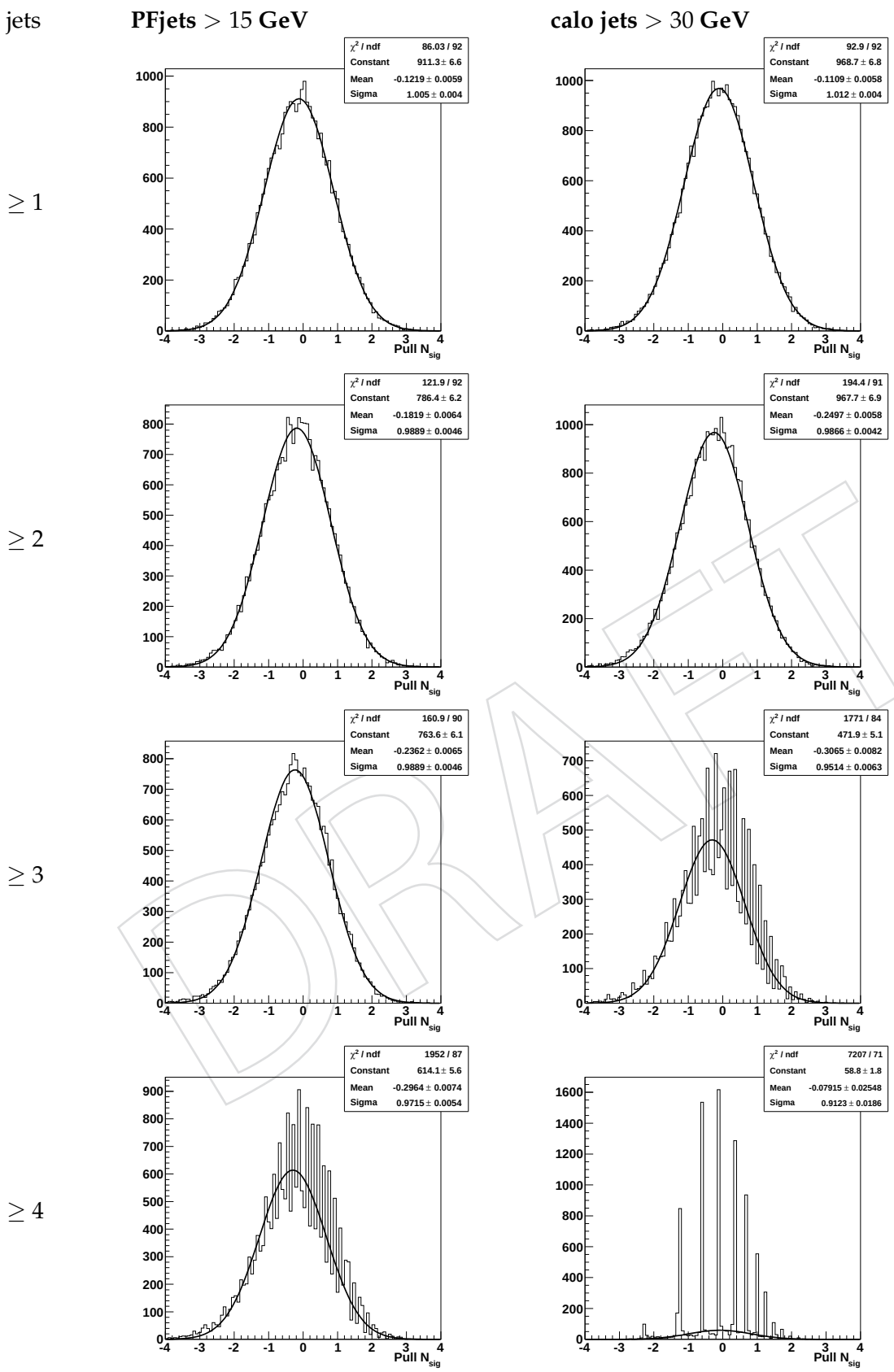


Figure 3: Pull distributions for N_{signal} toy fits using the polynomial background model (equation 2). The deviations from Gaussian behavior for four jets is caused by discreteness effects due to the low number of events in this bin.

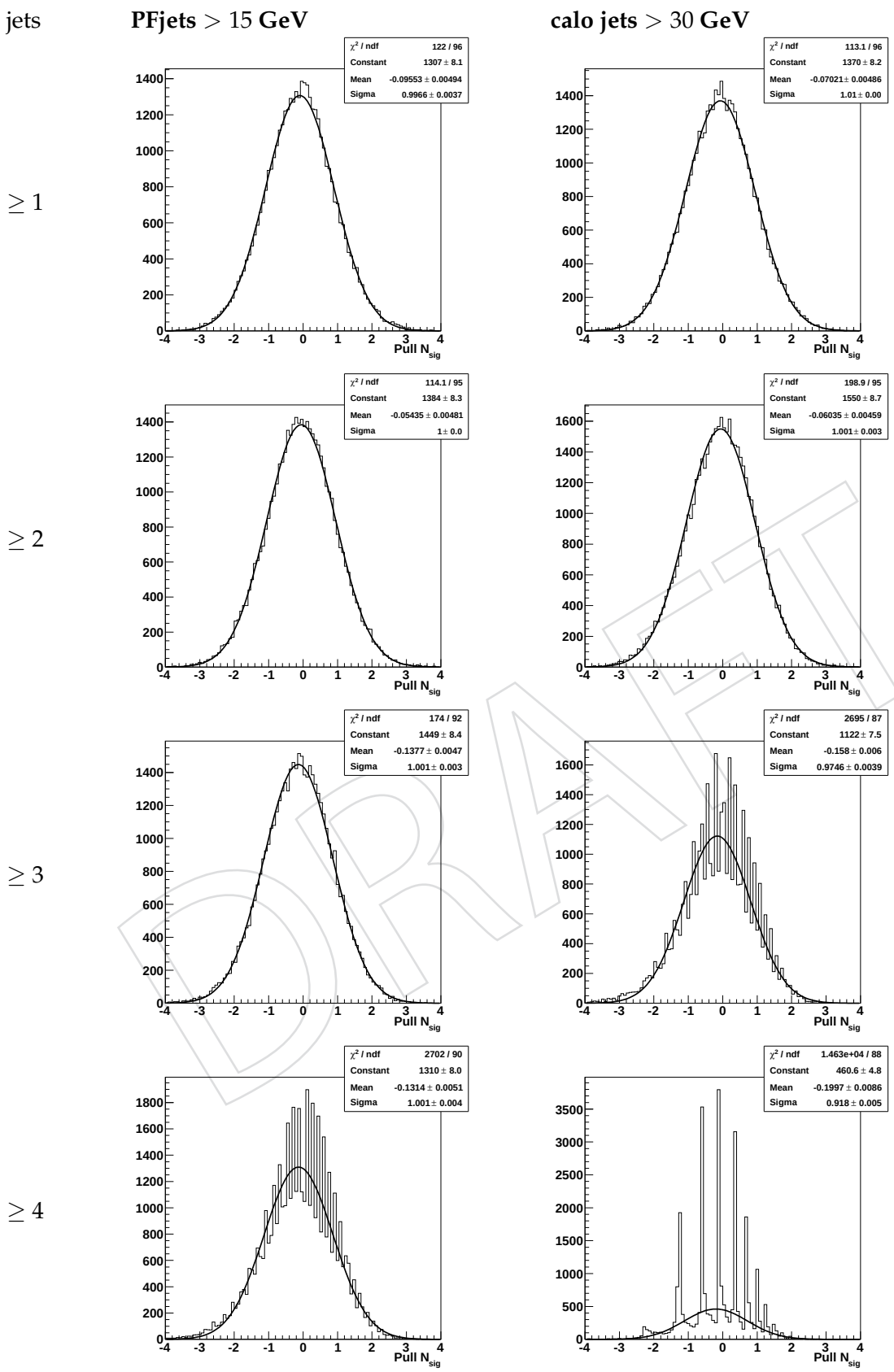


Figure 4: Pull distributions for N_{signal} toy fits using the modified exponential background model (equation 3). The deviations from Gaussian behavior for four jets is caused by discreteness effects due to the low number of events in this bin.

Table 12: m, σ, and expected error as a function of the jet definition and multiplicity, as obtained from a set of toy Monte Carlo experiments using the modified exponential background model.

jets		$m(pull)$	$\sigma(pull)$	$<\sigma(N_i)>$	$<\sigma(N_i)/N_i>$	pull[events]
PFjets $>$ 15 **GeV**	≥ 1	-0.0955	0.997	44.5	0.033	4.25
	≥ 2	-0.0543	1.00	24.8	0.059	1.34
	≥ 3	-0.138	1.001	12.6	0.11	1.7
	≥ 4	-0.131	1.00	6.31	0.21	0.83
calo jets $>$ 30 **GeV**	≥ 1	-0.0702	1.01	32.6	0.045	2.29
	≥ 2	-0.0603	1.00	13.8	0.10	0.83
	≥ 3	-0.158	0.974	5.61	0.24	0.89
	≥ 4	-0.200	0.918	2.34	0.49	0.468

the m_{ee} distributions for each *exclusive* jet-count N_i^{ex}[3], such that each event enters the global fit exactly once.

As shown in section 4.2.2, the signal shape is constant as function of N_i beyond $i = 0$. Thus it is possible to use the same signal shape for all jet bins, reducing the fluctuations otherwise observed in the shape parameters determined in the high N_i, low statistics bins. This comes at the cost of not determining N_0. As it is not expected to follow the XXX-scaling, the $N_{jet} = 0$ bin is not used in the final analysis of the ratios in any case.

The global fit likelihood is written in terms of two signal species: i) the Z+jets signal, represented by a single Cruijff function for all bins ii) the background, which is allow to vary separately in each jet bin.

We write the likelihood as:

$$\mathcal{L} = \prod^{i} \frac{e^{(N^{ex}_{Z,i}+N^{ex}_{B,i})}}{(N^{ex}_{Z,i}+N^{ex}_{B,i})!} \prod_{j}^{events} \sum_{i} \delta(n_{jet,j}-i)[N^{ex}_{Z,i} \cdot P_Z(m_{ee,j}) + N^{ex}_{B,i} \cdot P^i_B(m_{ee_i,j})].$$

where $N^{ex}_{Z,i} = N_{Z,i} - N_{Z,i+1}$ (N^{ex}_B) are the number of signal (background) events for a given exclusive jet multiplicity.

The resulting fit behaves similar to the preliminary fits, showing similar uncertainties and pulls (see Figure 5).

4.3 Ratios

The scaling of the Z+jets yield as a function of the jet multiplicity is parametrized in terms of the quantity

$$C_n = \frac{N_n}{N_{n+1}}. \tag{5}$$

[3]For the highest jet bin under consideration the inclusive count is still used.

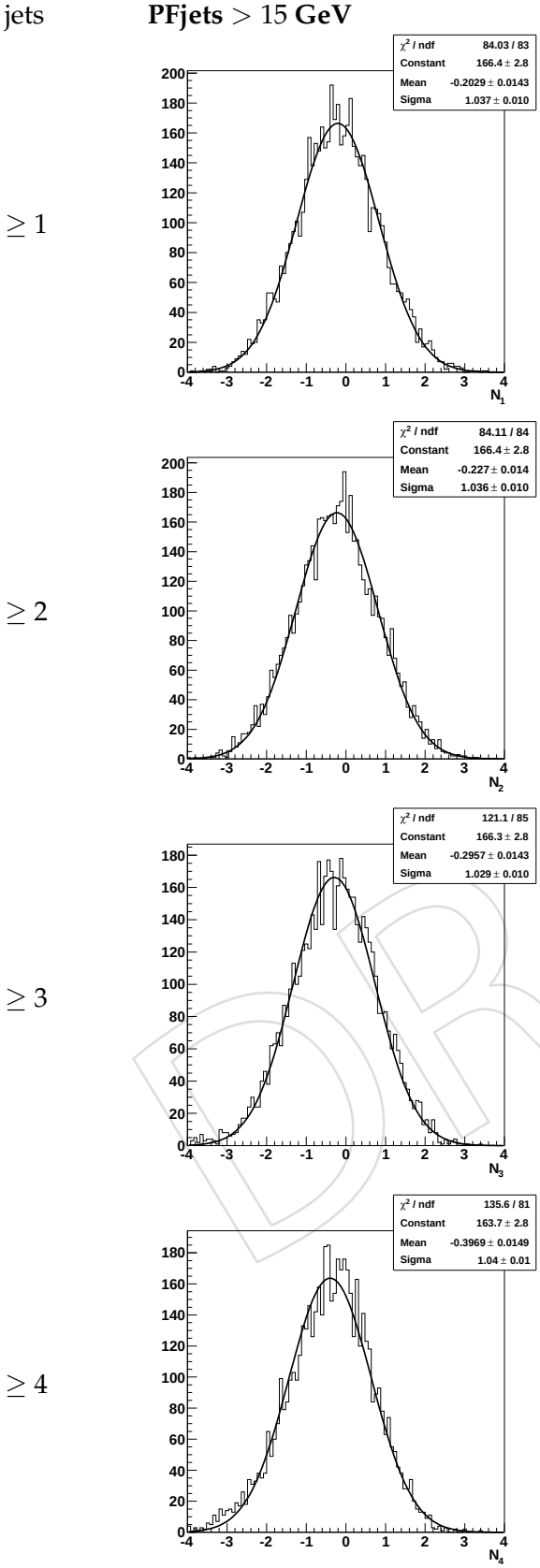


Figure 5: Pull distributions for N_{signal} toy fits using the modified exponential background model (equation 3) in the scenario of PFjets above 15 GeV. The deviations from Gaussian behavior for four jets is caused by discreteness effects due to the low number of events in this bin.

where $N_n = N(Z + n \text{ jets})$. In leading order QCD this ratio is expected to be constant as function of n_{jet} and proportional to α_S, with higher order contributions introducing a weak dependence on n_{jet}. This behavior has been observed at the Tevatron, where the C_n have been found to compatible with a linear rise as function of n_{jet} [**?**]. Figure 8 shows the expected distribution of C_n as expected from the Madgraph simulation.

4.3.1 Global Fit for Ratios

As the C_n are strongly correlated between neighboring bins, a simple χ^2 is not adequate to obtain a linear parametrization of the C_n and check the goodness of fit

Instead we modify our global fit function 5 to directly extract desired linear relationship by maximizing the likelihood with respect to the linear parametrization of the C_n, i.e. $N^{ex}_{Z,i} = N_{Z,i} - N_{Z,i+1} = N_{Z,tot}/\prod_j^i(\alpha + \beta \cdot j)^i - N_{Z,tot}/\prod_j^{i+1}(\alpha + \beta \cdot j)$, such that $C_n = \alpha + \beta * N$ is implicitly assumed. Using the covariance matrix determined in the fit, it is now possible to obtain uncertainties for the linear approximation that properly account for the correlations between the C_n.

Pseudoexperiments were performed under the same conditions as described above (see Figure 6), showing that this type of fit can reliably extract a linear parametrization for the C_n. The Gaussian profile of the pull distributions in particular shows that the statistical uncertainties are properly propagated.

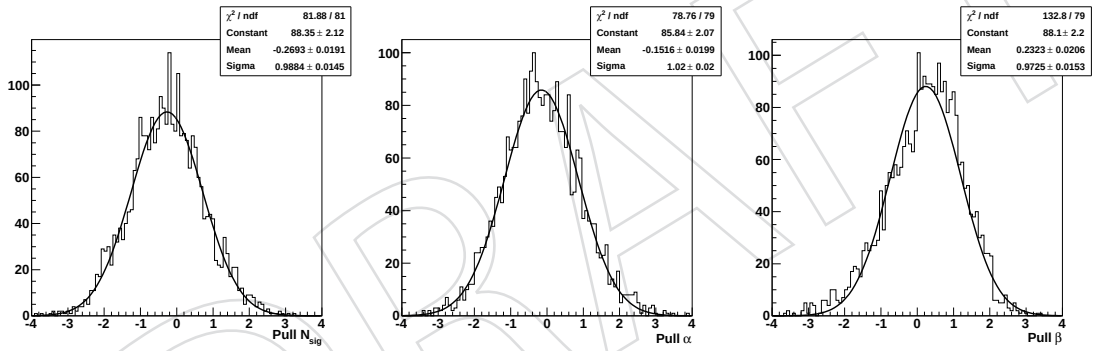


Figure 6: Pull distributions for toy fits using the modified exponential background model (equation 3) in the scenario of PFjets above 15 GeV.

4.3.2 Goodness of Fit Tests

As discussed above, χ^2 cannot be easily used to asses the linearity of the C_n due to correlations. However, the global fit offers new avenues to determine whether the linear approximation is compatible with the data. The test proceeds by recording the maximum likelihood of the nominal fit to the ratios as described above as well as the maximum likelihood where the C_n are described by a fourth order polynomial (equivalent to all the C_n being free). Wilks' Theorem [8] implies that $\Delta_{nll} = \ln \mathcal{L}_{nominal} - \ln \mathcal{L}_{polynomial}$ follows a chi^2 distribution with a number of degrees of freedom equal to the difference in the number of free parameters in the fits[4]. Δ_{nll} thus provides a suitable substitute for the traditional χ^2 often used to determine the goodness of fit. Figure 7 shows the distribution of Δ_{nll} overlaid with a chi^2 distribution of two degrees

[4]The often problematic requirement that the additional parameters need to be well defined under the null hypothesis is met in this case.

of freedom for a set of pseudoexperiments under the same assumptions as above, illustrating Wilks' Theorem.

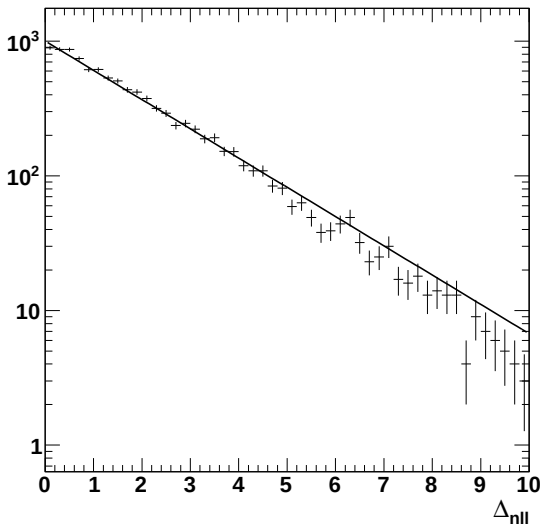


Figure 7: Distribution of Δ_{nll} for a set of pseudoexperiments superimposed on a χ^2 distribution for two degrees of freedom.

5 Systematic Uncertainties

5.1 Jet Energy Scale

The jet energy scale is evaluated following the recommendations of the CMS Jet-MET group [9]. Two uncertainties are considered, corresponding to the two corrections applied to the jet energies: an absolute scale uncertainty and an η-dependent inter-calibration uncertainty. These two uncertainties are treated as independent (and thus added in quadrature), while each of is assumed to be correlated. The absolute scale uncertainty is evaluated by repeating the complete analysis with jet energies adjusted by $\pm 10\%(\pm 5\%)$ for calojets (PFjets) respectively. The difference between the nominal and shifted yields is taken to be the uncertainty. The uncertainty on the η-intercalibration is evaluated by repeating the complete analysis with all jet energies by $\pm 2\% \cdot |\eta|$.

As expected, the jet energy scale uncertainty on the event yield rises roughly exponentially with the number of jets (see Figure 8). However, the jet energy scale uncertainty on the $N/N+1$ jet ratio remains reasonably constant as the uncertainty from the first N jets cancels in the ratio. Further cancellation can be achieved if the double ratio of $(N/N+1)_Z/(N/N+1)_W$ cross sections is studied, where all jet energy scale uncertainties approximately cancel.

5.2 Background Shape Uncertainty

As the parameters of the background parametrization are floated in the fit, there is no direct uncertainty connected to the background shape. It is, however, necessary to ensure that the functional form of the parametrization is able to properly describe the real background shape. To check that this is the case we investigate a sample enriched in background events by using the same event selection as in the nominal analysis, but inverting the cut on track isolation.

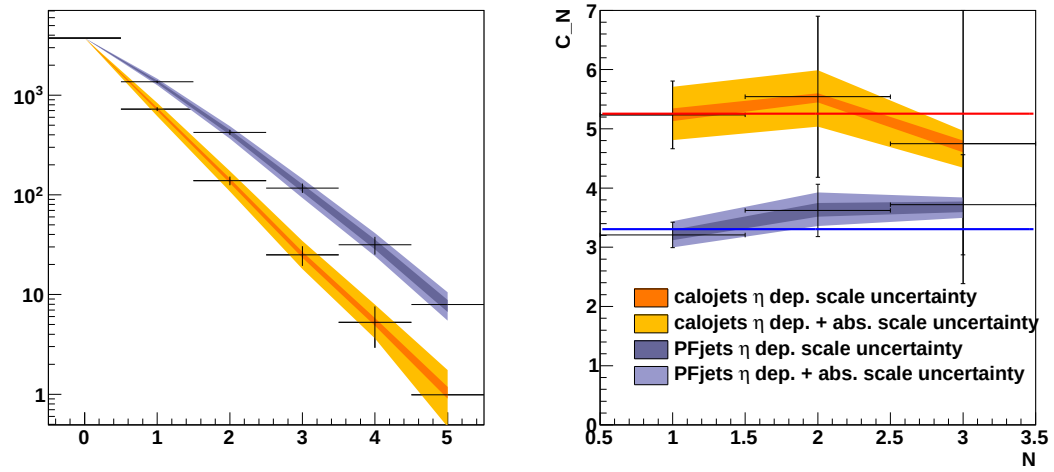


Figure 8: N_jet distribution (left) and $N/N+1$ jet ratios for the signal simulation. The statistical uncertainties are scale to 10 pb^{-}1, the error bands show the η-dependent jet energy scale uncertainty (inner band) and quadratic sum of η-dependent and absolute jet energy scale uncertainty (outer band). The horizontal lines indicate the weighted average of the $N/N+1$ jet ratios.

Figure 9 shows that the shapes of the background simulation agree well with the shape of the background enriched sample in simulation as well as data.

The data distribution may be used as a template for the expected background if the simulation shows that the simulated background spectrum in the nominal selection agrees with the simulated spectrom in the inverted selection. The spectra of the two selections agree reasonably well, however, the low statistics of the simulated background doen't allow strong conclusions. Thus it is important to let the background shape vary freely in the fit.

5.3 Signal Shape Uncertainty

The core of the Z-peak is well described by the Gaussian part of the signal-function and no systematic uncertainty needs to be applied. The two parameters α_L and α_R, however, are fixed in the fit and thus systematic uncertainties in the determination of $\alpha_{L/R}$ need to be propagated to the final result. We fix $\alpha_{L/R}$ to the values expected from the signal simulation. To check the validity of the simulation in general and $\alpha_{L/R}$ in particular, we use a data sample where both electrons are required to pass a much stricter set of selection criteria ("WP80" 1), for which a background contribution of less then 1% is expected. Due to the limited number of Z+jets events in the data, the detailed comparison is only possible in the Z+1 jet sample. However, as shown in table 10, $\alpha_{L/R}$ does not significantly depend on the jet multiplicity, so that an extrapolation to higher jet bins appears valid.

Comparing α_R for data and simulation with WP80 selection as well as the simulation with WP95 selection shows values well in agreement (simulation:$alpha_{R,WP95} = 0.202 + / - 0.0050$, $alpha_{R,WP80} = 0.203 \pm 0.0056$; data: $alpha_{R,WP80} = 0.20 \pm 0.028$). The comparison of α_L, however, shows significant differences between selections in the simulation (simulation:$alpha_{L,WP95} = 0.240 \pm 0.0037$, $alpha_{L,WP80} = 0.224 \pm 0.0040$; data: $alpha_{L,WP80} = 0.246 + / - 0.025$)). The reasonable agreement between data and MC for the tight selection inspires enough confidence in the simulation to use the value to use the simulated value of α_L for the signal extraction with the loose selection.

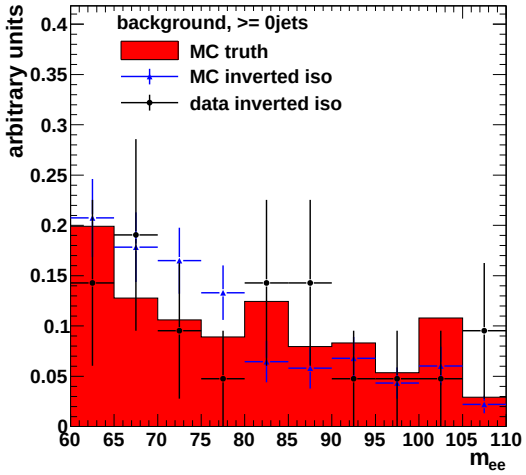


Figure 9: Invariant di-electron mass distribution for the sample with inverted track isolation in data and Simulation, compared to the true simulated background distribution. The plots are normalized. XXX REPLACE WITH MORE DATA XXX

Two options are explored for the treatment of the remaining uncertainty due to α_L/R: First, we check the dependence of measured yields on the event yields by redoing the fit with the parameters for $\alpha_{L/R}$ varied by $\pm 20\%$. Figure 10 shows the change in the observed yields as a function of the realtive change of α_L int the Z+1 jet bin.

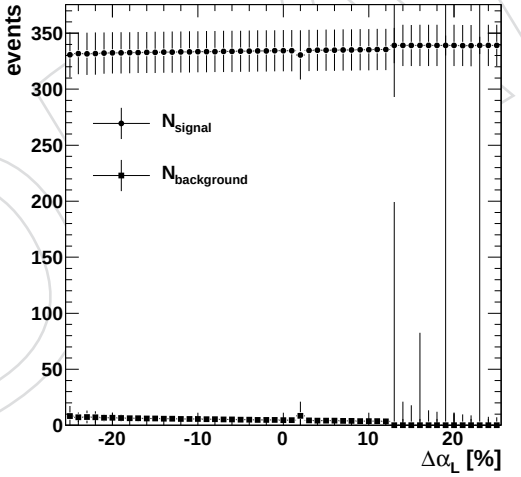


Figure 10: Signal and background yields for Z+1 or more jets as function of the variation of α_L.

Second, to ensure a proper combination with the other uncertainties, α_L/R are included in the global fit as "nuisance parameters": they are allowed to vary in the fit but are constrained by Gaussian PDFs, centered on the value obtained from simulation and a width of 20%.

5.4 Selection Efficiencies

To check that the selection efficiencies obtained from the simulation are reasonable, two control samples are obtained: In one sample the first leg of the Z is required to pass WP95 cuts while the

second leg must pass WP95 cuts, except track-, ecal- and hcal isolation. In the second sample the first leg of the Z is required to pass WP95 cuts while the second leg must pass WP95 cuts, except the ones on cluster shape, track-matching and H/E. These samples are used with the s-plot technique [10] to obtain the distributions of all the electron identification variables for the signal from the data itself. Comparing the result to the simulation (see figures 11 and 12) and previous results [6].

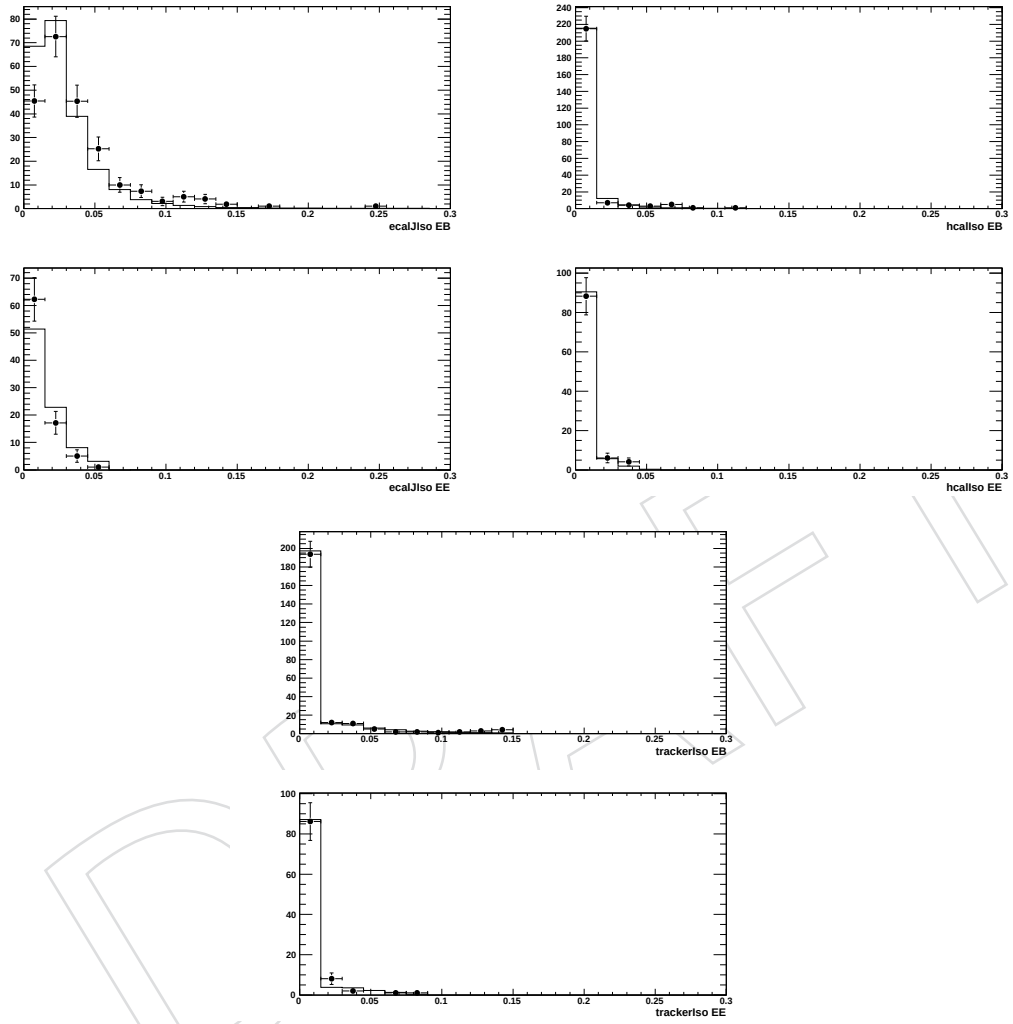


Figure 11: Splots for the isolation variables for Z plus one or more jet events (points), compared to the signal simulation (line). The simulation is normalized to the data.

6 Results from Early Data

Although the integrated luminosity collected at this point in time is much lower then the advertised in the Monte Carlo study, it is nevertheless instructive to apply the analysis to the sample of already collected data. Due to the low accumulated statistics, only results for PFjets$>$ 15 GeV are shown (XXX show 30GeV with final dataXXX).

Figures 13 and 14 show the raw invariant mass and N_{jet}^{ex} spectra of the data. The Simulation is in reasonable agreement with the data, though there are notable differences as well: The

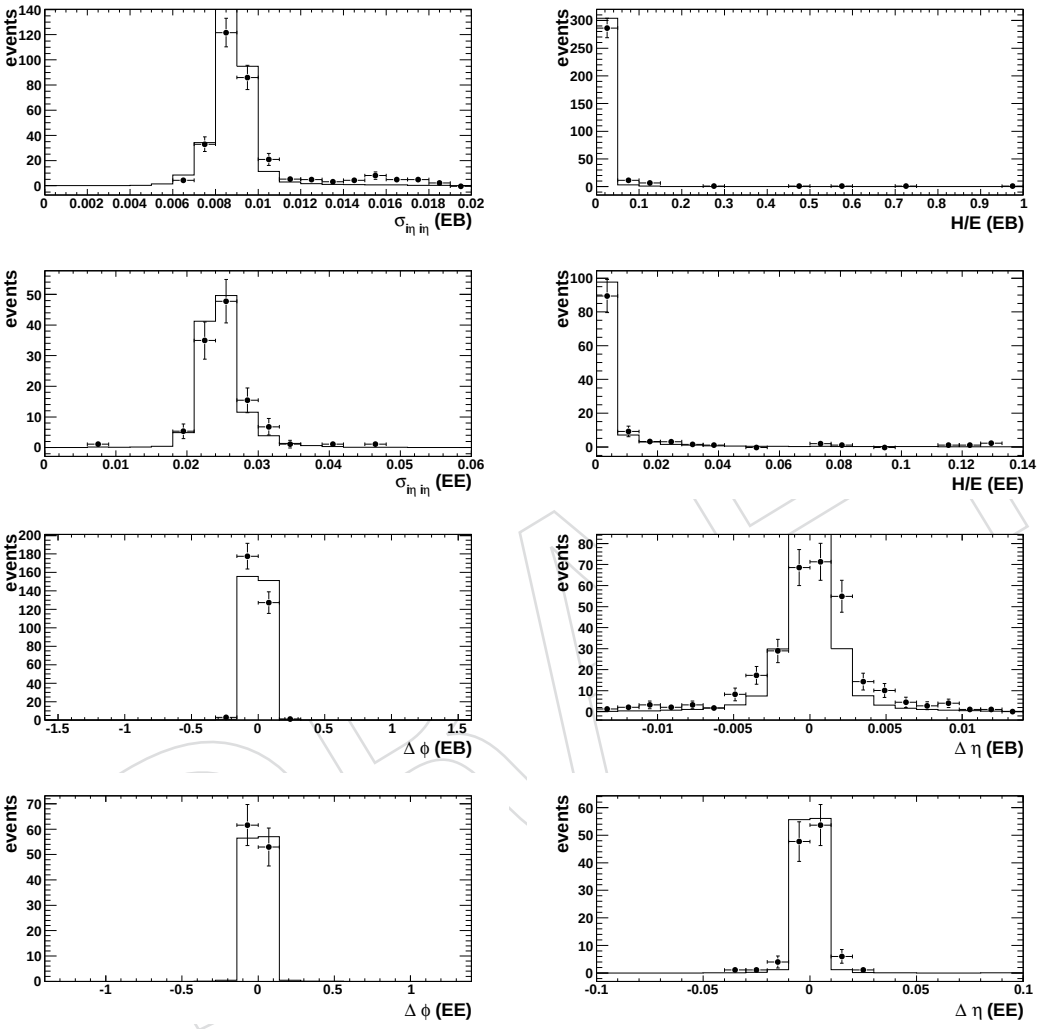


Figure 12: Splots for the electron identification variables for Z plus one or more jet events (points), compared to the signal simulation (line). The simulation is normalized to the data.

Mass peak of the Z is noticeably shifted to low masses, indicating imperfect ECAl calibration. Additionally, the total yield of events is overestimated in the simulation. As neither the total efficiency, nor the exact energy calibration have a large impact on the measurement of the ratios C_N, they are not discussed here in detail.

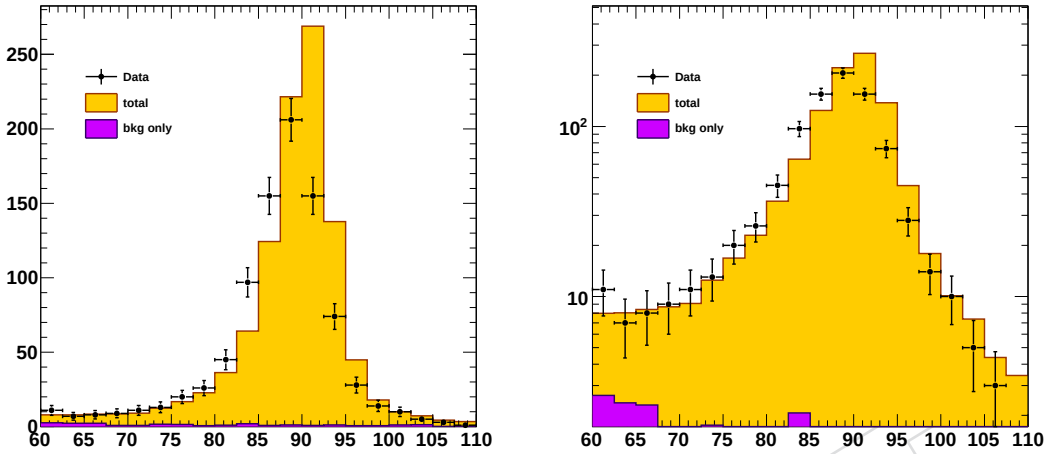


Figure 13: Di-electron mass spectrum of the selected events and in linear (left) and logarithmic scales (right).

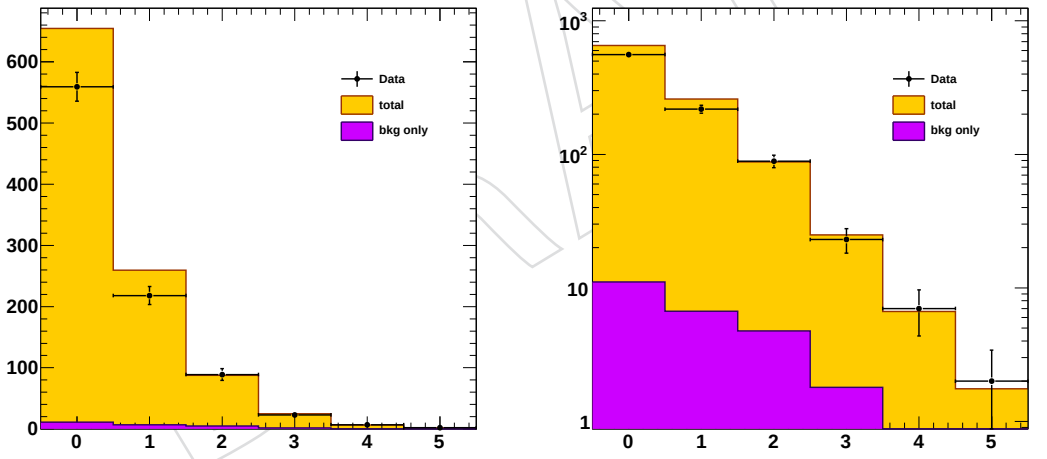


Figure 14: N_{jet}^{ex} mass spectrum of the selected events and in linear (left) and logarithmic scales (right).

Figure 15 shows the result of a differential fit to the yields discussed in section 4.2.4, with the resulting fit parameters listed in table 6. The result are reasonably consistent with expectations. The corresponding distribution of N_{jet} is shown in Figure 16, highlighting the different options to treat the uncertainties for $\alpha_{L/R}$.

The result of the linearly constrained fit as discussed in section 4.3.1 is shown in Figure 17. The datapoints for C_N are derived from the differential fit to the yields discussed above. The resulting linear relation is described by $2.84 \pm 0.27 + 0.20 \pm 0.17 \cdot N_{jet}$. (XXX 11pb number, up to 5 jets,

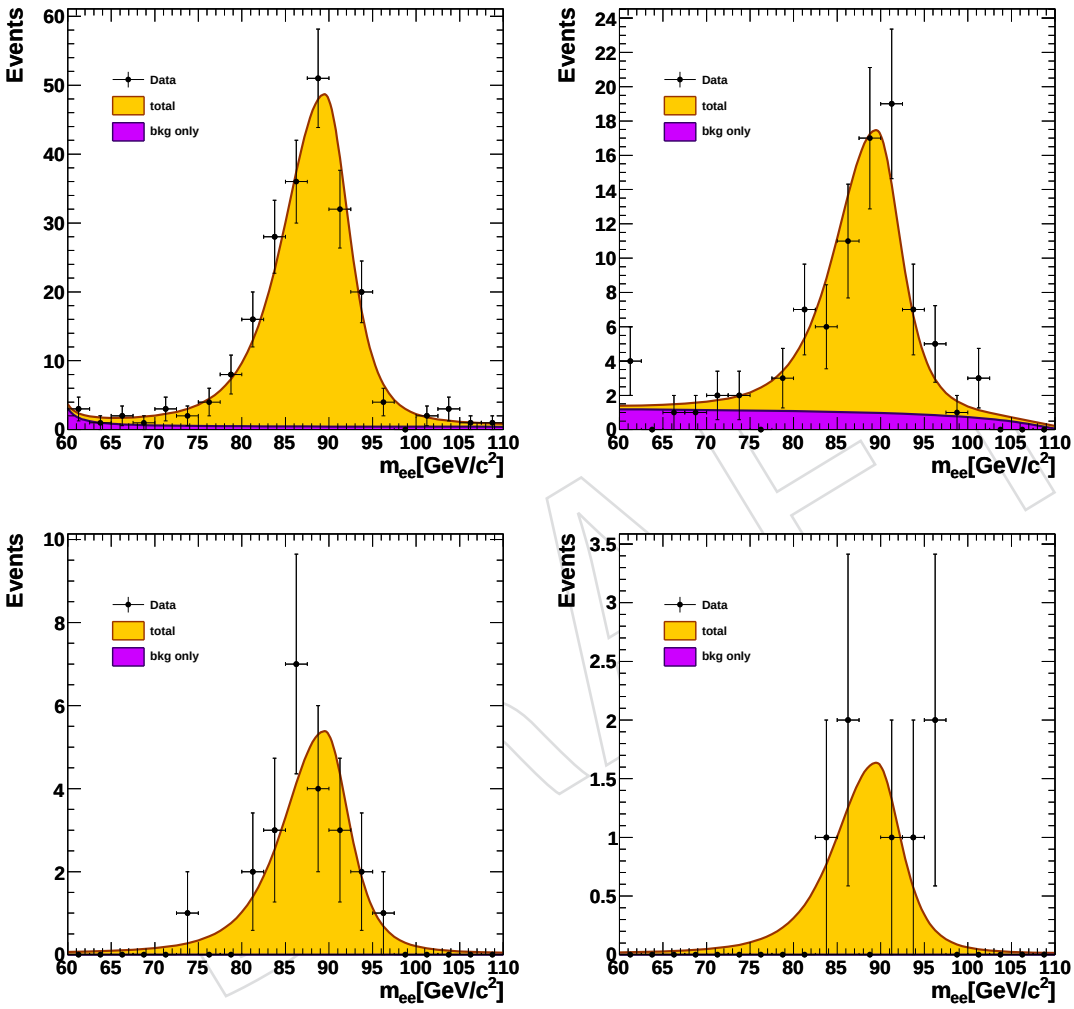


Figure 15: Fit result of the global fit in the first 4 bins of N_{jet}^{ex}.

Table 13: Parameters resulting from the global fit to inclusive yields (see section 4.2.4 and equations 1
and 3 for the definition of the parameters).

Signal	$N_{sig,1}$	309	$\pm$	22
	$N_{sig,2}$	103	$\pm$	12
	$N_{sig,3}$	32	$\pm$	5.7
	$N_{sig,4}$	9.0	$\pm$	3.0
	$N_{sig,5}$	2.0	$\pm$	1.4
	m	89.5	$\pm$	0.7
	σ_L	4.35	$\pm$	0.60
	σ_R	2.64	$\pm$	0.49
Background	a_1	-1.52313	$\pm$	3.5452
	b_1	0.30509	$\pm$	0.68133
	a_2	-343.634	$\pm$	325.14
	b_2	-0.0190000	$\pm$	0.0024061
	a_3	-35.0710	$\pm$	2282.4
	b_3	0.061001	$\pm$	229.82
	a_4	-40.1428	$\pm$	2280.0
	b_4	0.049183	$\pm$	229.82
	a_5	-41.8017	$\pm$	2279.3
	b_5	0.044838	$\pm$	229.82
	$N_{bkg,1}$	11.9	$\pm$	11.1
	$N_{bkg,2}$	18.5	$\pm$	7.7
	$N_{bkg,3}$	0.0	$\pm$	1.2
	$N_{bkg,4}$	0.0	$\pm$	1.1
	$N_{bkg,5}$	0.0	$\pm$	0.63

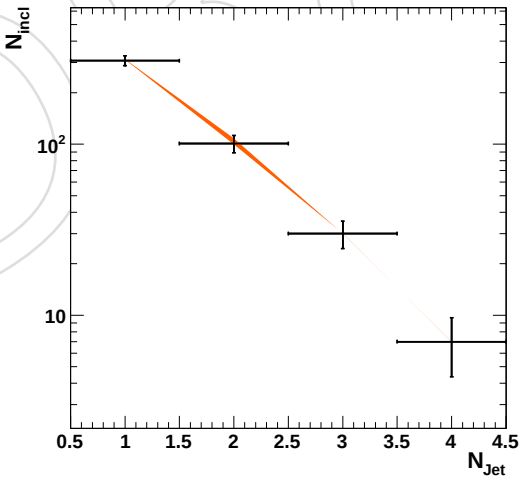


Figure 16: Fit result of the global fit in the first 4 bins of N_{jet}. The data-points and corresponding outer error bars show the results of the global fit, treating $\alpha_{L/R}$ as nuisance parameters. Fixing $alpha_{L/R}$ to this result yields the inner error-bars, showing the the contribution of the uncertainty due to $\alpha_{L/R}$ compared to the other uncertainties. The orange band represents the results of the global fit where $\alpha_{L/R}$ are fixed to the simulated values $\pm$20%.

harmonize XXX) The log-likelihood difference between the constrained und unconstrained fits is 1.04, indicating that the data are compatible with a linear realtionship with a probability of 59%. The Complete result in comparison to the simualtion is summarized in Figure 18.

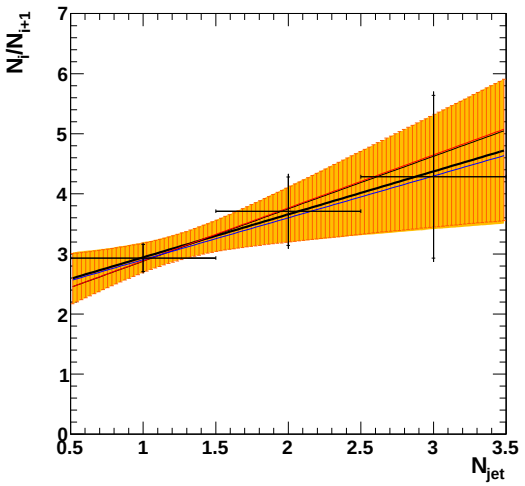


Figure 17: Fit result of the global fit of the ratios (thick black line). The datapoints are derived from the fit results in Figure 16. Note that the the linear fit and the datapoints are not directly related, but the results of different fits. The yellow errorband represents the full uncertainty on the global fit. The orange shaded region shows the uncertainty excluding the uncertainty due to $\alpha_{L/R}$ (obtained as the inner error bars in Figure 16). Fixing $\alpha_{L/R}$ ($\pm$20%) to the simulated vale yields the thin black (red/blue) line.

7 Conclusion

The observed values for the event yields in the $Z \rightarrow ee$ + jets channel have been observed to be compatible with expectations. The ratios of the $N/N+1$ jet yields are well described by a linear relation, similar to the observations in previous experiments. This suggests that the $Z \rightarrow ee$ + jet sample presented here will prove essential to estimate backgrounds in various searches with final states including electrons and jets as well as for detector calibration.

References

[1] CDF - Run II Collaboration, "Measurement of inclusive jet cross-sections in $Z/\gamma* \rightarrow e^+e^-$ + jets production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev. Lett.* **100** (2008) 102001, arXiv:0711.3717.

[2] D0 Collaboration, "Measurement of the ratios of the Z/G* + ¿= n jet production cross sections to the total inclusive Z/G* cross section in ppbar collisions at sqrt(s) = 1.96 TeV", *Phys. Lett.* **B658** (2008) 112–119, arXiv:hep-ex/0608052. doi:10.1016/j.physletb.2007.10.046.

[3] M. Bona, E. D. Marco, J. Lykken et al., "The Z+jets "candle" in dielectron+jets and dimuon+jets final states at CMS with pp collisions at center-of-mass energy 10 TeV.", *CMS Note* (2006).

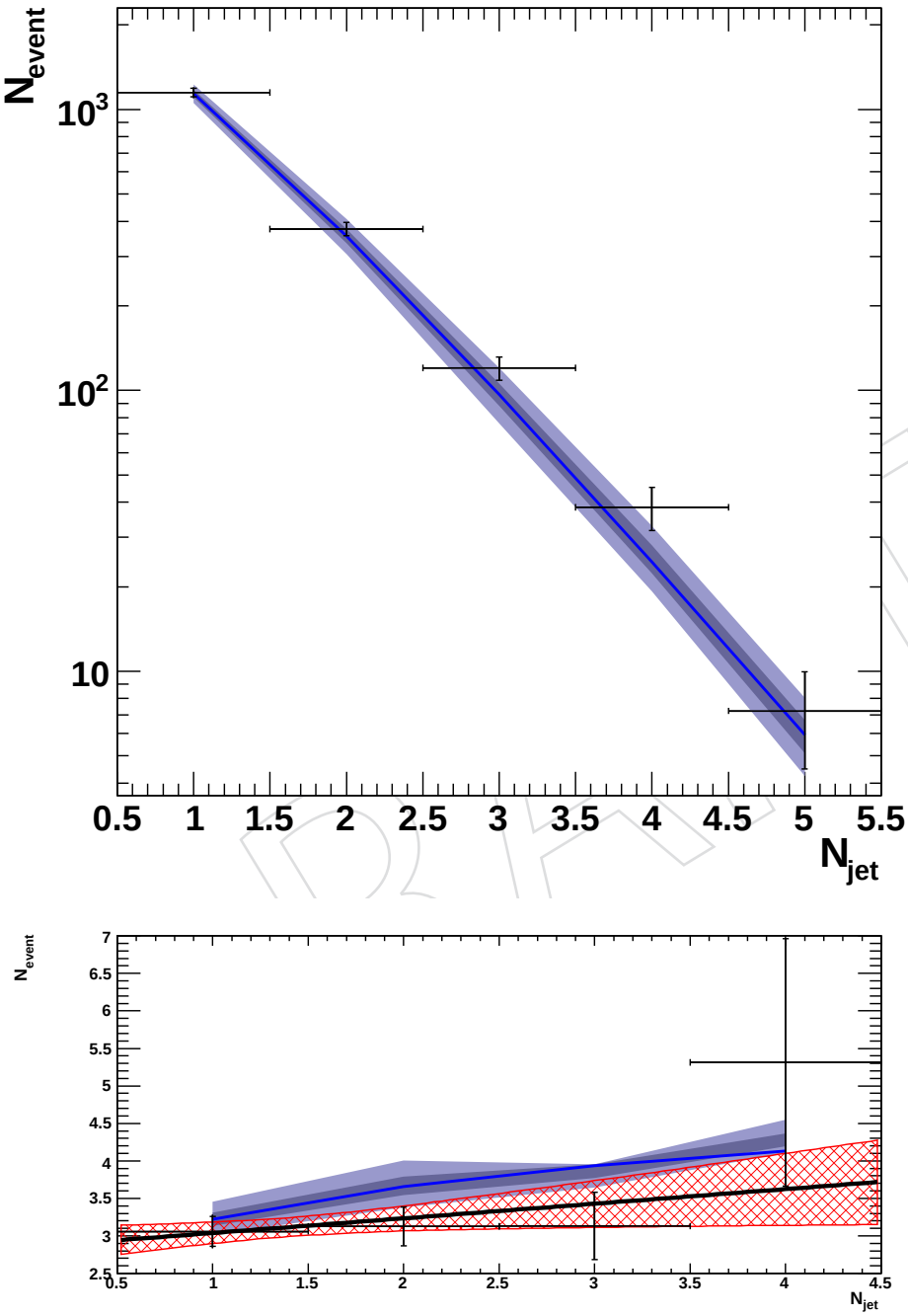


Figure 18: Fit result of the global fit of the yields (datapoints) compared to the simulation (blue line) and the linear constrained fit (black line). The blue bands indicate the jet energy scale uncertainty (inner band: η-dependent uncertainty, outer band:η-dependent and global uncertainty added in square). The red hatched band indicates the total uncertainty of the lionear constraint fit.

[4] W. Adam, S. Baffioni, F. Beaudette et al., "Electron Reconstruction in CMS", *CMS Note* **AN-2009-164** (2009).

[5] G. Daskalakis, N. Rompotis, and C. Seez, "Data driven selection cut tuning for electrons", *CMS Note* **CMS AN AN-09-108** (2009).

[6] T. V. B. T. F. of the CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s}$=7 TeV", *CMS Note* **CMS AN AN-10-116** (2010).

[7] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[8] S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses", *Ann. Math. Stat.* **9** (1938) 60–62.

[9] T. C. Collaboration, "Jet Performance in pp Collisons at $\sqrt{s} = 7$ TeV", *CMS PAS JME-10-003* (2010).

[10] M. Pivk and F. R. Le Diberder, "sPlot: a statistical tool to unfold data distributions", *Nucl. Instrum. Meth.* **A555** (2005) 356–369, arXiv:physics/0402083. doi:10.1016/j.nima.2005.08.106.