CMS Physics Analysis Summary

Contact: cms-pag-conveners-fsq@cern.ch 2013/09/17

Measurement of pure electroweak production of a Z boson in association with forward/backward jets in proton-proton collisions at $\sqrt{s}$=8 TeV

The CMS Collaboration

Abstract

The measurement of the electroweak production cross section of a Z-boson with two foward/backward jets (EWK Zjj) in proton-proton collisions at $\sqrt{s} = 8$ TeV is presented based on a data sample recorded by the CMS experiment at the LHC with an integrated luminosity of 19.7 fb^{-1}. Different methods are used to extract the signal, including a data-driven approach that models the QCD Zjj background based on γjj events. The cross section for this process is measured in di-electron and di-muon final states in the kinematic region $m_{\ell\ell} > 50$ GeV, $m_{\mathrm{jj}} > 120$ GeV, transverse momenta $p_{\mathrm{T}j} > 25$ GeV and pseudorapidity $|\eta_j| < 5$. The measurement, combining different methods and channels, yields $\sigma = 226 \pm 26_{\mathrm{stat}} \pm 35_{\mathrm{syst}}$ fb, in agreement with the theoretical cross section. The hadronic activity in events with Z-boson production in association with jets is also studied, in particular in the rapidity interval between the jets.

1 Introduction

In proton proton (pp) collisions at the Large Hadron Collider (LHC) the dominant source of production of a Z boson followed by a leptonic decay $Z \to \ell\ell$ in association with two jets is through mixed electroweak (EW) and strong (QCD) processes of order $\mathcal{O}(\alpha_{EW}^2\alpha_{QCD}^2)$, commonly known as Drell-Yan (DY) plus jets processes. Pure electroweak productions of the $\ell\ell jj$ final state, of order $\mathcal{O}(\alpha_{EW}^4)$, are more rare at the LHC [1], but are expected to carry a distinctive hallmark which can be explored experimentally: two jets of very high energy, well separated in pseudo-rapidity and with a large invariant mass, are expected to be produced in association with the dilepton pair [2, 3]. In the following, we refer to these jets that originate from the fragmentation of the outgoing quarks in EW processes as "tagging jets" and to the the process which originated them as EWK Zjj. Figure 1 shows representative Feynman diagrams for the EW production of dilepton pairs in association with two jets including (i) Vector Boson Fusion (VBF) processes, (ii) Z-bremsstrahlung processes, and (iii) multiperipheral processes, Detailed calculations reveal the presence of large negative interference between the pure VBF process and the other two categories [1, 3].

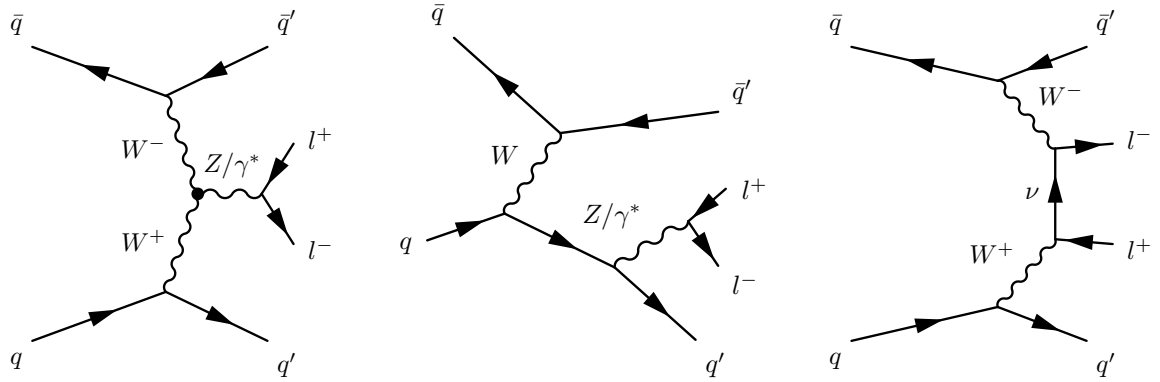

Figure 1: Representative diagrams for dilepton production in association with two jets from pure electroweak processes. Vector boson fusion (left), Bremmstrahlung-like (center), and multiperipheral (right) productions.

The study of these processes establish an important foundation for the more general study of vector boson fusion processes which include the Higgs boson [4, 5] as well as searches for beyond the standard model physics [6, 7]. If isolated successfully from the remaining backgrounds the properties of EWK Zjj processes can be compared with the standard model (SM) predictions. The selected events furthermore be used as a probe for anomalous triple-gauge-boson couplings [8] and for the investigation of the hadronic activity in the region defined in between the tagging jets [9, 10] to understand the possible vetoing of additional parton radiation between forward-backward jets in VBF processes.

At the LHC the EWK Zjj process has been measured by the Compact Muon Solenoid (CMS) experiment using proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ [11]. The result has been found to be in good agreement with the SM prediction. This work presents the preliminary measurement carried out using proton-proton collision data collected in 2012 at $\sqrt{s} = 8\,\text{TeV}$ by CMS, corresponding to an integrated luminosity of $19.7\,\text{fb}^{-1}$. As the signal-to-background ratio for the cross section measurement is small, two methods of signal extraction are employed to confirm and cross-check the presence of the signal. A multivariate analysis similar to the one used to analyse 7 TeV data [11] is repeated with the 8 TeV data. In addition a new method, based on data to model the main background, is proposed and its limitations are discussed. The two results are combined to yield the final measurement.

We also present a separate study of the hadronic activity in DY plus two jets events, including the level of hadronic activity in the rapidity interval between the two tagging jets.

2 CMS detector, event simulation, reconstruction, and selection

A detailed description of CMS can be found in Ref. [12]. The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point, the x-axis pointing to the center of the LHC ring, the y-axis pointing up, and the z-axis along the counterclockwise-beam direction as viewed from above. The polar angle θ is measured from the positive z-axis and the azimuthal angle ϕ is measured in the x–y plane. The pseudorapidity η is defined as $-\ln(\tan(\theta/2))$, which equals the rapidity $y = \ln((E + p_z)/(E - p_z))$ for massless particles.

In this analysis we make use of reconstructed charged leptons (electrons or muons) and jets. Muons are reconstructed [13] by fitting trajectories based on hits in the silicon tracker and the muon system. Electrons are reconstructed [14] from clusters of energy deposits in the electromagnetic calorimeter matched to tracks in the silicon tracker. Two different types of jets are used in the analysis: jet-plus-track (JPT) and particle-flow (PF) jets [15]. In both cases the anti-k_T algorithm with a distance parameter of 0.5 is used [16]. The JPT jets are reconstructed from calorimeter jets whose energy response and resolution are improved by incorporating tracking information according to the JPT algorithm [17]. The PF algorithm [18, 19] combines the information from all relevant CMS sub-detectors to identify and reconstruct particle candidates in the event: muons, electrons, photons, charged hadrons, and neutral hadrons. The PF jets are reconstructed by clustering these particle candidates.

Signal events are simulated with the MADGRAPH (v5.1.3.30) generator [20], interfaced with PYTHIA 6.4.26 [21] for parton showering (PS) and hadronization. The CTEQ6L1 parton distribution functions [22] are used in the event generation.

The EWK $\ell\ell$jj leading order (LO) cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV evaluated by MADGRAPH is $\sigma^{\mathrm{LO}}_{\mathrm{EWK}\ \ell\ell\mathrm{jj}} = 213$ fb for an invariant mass of the two leptons $M_{\ell\ell} > 50$ GeV, an invariant mass of the dijet $M_{\mathrm{jj}} > 120$ GeV, a transverse momentum of the tag jets $p_{\mathrm{T}j} > 25$ GeV and pseudo-rapidity $|\eta_j| < 5$. In the definition of the phase space the dijet separation is required to be $\Delta R_{jj} = \sqrt{\Delta\eta_{jj}^2 + \Delta\phi_{jj}^2} > 0.5$, where $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ are the pseudo-rapidity distance and the difference in azimuthal angle of the tag jets, respectively. The leading order (LO) cross section is verified to agree with a computation performed using the VBFNLO (v.2.6.3) generator [23–25]. Based on the latter we estimate the next-to-leading order (NLO) to LO scale factor to be 1.12, mostly due to real additional QCD radiation emissions.

Background from DY processes, that include the production of two isolated leptons with large transverse momentum p_{T} and extra jets produced from QCD interactions, are generated with MADGRAPH via a matrix element (ME) calculation that includes up to four jets at parton level. The ME and parton shower (ME-PS) matching is performed following the ktMLM prescription [20]. The generation of the DY $\ell\ell jj$ background does not include the signal electroweak production of the Z boson with two jets. The production cross section is normalized to 4.393 nb as computed at next-to-next-leading order (NNLO) with FEWZ [26], for dilepton invariant masses above 10 GeV. Dynamical scale factors, obtained with MCFM [27], are used to re-weight the simulation to the NLO prediction. The re-weighting procedure is observed to improve greatly the comparison with the data and will be discussed in more detail in Section 3.

The background from top pair production ($\mathrm{t\bar{t}}$), are also generated with MADGRAPH and normalized to the NNLO+NNLL (next-to-next-to-leading-logarithmic) total cross section of 245.8 pb

[28]. Diboson production processes WW, WZ, and ZZ are generated with PYTHIA and normalized to the NLO cross section computed with MCFM of 54.8 pb, 33.2 pb and 17.7 pb, respectively. Single-top processes are generated with POWHEG (v1.0, r1380) [29], and normalized to their approximate NNLO cross sections of 22.2 pb, 87.1 pb and 5.55 pb, respectively for the tW, t and s production channels [30].

The production of a W boson in association with jets is also generated with MADGRAPH and normalized to a total cross section of 36.3 nb computed at NNLO with FEWZ. Multijet QCD processes are also simulated with MADGRAPH, but are found to yield negligible contributions to the event selections.

Where available, the theoretical prediction for each process listed above has been verified to be in agreement with the experimental measurements performed by the CMS experiment [31–34]. The theoretical predictions are however assumed in our analysis given they have smaller uncertainty with respect to the experimental measurements.

Generated events are processed through the full CMS detector simulation based on GEANT4 [35, 36], followed by a detailed trigger emulation, and the standard event reconstruction. Minimum-bias events are superimposed upon the hard interaction to simulate the effects of additional interactions per beam crossing (pileup). The multiplicity distribution of the pileup events in the simulation is matched with that observed in data. An average pileup rate of $\approx$23 additional collisions is estimated from data.

The event selection is optimized to identify the dilepton final states, with two prompt isolated leptons and at least two jets. Double lepton triggers are used to acquire the data samples where a minimum transverse momentum (p_{T}) of 8 and 17 GeV are required for each lepton. Electron-based triggers include additional isolation requirements, both in the tracker and calorimeter detectors. A single isolated muon trigger with a p_{T} >24 GeV is also used to acquire the data.

Reconstructed electron or muon candidates are required to have p_{T} >20 GeV. Muon candidates are required to be reconstructed in the fiducial region $|\eta| \leq 2.4$ of the tracker system and electron candidates are required to be reconstructed within $|\eta| \leq 2.5$ with the exception of the barrel-to-endcap transition of the electromagnetic calorimeter [12]. The track associated to each lepton candidate is required to have an impact parameter compatible with prompt production from the primary vertex. The primary vertex (PV) used as a reference for each event is chosen to be the one which has the largest $\sum p_{\mathrm{T}}^2$ where the sum runs over all the tracks used in the vertex fit. A particle-based relative isolation is computed for each lepton and it is corrected on an event-per-event basis for the contribution from pileup events. We require that the scalar sum of the p_{T} of all particle candidates reconstructed inside a cone with radius R = $\sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ around the lepton's thrust axis is less than 10% (12%) of the electron (muon) transverse momentum. In the case more than two isolated leptons are found, the ambiguity is resolved by selecting the dilepton pair with opposite electric charge and highest sum of transverse momenta. Same-flavor dileptons (ee or $\mu\mu$) compatible with $Z \to \ell\ell$ decays are selected by requiring $|M_Z - M_{\ell\ell}|$ <15 GeV.

For PF-jets the momentum is defined as the vectorial sum of the momenta of all particle candidate clustered in the jet. For JPT jets the associated tracks are classified as in-cone (out-of-cone) if they point inwards (outwards) the jet cone at the calorimeter surface, and the direction of the jet is re-calculated with the tracks. As a result of the JPT algorithm, both the energy and the direction of the jet are improved. In both cases an offset correction is applied to take into account the extra energy clustered in jets due to additional proton-proton interactions within the same bunch crossing [37, 38].

In the offset subtraction we also include charged particle candidates associated to secondary vertices reconstructed in the event. Jet energy-scale corrections are derived from the simulation, and are confirmed with in situ measurements of the energy balance of dijet and photon+jet events. Additional selection criteria are applied to each event to remove spurious jet-like features originating from isolated noise patterns in certain HCAL regions. Jet identification criteria are furthermore applied to remove contribution from pileup jets. These criteria are described in more detail in [39]. The data-to-simulation ratio of the efficiency of these algorithms is measured in data and observed to be close to 1 accross the full rapidity range used in the analysis.

In the pre-selection of the events we consider events with at least two jets with p_{T} >30 GeV and $|\eta|$ ≤4.7 and the two leading p_{T} jets are chosen as the tagging jets. For the measurement of the cross section we require the leading jet to have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$.

3 Control regions and background modeling

As specified previously in Section 2, we define our signal region after the selection of a Z boson candidate produced in association with two jets. Control regions are used in our analysis to validate the jet calibrated energy response and efficiency, to derive data-driven background templates and finally to constraint partially the rate and uncertainties affecting the background estimation.

3.1 Jet identification and response

Events with a reconstructed $\mathrm{Z} \to \mu\mu$ or photon candidates produced in association with a single jet with p_{T} >30 GeV are used as a control region in our analysis. The boson candidate and the jet are required to be found in a back-to-back configuration with $|\Delta\phi(\text{jet}, \mathrm{Z} \text{ or } \gamma)| >$ 2.7 rad and two measurements are performed on the reliability of the simulation regarding the efficiency of the algorithms used to reject calorimetric noise and pileup-induced jets and the jet energy response with respect to the reconstructed transverse momentum of the boson candidate.

The jet identification criteria is based on the fraction of energy deposited in the different calorimeters or different calorimeter segmentations. A multivariate analysis based on the kinematics of the jet, on the topological configuration of the energy deposits as well as on the tracks which can be associated to the jet and to a secondary vertex of the event is furthermore used to discriminate pileup-induced jets [39]. The inclusive efficiency measured in the control sample is >98% for both types of jets and its' dependency on the pseudo-rapidity of the jet is observed to be within 1% of the one predicted by simulation. The η-dependent differences are assigned as a systematic uncertainty in the signal selection

The same control region is furthermore used to verify the standard jet energy response [40]. The jet response is defined as $R = p_{\mathrm{T}}(\text{jet})/p_{\mathrm{T}}(\mathrm{Z} \text{ or } \gamma)$ and from the double ratio $R_{\text{data}}/R_{\text{simulation}}$ we extract a residual uncertainty which is assigned as a systematic uncertainty. Although partially covered by the standard jet energy scale uncertainties, this procedure allows us to consider possible residual uncertainties stemming from the particular phase space region being selected in our analysis. This is particularly crucial in the forward region, where the uncertainties in response are larger due to instrumental sources of uncertainty. Overall the double ratio $R_{\text{data}}/R_{\text{simulation}}$ is close to 1 with a small loss of ≈ 5% in the region where the tracker looses its acceptance (2.7< $|\eta|$ <3.2).

3.2 Background modeling

The main background to the EWK Zjj process is from QCD Zjj production. Two alternative background models are explored.

3.2.1 Method I: simulation-based prediction

The standard simulation of QCD Zjj is based on MADGRAPH and lacks therefore possible virtual corrections at higher orders then the Born level. We have studied these corrections by comparing the predictions using MCFM at parton-level with the ones obtained from MADGRAPH after showering with PYTHIA. The choice is dictated by the fact that the jet-parton matching (CKKW reweighting [41]) needs to be performed in order to include the NLO effects for real emmissions. Based on this study we derive an dynamical LO to NLO k-factor based on M_{jj} and on the rapidity of the Z boson in the dijet rest frame (y^*). The k-factor is observed to have a small dependency on M_{jj}, but to increase steeply with y^*. A correction >20% is obtained for $y^* > 2.5$. The k-factor is used to re-weight the simulation on an event-per-event basis and an uncertainty. Half the difference between the simulated prediction with and without the re-weighting is assigned as systematic uncertainty to the method. Figure 2 compares the final distribution of the dijet invariant mass and the y^* variable after the re-weighting procedure is applied. A fair agreement is observed between data and simulation. We notice that the background prediction for the last bin in the M_{jj} distribution, although not represented due to limited statistics in simulation, is estimated to be ≈0.5 events.

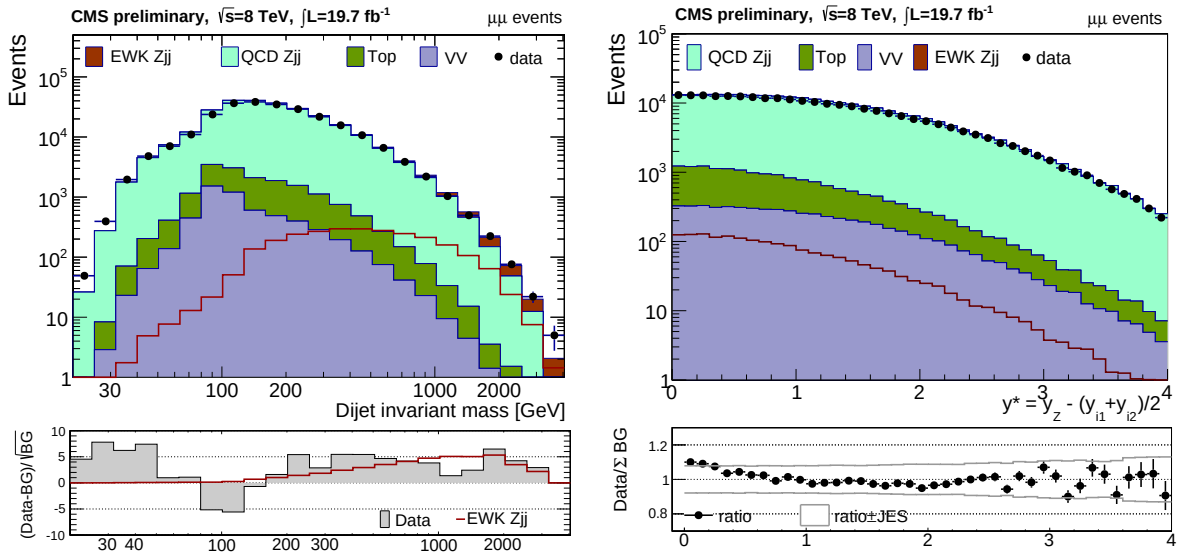

Figure 2: Distribution for the dijet invariant mass (*left*) and y^* (*right*) variables in di-muon+2 JPT jet events. The contributions from the different background sources and the signal are shown stacked and data is super-imposed. The bottom panels show the ratio between the data residuals after subtracting the background expectations and the square root of the number of background events (*left*) or the ratio of data to total events predicted (*right*). In the latter the jet energy scale uncertainty envelope is shown.

3.2.2 Method II: data-driven prediction

We select in data a γ+2 jets sample in similar way as the Z+2 jets sample. We expect the QCD production of γ+2 jets resembles the production of QCD Zjj yielding therefore a possible data-based model for the shapes of the kinematics of the tag jets. The differences which may be

induced by the specific Z or γ sample selections are mostly mitigated by re-weighting the p_{T} of the photon to the transverse momentum of the Z candidate. The difference in the mass is not observed, in simulation, to entail a difficulty in the identification of the dijet kinematics in both samples for $M_{\mathrm{jj}} > 2M_{\mathrm{Z}}$. Given however that at low p_{T} the photon sample is non-negligibly affected by multijets production and given that the selection of the low p_{T} region in data is also affected by very high trigger prescale we impose tighter kinematics constraints on the reconstructed boson. To fully match the Z and γ kinematics we require $p_{\mathrm{T}}(\mathrm{Z\ or\ }\gamma) > 50\,\mathrm{GeV}$ and $|y(\mathrm{Z\ or\ }\gamma)| < 1.442$. In the massless particle limit, the cut on the pseudo-rapidity corresponds to the physical boundary of the central (barrel) region of the electromagnetic calorimeter of CMS [12]. These requirements are furthermore propagated to the analysis of dilepton events in Method II.

The γ+jets method is firstly tested in simulation by categorizing the Z or γ+2 jets samples in different regions and comparing the target and data-driven templates. An overall good agreement is found for the different tag jet kinematics predictions as well as for global event observables such as energy fluxes, angular correlations, amongst others. Figure 3 illustrates the result of one of the compatibility tests performed in simulation for events with a high dijet invariant mass. We illustrate the good agreement found for the description of the $\Delta\eta_{\mathrm{jj}}$, $\Delta\phi_{\mathrm{jj}}$ and the relative p_{T} the dijet system with respect to the individual jets p_{T} (i.e. $\Delta^{\mathrm{rel}}_{\mathrm{p_T}} = p_{\mathrm{T}}(j_1, j_2)/[p_{\mathrm{T}}(j_1) + p_{\mathrm{T}}(j_2)]$). In this analysis PF jets are used.

In an ideal situation, we could rely on the results of the compatibility tests described above to derive a correction factor to be applied to the estimate of the QCD Zjj production shapes to be perfomed in data. In practice, due to the limited amount of MC events in the phase space of interest and also due to significant uncertainties on handling the contamination from multijets events in data, we prefer to use the simulation-based compatibility test results to assign an uncertainty on the final shape. We consider therefore the difference in the compatibility tests with respect to a pure prompt photon scenario as one of the systematic uncertainties and the stability of the compatibility tests versus PDF variations as another source of uncertainty. In data we use furthermore the difference between a tight and a loose photon selection to build the uncertainty envelope which is used to cover the effect of the contamination of multijets.

The final shape for QCD Zjj, derived from data, is obtained after subtracting a residual contamination from EWK production of a γ with two jets (EWK γjj) [42]. To estimate the fraction of such events we have used the MADGRAPH generator. For a generator phase-space defined with $m_{jj} > 120\,\mathrm{GeV}$, $p_{\mathrm{T}j} > 30\,\mathrm{GeV}$, $|\eta_j| < 5$, $p_{\mathrm{T}\gamma} > 50\,\mathrm{GeV}$ and $|\eta_\gamma| < 1.5$ the cross section is expected to be 2.72 pb. After the standard event reconstruction and selection, we estimate that the ratio between the number of EWK γjj candidate events to the total number of photon events selected in data, is $\approx$5 times smaller than the ratio between the expected EWK Zjj and QCD Zjj yields. This ratio is furthermore expected to be stable, independently of the dijet invariant mass. In the subtraction procedure, besides taking into account the simulation statistics, a 30% uncertainty normalization uncertainty is assigned to this process. This corresponds to approximately twice the envelope of variations obtained after computing the cross section for this process at NLO with VBFNLO and after tightening the selection cuts or varying the factorization and renormalization scales.

After the pre-selection of Z+2 jets events, the events are used as a control region, as in Method I. Before looking for the discrimination of signal from the background, we compare the observed distributions for the kinematics of the tag jets with the predictions either from simulation or from the photon-based templates. Figure 4 summarizes the reconstructed kinematics of the tag jets. A fair agreement is observed for the kinematics as described by the photon-based predic-

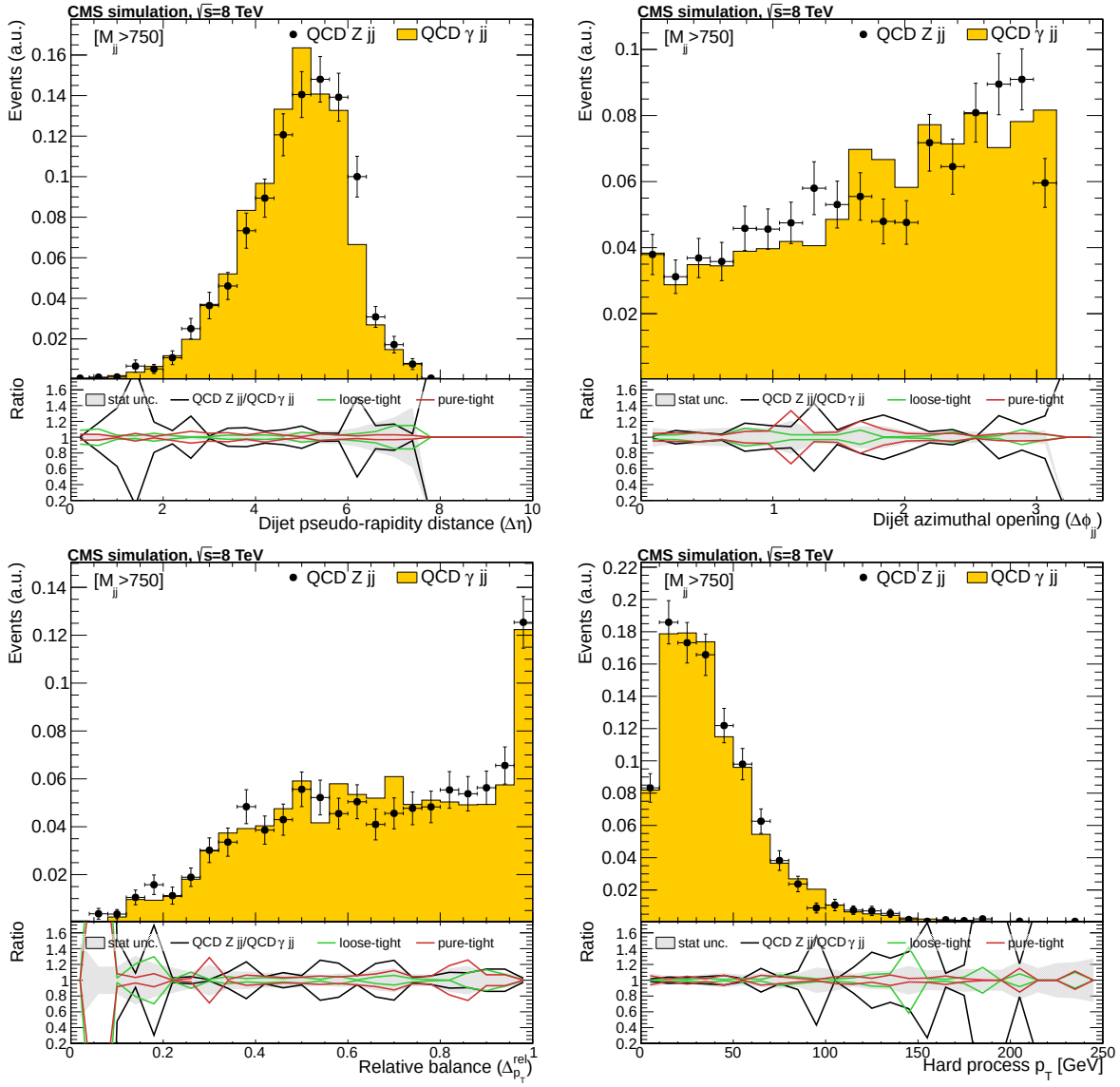

Figure 3: Simulation-based compatibility tests between the photon-based data-driven prediction for events with $M_{jj} > 750\,\text{GeV}$ using the PF jet algorithm. *Top*: the dijet rapidity distance is shown on the *left* while the difference in azimuthal angle is shown on the *right*. *Bottom*: the relative p_T balance of the jets is shown on the *left* while the transverse momentum of the hard process (dijet+Z system) is shown on the *right*. The QCD γjj distribution contains the contribution from prompt and fake photons as estimated from simulation and it is compared to the simulated QCD Zjj sample. The bottom panel shows the ratio between the target distribution and the photon-based prediction. The estimated contributions to the uncertainty on the γ-based shape are shown separately in the bottom panel.

tion in Method II. Good agreement is also observed in different dijet invariant mass categories but omitted for simplicity.

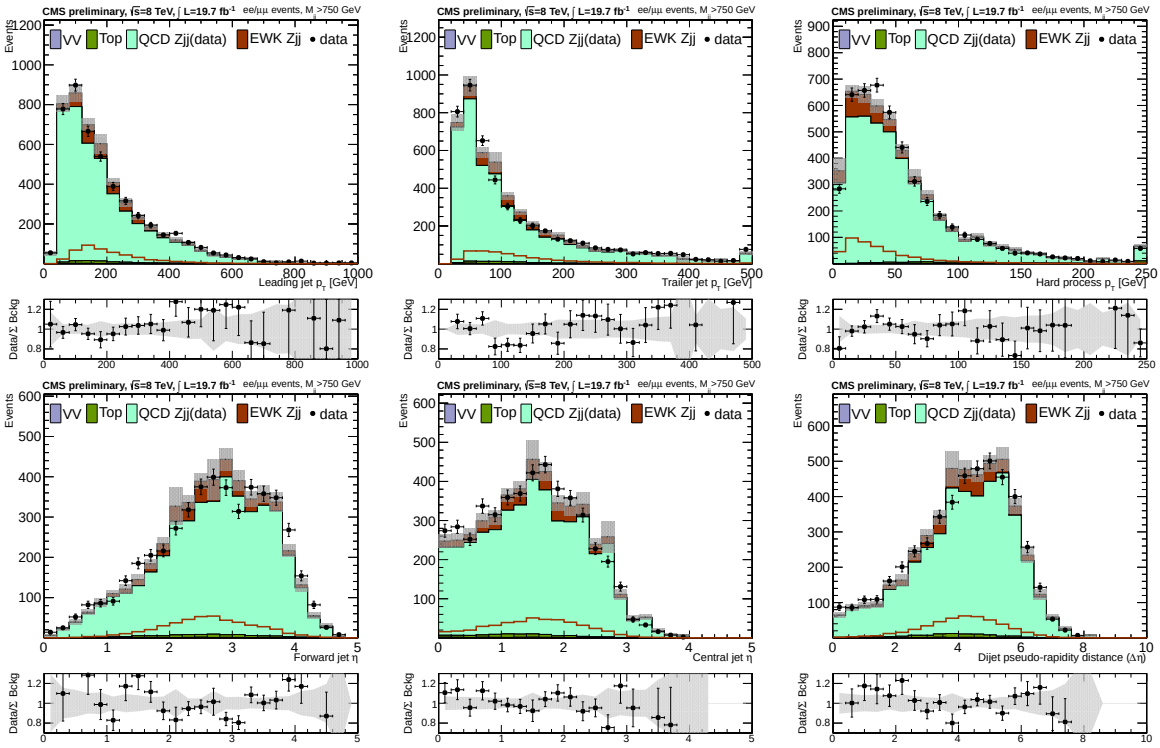

Figure 4: Distributions for the tag jets at pre-selection level. Both di-muon and di-electron plus two PF jets with M_{jj} >750 GeV events are used. *Top* - transverse momentum of the leading (*left*) and sub-leading (*center*) PF jets and of the hard process (dijet+Z system) (*right*). *Bottom* - pseudo-rapidity of the most forward (*left*) and most central (*center*) jets. The rapidity distance of the jets is shown on the *right*. The contributions from the different background sources and the signal are shown stacked and data is super-imposed. The bottom panels show the ratio of data to total prediction. The total uncertainty from the data-driven background prediction obtained from the photon control sample is shown as a shaded band in both the distribution and the bottom panel.

4 Signal discrimination

The EWK Zjj signal is characterized by a large jet separation ($\Delta\eta_{jj}$) which stems from the small-angle scattering of the two initial partons. Owing to both the topological configuration and the large transverse momentum of the outgoing partons, M_{jj} is also expected to be large. These characteristics are expected to yield the best separation power with respect to QCD Zjj. Besides these one can furthermore explore the fact that the Z boson candidate is expected to be produced centrally in the rapidity gap distance region. Defining $y^* = y_Z - \frac{1}{2}(y_{j_1} + y_{j_2})$ as the rapidity of the Z boson candidate in the dijet rest frame we expect the signal to be found with lower values of y^* with respect to the main background.

Two different analyses, exploring distinct discrimination aspects mentioned above, are carried out with the purpose of cross-checking independently the final result.

The first analysis (Method I) follows the procedure previously adopted for the 7 TeV measurement [11]. It uses the JPT-jet approach and builds a multivariate discriminator [43] based on

simple discriminating variables which reflect mostly the kinematics of the tag jets and of the Z-boson. The variables used to train a Boosted Decision Tree (BDT) in the discrimination of the signal are: the transverse momentum of the tag jets and of the Z-boson candidate; the dijet invariant mass, rapidity distance, rapidity sum and azimuthal angle difference; the difference in rapidity between each tag jet and the Z-boson candidate; the rapidity of the Z-boson measured in the laboratory and in the dijet rest frame. The description of each variable is verified by comparing data with the simulation-based expectations in the pre-selection control regions. In this analysis all processes are modeled from simulation and only the di-muon final state is considered.

The second analysis (Method II) explores both $Z \to ee$ and $Z \to \mu\mu$ decay channels, makes use of the PF-jet approach and a data-driven prediction of the main background based on the photon control sample. As explained in Section 3, in this case we tighten the selection of the Z-boson kinematics to obtain a better equivalence in the γ+jets sample. A linear Fisher discriminator [44] based on the dijet kinematics is used. The discriminator makes use of $\Delta\eta_{jj}$, M_{jj} and $\Delta^{rel}_{p_T}$.

Figure 5 shows, at pre-selection level, the distributions of the discriminators used in the two different analyses. A reasonable agreement is observed overall.

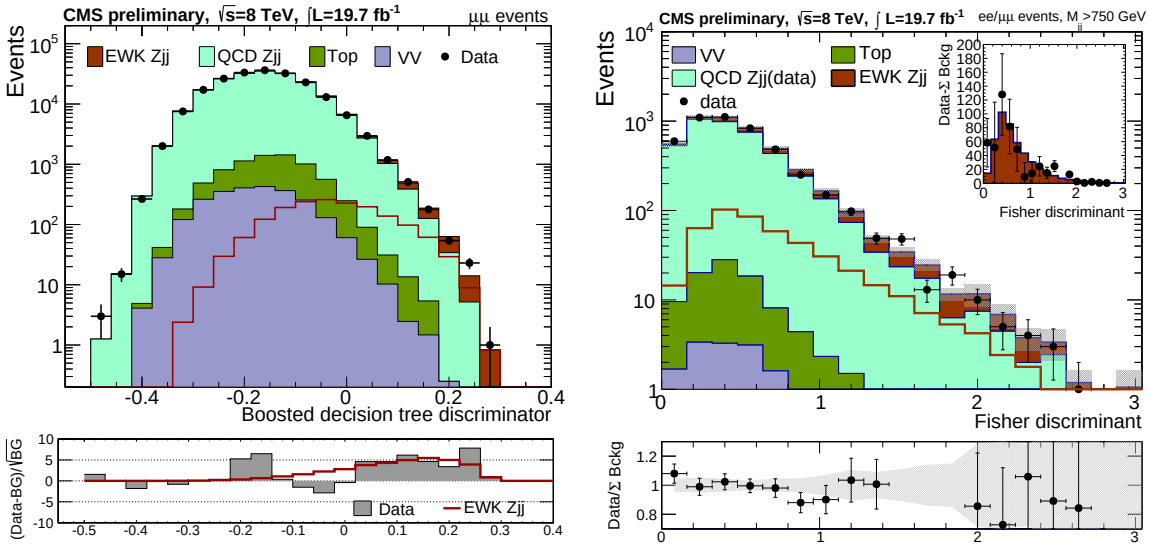

Figure 5: Shape discriminators used in the two different analysis described in the text: full kinematics BDT (*left*) and dijet Fisher (*right*) discriminators. The *right* figure depicts the discriminator shape in the most sensitive region (M_{jj} >750 GeV) and the inset displays the result of a bin-by-bin background subtraction. The total uncertainty from the data-driven background prediction obtained from the photon control sample is shown as a shaded band in both the distribution and the bottom panel. Both dilepton channels are included in the Fisher discriminator while only di-muon final states are represented for the BDT discriminator.

The discriminator shapes are used to extract the signal strength, i.e. $\mu = \sigma/\sigma_{th}$ from a templated fit to the data. In the case of Method I a χ^2 fit is performed to the inclusive data and the systematic uncertainties are evaluated separately. In the case of method II the systematic uncertainties affecting the rate and the shapes are parametrized as function of nuisance parameters and a profile likelihood ratio method is used to extract the signal strength [45, 46]. In this case the fit is performed simultaneously to 2×6 categories corresponding to the two dilepton channels and six dijet invariant mass categories. The technique is used to constrain partially the systematic uncertainties affecting the measurement.

In the next section we discuss in detail the systematic uncertainties considered in our analysis.

5 Systematic uncertainties

In this section we describe the main systematic uncertainties affecting our measurement. Besides the statistical uncertainty of the fits to be performed, one has to consider the unknown aspects which may influence both the rate and the kinematics of the predictions for the signal and backgrounds.

5.1 Experimental uncertainties

The following experimental uncertainties are considered:

Luminosity - a 2.6% uncertainty on the integrated luminosity of the analyzed data is assigned based on [47].

Trigger and selection efficiencies we assign a total 2% (3.5%) uncertainty on the total trigger and selection efficiency in the di-electron (di-muon) channel.

Jet energy scale and resolution the energy of the jets enters in our analysis not only at selection level but also in the computation of the kinematics variables used for discrimination. Thus, the jet energy scale uncertainty (JES) affects both the event yield by migration of events and the final discriminator shapes. The effect of the jet energy scale uncertainty is studied by rescaling up and down the reconstructed jet energy by a p_{T} and η dependent scale factor [40]. Similar approach is used for the jet energy resolution (JER).

Residual uncertainty on jet response the residual difference in the response observed in the balancing of a Z or γ candidate with a jet discussed in Section 3 is assigned as a conservative estimate on top of the standard jet energy scale uncertainties.

Pileup pileup is not expected to affect the identification and isolation of the leptons neither the corrected energy of the jets. It may however induce different dijet topologies due to the contamination of tracks and calorimetric deposits when running a clustering algorithm. We evaluate this uncertainty by generating two alternative shapes by re-weighting the observed pileup by 6%.

Simulation statistics for signal and backgrounds assumed from simulation we build a conservative envelope for the shapes by shifting each bin simultaneously by $\pm$1-σ_{stat}. This generates two alternatives to the nominal shape to be analyzed.

5.2 Theoretical uncertainties

We have considered the following theoretical uncertainties in the analysis:

PDF PDF uncertainties are evaluated by considering the diagonalized uncertainty sources of the CTEQ6L1 PDF set [22]. For each source a new weight is derived event by event and used to generate an alternative signal shape. The difference with respect to the nominal prediction is maximized (minimized) for each independent 90% CL variation and added in quadrature to estimate the final uncertainty.

Factorization and renormalization scale unlike the main background, the two jets in the signal have its origin on an electroweak vertex. Therefore changing the QCD scales is not expected to have a great impact on the final cross section. In VBNLO the default factorization (μ_F) and renormalization (μ_R) scales are set to . We change the default choice by rescaling $Q = \mu_R = \mu_F$ by a factor of 2 and 1/2. This variation translates to a change in the absolute cross section of $^{+4.0}_{-3.8}$% at LO and $^{+0.2}_{-0.5}$% at NLO. The variation of the scale can also change slightly the observed kinematics. We choose to parametrize this change using the generator level dilepton p_{T}. Given that our model signal is generated at LO we use the varied Q^2 samples at LO to parametrize the uncertainty envelope due to the choice of these quantities. The impact is propagated by re-weighting the generator level dilepton p_{T} on an event-by-event basis.

Background normalization Di-boson and top processes are modeled based on simulation. We assign therefore an intrinsic normalization on their uncertainty due to the PDF+QCD scale uncertainties. The uncertainties are assigned based on [27, 28, 30]. In practice the contribution from these processes is very small and therefore we assume other uncertainties affecting their kinematics can be neglected.

5.3 Uncertainty on QCD Zjj

The uncertainty on the prediction of the main background (QCD Zjj) deserves a separate discussion depending on whether it has been modeled from data or simulation.

For the simulation-driven modeling of this background in Method I, we consider the instrumental variations described previously (namely JES, JER and pileup). On top of these variations we also consider the uncertainty in the reweighting to NLO by considering alternative sets of weights: either by varying the nominal weights by their uncertainty either by varying simultaneously the factorization and renormalization scales by a factor of $\frac{1}{2} - 2$ at NLO.

For the data-driven modeling used in Method II, we consider the effect on the shapes from four distinct sources: statistical size of the photon sample; difference in a tight-photon selection to a pure photon case obtained in simulation; envelope of the PDF variations in the simulation compatibility tests; difference observed in data selected with a loose-photon selection with respect to the data selected with a tight-photon selection case. For simplicity we consider only two alternative sets of shapes: the one stemming from the bin-by-bin variation of the $\pm 1 - \sigma_{\mathrm{stat}}$ and the one stemming from the variation of the quadratic sum of the differences in compatibility tests with the loose-to-tight selection in data. The shape might be affected as well by residual EWK γjj events selected in the photon sample. We furthermore assign a 30% normalization uncertainty to the EWK γjj normalization. Both this and the statistical uncertainty of the EWK γjj used to derive the bin-by-bin contamination of the shapes are taken into account in the subtraction procedure. Besides the uncertainty in the shape just described, we also assign an extra normalization uncertainty to the initial estimate of QCD Zjj. The yields are allowed to float freely in the range $[\frac{1}{2}, 2]$ by assigning a nuisance parameter log-uniform distibuted.

5.4 Summary of systematic uncertainties

Table 1 summarizes the systematic uncertainties described above, and their impact on the normalization. The table also includes how the uncertainties are treated in the fitting procedure with respect to normalization and shape variations and with respect to correlation amongst channels.

Table 1: Summary the relative systematic uncertainties considered in the different analyses. A • (○) signals if the uncertainty affects the shape (rate) of a process in the fit.

Source	Source	Shape	Method I	Method II
Experimental	Luminosity	○	0.026	
	Trigger/selection	○	0.02-0.03	
	JES	•	0.01-0.1	
	JER	•	0.06-0.15	
	Residual jet response	•	0.01-0.05	
	Pileup	•	0.06	
	Simulation statistics	•	variable	
	QCD Z+2 jets shape (data)	•	-	variable
Theoretical	PDF (signal)	•	variable	
	Q^2 choice (signal)	•	$\frac{1}{2}$-1	
	$t\bar{t}$, single top normalization	○	0.07	
	WW/WZ/ZZ normalization	○	0.1	
	QCD Zjj normalization (data)	○	-	$\frac{1}{2}$-2
	QCD Zjj Q^2 choice (MC)	•/○	$\frac{1}{2}$-1	-
	QCD Zjj shape (MC)	•	variable	-

6 Measurement of the inclusive production cross section

The signal strength is extracted from the fit to the discriminator shapes as discussed in Section 4. Table 2 reports the final event yield obtained for the high score region of the BDT discriminator used in Method I while Tab. 3 reports the final event yield obtained in each category after the fit is performed with Method II. A fair agreement between data and expectations is observed for both methods, in all categories.

Table 2: Event yields expected after the fit for background and signal processes using Method I after requiring BDT>0.02. The yields are compared to the data observed in the different channels and categories. The uncertainties quoted for signal, di-bosons (VV) and processes with top quarks ($t\bar{t}$ and single-top quark) reflect the limited amount of MC events. For QCD Zjj we quote, as well, the total systematic uncertainty.

Channel	VV	Top	QCD Zjj	Total	EWK Zjj	Data
$\mu\mu$	41 ± 1	99 ± 8	$4213 \pm 48 \pm 450$	$\mathbf{4353 \pm 49 \pm 450}$	551 ± 4	**4902**

Table 4 summarizes the results obtained for the fit of the signal strength with each method. The results obtained are compatible amongst the different dilepton channels and methods within uncertainty and compatible with the NLO prediction. Using the profile likelihood used in the fit in Method II, we estimate that the observed significance with respect to a background only hypothesis is of 5.0, while the SM expectation is 5.9.

The result of the fits can be used to extrapolate the signal cross section corresponding to the generator-level phase space region. With method I we measure

$$\sigma(\text{EWK } \ell\ell + \text{jj})_{\text{I}} = 191 \pm 29_{\text{stat}} \pm 39_{\text{syst}} \text{ fb},$$

while with method II we measure

Table 3: Event yields expected after the fit for background and signal processes using Method II. The yields are compared to the data observed in the different channels and categories. The uncertainties quoted for signal, di-bosons (VV) and processes with top quarks ($t\bar{t}$ and single-top quark) reflect the limited amount of MC events. For QCD Zjj we quote, as well, the total systematic uncertainty.

M_{jj}	channel	VV	Top	QCD Zjj	Total	EWK Zjj	Data
$250-350$	ee	92 ± 2	423 ± 7	$(339 \pm 6 \pm 7) \times 10^2$	$\mathbf{(344 \pm 6 \pm 7) \times 10^2}$	178 ± 4	**38885**
	$\mu\mu$	126 ± 2	574 ± 7	$(477 \pm 8 \pm 9) \times 10^2$	$\mathbf{(479 \pm 8 \pm 9) \times 10^2}$	252 ± 5	**55172**
$350-450$	ee	42 ± 1	151 ± 4	$(140 \pm 6 \pm 3) \times 10^2$	$\mathbf{(142 \pm 6 \pm 3) \times 10^2}$	135 ± 3	**14001**
	$\mu\mu$	58 ± 1	209 ± 5	$(194 \pm 10 \pm 4) \times 10^2$	$\mathbf{(197 \pm 9 \pm 4) \times 10^2}$	178 ± 4	**19416**
$450-550$	ee	20 ± 0.8	70 ± 3	$5973 \pm 363 \pm 119$	$\mathbf{6063 \pm 363 \pm 119}$	92 ± 3	**5786**
	$\mu\mu$	25 ± 0.9	94 ± 3	$8516 \pm 517 \pm 170$	$\mathbf{8634 \pm 517 \pm 170}$	124 ± 3	**8329**
$550-750$	ee	15 ± 0.7	55 ± 2.2	$4121 \pm 159 \pm 82$	$\mathbf{4192 \pm 159 \pm 83}$	110 ± 3	**4110**
	$\mu\mu$	19 ± 0.8	70 ± 2.7	$5560 \pm 204 \pm 111$	$\mathbf{5650 \pm 204 \pm 111}$	153 ± 4	**5637**
$750-1000$	ee	5 ± 0.4	22 ± 1	$1335 \pm 38 \pm 27$	$\mathbf{1361 \pm 38 \pm 27}$	69 ± 2	**1379**
	$\mu\mu$	7 ± 0.5	24 ± 2	$1871 \pm 69 \pm 37$	$\mathbf{1905 \pm 69 \pm 37}$	92 ± 3	**1911**
> 1000	ee	3 ± 0.3	9 ± 1	$641 \pm 25 \pm 13$	$\mathbf{654 \pm 25 \pm 13}$	81 ± 3	**679**
	$\mu\mu$	4 ± 0.3	13 ± 1	$864 \pm 33 \pm 17$	$\mathbf{880 \pm 33 \pm 17}$	108 ± 3	**920**

Table 4: Signal strength fit in the different analysis and channels. The first row quotes the total uncertainty attained. The breakup of the uncertainty in given in the subsequent rows. The last column contains the result of the combination for Method II.

	Analysis			
	Method I	Method II		
	$\mu\mu$	ee	$\mu\mu$	Combination
$\mu = \sigma/\sigma_{\text{th}}$	0.80 ± 0.20	0.82 ± 0.37	1.30 ± 0.30	1.27 ± 0.27
Statistical uncertainty	0.12	0.23	0.16	0.12
Systematic uncertainty	0.16	0.28	0.25	0.24
Luminosity	0.03	0.03	0.03	0.03
Trigger/lepton selection	0.01	0.03	0.03	0.02
JES+residual response	0.05	0.05	0.04	0.04
JER	0.02	0.02	0.02	0.02
Pileup	0.06	0.03	0.03	0.02
QCD Zjj	0.13	0.26	0.23	0.22
Top, dibosons	0.01	0.04	0.03	0.02
Signal	0.05	0.08	0.07	0.06

$$\sigma(\text{EWK } \ell\ell + \text{jj})_{\text{II}} = 303 \pm 29_{\text{stat}} \pm 57_{\text{syst}} \text{ fb}.$$

The individual results can be furthermore combined. For the combination we make use of the Best Linear Unbiased Estimator technique [48, 49] assuming that the measurements are statistically correlated in the di-muon channel, i.e. we assign a statistical correlation factor of 1.0. The di-electron channel, which is solely used in Method II, enters with a statistical correlation factor of 0 in the combination. All systematic uncertainty sources are assumed to be fully correlated, i.e. with a correlation factor of 1.0, except the one affecting the estimation of the QCD Zjj background for which we assign a correlation factor 0. Under this assumption we obtain the following result

$$\sigma(\text{EWK } \ell\ell\text{jj}) = 226 \pm 26_{\text{stat}} \pm 35_{\text{syst}} \text{ fb},$$

corresponding to a χ^2/dof=2.8/1 with a 9.4% probability. The relative weight assigned to method I is 68%. The results obtained in individually by each method are found to be in agreement within total uncertainty. Both the individual results and the final combination are found to be in good agreement with the SM prediction.

7 Study of the hadronic activity of the selected events

After establishing the signal, we study the properties of the hadronic activity in the selected events. The study is divided in three sections: radiation patterns in Z+multijet events, the profile of the charged hadronic activity as function of several kinematics variables, the production of extra jets in a high purity region ($M_{\text{jj}} > 1250\,\text{GeV}$). Overall we expect a significant suppression of the hadronic activity for the signal owing to the fact that the final state objects have their origin in pure electroweak interactions in contrast with the production of Z+2 jets via QCD. In these studies we do not yet perform an unfolding of the observed data. The reconstructed quantities are compared directly to the prediction obtained with a full simulation of the CMS detector.

7.1 Measurements of the radiation patterns in multijet events in association with a Z boson

In hard multijet events in association with a Z boson, the observables referred to as "radiation patterns" are:

- the number of jets, N_j;
- the total scalar sum of the transverse momenta of jets reconstructed within $|\eta| < 4.7$, H_{T};
- $\Delta\eta_{j_1 j_2}$ between the two most forward-backward jets (which are not necessarily the two highest–p_{T} jets);
- the cosine of the azimuthal angle difference, $\cos|\phi_{j_1} - \phi_{j_2}| = \cos\Delta\phi_{j_1 j_2}$, between the two most forward-backward jets.

These observables are investigated following the prescriptions and suggestions from [50], where the model dependence is estimated by comparing the predictions from MCFM [27], PYTHIA [21], ALPGEN [51]+PYTHIA, and the HEJ [52] programs.

The observables N_j, H_T, $\Delta\eta_{j_1j_2}$, and $\cos\Delta\phi_{j_1j_2}$ are measured for jets with $p_T > 40\,\text{GeV}$. The events are required to satisfy the $Z \to \mu\mu$ and $Z \to ee$ selection criteria. Figures 6 and 7 show the average number of jets and the average $\cos\Delta\phi_{j_1j_2}$ as a function of the total H_T and $\Delta\eta_{j_1j_2}$. For low values of $\Delta\eta_{j_1j_2}$ we note a change in the behavior mostly induced by the onset of jet merging from gluon splitting. In all cases the MADGRAPH + PYTHIA (ME-PS) predictions are found to be in good agreement with the data.

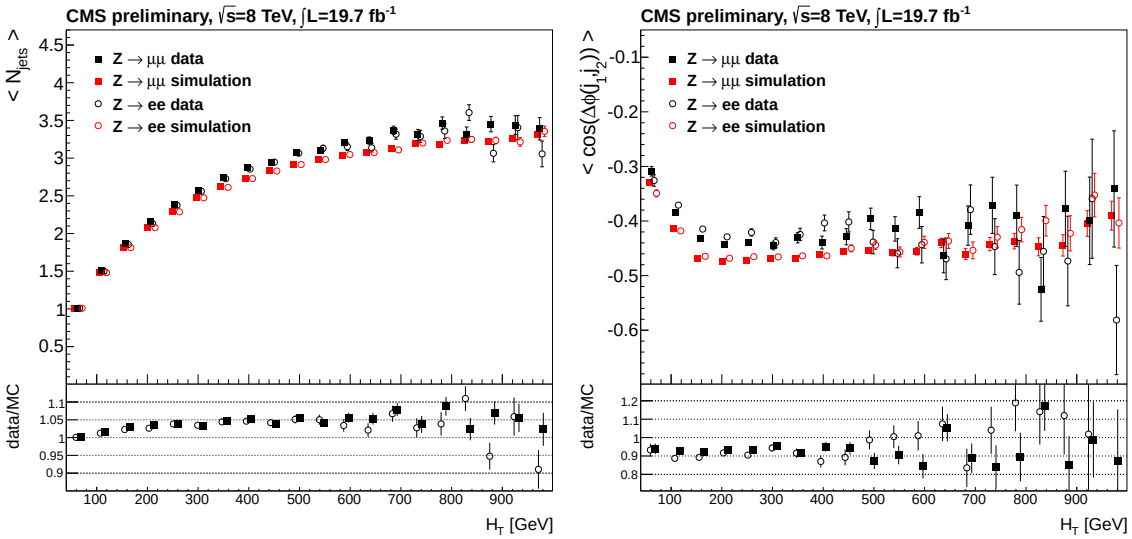

Figure 6: Average number of jets with $p_T > 40\,\text{GeV}$ as a function of the their total H_T in Z plus at least one jet events (*left*) and average $\cos\Delta\phi_{j_1j_2}$ as a function of the total H_T in DY Zjj events (*right*). The data points and the points from simulation are shown with the statistical uncertainties.

7.2 Study of the charged hadronic activity

The hadronic activity measurements of selected Z+2 jets events performed with proton-proton collision data at $\sqrt{s} = 7$ TeV [11], have been repeated with the 8 TeV data samples, and the results are given below. These include the level of hadronic activity in the rapidity interval between the two tagging jets and the properties of multi-jets in events with a Z boson.

For this study a collection of tracks is built with reconstructed high-purity tracks [53] with $p_T > 300\,\text{MeV}$ that are uniquely associated with the main primary vertex in the event. Tracks associated with the two leptons or with the tagging jets are not included. The association between the tracks and the reconstructed primary vertices is carried out by minimizing the longitudinal distance between the PV and the point of closest approach of the track helix to the PV,i.e. $d_z(\text{PV})$. The association is required to satisfy $d_z(\text{PV}) < 2\,\text{mm}$ and $d_z(\text{PV}) < 3\delta d_z(\text{PV})$, where $\delta d_z(\text{PV})$ is the uncertainty on $d_z(\text{PV})$.

A collection of "soft track jets" is built by clustering the tracks with the anti-k_T clustering algorithm [16] with a distance parameter of 0.5. The use of track jets represents a clean and well understood method [54] to reconstruct jets with energy as low as a few GeV. Crucially, these jets are not affected by pileup because of the association of their tracks with the hard-scattering vertex [55].

For the purpose of studying the central hadronic activity between the tagging jets, only soft track jets with pseudorapidity $\eta_{\text{min}}^{\text{tag jet}} + 0.5 < \eta < \eta_{\text{max}}^{\text{tag jet}} - 0.5$ are considered. We use up to

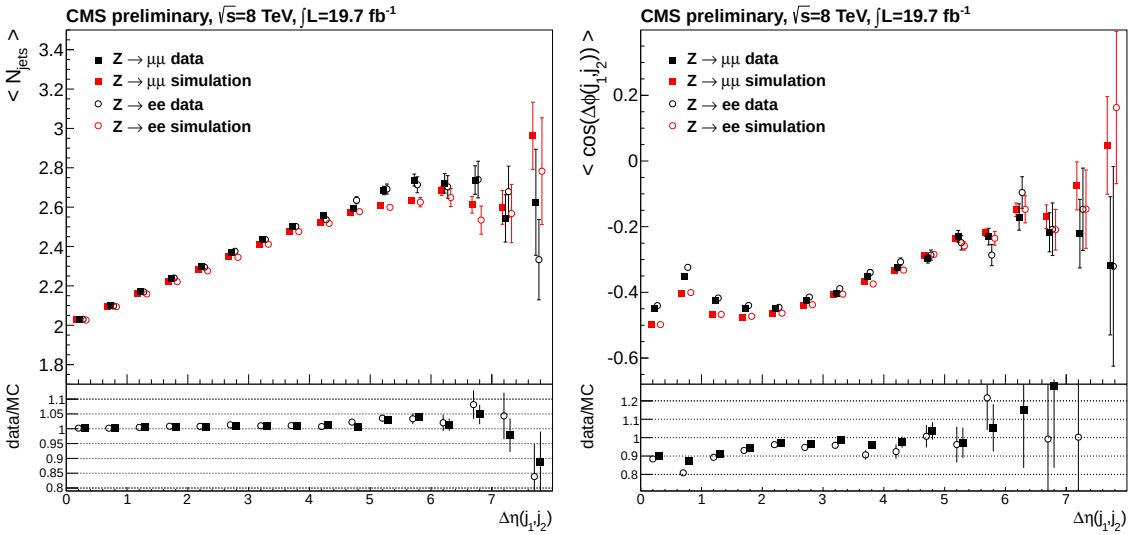

Figure 7: Average number of jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ as a function of $\Delta\eta_{j_1 j_2}$ (*left*) and average $\cos\Delta\phi_{j_1 j_2}$ as a function of $\Delta\eta_{j_1 j_2}$ separation (*right*) in DY Zjj events. The data points and the points from simulation are shown with the statistical uncertainties.

three leading-p_{T} soft track jets to compute the soft $H_{\mathrm{T}}(3)$ variable as the scalar sum of jets' p_{T}. The soft $H_{\mathrm{T}}(3)$ variable is chosen to monitor the hadronic activity in the rapidity interval between the two jets.

The evolution of the average soft $H_{\mathrm{T}}(3)$ for the Z+2 jets event candidates as a function of M_{jj} and $\Delta\eta_{\mathrm{jj}}$ is shown in Fig. 8. Good agreement is observed between the simulation and the data for the different mass and pseudorapidity intervals. At extreme values of both variables, i.e. $M_{\mathrm{jj}} > 1\,\mathrm{TeV}$ and $\Delta\eta_{\mathrm{jj}} > 6$, we observe that the simulation tends to predict higher values of the average soft $H_{\mathrm{T}}(3)$ than the ones reconstructed in data.

7.3 Study of the third jet properties in a high purity region

We conclude this section with a study of the emission third (and extra) jets in a high purity region for which the dijet invariant mass is greater than 1250 GeV. In this study we consider, besides the two tag jets used in the pre-selection, all PF-based jets with a $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ found within the pseudo-rapidity distance defined by the dijet system, as in Section 7.2. The background prediction is modeled from the photon control sample.

Figure 9 shows the counting of jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, the scalar sum of the p_{T} of all jets in the event (soft H_{T}) and two properties of the leading third jet in the event: p_{T} and pseudo-rapidity measured in the dijet rest frame, i.e. $y^{*}_{\mathrm{j3}} = y_{\mathrm{j3}} - (y_{\mathrm{j1}} + y_{\mathrm{j2}})/2$. Good agreement is observed in all cases between data and the prediction for the most variables. The selected third jets are however observed to be slightly more central than expected. The properties of the signal can be isolated by applying a direct subtraction of the estimated background contribution. The observed significance is sufficient to overcome most of the statistical uncertainties stemming from this procedure. As expected we observe a suppression of the emission of a third jet in the events when we only take into account the background prediction.

The distributions above can be used to compute the efficiency of an hadronic veto either based on the p_{T} of the third jet or on the soft H_{T} variable. Figure 10 shows the gap fractions for these two variables. The gap fractions correspond to the fraction of events which do not have

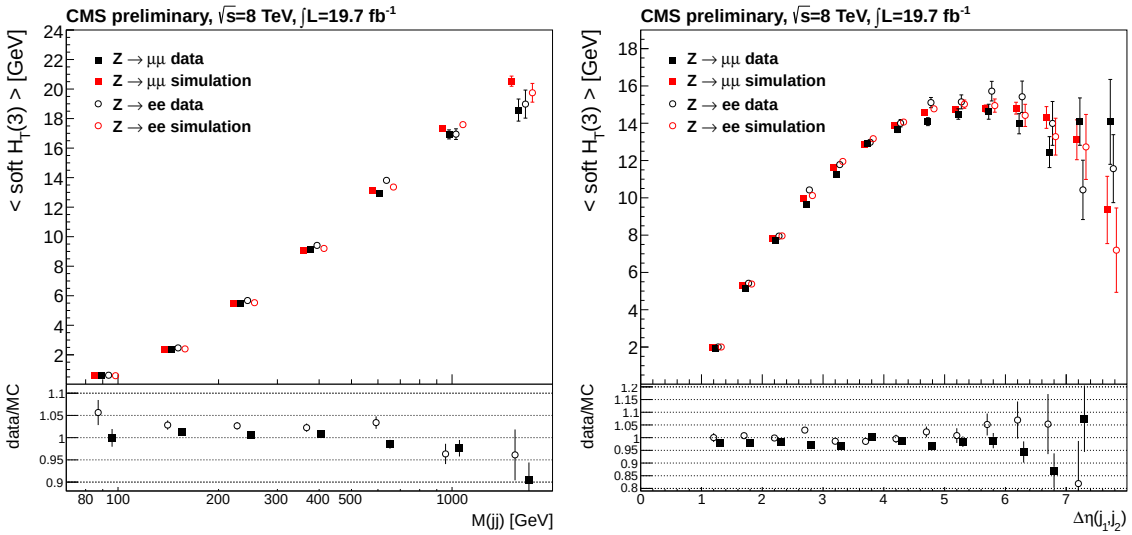

Figure 8: Average H_{T} of the three leading soft track jets in the pseudorapidity gap between the tagging jets for $p_{\mathrm{T}}^{j_1,j_2} > 50,\,30\,\mathrm{GeV}$ as a function of the dijet invariant mass (*left*) and the dijet $\Delta\eta_{j_1 j_2}$ separation (*right*) for both the dielectron and dimuon channels in DY Zjj events. The data points and the points from simulation are shown with the statistical uncertainties. At each measured point the marker symbols have been slightly displaced along the x-axis.

a third jet with a p_{T} above a given threshold or with soft H_{T} above a given threshold. Besides comparing the gap fraction measured in data with the data-driven background prediction we also include the MC-based prediction. All are in agreement within the uncertainties.

8 Summary

We have measured the cross section for the production of same flavor dilepton pairs in association with two jets from pure electroweak processes in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. Two different approaches have been considered and lead, after combination, to

$$\sigma(\mathrm{EWK}\ \ell\ell\mathrm{jj}) = 226 \pm 26_{\mathrm{stat}} \pm 35_{\mathrm{syst}}\ \mathrm{fb}.$$

The result is in agreement with the theoretical cross section at NLO which is 239 fb. Preliminary studies of the hadronic activity of the events selected in the high purity categories have also been presented including a measurement of the gap fractions for soft-hadronic vetoes. A good agreement is found between data and different predictions.

References

[1] C. Oleari and D. Zeppenfeld, "Next-to-leading order QCD corrections to W and Z production via vector-boson fusion", *Phys. Rev. D* **69** (2004) 093004, doi:10.1103/PhysRevD.69.093004, arXiv:hep-ph/0310156.

[2] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color singlet exchange in Z + two jet events at the CERN LHC", *Phys.Rev.* **D54** (1996) 6680–6689, doi:10.1103/PhysRevD.54.6680, arXiv:hep-ph/9605444.

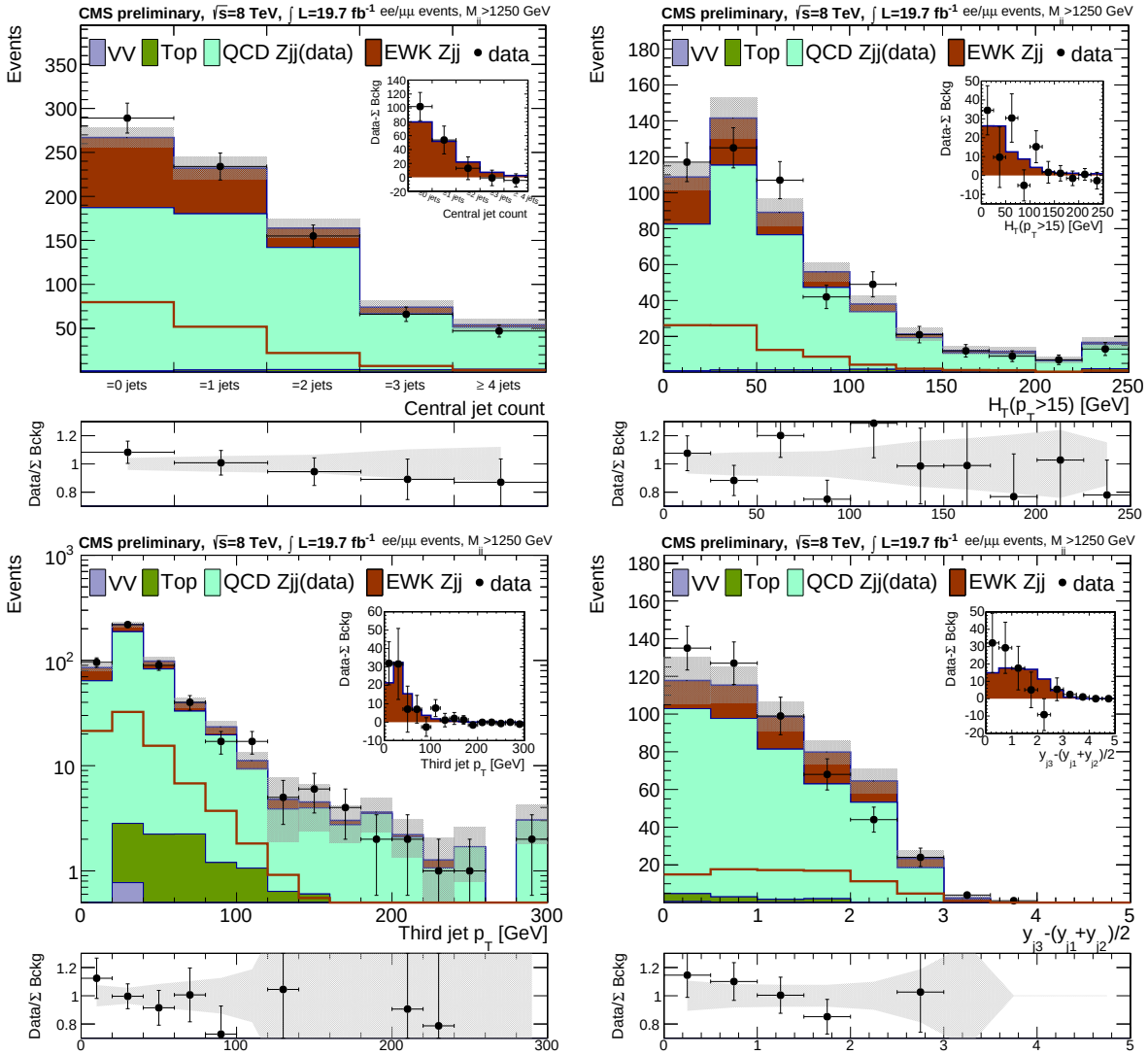

Figure 9: Control distributions for the hadronic activity of events with $M_{jj} > 1250\,\text{GeV}$. *Top* : the jet multiplicity is shown on the *left* and the scalar sum of the p_T of all the extra jets is shown on the *right*. *Bottom*: third jet properties - the p_T is shown on the *left* while the pseudo-rapidity in the dijet rest frame is shown on the *right*. For each distribution the insets show the result of the direct bin-by-bin background subtraction. The grey band in the bottom panels represents the total uncertainty assigned to the background.

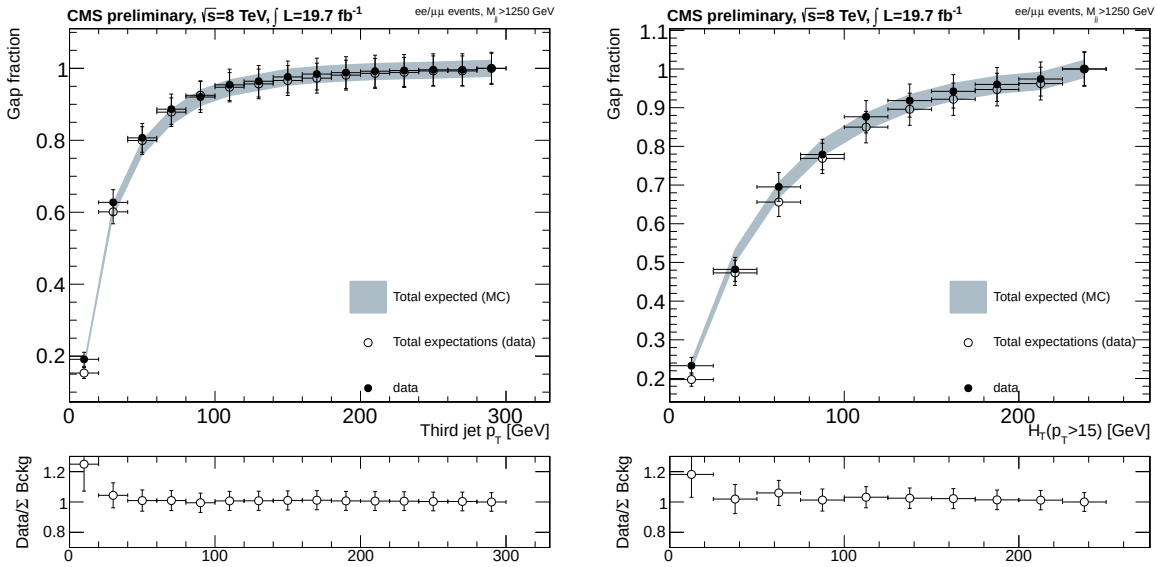

Figure 10: Gap fraction for the p_{T} of the third jet (*left*) and scalar sum of the p_{T} of all the jets (*right*). The observed fraction is compared to both data-driven and simulated predictions. The bottom panel shows the ratio between the observed data and the data-driven background prediction.

[3] V. Khoze, M. Ryskin, W. Stirling, and P. Williams, "A Z monitor to calibrate Higgs production via vector boson fusion with rapidity gaps at the LHC", *Eur.Phys.J.* **C26** (2003) 429–440, doi:10.1140/epjc/s2002-01069-2, arXiv:hep-ph/0207365.

[4] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[6] G.-C. Cho et al., "Weak boson fusion production of supersymmetric particles at the CERN LHC", *Phys.Rev.* **D73** (2006) 054002, doi:10.1103/PhysRevD.73.054002, arXiv:hep-ph/0601063.

[7] B. Dutta et al., "Vector Boson Fusion Processes as a Probe of Supersymmetric Electroweak Sectors at the LHC", *Phys.Rev.* **D87** (2013) 035029, doi:10.1103/PhysRevD.87.035029, arXiv:1210.0964.

[8] U. Baur and D. Zeppenfeld, "Measuring three vector boson couplings in qq→ qq W at the SSC", *procs. of Workshop on Physics at Current Accelerators and the Supercollider* (1993) arXiv:hep-ph/9309227.

[9] J. Bjorken, "Rapidity gaps and jets as a new physics signature in very high-energy hadron hadron collisions", *Phys.Rev.* **D47** (1993) 101–113, doi:10.1103/PhysRevD.47.101.

[10] F. Schissler and D. Zeppenfeld, "Parton Shower Effects on W and Z Production via Vector Boson Fusion at NLO QCD", *JHEP* **1304** (2013) 057, doi:10.1007/JHEP04(2013)057, arXiv:1302.2884.

[11] CMS Collaboration, "Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at $\sqrt{s}$ = 7 TeV", arXiv:1305.7389.

[12] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[13] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[14] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s}$ = 7 TeV", CMS Physics Analysis Summary CMS-PAS-EGM-10-004, CERN, Geneva, (2010).

[15] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[16] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k(t) jet clustering algorithm", *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[17] CMS Collaboration, "The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS", CMS Physics Analysis Summary CMS-PAS-JME-09-002, CERN, Geneva, (2009).

[18] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, Geneva, (2009).

[19] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in minimum bias and jet events from pp collisions at 7 TeV", CMS Physics Analysis Summary CMS PAS PFT-10-002, CERN, Geneva, (2010).

[20] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[21] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[22] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **0207** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.

[23] K. Arnold et al., "VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons", *Comput.Phys.Commun.* **180** (2009) 1661–1670, doi:10.1016/j.cpc.2009.03.006, arXiv:0811.4559.

[24] K. Arnold et al., "VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.6.0", arXiv:1107.4038.

[25] K. Arnold et al., "Release Note – VBFNLO-2.6.0", arXiv:1207.4975.

[26] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $O(\alpha_S^2)$", *Phys. Rev.* **D74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[27] J. M. Campbell and R. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10–15, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[28] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $\mathcal{O}(\alpha_S^4)$", *Phys.Rev.Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[29] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **1006** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[30] N. Kidonakis, "Differential and total cross sections for top pair and single top production", doi:10.3204/DESY-PROC-2012-02/251, arXiv:1205.3453.

[31] CMS Collaboration, "Top pair cross section in dileptons", CMS Physics Analysis Summary CMS-PAS-TOP-12-007, CERN, Geneva, (2012).

[32] CMS Collaboration, "Measurement of the single-top t-channel cross section in pp collisions at centre-of-mass energy of 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-011, CERN, Geneva, (2012).

[33] CMS Collaboration, "Inclusive W/Z cross section at 8 TeV", CMS Physics Analysis Summary CMS-PAS-SMP-12-011, CERN, Geneva, (2012).

[34] CMS Collaboration, "Measurement of W+W- and ZZ production cross sections in pp collisions at sqrt(s) = 8 TeV", *Phys.Lett.* **B721** (2013) 190–211, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698.

[35] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[36] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[37] M. Cacciari, G. Salam, and G. Soyez, "The catchment area of jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/05, arXiv:0802.1188.

[38] M. Cacciari and G. Salam, "Pileup subtraction using jet areas", *Phys. Lett.* **B659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[39] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005, CERN, Geneva, (2013).

[40] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (November, 2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[41] S. Catani, F. Krauss, R. Kuhn, and B. Webber, "QCD matrix elements + parton showers", *JHEP* **0111** (2001) 063, doi:10.1088/1126-6708/2001/11/063, arXiv:hep-ph/0109231.

[42] B. Jager, "Next-to-leading order QCD corrections to photon production via weak-boson fusion", *Phys.Rev.* **D81** (2010) 114016, doi:10.1103/PhysRevD.81.114016, arXiv:1004.0825.

[43] A. Hoecker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[44] R. Fisher, "The use of multiple measurements in taxonomic problems", *Annals of Eugenics* **7** (1936), no. 2, 179–188, doi:10.1111/j.1469-1809.1936.tb02137.x.

[45] L. Moneta et al., "The RooStats project", *PoS* **ACAT2010** (2010) 057, arXiv:1009.1003.

[46] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in summer 2011", ATLAS / CMS note, ATL-PHYS-PUB-2011-011 / CMS-NOTE-2011-005, CERN, Geneva, (2011).

[47] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CERN, Geneva, (2013).

[48] L. Lyons, D. Gibaut, and P. Clifford, "How to combine correlated estimates of a single physics quantity", *Nucl.Instrum.Meth.* **A270** (1988) 110, doi:10.1016/0168-9002(88)90018-6.

[49] L. Lyons, A. J. Martin, and D. H. Saxon, "On the determination of the B lifetime by combining results of different experiments", *Phys.Rev.* **D41** (1990) 982, doi:10.1103/PhysRevD.41.982.

[50] SM and NLO Multileg Working Group, "The SM and NLO Multileg Working Group: Summary report", arXiv:1003.1241.

[51] M. L. Mangano et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* **0307** (2003) 001, doi:10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293.

[52] J. R. Andersen and J. M. Smillie, "Multiple Jets at the LHC with High Energy Jets", *JHEP* **1106** (2011) 010, doi:10.1007/JHEP06(2011)010, arXiv:1101.5394.

[53] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-005, CERN, Geneva, (2010).

[54] CMS Collaboration, "Commissioning of TrackJets in pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-006, CERN, Geneva, (2010).

[55] CMS Collaboration, "Performance of Jet Reconstruction with Charged Tracks only", CMS Physics Analysis Summary CMS-PAS-JME-08-001, CERN, Geneva, (2009).