Measurement of electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 8$ TeV

CMS Collaboration

CERN, 1211 Geneva 23, Switzerland

Received: 12 October 2014 / Accepted: 15 December 2014 / Published online: 10 February 2015

© CERN for the benefit of the CMS collaboration 2015. This article is published with open access at Springerlink.com

Abstract  The purely electroweak (EW) cross section for the production of two jets in association with a Z boson, in proton–proton collisions at $\sqrt{s} = 8$ TeV, is measured using data recorded by the CMS experiment at the CERN LHC, corresponding to an integrated luminosity of 19.7 fb$^{-1}$. The electroweak cross section for the $\ell\ell jj$ final state (with $\ell = e$ or $\mu$ and $j$ representing the quarks produced in the hard interaction) in the kinematic region defined by $M_{\ell\ell} > 50$ GeV, $M_{jj} > 120$ GeV, transverse momentum $p_Tj > 25$ GeV, and pseudorapidity $|\eta_j| < 5$, is found to be $\sigma_W(\ell\ell jj) = 174 \pm 15 \text{(stat)} \pm 40 \text{(syst)} \text{fb}$, in agreement with the standard model prediction. The associated jet activity of the selected events is studied, in particular in a signal-enriched region of phase space, and the measurements are found to be in agreement with QCD predictions.

1 Introduction

The production of a Z boson in association with two jets in proton–proton (pp) collisions is dominated by a mixture of electroweak (EW) and strong processes of order $\alpha_S^2$. For $Z \rightarrow \ell\ell$ leptonic decays, such events are referred to as “Drell–Yan (DY) + jets” or DY $Zjj$ events.

Purely electroweak $\ell\ell jj$ production contributing to the same final state is expected at order $\alpha_W^2$, resulting in a comparatively small cross section [1]. This process is however predicted to have a distinctive signature of two jets of very high energy and large $jj$ invariant mass, $M_{jj}$, separated by a large rapidity interval that can be occupied by the two charged leptons and where extra gluon emission is suppressed [2,3]. We refer to jets produced through the fragmentation of the outgoing quarks in pure EW processes as “tagging jets”, and to the process from which they originate as “EW Zjj”. Figure 1 shows representative Feynman diagrams for the EW Zjj processes, namely (left) vector boson fusion (VBF), (middle) bremsstrahlung-like, and (right) multiperipheral production. Detailed calculations reveal the presence of a large negative interference between the pure VBF process and the two other categories [1,3]. These diagrams represent the signal (S) in the data.

For inclusive $\ell\ell jj$ final states, some of the diagrams with same initial- and final-state particles and quantum numbers can interfere, even if they do not involve exclusively EW interactions. Figure 2 (left) shows one example of order $\alpha_S^2$ corrections to DY production that have the same initial and final state as those in Fig. 1. A different order $\alpha_S^2$ correction that does not interfere with the EW signal, is shown in Fig. 2 (right).

The study of EW $Zjj$ processes is part of a more general investigation of standard model (SM) vector boson fusion and scattering processes that include the Higgs boson [4–6] and searches for physics beyond the standard model [7,8]. When isolated from the backgrounds, the properties of EW Zjj events can be compared with SM predictions. Probing the jet activity in the selected events in particular can shed light on the selection (or vetoing) of additional parton radiation to the tagging jets [9,10].

At the CERN LHC, the EW Zjj process was first measured by the CMS experiment using pp collisions at $\sqrt{s} = 7$ TeV [11], and more recently by the ATLAS experiment at $\sqrt{s} = 8$ TeV [12]. Both results have been found to agree with the expectations of the SM. Our present work reflects the measurement at CMS using pp collision data collected at $\sqrt{s} = 8$ TeV during 2012 that correspond to an integrated luminosity of 19.7 fb$^{-1}$. As the signal-to-background ratio for the measurement is small, different methods are used to enhance the signal fraction, to confirm the presence of the signal, and to measure the cross section. Besides the two multivariate analyses, based on the methods developed for the 7 TeV analysis [11], a new method is presented, using a model of the main background based on real pp collisions. The analysis of the 8 TeV data, offers the opportunity of reducing the uncertainties of the 7 TeV measurements, given the larger integrated luminosity, and to add robustness to the results with the new data-based method.
Fig. 1 Representative Feynman diagrams for dilepton production in association with two jets from purely electroweak contributions: (left) vector boson fusion, (middle) bremsstrahlung-like, and (right) multiperipheral production

Fig. 2 Representative diagrams for order $\alpha_s^2$ corrections to DY production that comprise the main background (B) in this study

This paper is organised as follows: Sect. 2 describes the experimental apparatus and Sect. 3 the simulations. Event selection procedures are described in Sect. 4, and Sect. 5 discusses the selection efficiencies and background models in control regions. Section 6 details the strategies adopted in our analysis to extract the signal from the data, and the corresponding systematic uncertainties are summarised in Sect. 7. The results obtained are presented in Sect. 8, and we conclude with a study of jet properties in a DY $Zjj$-dominated control region, as well as in a high-purity, EW $Zjj$-enriched region in Sect. 9. Finally, a brief summary of the results is given in Sect. 10.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. The solenoid volume contains a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass/scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Muons are measured in gas-ionisation tracking detectors embedded in the steel flux-return yoke outside the solenoid. Extensive forward calorimetry complements the coverage provided by the barrel and endcap detectors.

The silicon tracker consists of 1440 silicon pixel modules and 15 148 silicon strip detector modules, located in the field of the superconducting solenoid. It measures charged particles within $|\eta| < 2.5$, providing an impact parameter resolution of $\approx 15 \mu m$ and a transverse momentum ($p_T$) resolution of about 1.5% for $p_T = 100$ GeV particles.

The energy of electrons is measured after combining the information from the ECAL and the tracker, whereas their direction is measured by the tracker. The invariant mass resolution for $Z \rightarrow ee$ decays is 1.6% when both electrons are in the ECAL barrel, and 2.6% when both electrons are in the ECAL endcap [13]. Matching muons to tracks measured in the silicon tracker yields a $p_T$ resolution between 1 and 10%, for $p_T$ values up to 1 TeV. The jet energy resolution (JER) is typically $\approx 15$% at 50 GeV, 8% at 100 GeV, and 4% at 1 TeV [14].
3 Simulation of signal and background events

Signal events are simulated at leading order (LO) using the MadGraph (v5.1.3.30) Monte Carlo (MC) generator [15,16], interfaced to PYTHIA (v6.4.26) [17] for parton showering (PS) and hadronisation. The CTEQ6L1 [18] parton distribution functions (PDF) are used to generate the event, the factorisation ($\mu_F$) and renormalisation ($\mu_R$) scales being both fixed to be equal to the $Z$-boson mass [19]. The underlying event is modelled with the so-called Z2* tune [20]. The simulation does not include the generation of extra partons at matrix-element level. In the kinematic region defined by dilepton mass $M_{\ell\ell} > 50$ GeV, parton transverse momentum $p_{Tj} > 25$ GeV, parton pseudorapidity $|\eta_j| < 5$, diparton mass $M_{jj} > 120$ GeV, and angular separation $\Delta R_{jj} = \sqrt{(\Delta \eta_{jj})^2 + (\Delta \phi_{jj})^2} > 0.5$, where $\Delta \eta_{jj}$ and $\Delta \phi_{jj}$ are the differences in pseudorapidity and azimuthal angle between the tagging partons, the cross section in the $\ell\ell_j j final state (with $\ell = e$ or $\mu$) is expected to be $\sigma_{\text{LO}}(\text{EW } \ell\ell_j j) = 208^{+3}_{-1}(\text{scale}) \pm 7$ (PDF) fb, where the first uncertainty is obtained by changing simultaneously $\mu_F$ and $\mu_R$ by factors of 2 and 1/2, and the second from the uncertainties in the PDFs which has been estimated following the PDF4LHC prescription [18,21–24]. The LO signal cross section and kinematic distributions estimated with MadGraph are found to be in good agreement with the LO predictions of the VBFNLO generator (v.2.6.3) [25–27].

Background DY events are also generated with MadGraph using a LO matrix element (ME) calculation that includes up to four partons generated from quantum chromodynamics (QCD) interactions. The ME-PS matching is performed following the ktMLM prescription [28,29]. The dilepton DY production for $M_{\ell\ell} > 50$ GeV is normalised to $\sigma_{\text{th}}(\text{DY}) = 3.504$ nb, as computed at next-to-next-leading order (NNLO) with FEWZ [30].

The evaluation of the interference between EW Zjj and DY Zjj processes relies on the predictions obtained with MadGraph. Three samples, one of pure signal, one pure background, and one including both $\alpha_s^{\text{EW}}$ and $\alpha_s^{\text{QCD}}$ contributions are generated for this purpose. The differential cross sections are compared and used to estimate the expected interference contributions at the parton level.

Other residual background is expected from events with two leptons of same flavour with accompanying jets in the final state. Production of tt events is generated with MadGraph, including up to three extra partons, and normalised to the NNLO with next-to-next-leading-logarithmic corrections to an inclusive cross section of 245.8 pb [31]. Single-top-quark processes are modelled at next-to-leading order (NLO) with POWHEG [32–36] and normalised, respectively, to cross sections of $22 \pm 2$, $86 \pm 3$, and $5.6 \pm 0.2$ pb for the tW, t-, and s- channel production [37,38]. Diboson production processes WW, WZ, and ZZ are generated with MadGraph and normalised, respectively, to the cross sections of 59.8, 33.2, and 17.7 pb, computed at NNLO [39] and with MCFM [40]. Throughout this paper we use the abbreviation VV when referring to the sum of the processes which yield two vector bosons.

The production of a W boson in association with jets, where the W decays to a charged lepton and a neutrino, is generated with MadGraph, and normalised to a total cross section of 36.3 nb, computed at NNLO with FEWZ. Multijet QCD processes are also studied in simulation, but are found to yield negligible contributions to the selected events.

A detector simulation based on GEANT4 (v.9.4p03) [41, 42] is applied to all the generated signal and background samples. The presence of multiple pp interactions in the same beam crossing (pileup) is incorporated by simulating additional interactions (both in-time and out-of-time with the collision) with a multiplicity that matches the one observed in data. The average number of pileup events is estimated as $\approx 21$ interactions per bunch crossing.

4 Reconstruction and selection of events

The event selection is optimised to identify dilepton final states with two isolated, high-$p_T$ leptons, and at least two high-$p_T$ jets. Dilepton triggers are used to acquire the data, where one lepton is required to have $p_T > 17$ GeV and the other to have $p_T > 8$ GeV. Electron-based triggers include additional isolation requirements, both in the tracker detectors and in the calorimeters. A single-isolated-muon trigger, with a requirement of $p_T > 24$ GeV, is used to complement the dimuon trigger and increase the efficiency of the selection.

Electrons are reconstructed from clusters of energy deposits in the ECAL that match tracks extrapolated from the silicon tracker [43]. Muons are reconstructed by fitting trajectories based on hits in the silicon tracker and in the outer muon system [44]. Reconstructed electron or muon candidates are required to have $p_T > 20$ GeV. Electron candidates are required to be reconstructed within $|\eta| \leq 2.5$, excluding the CMS barrel-to-endcap transition region of the ECAL [45], and muon candidates are required to be reconstructed in the fiducial region $|\eta| \leq 2.4$ of the tracker system. The track associated to a lepton candidate is required to have both its transverse and longitudinal impact parameters compatible with the position of the primary vertex (PV) of the event. The PV for each event is defined as the one with the largest $\sum p_T^2$, where the sum runs over all the tracks used to fit the vertex. A particle-based relative isolation parameter is computed for each lepton, and corrected on
an event-by-event basis for contributions from pileup. The particle candidates used to compute the isolation variable are reconstructed with the particle flow algorithm which will be detailed later. We require that the sum of the scalar \( p_T \) of all particle candidates reconstructed in an isolation cone with radius \( R = \sqrt{\Delta \eta^2 + (\Delta \phi)^2} < 0.4 \) around the lepton’s momentum vector is <10 or <12% of the electron or muon \( p_T \) value, respectively. The two leptons with opposite electric charge and with highest \( p_T \) are chosen to form the dilepton pair. Same-flavoured dileptons (ee or \( \mu \mu \)) compatible with \( Z \rightarrow \ell \ell \) decays are then selected by requiring \( |M_{Z} - M_{\ell \ell}| < 15 \text{ GeV} \), where \( M_{Z} \) is the mass of the \( Z \) boson [19].

Two types of jets are used in the analysis: “jet-plus-track” (JPT) [46] and particle-flow (PF) [14] jets. Both cases use the anti-\( k_T \) algorithm [47,48] with a distance parameter of 0.5 to define jets. The information from the ECAL, HCAL and tracker are used by both algorithms in distinct ways. The JPT algorithm improves the energy response and resolution of calorimeter jets by incorporating additional tracking information. For JPT jets the associated tracks are classified as in-cone or out-of-cone if they point to within or outside the jet cone around the jet axis at the surface of the calorimeter. The momenta of both in-cone and out-of-cone tracks are then added to the energy of the associated calorimeter jet and for in-cone tracks the expected average energy deposition in the calorimeters is subtracted based on the momentum of the track. The direction of the jet axis is also corrected by the algorithm. As a result, the JPT algorithm improves both the energy and the direction of the jet. The PF algorithm [49,50] 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 constructed by clustering these particle candidates and the jet momentum is defined as the vectorial sum of the momenta of all particle candidates. An area-based correction is applied to both JPT and PF jets, to account for the extra energy that is clustered on the fraction of tracks in the jet, associated to other reconstructed jets. Such pileup jets can be rejected through a multivariate analysis based on the kinematics of the jet, on the topological configuration of its constituents, and on the fraction of tracks in the jet, associated to other reconstructed PVs in the same event [53]. The efficiency of both jet identification and pileup rejection is measured in the control sample, and determined to be >98% for both JPT and PF jets. The dependence of this efficiency on \( \eta \) agrees with that predicted in MC simulation. The residual \( \eta \)-dependent difference is used to assign a systematic uncertainty in the selected signal.

In the preselection of events we require at least two jets with \( p_T > 30 \text{ GeV} \) and \( |\eta| \leq 4.7 \). The two jets of highest \( p_T \) jets are defined as the tagging jets. For the measurement of the cross section, we require the leading jet to have \( p_T > 50 \text{ GeV} \) and the dijet invariant mass \( M_{jj} > 200 \text{ GeV} \). Other selection requirements will be described below, as they depend on the analysis.

### 5 Control regions for jets and modelling of background

In our analysis, we select control regions for different purposes: to validate the calibrated jet energy response and efficiencies of jet-identification criteria, to estimate the backgrounds and to verify the agreement between data and estimates of background. The following details the result of these cross-checks.

#### 5.1 Jet identification and response

Events with either a \( Z \rightarrow \mu \mu \) or a photon candidate, produced in association with a single jet with \( p_T > 30 \text{ GeV} \), are used as one of the control samples in this analysis. The \( Z \) candidate or the photon, and the associated jet are required to have \( |\Delta \phi (\text{jet, Z or } \gamma)| > 2.7 \text{ rad} \). These events enable a measure of the efficiency of the algorithms used to reject calorimeter noise and pileup-induced jets, and to check the jet energy response.

The jet identification criteria are based on the fractions of the jet energy deposited in different calorimeter elements [14]. Besides calorimetric noise, pileup events result in additional reconstructed jets. Such pileup jets can be rejected through a multivariate analysis based on the kinematics of the jet, on the topological configuration of its constituents, and on the fraction of tracks in the jet, associated to other reconstructed PVs in the same event [53]. The efficiency of both jet identification and pileup rejection is measured in the control sample, and determined to be >98% for both JPT and PF jets. The dependence of this efficiency on \( \eta \) agrees with that predicted in MC simulation. The residual \( \eta \)-dependent difference is used to assign a systematic uncertainty in the selected signal.

The same control sample is also used to verify the jet energy response [14], which is defined from the ratio \( \left[ \frac{p_T(\text{jet})}{p_T(Z \text{ or } \gamma)} \right] \). The double ratio of the response in data and in simulation, i.e. \( \left[ \frac{p_T(\text{jet})}{p_T(Z \text{ or } \gamma)} \right]_{\text{data}}/\left[ \frac{p_T(\text{jet})}{p_T(Z \text{ or } \gamma)} \right]_{\text{MC}} \), provides a residual uncertainty that is assigned as a systematic source of uncertainty to the measurement. Although partially covered by the JES uncertainties, this procedure considers possible residual uncertainties in the particular phase-space regions selected in our analysis. This evaluation is crucial for the most forward region of \( \eta \), where the uncertainties in response are large. The double
5.2 Discriminating gluons from quarks

Jets in signal events are expected to originate from quarks while for background events it is more probable that jets are initiated by a gluon emitted from a radiative QCD process. A quark–gluon (q/g) discriminant [11] is evaluated for the events, with $R_T^\text{hard} \geq 0.14$. The $z^*$ distribution (right) is shown for the same category of events. The panels below the distributions show the ratio between the data and expectations as well as the uncertainty envelope for the impact of the uncertainty of the JES.

In addition the q/g discriminant makes use of the following variables, computed using the weighted $p_T^3$-sum of the particles inside a jet: the jet constituents’ major root-mean-square (RMS) distance in the $\eta$–$\phi$ plane, the jet constituents’ minor RMS distance in the $\eta$–$\phi$ plane, and the jet asymmetry pull. Further details can be found in [54, 55].

The variables are used as an input to a likelihood-ratio discriminant that is trained using the TMVA package [56] on gluon and quark jets from simulated dijet events. To improve the separation power, all variables are corrected for their pileup contamination using the same estimator for the average energy density from pileup interactions [51, 52], as previously defined in Sect. 4. The performance of the q/g discriminant has been evaluated and validated using independent, exclusive samples of $Z$+jet and dijet data [54]. The use of the gluon–quark likelihood discriminator leads to a decrease of the statistical uncertainty of the measured signal by about 5%.
5.3 Modeling background

Alternative background models are explored for the dominant DY Zjj background. Given that the majority of the $\ell\ell jj$ final states are produced through DY Zjj processes it is crucial to have different handles on the behavior of this process, in particular, in the signal phase space region.

Simulation-based prediction for background

The effect of virtual corrections to the MADGRAPH-based (Born-level) description of DY Zjj is studied using MCFM. Comparisons are made between the predictions of MCFM parton-level distributions with NLO and LO calculations and these studies provide a dynamic NLO to LO scale factor (K-
Fig. 6 Distributions for the tagging jets for $M_{jj} > 750$ GeV in the combined dielectron and dimuon event sample: (upper left) $p_T$ of the leading jet, (upper right) $p_T$ of the sub-leading jet, (middle left) hard process $p_T$ (dijet+Z system), (middle right) $\eta$ of the most forward jet, (lower left) $\eta$ of the most central jet and (lower right) $\Delta \eta_{jj}$ of the tagging jets. In the top panels, the contributions from the different background sources and the signal are shown stacked being data superimposed. In all plots the signal shape is also superimposed separately as a thick line. The bottom panels show the ratio between data and total prediction. The total uncertainty assigned to the DY Zjj background estimate from $yjj$ control sample in data is shown in all panels as a shaded grey band.
factor) as a function of $M_{jj}$ and of the difference between the rapidity of the $Z$ boson and the average rapidity of the two tagging jets, i.e.

$$y^* = yZ - \frac{1}{2}(y_{j1} + y_{j2}).$$

(1)

The $K$-factor is observed to have a minor dependence on $M_{jj}$, but to increase steeply with $|y^*|$, and a correction greater than 10% relative to the signal, is obtained for $|y^*| > 1.2$. As a consequence, an event selection of $|y^*| < 1.2$ is introduced in the DY $Zjj$ simulation-based analyses. Finally, the difference between the nominal MADGRAPH prediction and the one obtained after reweighting it with the dynamic $K$-factor, on an event-by-event basis, is assigned as a systematic uncertainty for the DY $Zjj$ background prediction from simulation.

For the selection of the signal-region in the analysis where DY $Zjj$ is based on simulation we make use of an event balance variable, $R_{p_T}^{hard}$, defined as

$$R_{p_T}^{hard} = \frac{|p_{Tj1} + p_{Tj2} + p_{TZ}|}{|p_{Tj1}| + |p_{Tj2}| + |p_{TZ}|} = \frac{|p_{Tj1} + p_{Tj2} + p_{TZ}|}{|p_{Tj1}| + |p_{Tj2}| + |p_{TZ}|},$$

(2)

where the numerator is the estimator of the $p_T$ for the hard process, i.e. $p_{Tj1}^{hard}$. The distribution of the $R_{p_T}^{hard}$ variable is shown in Fig. 3 (left), where data and simulation are found to be in agreement with each other. It can be seen, from the same figure, that the variable is robust against the variation of JES according to its uncertainty. We apply a requirement of $R_{p_T}^{hard} < 0.14$ to select the signal region and the events failing this requirement are used as a control region for the analyses. The cut is motivated by the fact that the signal is expected to have the $Z$ boson balanced with respect to the dijet system in the transverse plane. The events which fail this requirement are used as control region for the modelling of the background. The $M_{jj}$ distribution in dimuon events for the signal and control regions is shown in Fig. 3, (middle) and (right), correspondingly. The reweighting of the DY $Zjj$ background is applied to the simulation, as described above. Data and predictions are found to be in agreement with each other.

Figure 4 shows distributions for angle-related variables. Fair agreement is observed for the absolute differences in the azimuthal angle ($\Delta \phi_{jj}$) and in the pseudorapidity ($\Delta \eta_{jj}$) of the tagging jets which are shown on the left and middle, respectively. The $z^*$ variable [10] is shown in Fig. 4 (right), and it is defined as

$$z^* = \frac{y^*}{\Delta y_{jj}}.$$  

(3)

Data is verified to be in good agreement with the prediction for the distribution in $z^*$ variable.

<table>
<thead>
<tr>
<th>Channel</th>
<th>ee, $\mu$$\mu$</th>
<th>$\mu$$\mu$</th>
<th>ee, $\mu$$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channels</td>
<td>ee, $\mu$$\mu$</td>
<td>$\mu$$\mu$</td>
<td>ee, $\mu$$\mu$</td>
</tr>
<tr>
<td>Variables</td>
<td>$M_{jj}$</td>
<td>$p_{Tj1}$, $p_{Tj2}$</td>
<td>$\eta_{j1}$, $\eta_{j2}$</td>
</tr>
<tr>
<td>Jets</td>
<td>PF</td>
<td>JPT</td>
<td>PF</td>
</tr>
</tbody>
</table>

| DY $Zjj$ model | MC-based | MC-based | From data |

**Data-based prediction for background**

The diagrams contributing to the production of a photon and two jets ($\gamma jj$) are expected to resemble those involved in the production of DY $Zjj$ (see Fig. 2). Thus, we build a data-based model for the shapes of the distributions of the kinematic observables of the tagging jets from $\gamma jj$ events selected in a similar way as the $Zjj$ ones. The differences, specific to the $Z$ or photon-sample, are expected to be mitigated by reweighting the $p_T$ of the photons to the $p_T$ of the $Z$ candidates. From simulation, we expect that the differences between the $\gamma$ and $Z$ masses do not contribute significantly when matching the dijet kinematics between the two samples after $M_{jj} > 2M_Z$ is required. Given that the photon sample is affected by multijet production, and that the selection of the low-$p_T$ region in data is also affected by very large prescaling at the trigger stages, we impose tighter kinematic constraints on the reconstructed boson, with respect to the ones applied at pre-selection (Sect. 4). To match effectively the $Z$ and photon kinematics, we require $p_T(Z \text{ or } \gamma) > 50$ GeV and rapidity $|y(Z \text{ or } \gamma)| < 1.44$. The rapidity requirement corresponds to the physical boundary of the central (barrel) region of the CMS ECAL [45].
The method is checked in simulation by characterising the DY Zjj or direct photon events in different physical regions defined according to the reconstructed $M_{jj}$ and comparing both distributions. Figure 5 illustrates the compatibility of simulated events with a high dijet invariant mass. Good agreement is found for the $\eta$ of the most forward jet, the $\Delta\eta_{jj}$ variable and the ratio between the $p_T$ of the dijet system to the scalar sum of the tagging jets’ $p_T$.

The smallest of the quark/gluon discriminant value among the tagging jets is also found to be in agreement — Fig. 5 (top right). In general, the kinematics of the tagging jets predicted from the photon sample are found to be in agreement with those observed in DY Z events also for lower $M_{jj}$ values. A
similar conclusion holds for other global event observables inspected in the simulation, such as energy fluxes and angular correlations.

The result of the compatibility tests described above have the potential to yield a correction factor to be applied to the DY \( Zjj \) prediction from the photon data. However due to the limited statistics in our simulation and due to uncertainties in handling the simulation of residual background from multijet events in data, we have opted to use the simulation-based compatibility test results to assign, instead, an uncertainty in the final shape. We assign the difference in the compatibility tests relative to a pure prompt-photon possibility as one of the systematic uncertainties. The changes observed in the compatibility test, obtained after varying the PDF by its uncertainties synchronously in the two samples is also assigned as a source of uncertainty. In data, the difference between a “tight” and a “loose” photon selections is, furthermore, assigned as an extra source of systematic uncertainty. The selection is tightened by applying stricter requirements on the photon identification and isolation requirements. This prescription is adopted to cover possible effects from the contamination of multijet processes.

The final distributions for DY \( Zjj \) events are obtained after subtracting a residual contamination from pure EW production of a photon in association with two jets (EW \( \gamma jj \)) [57]. The diagrams for the latter process are similar to the ones of Fig. 1 (left) and (middle), where the \( Z/\gamma^* \) is now a real photon. For a fiducial phase space defined by \( M_{jj} > 120 \, \text{GeV} \), \( p_T^{Tj} > 30 \, \text{GeV} \), \( |\eta_j| < 5 \), \( p_T^{\gamma} > 50 \, \text{GeV} \) and \( |\eta_{jj}| < 1.5 \), the production cross section of EW \( \gamma jj \) process is expected to be 2.72 pb, based on the MADGRAPH generator. After event reconstruction and selection, we estimate the ratio of the number of EW \( \gamma jj \) candidate events to the total number of photon events selected in data to be a factor of \( \approx 5 \) times smaller than the ratio between the expected EW \( Zij \) and DY \( Zjj \) yields. From simulations this ratio is expected to be independent of \( M_{ij} \). In the subtraction procedure, a 30% normalisation uncertainty is assigned to this residual process, which corresponds to approximately twice the envelope of variations obtained for the cross section at NLO with VBFNLO, after tightening the selection criteria and changing the factorisation and renormalisation scales.

The results obtained when the data-based prediction, used to characterise the DY \( Zjj \) contribution to the reconstructed kinematics of the tagging jets in data, show a good agreement for different dijet invariant mass categories. Figure 6 illustrates the agreement observed for \( M_{jj} > 750 \, \text{GeV} \) in the distribution of different variables: (upper left) \( p_T \) of the leading jet, (upper right) \( p_T \) of the sub-leading jet, (middle left) hard process \( p_T \) (dijet+Z system), (middle right) \( \eta \) of the most forward jet, (lower left) \( \eta \) of the most central jet and (lower right) \( \Delta \eta_{jj} \) of the tagging jets.

![Fig. 8 Distributions for the BDT discriminants in \( \mu\mu \) events, for the control region (top row) and signal region (bottom row), used by analysis B. The ratio for data to MC simulations is given in the bottom panel on the left, showing the impact of changes in JES by \( \pm 1 \) SD. The bottom panel on the right shows the difference between data or the expected EW \( Zjj \) contribution with respect to the background (BG) ](image)

6 Signal discriminants and extraction procedure

We use a multivariate analysis technique that provides separation of the DY \( Zjj \) and EW \( Zjj \) components of the inclusive \( \ell\ell jj \) spectrum. As discussed previously, the EW \( Zjj \) signal is characterised by a large \( \Delta \eta_{jj} \) jet separation that stems from the small-angle scattering of the two initial partons. Owing to both the topological configuration and the large \( p_T \) of the
outgoing partons, the $M_{jj}$ variable is also expected to be large. The evolution of $\Delta y_{jj}$ with $M_{jj}$ is expected to be different in signal and background events and therefore these characteristics are expected to yield the best separation power between the EW $Zjj$ and the DY $Zjj$ productions. In addition, one can exploit the fact that the $Z$-boson candidate is expected to be produced centrally in the rapidity region defined by the two tagging jets and that the $Zjj$ system is approximately balanced in the transverse plane. As a consequence, we expect the signal to be found with lower values of both $y^*$ and $p_T^{\text{hard}}$, compared to the DY background. Other variables which can be used to enhance the separation are related to the kinematics of the event ($p_T$, rapidity, and distance between the jets and/or the $Z$ boson) or to the properties of the jets that are expected to be initiated by quarks. We combine these variables using three alternative multivariate analyses with the goal of cross-checking the final result. All three analyses make use of boosted decision tree (BDT) discriminators implemented using TMVA package [56] to achieve the best expected separation between the EW $Zjj$ and DY $Zjj$ processes.

Analysis A expands one of the procedures previously adopted for the 7 TeV measurement [11]. It uses both dimuon and dielectron final states and PF jet reconstruction. A multivariate discriminator making use of the dijet and $Z$ boson kinematics is built. A choice is made for variables which are robust against JES uncertainties. Extra discrimination information, related to the $q/g$ nature of the jet, is included. All processes are modelled from simulation, and the description of each variable is verified by comparing data with the simulation-based expectations in control regions.

Analysis B uses only the dimuon final state and the JPT jet reconstruction approach. It builds a discriminator which tries to profit from the full kinematics of the event including the tagging jets and the $Z$ boson. Similarly to analysis A it expands one of the cross-check procedures previously adopted for the 7 TeV measurement [11] and relies on simulation-based prediction of the backgrounds.

Analysis C uses solely dijet-related variables in the multivariate discriminator and selects both the dimuon and dielectron final states with PF jets. Lepton-related selection variables are not used as the main background is derived from the photon control sample. In this analysis events are split in four categories for $M_{jj}$ values in the intervals 450–550 GeV, 550–750 GeV, 750–1,000 GeV, and above 1,000 GeV, which have been chosen to have similar numbers of expected signal events.

Table 1 compares in more detail the three independent analyses A, B and C. From simulation, the statistical correlation between the analyses, if performed with the same final state, is estimated to be $\approx 60\%$. 

Figures 7, 8 and 9 show the distributions of the discriminants for the three analyses. Good agreement is observed overall in both the signal and in the control regions which are defined according to the value of the $R_{T}^{\text{hard}}$ or $M_{jj}$ variables (see Sect. 5.3).

Each analysis has a binned maximum likelihood formed from the expected rates for each process, as function of the value of the discriminant, which is used to fit simultaneously across the control and signal categories the strength modifiers for the EW $Zjj$ and DY $Zjj$ processes, $\mu = \sigma(\text{EW } Zjj)/\sigma_{\text{LO}}(\text{EW } \ell\ell jj)$ and $\nu = \sigma(\text{DY})/\sigma_{\text{NLO}}(\text{DY})$. Nuisance parameters are added to modify the expected rates and shapes according to the estimate of the systematic uncertainties affecting the analysis and are mostly assumed to have a log-normal distribution.

The interference between the EW $Zjj$ and the DY $Zjj$ processes is taken into account in the fitting procedure. A parameterisation of the interference effects, as a function of the parton-level $M_{jj}$ variable, is derived from the MADGRAPH simulation described in Sect. 3. The matrix elements for the EW $Zjj$ and DY $Zjj$ processes provide the total yields for the $\ell\ell jj$ final state as

$$\hat{N}_{\ell\ell jj}(\mu, \nu) = \mu N_{\text{EW } Zjj} + \nu N_{\text{DY } Zjj}, \quad (5)$$

where $N_{\text{EW } Zjj}, N_{\text{DY } Zjj}$ are the yields for the EW $Zjj$ and DY $Zjj$ processes, $N_1$ is the expected contribution from the interference to the total yield, and $\mu$ and $\nu$ are the strength factors that modify the SM predictions. In the absence of signal (or background) the contribution from the interference term vanishes in Eq. (5).

The parameters of the model ($\mu$ and $\nu$) are determined maximising a likelihood ($L$). Systematic uncertainties are incorporated in the fit by scanning the profile likelihood ratio $\lambda$, defined as

$$\lambda(\mu, \nu) = \frac{L(\mu, \nu, \hat{\theta})}{L(\hat{\mu}, \hat{\nu}, \hat{\theta})}, \quad (6)$$

where the denominator has estimators $\hat{\mu}, \hat{\nu}$ and $\hat{\theta}$ that maximise the likelihood, and the numerator has estimators $\hat{\theta}$ that maximise the likelihood for the specified $\mu$ and $\nu$ strengths. The statistical methodology used is similar to the one used in the CMS Higgs analysis [5] using asymptotic formulas [58]. In this procedure some of the systematic uncertainties affecting the measurement of the signal strength are partially constrained. The DY $Zjj$ strength is constrained by the uncertainties in analyses A and B and is free to change in C. In all cases the difference of the result relative to the one that would have been obtained without taking the interference term into account, is assigned as a systematic uncertainty of the measurement. This shall be discussed in more detail in the next section where the systematic uncertainties affecting our analysis are summarised.
Fig. 9 Distributions for the BDT discriminants in ee+μμ events for different $M_{jj}$ categories, used in analysis C. The ratios at the bottom each subfigure of the top row gives the results of data to expectation for the two control regions of $M_{jj}$. The lower panel of the bottom subfigure shows the difference between data or the expected EW Zjj contribution with respect to the background (BG).

7 Systematic uncertainties

The main systematic uncertainties affecting our measurement are classified into experimental and theoretical sources.

7.1 Experimental uncertainties

The following experimental uncertainties are considered:

Luminosity—A 2.6% uncertainty is assigned to the value of the integrated luminosity [59].

Trigger and selection efficiencies—We assign total 2 and 3% uncertainties on the total trigger and selection efficiencies in the ee and $\mu\mu$ channels, respectively. These uncertainties have been estimated by comparing the lepton efficiencies expected in simulation and measured in data with a “tag-and-probe” method [60].

Jet energy scale and resolution—The energy of the jets enters in our analysis not only at the selection level but also in the computation of the kinematic variables used in forming discriminants. The uncertainty on JES affects
therefore both the expected event yields, through the migration of events to different bins, and the final distributions. In addition to the standard JES uncertainty, the residual difference in the response observed in the balancing of a Z or γ candidate with a jet, discussed in Sect. 5, is assigned as a systematic uncertainty. The effect of the JES uncertainty is studied by rescaling up and down the reconstructed jet energy by a \( p_T \) - and \( \eta \)-dependent scale factor \[14\]. An analogous approach is used for the JER.

### Table 2

Summary of the relative variation of uncertainty sources (in \%) considered for the evaluation of the systematic uncertainties in the different analyses. A filled or open circle signals whether that uncertainty affects the distribution or the absolute rate of a process in the fit, respectively. For some of the uncertainty sources “variable” is used to signal that the range is not unambiguously quantifiable by a range, as it depends on the value of the discriminants, event category and may also have a statistical component.

<table>
<thead>
<tr>
<th>Source</th>
<th>Shape</th>
<th>Methods A, B</th>
<th>Method C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminosity</td>
<td>○</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Trigger/selection</td>
<td>○</td>
<td>2–3</td>
<td></td>
</tr>
<tr>
<td>JES and residual jet response</td>
<td>●</td>
<td>1–10</td>
<td></td>
</tr>
<tr>
<td>JER</td>
<td>●</td>
<td>6–15</td>
<td></td>
</tr>
<tr>
<td>Pileup</td>
<td>●</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Simulation statistics</td>
<td>●</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>DY Zjj distribution (data)</td>
<td>●</td>
<td>– Variable</td>
<td></td>
</tr>
</tbody>
</table>

| Theoretical                   |       |              |          |
| PDF                           | ●     | Variable     |          |
| \( \mu_R/\mu_F \) (signal)    | ●     | Variable     |          |
| DY Zjj shape (MC)             | ●     | Variable     |          |
| DY Zjj shape (PDF and EW \( \gamma \)jj contribution) | ● | – | Variable |
| Interference                  | ●     | 100          |          |
| Normalisation of top-quark and diboson processes | ○ | 7–10 |

### Table 3

Event yields expected after fits to background and signal processes in methods A or B, using the initial selections (summarised in Table 1), and requiring \( S/B > 10 \% \). The yields are compared to the data observed in the different channels and categories. The total uncertainties quoted for signal, DY Zjj, dibosons (VV), and processes with top quarks (if and single top quarks) are dominated by JES uncertainties and include other sources, e.g., the statistical fluctuations in the MC samples.

<table>
<thead>
<tr>
<th>Selection</th>
<th>Channel</th>
<th>VV</th>
<th>Top quark</th>
<th>DY Zjj</th>
<th>Total backgrounds</th>
<th>EW Zjj</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>ee (A)</td>
<td>255±14</td>
<td>314±15</td>
<td>20,083±857</td>
<td>20,652±857</td>
<td>659±16</td>
<td>20,752</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu ) (A)</td>
<td>355±15</td>
<td>456±16</td>
<td>30,042±1,230</td>
<td>30,853±1,230</td>
<td>925±22</td>
<td>30,306</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu ) (B)</td>
<td>226±13</td>
<td>295±12</td>
<td>25,505±1,735</td>
<td>26,026±1,735</td>
<td>833±14</td>
<td>26,651</td>
</tr>
<tr>
<td>BDT&gt;0.05</td>
<td>ee (A)</td>
<td>56±6</td>
<td>50±7</td>
<td>3,541±169</td>
<td>3,647±169</td>
<td>427±12</td>
<td>3,979</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu ) (A)</td>
<td>38±5</td>
<td>36±5</td>
<td>2,867±135</td>
<td>2,941±135</td>
<td>459±14</td>
<td>3,182</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu ) (B)</td>
<td>36±3</td>
<td>35±3</td>
<td>3,871±273</td>
<td>3,942±273</td>
<td>514±12</td>
<td>4,312</td>
</tr>
</tbody>
</table>

In both cases the uncertainties are derived separately of PF and JPT jets.

\( q/g \) discriminator—The uncertainty on the performance of the \( q/g \) discriminator has been measured using independent Z+jet and dijet data, after comparing with the corresponding simulation predictions \[54\]. The parametrization of the estimated uncertainty is used on an event-per-event basis to derive alternative predictions for the signal and background which are profiled in the fit for the signal.

Pileup—Pileup is not expected to affect the identification and isolation of the leptons or the corrected energy of the jets. When the jet clustering algorithm is run, pileup can, however, induce a distortion of the reconstructed dijet system due to the contamination of tracks and calorimetric deposits. We evaluate this uncertainty by generating two alternative distributions after changing the number of pileup interactions by \( \pm 5 \% \), according to the uncertainty on the inelastic pp cross section at \( \sqrt{s} = 8 \) TeV.

Statistics of simulation—For signal and backgrounds which are estimated from simulation we form envelopes for the distributions by shifting all bin contents simultaneously up or down by its statistical uncertainty. This generates two alternatives to the nominal shape to be analysed. However, when a bin has an uncertainty which is \( >10 \% \), we assign an additional, independent uncertainty to it in the fit in order to avoid overconstraining a specific background from a single bin in the fit.

### 7.2 Theoretical uncertainties

We have considered the following theoretical uncertainties in the analysis:

- PDF—The PDF uncertainties are evaluated by considering the pdf4lhc prescription \[18,21–24\], where for each source a new weight is extracted event-by-event and used to generate an alternative signal distribution. The up and down changes relative to the nominal prediction for each
Table 4 Event yields expected before the fit to background and signal processes in method C. The yields are compared to the data observed in the different channels and categories. The total systematic uncertainty assigned to the normalisation of the processes is shown.

<table>
<thead>
<tr>
<th>( M_T ) (GeV)</th>
<th>Channel</th>
<th>VV</th>
<th>Top quark</th>
<th>DY ( Zjj )</th>
<th>Total backgrounds</th>
<th>EW ( Zjj )</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>450–550</td>
<td>ee</td>
<td>20±2</td>
<td>68±4</td>
<td>5.438±731</td>
<td>5.526±731</td>
<td>94±6</td>
<td>5.809</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu )</td>
<td>27±2</td>
<td>96±4</td>
<td>7.325±983</td>
<td>7.448±983</td>
<td>128±8</td>
<td>8.391</td>
</tr>
<tr>
<td>550–750</td>
<td>ee</td>
<td>16±1</td>
<td>56±3</td>
<td>3.802±496</td>
<td>3.874±664</td>
<td>112±7</td>
<td>4.139</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu )</td>
<td>30±2</td>
<td>69±4</td>
<td>5.234±683</td>
<td>5.333±896</td>
<td>155±10</td>
<td>5.652</td>
</tr>
<tr>
<td>750–1,000</td>
<td>ee</td>
<td>5.4±0.5</td>
<td>20±2</td>
<td>1.300±188</td>
<td>1.325±236</td>
<td>73±5</td>
<td>1.384</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu )</td>
<td>7.5±0.6</td>
<td>26±2</td>
<td>1.846±262</td>
<td>1.880±313</td>
<td>98±6</td>
<td>1.927</td>
</tr>
<tr>
<td>&gt;1,000</td>
<td>ee</td>
<td>2.7±0.4</td>
<td>10.2±0.8</td>
<td>600±84</td>
<td>613±90</td>
<td>84±6</td>
<td>684</td>
</tr>
<tr>
<td></td>
<td>( \mu\mu )</td>
<td>4.2±0.4</td>
<td>13±1</td>
<td>913±127</td>
<td>930±122</td>
<td>114±8</td>
<td>923</td>
</tr>
</tbody>
</table>

Table 5 Fitted signal strengths in the different analyses and channels including the statistical and systematic uncertainties. For method C, only events with \( M_T > 450 \) GeV are used. The breakup of the systematic components of the uncertainty is given in detail in the listings.

<table>
<thead>
<tr>
<th>Analysis A</th>
<th>Analysis B</th>
<th>Analysis C</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
</tr>
<tr>
<td>ee</td>
<td>ee</td>
<td>ee</td>
</tr>
<tr>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
</tr>
<tr>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
<td>( \mu\mu )</td>
</tr>
</tbody>
</table>

Independent variable and are added in quadrature to estimate the final uncertainty.

Factorisation and renormalisation scales—In contrast to the main background, the two signal process partons originate from electroweak vertices. Changing the QCD factorisation and renormalisation scales is therefore not expected to have a large impact on the final cross section. The renormalisation scale, in particular, is not expected to have any impact at LO. Changing the values of \( \mu_F \) and \( \mu_R \) from their defaults by 2 or 1/2 we find a variation of \( \approx 4\% \) in MadGraph and in VBFNLO. As the change in the scales can also affect the expected kinematics, we use the altered \( \mu_R/\mu_F \) samples to extract a weight that is applied at the generator level on an event-by-event basis. The parameterisation is done as function of the dilepton \( p_T \). The changes induced in the form of the discriminant at the reconstruction level are assigned as systematic uncertainties.

DY \( Zjj \) prediction—for the modelling of the DY \( Zjj \) background from simulation, as we indicated previously, we consider the full difference between the Born-level MadGraph prediction and the NLO prediction based on MCQ as a systematic uncertainty. The differences are particularly noticeable at very large \( M_T \) and at large \( y^* \). For the data-based modelling of DY \( Zjj \) we consider the effect induced on the discriminant functions from five distinct sources. Not all are of theoretical nature, nevertheless, we list them here for simplicity. We consider not only the statistical size of the photon sample but also the difference observed in data selected with a loose-photon selection relative to the data selected with a tight-photon selection. From simulation, the expected difference, between the tight-photon selection and a pure photon sample is also considered, and added in quadrature to the previous. Furthermore, we consider the envelope of the PDF changes induced in the simulated compatibility tests, and the contamination from
residual EW γjj events in the photon sample. For the latter, we assign a 30% uncertainty to the EW γjj contribution, which is added in quadrature to the statistical uncertainty in the simulated events for this process.

Normalisation of residual backgrounds—Diboson and top-quark processes are modelled with a MC simulation. Thus, we assign an intrinsic uncertainty in their normalisation according to their uncertainty which arises from the PDF and factorisation/renormalisation scales. The uncertainties are assigned based on [31,37,40].

Interference between EW Zjj and DY Zjj—The difference observed in the fit when the interference term is neglected relative to the nominal result is used to estimate the uncertainty due to the interference of the signal and the background.

7.3 Summary of systematic uncertainties

Table 2 summarises the systematic uncertainties described above. We give their magnitudes at the input level, and whether they are treated as normalisation uncertainties or uncertainties in the distributions used to fit the data. The uncertainties are organised according to their experimental or theoretical nature.

8 Measurement of the EW Zjj production cross section

Table 3 reports the expected and observed event yields after imposing a minimum value for the discriminators used in methods A and B such that \( S/B > 10 \% \). Table 4 reports the event yields obtained in each category for method C. Fair agreement is observed between data and expectations for the sum of signal and background, for both methods, in all categories.

The signal strength is extracted from the fit to the discriminator shapes as discussed in Sect. 6. Table 5 summarises the results obtained for the fits to the signal strengths in each method. The results obtained are compatible among the dilepton channels and different methods, and in agreement with the SM prediction of unity. Methods A and B are dominated by the systematic uncertainty stemming from the modelling of the DY Zjj background and the interference with the EW Zjj signal. Method C is dominated by the statistical uncertainty in the fit and, due to tighter selection criteria, is expected to be less affected by the modelling of the interference. In method C, the DY Zjj modelling uncertainty is partially due to the statistics of the photon sample. With the exception of jet energy resolution, which has a larger impact in method C due to its tighter \( M_{jj} \) selection, all other uncertainties are of similar magnitude for the different methods.

For the results from method C, the 68 and 95% confidence levels (CL) obtained for the combined fit of the EW Zjj and DY Zjj strengths are shown in Fig. 10. Good agreement is found with the SM prediction for both components, as well as with the expected magnitude of the CL intervals. The DY Zjj strength is measured to be \( 0.978 \pm 0.013 \) (stat) \( \pm 0.036 \) (syst) in the ee channel, \( 1.016 \pm 0.011 \) (stat) \( \pm 0.034 \) (syst) in the \( \mu \mu \) channel, and \( 0.996 \pm 0.008 \) (stat) \( \pm 0.025 \) (syst) after the combination of the previous two.

From the combined fit of the two channels in analysis A we obtain the signal strength \( \mu = 0.84 \pm 0.07 \) (stat) \( \pm 0.19 \) (syst) = 0.84 \( \pm 0.20 \) (total), corresponding to a measured signal cross section

\[
\sigma(\text{EW } \ell \ell jj) = 174 \pm 15 \text{ (stat)} \pm 40 \text{ (syst)} \text{ fb}
\]

\[
= 174 \pm 42 \text{ (total)} \text{ fb},
\]

in agreement with the SM prediction \( \sigma_{\text{LO}}(\text{EW } \ell \ell jj) = 208 \pm 18 \text{ fb} \). Using the same statistical methodology, as described in Sect. 6, the background-only hypothesis is excluded with a significance greater than 5\( \sigma \).

9 Study of the hadronic and jet activity in Z+jet events

After establishing the signal, we examine the properties of the hadronic activity in the selected events. Radiation patterns and the profile of the charged hadronic activity as a function of several kinematic variables are explored in a region dominated by the main background, DY Zjj; these studies are pre-
Fig. 11 (Left) The average number of jets with $p_T > 40$ GeV as a function of the total $H_T$ in events containing a Z and at least one jet, and (right) average $\cos \Delta \phi_{jj}$ as a function of the total $H_T$ in events containing a Z and at least two jets. The ratios of data to expectation are given below the main panels. At each ordinate, the entries are separated for clarity. The expectations for EW Zjj are shown separately. The data and simulation points are shown with their statistical uncertainties presented in Sects. 9.1 and 9.2. The production of additional jets in a region with a larger contribution of EW Zjj processes is furthermore pursued in Sect. 9.3. We expect a significant suppression of the hadronic activity in signal events because the final-state objects have origin in purely electroweak interactions, in contrast with the radiative QCD production of jets in DY Zjj events. The reconstructed distributions are compared directly to the prediction obtained with a full simulation of the
CMS detector (see Sect. 3) and extends the studies reported in [61] to the phase space region of interest for the study of the EW Zjj process.

9.1 Jet radiation patterns

For the Z+jets events, the observables referred to as “radiation patterns” correspond to: (i) the number of jets, \( N_j \), (ii) the total scalar sum of the transverse momenta of jets reconstructed within \( |\eta| < 4.7 \), \( H_T \), (iii) \( \Delta \eta_{jj} \) between the two jets with \( p_T > 40 \text{ GeV} \) which span the largest pseudorapidity gap in the event (not required to be the two leading-\( p_T \) jets), and (iv) the cosine of the azimuthal angle difference, \( \cos|\phi_{j1} - \phi_{j2}| = \cos \Delta \phi_{jj} \), for the two jets with criterion (iii). These observables are measured using events that are required to satisfy the \( Z \rightarrow \mu \mu \) and \( Z \rightarrow ee \) selection criteria of analyses A and B. These observables are investigated following the prescriptions and suggestions from Ref. [62], where the model dependence is estimated by comparing different generators.

Figures 11 and 12 show the average number of jets and the average \( \cos \Delta \phi_{jj} \) as a function of the total \( H_T \) and \( \Delta \eta_{jj} \). The MadGraph + Pythia (ME-PS) predictions are in good agreement with the data, even in the regions of largest \( H_T \) and \( \Delta \eta_{jj} \). In both cases we estimate that the contribution from EW Zjj is \(< 1\%\). Jet multiplicity increases both as function of \( H_T \) and \( \Delta \eta_{jj} \). The increase of \( H_T \) and \( \Delta \eta_{jj} \) induces, in average, an increase of jet multiplicity and leads to different dijet configurations in the azimuthal plane. In average the two selected jets are separated by \( 120^\circ \) deg, independently of \( H_T \). This separation tends to decrease for larger \( \Delta \eta_{jj} \) separation. The behavior observed for \( \cos \Delta \phi_{jj} \) when \( \Delta \eta_{jj} < 0.5 \) is related to the jet distance parameter used in the reconstruction (\( R = 0.5 \)). In data, the separation of the jets in the \( \cos \Delta \phi_{jj} \) variable, is observed to be \(< 5\% \) smaller with respect to the simulation.

9.2 Study of the charged hadronic activity

For this study, a collection is formed of high-purity tracks \([63]\) with \( p_T > 0.3 \text{ GeV} \), uniquely associated with the main PV in the event. Tracks associated with the two leptons or with the tagging jets are excluded from the selection. The association between the selected tracks and the reconstructed PVs is carried out by minimising the longitudinal impact parameter which is defined as the \( z \)-distance between the PV and the point of closest approach of the track helix to the PV, labeled \( d_{zPV} \). The association is required to satisfy the conditions \( d_{zPV} < 2 \text{ mm} \) and \( d_{zPV} < 3 \delta d_{zPV} \), where \( \delta d_{zPV} \) is the uncertainty on \( d_{zPV} \).

A collection of “soft track-jets” is defined by clustering the selected tracks using the anti-\( k_T \) clustering algorithm \([47]\) with a distance parameter of \( R = 0.5 \). The use of track jets represents a clean and well-understood method \([64]\) to reconstruct jets with energy as low as a few GeV. These jets are not affected by pileup, because of the association of their tracks with the hard-scattering vertex \([65]\).
To study the central hadronic activity between the tagging jets, only track jets of low $p_T$, and within $\eta_{\text{min}}^{\text{tag jet}} + 0.5 < \eta < \eta_{\text{max}}^{\text{tag jet}} - 0.5$ are considered. For each event, we compute the scalar sum of the $p_T$ of up to three leading-$p_T$ soft-track jets, and define it as the soft $H_T$ variable. This variable is chosen to monitor the hadronic activity in the rapidity interval between the two jets.

The dependence of the average soft $H_T$ for the Zjj events as a function of $M_{jj}$ and $\Delta \eta_{jj}$ is shown in Fig. 13. Inclusively, the contribution from EW Zjj is estimated to be at the level of 1 %, but it is expected to evolve as function of the different variables, being 5 % (20 %) for $|\Delta \eta_{jj}| > 4$ ($M_{jj} > 1$ TeV). Overall, good agreement is observed between data and the simulation. The average value of the soft $H_T$ is observed...
to increase linearly with $M_{jj}$, and to saturate its value for $\Delta \eta_{jj} > 5$, as a consequence of the limited acceptance of the CMS tracker.

9.3 Jet activity studies in a high-purity region

The evidence for EW production of $\ell\ell jj$ final states can also be supported through a study of the emission of a third and other extra jets in a region of high signal purity, i.e. for large $M_{jj}$. In this study, we compare two regions, one with $M_{jj} > 750$ GeV and another with $M_{jj} > 1,250$ GeV. Aside from the two tagging jets used in the preselection, we use all PF-based jets with a $p_T > 15$ GeV found within the $\Delta \eta_{jj}$ of the tagging jets. The background is modelled from the photon control sample (analysis C), and uses the normalisations obtained from the fit discussed in Sect. 8. Where relevant we also compare the results using the MC-based modelling of the background.

The number of extra jets, as well as their scalar $p_T$ sum ($H_T$), are shown in Fig. 14. Data and expectations are generally in good agreement for both distributions in the two $M_{jj}$ regions. A clear suppression of the emission of a third jet is observed in data, when we take into account the background-only predictions. After subtraction of the background, which
is shown as an inset in the different figures, we observe that slightly less extra jets tend to be counted in data with respect to the simulated signal. Notice that in the simulation of the signal, the extra jets have their origin in a parton-shower approach (see Sect. 3).

The $p_T$ values and the pseudorapidities relative to the average of the two tagging jets, i.e. $\eta_{j3} = \eta_{j3} - (\eta_{j1} + \eta_{j2})/2$, of the third leading-$p_T$ jet in the event, are shown in Fig. 15. There are some deviations of the data observed relative to the predictions. In particular, the third jet is observed to be slightly more central than expected. The poor statistical and other uncertainties prevent us, however, from drawing further conclusions.

The above distributions can be used to compute gap fractions. We define a gap fraction as the fraction of events which do not have reconstructed kinematics above a given threshold. The most interesting gap fractions can be computed for the $p_T$ of the leading additional jet, and the $H_T$ variable.

Fig. 16 Gap fractions for: (top row) $p_T$ of leading additional jet, (bottom row) the $H_T$ variable within the $\Delta\eta_{jj}$ of the two tagging jets in events with $M_{jj} > 750$ GeV (left) or $M_{jj} > 1250$ GeV (right). The observed gap fractions in data are compared to two different signal plus background predictions where DY $Zjj$ is modelled either from $\gamma jj$ data or from simulation. The bottom panels show the ratio between the observed data and different predictions.
These gap fractions are, in practice, measurements of the efficiency of extra jet veto in VBF-like topologies. By comparing different expectations with the observed data we can quantify how reliable is the modelling of the extra jet activity, in particular in a signal-enriched region. Figure 16 shows the gap fractions expected and observed in data. Two expectations are compared: the one using a full MC approach and the one where the DY $Zjj$ background is predicted from the $\gamma jj$ data. Both predictions are found to be in agreement with the data for the $p_T$ of the leading additional jet and the soft $H_T$ variable.

10 Summary

The cross section for the purely electroweak production of a $Z$ boson in association with two jets in the $\ell\ell jj$ final state, in proton–proton collisions at $\sqrt{s} = 8$ TeV has been measured to be

$$\sigma (\text{EW} \, \ell\ell jj) = 174 \pm 15 \text{ (stat)} \pm 40 \text{ (syst)} \, \text{fb},$$

in agreement with the SM prediction. Aside from the two analyses previously used to determine the cross section of this process at 7 TeV [11], a new analysis has been implemented using a data-based model for the main background. The increased integrated luminosity recorded at 8 TeV, an improved selection method, and more precise modelling of signal and background processes have allowed us to obtain a more precise measurement of the EW $Zjj$ process relative to the 7 TeV result.

Studies of the jet activity in the selected events show generally good agreement with the MadGraph+Pythia predictions. In events with high signal purity, the additional hadron activity has also been characterised, as well as the gap fractions. Good agreement has been found between data and QCD predictions.

Acknowledgments. We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centres and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); MoER, IUT and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF and WCU (Republic of Korea); LAS (Lithuania); MOE and UM (Malaysia); CINVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS and RFBR (Russia); MESTD (Serbia); SEIDI and CPAN (Spain); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA). Individuals have received support from the Marie-Curie programme and the European Research Council and EPLANET (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS programme of Foundation for Polish Science, cofinanced from European Union, Regional Development Fund; the Compagnia di San Paolo (Torino); the Consorzio per la Fisica (Trieste); MIUR project 2010XT4TXM (Italy); the Thalis and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF; and the National Priorities Research Program by Qatar National Research Fund.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited. Funded by SCOAP3 / License Version CC BY 4.0.

References

11. CMS Collaboration, Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the
University of Debrecen, Debrecen, Hungary
P. Raics, Z. L. Trocsanyi, B. Ujvari

National Institute of Science Education and Research, Bhubaneswar, India
S. K. Swain

Panjab University, Chandigarh, India

University of Delhi, Delhi, India
Ashok Kumar, Arun Kumar, S. Ahuja, A. Bhardwaj, B. C. Choudhary, A. Kumar, S. Malhotra, M. Naimuddin, K. Ranjan, V. Sharma

Saha Institute of Nuclear Physics, Kolkata, India

Bhabha Atomic Research Centre, Mumbai, India
A. Abdulsalam, D. Dutta, S. Kailas, V. Kumar, A. K. Mohanty, L. M. Pant, P. Shukla, A. Topkar

Tata Institute of Fundamental Research, Mumbai, India

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
H. Bakhshiansohi, H. Behnamian, S. M. Etesami, A. Fahim, R. Goldouzian, A. Jafari, M. Khakzad, M. Mohammadi Najafabadi, M. Naseri, S. Paktinat Mehdiabadi, B. Safarzadeh, M. Zeinali

University College Dublin, Dublin, Ireland
M. Felcini, M. Grunewald

INFN Sezione di Bari, Università di Bari, Politecnico di Bari, Bari, Italy

INFN Sezione di Bologna, Università di Bologna, Bologna, Italy

INFN Sezione di Catania, Università di Catania, CSFSNMS, Catania, Italy
S. Albergo, G. Cappello, M. Chiorboli, S. Costa, F. Giordano, R. Potenza, A. Tricomi, C. Tuveri

INFN Sezione di Firenze, Università di Firenze, Firenze, Italy
G. Barbaglio, V. Ciulli, C. Civinini, R. D’Alessandro, E. Focardi, E. Gallo, S. Gonzi, V. Gorì, P. Lenzi, M. Meschini, S. Paoletti, G. Sguazzoni, A. Tropiano

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, F. Fabbrì, D. Piccolo

INFN Sezione di Genova, Università di Genova, Genoa, Italy
F. Ferro, M. Lo Vetere, E. Robutti, S. Tosi

INFN Sezione di Milano-Bicocca, Università di Milano-Bicocca, Milan, Italy
The University of Iowa, Iowa City, USA

Johns Hopkins University, Baltimore, USA

The University of Kansas, Lawrence, USA

Kansas State University, Manhattan, USA

Lawrence Livermore National Laboratory, Livermore, USA
J. Gronberg, D. Lange, F. Rebassoo, D. Wright

University of Maryland, College Park, USA

Massachusetts Institute of Technology, Cambridge, USA

University of Minnesota, Minneapolis, USA

University of Mississippi, Oxford, USA
J. G. Acosta, S. Olivers

University of Nebraska-Lincoln, Lincoln, USA

State University of New York at Buffalo, Buffalo, USA
J. Dolen, A. Godshalk, I. Iashvili, A. Kharchilava, A. Kumar, S. Rappoccio

Northeastern University, Boston, USA

Northwestern University, Evanston, USA

University of Notre Dame, Notre Dame, USA

The Ohio State University, Columbus, USA

Springer
Princeton University, Princeton, USA

University of Puerto Rico, Mayaguez, USA
E. Brownson, H. Mendez, J. E. Ramirez Vargas

Purdue University, West Lafayette, USA

Purdue University Calumet, Hammond, USA
N. Parashar, J. Stupak

Rice University, Houston, USA
A. Adair, B. Akgun, K. M. Ecklund, F. J. M. Geurts, W. Li, B. Michlin, B. P. Padley, R. Redjimi, J. Roberts, J. Zabel

University of Rochester, Rochester, USA
B. Betchart, A. Bodek, R. Covarelli, P. de Barbaro, R. Demina, Y. Eshaq, T. Ferbel, A. Garcia-Bellido, P. Goldenzweig, J. Han, A. Harel, A. Khukhunaishvili, G. Petrillo, D. Vishnevskiy

The Rockefeller University, New York, USA
R. Ciesielski, L. Demortier, K. Goulianos, G. Lungu, C. Mesropian

Rutgers, The State University of New Jersey, Piscataway, USA

University of Tennessee, Knoxville, USA
K. Rose, S. Spanier, A. York

Texas A&M University, College Station, USA

Texas Tech University, Lubbock, USA
N. Akchurin, C. Cowden, J. Damgov, C. Dragoiu, P. R. Dudero, J. Faulkner, K. Kovitanggoon, S. Kunori, S. W. Lee, T. Libeiro, I. Volobouev

Vanderbilt University, Nashville, USA

University of Virginia, Charlottesville, USA

Wayne State University, Detroit, USA
C. Clarke, R. Harr, P. E. Karchin, C. Kottachchi Kankanamge Don, P. Lamichhane, J. Sturdy

University of Wisconsin, Madison, USA

† Deceased
1: Also at Vienna University of Technology, Vienna, Austria
2: Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
3: Also at Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, Université de Haute Alsace Mulhouse, CNRS/IN2P3, Strasbourg, France
4: Also at National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
5: Also at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
6: Also at Universidade Estadual de Campinas, Campinas, Brazil
7: Also at Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France
8: Also at Joint Institute for Nuclear Research, Dubna, Russia
9: Also at Suez University, Suez, Egypt
10: Also at Cairo University, Cairo, Egypt
11: Also at Fayoum University, El-Fayoum, Egypt
12: Also at British University in Egypt, Cairo, Egypt
13: Now at Ain Shams University, Cairo, Egypt
14: Also at Université de Haute Alsace, Mulhouse, France
15: Also at Brandenburg University of Technology, Cottbus, Germany
16: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
17: Also at Eötvös Loránd University, Budapest, Hungary
18: Also at University of Debrecen, Debrecen, Hungary
19: Also at University of Visva-Bharati, Santiniketan, India
20: Now at King Abdulaziz University, Jeddah, Saudi Arabia
21: Also at University of Ruhuna, Matara, Sri Lanka
22: Also at Isfahan University of Technology, Isfahan, Iran
23: Also at Sharif University of Technology, Tehran, Iran
24: Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran
25: Also at Laboratori Nazionali di Legnaro dell’INFN, Legnaro, Italy
26: Also at Università degli Studi di Siena, Siena, Italy
27: Also at Centre National de la Recherche Scientifique (CNRS)-IN2P3, Paris, France
28: Also at Purdue University, West Lafayette, USA
29: Also at Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mexico
30: Also at Institute for Nuclear Research, Moscow, Russia
31: Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia
32: Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia
33: Also at Facoltà Ingegneria, Università di Roma, Rome, Italy
34: Also at Scuola Normale e Sezione dell’INFN, Pisa, Italy
35: Also at University of Athens, Athens, Greece
36: Also at Paul Scherrer Institut, Villigen, Switzerland
37: Also at Institute for Theoretical and Experimental Physics, Moscow, Russia
38: Also at Albert Einstein Center for Fundamental Physics, Bern, Switzerland
39: Also at Gaziosmanpasa University, Tokat, Turkey
40: Also at Adiyaman University, Adiyaman, Turkey
41: Also at Cag University, Mersin, Turkey
42: Also at Mersin University, Mersin, Turkey
43: Also at Izmir Institute of Technology, Izmir, Turkey
44: Also at Ozyegin University, Istanbul, Turkey
45: Also at Marmara University, Istanbul, Turkey
46: Also at Kafkas University, Kars, Turkey
47: Also at Rutherford Appleton Laboratory, Didcot, UK
48: Also at School of Physics and Astronomy, University of Southampton, Southampton, UK
49: Also at University of Belgrade, Faculty of Physics and Vinca Institute of Nuclear Sciences, Belgrade, Serbia
50: Also at Mimar Sinan University, Istanbul, Istanbul, Turkey
51: Also at Argonne National Laboratory, Argonne, USA
52: Also at Erzincan University, Erzincan, Turkey
53: Also at Yıldız Technical University, Istanbul, Turkey
54: Also at Texas A&M University at Qatar, Doha, Qatar
55: Also at Kyungpook National University, Daegu, Korea