CMS Physics Analysis Summary

Contact: cms-pag-conveners-smp@cern.ch 2016/07/31

Measurement of associated Z + charm production in pp collisions at $\sqrt{s} = 8$ TeV

The CMS Collaboration

Abstract

A study of the associated production of a Z boson and one charm-quark jet ($Z + c$) in pp collisions at a center-of-mass energy of 8 TeV is presented. The analysis is conducted with a data sample corresponding to an integrated luminosity of 19.7 fb^{-1}, collected by the CMS detector at the CERN LHC. The Z-boson candidates are identified through their decay into a pair of electrons or muons. Jets originating from heavy flavour quarks are identified using semileptonic decays of c- or b-flavoured hadrons and hadronic decays of charm hadrons. The measurements are performed for heavy flavour jets in the kinematic region $p_T^{jet} > 25$ GeV, $| \eta^{jet} | < 2.5$. The $Z + c$ production cross section is measured to be $\sigma(pp \to Z + c + X) = 8.6 \pm 0.5$ (stat.) ± 0.7 (syst.) pb. The relative production of a Z boson and at least one c- or b-quark jet is analysed in terms of cross sections ratio. The ratio of the $Z + c$ and $Z + b$ production cross sections is measured to be $\sigma(pp \to Z + c + X)/\sigma(pp \to Z + b + X) = 2.0 \pm 0.2$ (stat.) ± 0.2 (syst.). The $Z + c$ production cross section and the cross sections ratio are also measured differentially as a function of transverse momentum of the Z boson and of the heavy flavour jet. Measurements are compared with several theoretical predictions.

1 Introduction

The first run of the CERN Large Hadron Collider (LHC) has delivered a large sample of pp collisions containing events with production of a vector boson (V) accompanied by one or more jets in the final state. These events allow to test the predictions from perturbative QCD (pQCD) on event topologies never tested before. A moderate fraction of them involves the production of a vector boson in association with jets originated from heavy-flavour (HF) quarks and are used to study specific predictions of the strong sector of the standard model (SM).

Vector boson plus jets events constitute major backgrounds to many ongoing searches for extensions of SM. A proper characterization of these processes and validation of their description in the currently available physics simulation tools is essential for a reliable estimation of their contribution to the search regions. Events with a Z boson accompanied by at least one jet with charm-quark content, is a background for Flavour Changing Neutral Currents (FCNC) processes of the kind $t \to Z + c$ [1, 2]. Third-generation supersymmetric quarks decaying via charm quarks have been searched for in final states with a charm-quark jet and a large energy-momentum imbalance [3, 4]; $Z + c$ events with the Z boson decaying to neutrinos, thus invisibly, being one of the dominant backgrounds. An accurate estimation of the contribution from this process can thus be obtained from the study of the same reaction with the Z boson decaying to detectable leptons.

The possibility to reach experimental evidence of the existence of a nonperturbative intrinsic charm quark component in the nucleon has received renewed interest [5]. Recent publications [6–9] identify the associated production of vector bosons and charm-quark jets (V+c) as a suitable process to address this physics topic and provide some predictions of the impact of the existence of intrinsic charm in the proton on V+c processes at LHC. The main effect of an intrinsic-charm component inside the proton would be an enhancement of $Z + c$ production, mainly at large values of the transverse momentum of the charm-quark jet. New Parton Distribution Functions (PDF) sets have been recently released [10] where the charm PDF is no longer assumed to be perturbatively generated through pair production from gluons and light quarks, but it is parameterized and determined along with the light quark and gluon PDFs.

Production of a Z boson and a charm-quark jet has been studied in high energy hadron collisions by the D0 [11] and CDF [12] experiments at the Tevatron $p\overline{p}$ collider. More recently, the LHCb collaboration has measured the associated production of a Z boson and a D meson in the forward region in pp collisions at $\sqrt{s} = 7$ TeV [13].

We present in this paper a measurement of the production cross section of a Z boson and at least one jet originating from a c-quark ($Z + c$). In addition, the relative production of a Z boson and a jet originating from heavy quarks of different flavours (c or b) is quantified by the ratio of their production cross sections. The cross section $\sigma(pp \to Z + c + X)$ and the cross sections ratio $\sigma(pp \to Z + c + X)/\sigma(pp \to Z + b + X)$ are determined both inclusively and differentially as a function of the transverse momentum of the Z boson and the transverse momentum of the jet with heavy flavour content. The Z boson is identified through its decay into a pair of electrons or muons. Jets with HF-quark content are identified through the semileptonic decay of the c- or b-flavoured hadrons with a muon in the final state and using exclusive hadronic decays of charm hadrons. Cross section and cross sections ratio are measured at the level of stable particles. In order to minimize acceptance corrections, the measurements are restricted to a phase space close to the experimental fiducial volume with optimized sensitivity for these processes: two leptons with $p_T^{\ell} > 20$ GeV, pseudorapidity $|\eta^{\ell}| < 2.1$, and with the dilepton invariant mass consistent with the mass of the Z boson, $71 < m_{\ell\ell} < 111$ GeV, and a c- (b-) jet with $p_T^{jet} > 25$ GeV, pseudorapidity $|\eta^{jet}| < 2.5$ and separated from the leptons from the

Z-boson candidate by a distance $\Delta R(\text{jet}, \ell) = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.5$.

2 Data and simulated samples

The data analyzed for this paper are collected by the CMS experiment during the year 2012 at the pp centre-of-mass energy of 8 TeV and correspond to an integrated luminosity of $\mathcal{L} = 19.7 \pm 0.5\,\text{fb}^{-1}$.

Muon and electron candidates are reconstructed following standard CMS algorithms [14, 15]. Jets, missing transverse energy and related quantities are determined using the CMS particle-flow reconstruction algorithm [16, 17] that identifies and reconstructs all stable particles arising from a collision with an optimized combination of the signals measured from all sub-detectors. Jets are built from the particle-flow candidates using an anti-k_t clustering algorithm [18] with a size parameter of $R = 0.5$. The energy and momentum of the jets are corrected as a function of the jet p_T and η to account for the nonlinear response of the calorimeters [19, 20] and for the presence of additional pp interactions in the same bunch crossing (pileup). Jet energy corrections are derived using samples of simulated events and of real data from photon+jet and Z+jet processes.

Other physics processes produce events with the same final state topology than the signal of interest. The main background is the production of $\text{t}\bar{\text{t}}$ events. Smaller contributions are expected from the direct production of dibosons, WW, WZ, and ZZ.

Samples of simulated events are produced with Monte Carlo (MC) event generators, both for the signal process and for the main backgrounds. A sample of signal Z-boson events is generated with MADGRAPH v5.1.3.30 [21] event generator, interfaced with PYTHIA v6.4.26 [22] for parton showering and hadronization. The MADGRAPH generator produces parton-level events with a vector boson and up to four partons on the basis of a matrix-element calculation. The generation uses the PDF set CTEQ6L [23]. The MADGRAPH generator, interfaced with PYTHIA6 is also used to generate a sample of W events. For these two samples the matching scale is $m^2 = (10\ \text{GeV})^2$ and the factorization and normalization scales are set to $q^2 = M^2_{\text{Z/W}} + p^2_{\text{T,Z/W}}$.

A sample of $\text{t}\bar{\text{t}}$ events is generated with POWHEG v1.0 [24–27], interfaced with PYTHIA6 and using the CT10 [28] PDF set. The WW, WZ, ZZ processes are modelled with samples of events generated with PYTHIA6 and the PDF set CTEQ6L1 [23]. The PYTHIA6 parameters for the underlying event modelling in all the productions are set to the Z2* tune. The Z2* tune is derived from the Z1 tune [29], which uses the CTEQ5L PDF set, whereas Z2* adopts CTEQ6L. The Z2* tune is the result of retuning the PYTHIA parameters PARP(82) and PARP(90) by means of the automated PROFESSOR tool [30], yielding PARP(82)=1.921 and PARP(90)=0.227.

Generated events are processed through the full GEANT4-based [31] CMS detector simulation and trigger emulation. Simulated events are then reconstructed using the same algorithms that are used to reconstruct collision data. For electroweak processes the cross sections are normalized to the next-to-next-to-leading order (NNLO) calculation computed with FEWZ3.1 [32], using the PDF set MSTW2008NNLO [33]. The cross sections for dibosons production are evaluated at next-to-leading order (NLO) with MCFM [34] and the PDF set MSTW2008NLO [33]. The $\text{t}\bar{\text{t}}$ cross section is taken at NNLO from Ref. [35]. Predictions derived from the simulated samples are normalized to the integrated luminosity of the data sample. The simulated samples incorporate additional interactions per bunch crossing. Simulated events are weighted so that the pileup distribution matches the measured one, with an average of about 21 pileup interactions per bunch crossing.

Simulated samples are corrected for differences between data and MC description in lepton trigger, reconstruction and identification efficiencies (ϵ_ℓ). Lepton efficiencies are evaluated with clean samples of dilepton events in the Z mass peak with the "tag-and-probe" method [36], and a correction factor $\epsilon_\ell^{\text{data}}/\epsilon_\ell^{\text{MC}}$, binned in terms of transverse momentum and pseudorapidity of the leptons is computed. Correction factors are applied to the simulation as weigths on an event-by-event basis.

The simulated signal sample includes Z-boson events accompanied by jets originating from quarks of all flavours (b, c and light). Events are classified as $\text{Z}+\text{b}$, $\text{Z}+\text{c}$ or $\text{Z}+\text{light}$ according to the flavour of the generator level jets built from all showered particles after fragmentation and hadronisation, so stable particles, except neutrinos, and clustered with the same algorithm that is used to reconstruct the jets in the data. A generator level jet is defined to be b-flavoured if there is a b-meson ($500 < |\text{PDG}_{\text{id}}| < 600$) or b-baryon ($5000 < |\text{PDG}_{\text{id}}| < 6000$) among the particles generated in the event, within a cone of radius $\Delta R = 0.5$ of the jet axis. Similarly, a generator level jet is considered to be c-flavoured if there is a c-meson ($400 < |\text{PDG}_{\text{id}}| < 500$) or c-baryon ($4000 < |\text{PDG}_{\text{id}}| < 5000$) and no b-hadrons within a cone of $\Delta R = 0.5$ of the jet axis. A $\text{Z}+\text{jets}$ event is assigned as a $\text{Z}+\text{b}$ event if there is a generator level jet with $p_{\text{T}} > 15$ GeV identified as a b-flavoured jet, $\text{Z}+\text{c}$ if there is a c-flavoured generator level jet with $p_{\text{T}} > 15$ GeV and no b-flavoured generator level jets, and $\text{Z}+\text{light}$ otherwise.

3 Event selection

3.1 Selection of the Z+HF-jet events

Events with a pair of leptons are selected online by a trigger system that requires the presence of two lepton candidates of the same flavour with transverse energy or momentum exceeding 17 GeV and 8 GeV for the two lepton candidates. We follow closely the criteria used in the CMS $\text{Z} \to \text{e}^+\text{e}^-$ and $\text{Z} \to \mu^+\mu^-$ inclusive analyses [36] and require the presence of two opposite sign, high-p_{T} reconstructed leptons, the transverse momentum of the leptons, p_{T}^ℓ ($\ell = \text{e}, \mu$), has to be greater than 20 GeV. We restrict to the lepton pseudorapidity region $|\eta^\ell| < 2.1$ coherently with the sensitive region for the $\text{W}+\text{c}$ event selection described in Section 3.2.

An isolation variable, I_{comb}, is built to test the presence of additional activity around the selected leptons. It is defined as the sum of the transverse energy of neutral hadrons and photons and the momentum of charged particles in a cone of radius 0.3 (0.4) around the electron (muon) candidate, excluding the contribution from the lepton itself. The isolation variable is corrected for the contribution from particles in the cone of interest not originating from the primary vertex, but from other pileup vertices. The electron (muon) candidate is considered to be isolated if the isolation variable normalized to the lepton momentum is small, $I_{\text{comb}}/p_{\text{T}}^{\text{e}} < 0.15$ ($I_{\text{comb}}/p_{\text{T}}^{\mu} < 0.20$). Finally, we restrict to events with a dilepton invariant mass in the range $71 < m_{\ell\ell} < 111$ GeV.

A $\text{Z}+\text{jets}$ sample is selected by demanding the presence of at least one jet with $E_{\text{T}}^{\text{jet}} > 25$ GeV in the pseudorapidity range $|\eta^{\text{jet}}| < 2.5$. Jets with an angular separation of the jet axis with any of the selected leptons smaller than $\Delta R(\text{jet}, \ell) = 0.5$ are not considered. In order to reduce the contribution from $\text{t}\bar{\text{t}}$ events, we require the magnitude of the momentum imbalance in the transverse plane (missing transverse energy) to be smaller than 40 GeV.

Hadrons with c- or b-quark content decay weakly with lifetimes of the order of 10^{-12} s and decay lengths at the LHC energies > 100 μm. A secondary vertex well separated from their production vertex can be reconstructed from the tracks of their charged decay products. We

focus on the three following signatures to identify jets originating from a heavy flavour quark:

- A semileptonic decay of a heavy flavour hadron leading to a well identified muon participating in a displaced secondary vertex (we refer to events selected in this channel as *semileptonic mode*).
- A displaced secondary vertex with three tracks consistent with a $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ decay ($D^{\pm}$ *mode*).
- A displaced secondary vertex with two tracks consistent with a $D^0 \to K^-\pi^+$ ($\bar{D}^0 \to K^+\pi^-$) decay and associated to a previous $D^{*+}(2010) \to D^0\pi^+$ ($D^{*-}(2010) \to \bar{D}^0\pi^-$) decay at the primary vertex ($D^{*\pm}(2010)$ *mode*).

Displaced secondary vertices for the first two categories are formed with either the SSV [37] or the IVF [38, 39] CMS vertex algorithms; a different technique is employed for the third one.

Candidates to come from a $Z + c$ ($Z + b$) event with a semileptonic decay of the c (b) quark are selected by looking for a reconstructed muon (hereafter called *muon-inside-a-jet*) among the particles constituting a jet. This *muon-inside-a-jet* candidate has to be reconstructed in the pseudorapidity region $|\eta| < 2.5$ fulfilling the same tight quality criteria imposed to the muons from the Z decay. Its transverse momentum should be moderate $p_T^{\mu} < 25$ GeV, $p_T^{\mu}/p_T^{\text{jet}} < 0.6$, and it should not be isolated from hadron activity, the combined isolation variable has to be large, $I_{\text{comb}}/p_T^{\mu} > 0.2$. Furthermore the *muon-inside-a-jet* is required to participate in a secondary vertex, reconstructed either with the SSV or the IVF algorithms. We have selected 4145 events in the $Z \to e^+e^-$ channel and 5258 events in the $Z \to \mu^+\mu^-$ channel.

The selected sample is primarily composed of $Z + b$ events ($\approx 65\%$) with a significant fraction of $Z + c$ events ($\approx 25\%$) according to the expectations from the simulated samples. The contribution from $Z +$ light is very much suppressed ($\lesssim 5\%$), and the contribution from the other processes, $t\bar{t}$ and dibosons production is $\approx 5\%$.

Event candidates in the $D^{\pm}$ mode are selected by looking for secondary vertices made of three tracks and with a reconstructed invariant mass consistent with the $D^{\pm}$ mass: 1869.5 ± 0.4 MeV [40]. The sum of the charges of the tracks participating in the secondary vertex has to be ± 1. The kaon mass is assigned to the track with opposite sign to the total charge of the three-prong vertex, and the remaining tracks are assumed to have the mass of a charged pion. This assignment is correct in more than 99% of the cases, since the fraction of double Cabibbo-suppressed decays is extremely small: $\mathcal{B}(D^+ \to K^+\pi^+\pi^-)/\mathcal{B}(D^+ \to K^-\pi^+\pi^+) = 0.00577 \pm 0.00022$ [40].

The signal region is defined by the constraint $\Delta m(D^{\pm}) \equiv |m^{\text{rec}}(D^{\pm}) - 1.87 \text{ GeV}| < 0.05$ GeV, where $m^{\text{rec}}(D^{\pm})$ is the reconstructed mass of the $D^{\pm}$ candidate. The nonresonant background is subtracted from the events in the signal window by using the number of events selected in a control region away from the resonance, extending up to a window of 0.1 GeV width, $N[0.05 \text{ GeV} < \Delta m(D^{\pm}) < 0.10 \text{ GeV}]$. The number of selected events after background subtraction is 375 ± 44 in the $Z \to e^+e^-$ channel and 490 ± 48 in the $Z \to \mu^+\mu^-$ channel.

The charm fraction $\mathcal{B}(c \to D^{\pm})$ in the PYTHIA simulation ($19.44 \pm 0.02\%$) is lower than the value ($22.7 \pm 0.9 \pm 0.5\%$) obtained from a combination [41] of published measurements performed at LEP [42–44] and the branching fraction of the decay $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ ($7.96 \pm 0.03\%$), is also lower than the PDG value ($9.13 \pm 0.19\%$) [40]. Predicted event rates from the MC simulation are reweighted in order to match the experimental values. The final sample is enriched in $Z + c$ events ($\approx 60\%$) while the fraction of $Z + b$ events is $\approx 35\%$. The contribution from $Z +$ light events is fully suppressed, and the fraction of remaining $t\bar{t}$ and dibosons events is smaller than

5%.

Events with Z + jets candidates in the $D^{*\pm}(2010)$ mode are selected by requiring a vertex with two oppositely charged tracks among the tracks constituting the jet and assumed to be the decay products of a D^0. This two-track system is combined with a third track of the jet constituents assumed to be the *soft pion*, π_s, emitted in the strong decay $D^{*+}(2010) \to D^0\pi^+$, to build the $D^*(2010)^\pm$ candidate. Tracks candidates to be the *soft pion* should have a transverse momentum larger than 0.5 GeV and lay in a cone of radius $\Delta R(D^0, \pi_s) = 0.1$ around the line of flight of the D^0 candidate.

The track of the D^0 candidate with charge opposite to the charge of the *soft pion* is taken to be the kaon from the D^0 decay and is requested to have $p_T > 1.75$ GeV. The other track is assigned to be the pion and is required to have $p_T > 0.75$ GeV. Two-track combinations with an invariant mass different from the nominal D^0 mass (1864.86 ± 0.13 MeV) by less than 100 MeV are kept and a secondary vertex is built with them following the CMS KalmanVertexFitter algorithm [45]. The two-track system is kept as a valid D^0 candidate if the vertex probability is greater than 0.05.

To ensure a clean separation between the secondary and the primary vertices, the decay length significance in the transverse plane ($L_{xy}/\sigma(L_{xy})$) has to be larger than three. Furthermore, to guarantee that the reconstructed vertex corresponds to a two body decay of a hadron originating at the primary vertex, the momentum vector of the D^0 candidate has to be collinear with the line from the primary to the secondary vertex ($\cos\alpha_{xy} > 0.99$). Finally, events with a mass difference between the $D^{*\pm}(2010)$ and D^0 candidates within 5 MeV around the expected value (145.426 ± 0.002 MeV [40]) are selected.

The product of the branching fractions $\mathcal{B}(c \to D^{*+}(2010)) \times \mathcal{B}(D^{*+}(2010) \to D^0\pi^+) \times \mathcal{B}(D^0 \to K^-\pi^+)$ (+c.c.) in the PYTHIA simulation is $(0.741 \pm 0.005)\%$, which is about 15% larger than the estimation of the experimental value, $(0.622 \pm 0.020)\%$ [40, 41]. Expected event rates from the MC simulation are reweighted in order to match the experimental values.

The signal region is defined by the constraint $\Delta m(D^{*\pm}(2010)) \equiv |m^{rec}(D^{*\pm}(2010)) - 2.01 \text{ GeV}| < 0.04$ GeV, where $m^{rec}(D^{*\pm}(2010))$ is the reconstructed mass of the $D^{*\pm}(2010)$ candidate. The non resonant background contribution to the signal region is subtracted using the number of events selected in a control region away from the resonance, extending up to a window of $0.12(2 \times 0.06)$ GeV width, $N[0.04 \text{ GeV} < \Delta m(D^{*\pm}(2010)) < 0.10 \text{ GeV}]$ with the proper weight to account for the different width of the signal and control regions $(8/12)$. The number of selected events after background subtraction is 234 ± 22 in the $Z \to e^+e^-$ channel and 308 ± 24 in the $Z \to \mu^+\mu^-$ channel.

According to the predictions obtained with the simulated samples the fraction of Z + c events is high ($\approx$ 65%) and the contribution of Z + b events is $\approx$ 30%. Background from Z + light events is fully suppressed, while the fraction of remaining $t\bar{t}$ and dibosons events is smaller than 5%.

Systematic biases due to the assumed background subtraction are expected to be negligible compared to the statistical uncertainty, given the approximate agreement observed between data and MC.

3.2 Selection of the W+charm-jet events

An auxiliary sample consists of events originating from the associated production of a W boson and a jet originating from a c-quark (W + c). It is used to model characteristic distributions of

jets with c-quark content and to measure the c-tagging efficiency.

The production of a W boson in association with a c-quark proceeds at LO via the processes $sg \to W^- + c$ and $\bar{s}g \to W^+ + \bar{c}$. A key property of the $qg \to W + c$ reaction is the presence of a charm quark and a W boson with opposite-sign (OS) charges. Background processes deliver evenly OS and same-sign (SS) events, whereas $qg \to W + c$ is always OS. Therefore, distributions obtained after OS $-$ SS subtraction are largely dominated by the $W + c$ component, allowing for detailed studies of c-quark jets.

We select a sample of $W + c$ events following the criteria used in the analysis of W plus c-jet process reported in Ref. [46]. Candidate events are selected online using single-lepton triggers, which require at least one isolated electron (muon) with $p_T > 27\ (24)$ GeV in the pseudorapidity region $|\eta^\ell| < 2.1$. The lepton identification and isolation criteria are very similar to the ones used for $Z + c$ selection. The offline p_T threshold is increased to 30 (25) GeV for electrons (muons) due to the higher thresholds of the single-lepton triggers. The reconstructed transverse mass built from the missing transverse energy E_T^{miss}, and the lepton transverse energy and direction, $M_T = \sqrt{2\, p_T^\ell\, E_T^{miss}\, (1 - \cos(\phi^\ell - \phi^{E_T^{miss}}))}$ must be larger than 55 (50) GeV for events in the $W \to e\nu$ ($W \to \mu\nu$) channel.

Identification of jets originating from c-quarks proceeds exactly as it is described in Section 3.1. In all cases the charge of the c-quark is unequivocally known. In the semileptonic mode the charge of the muon determines the charge of the c-quark. In the other decay modes the charge of the $D^\pm$ and $D^{*\pm}(2010)$ candidates define the charge of the c-quark. Opposite-sign events can thus be defined when the muon ($D^\pm$ or $D^{*\pm}(2010)$ candidate) has opposite charge to the charge of the lepton from the W decay, and same-sign otherwise.

According to the expectations from the simulated samples, $W + c$ events are almost the only contributors to the distributions after OS $-$ SS subtraction, $\sim 90\%$ in the semileptonic decay modes and larger than 98% in the exclusive channels. Remaining backgrounds are subtracted according to their MC expectation.

3.3 Other data samples

A sample of $t\bar{t}$ events is selected using the leptonic decay modes of the W bosons from the $t\bar{t}$ pair, when they decay into leptons of different flavour. The $t\bar{t}$ production is a natural source of b-flavoured jets and allow to test the MC description of the relevant distributions for b-jets as well as the performance of the b-tagging method. This sample is also used to model the $t\bar{t}$ contribution in the variables of interest.

An $e\mu$-$t\bar{t}$ sample is selected online by a trigger path based on the presence of a electron-muon pair. The offline selection proceeds as for the Z+HF-jet events, except for the requirement that the two leptons are of the same flavour. After the full selection, contributions from processes other than $t\bar{t}$ production are negligible.

An additional $t\bar{t}$ enriched sample is used to estimate the normalization of the remaining $t\bar{t}$ background. Exactly the same selection as for the Z+HF-jet signal, i.e. with two leptons of the same flavour, ee or $\mu\mu$, is applied except for the $E_T^{miss} < 40$ GeV requirement. Instead, $E_T^{miss} > 80$ GeV is required. The small contribution from $Z +$ jets events in these samples is subtracted according to its MC expectation.

4 Measurement of the c- and b-quark tagging efficiencies

The accuracy of the c-tagging performance in the simulated samples is evaluated with a control sample of $W + c$ events with a well identified *muon-inside-a-jet*. The events are selected as described in Section 3.2 except for the requirement that the *muon-inside-a-jet* should come from a secondary vertex. The $W + c$ sample in the $W \to e\nu$ decay mode is employed to minimize background contribution from dimuon Drell-Yan events.

A $W + c$ event is defined to be "tagged" if there is a reconstructed secondary vertex in the jet and the *muon-inside-a-jet* is one of the tracks constituting the vertex. The c-jet tagging efficiency is defined as the fraction of "tagged" $W + c$ events, over all $W + c$ events, after OS $-$ SS subtraction:

$$\epsilon_c = \frac{N(W + c)^{OS-SS}(\text{SV-tagged})}{N(W + c)^{OS-SS}}$$

Efficiencies are obtained independently with the data and with the W+jets simulated samples and data-to-simulation scale factors, SF_c, are then computed as the ratio between the c-jet tagging efficiencies in data and simulation,

$$SF_c = \frac{\epsilon_c^{\text{data}}}{\epsilon_c^{\text{MC}}}$$

They are used to correct the simulated events as to better describe the efficiency of the CMS detector.

The c-jet tagging efficiencies and the scale factor SF_c, are computed both inclusively and as a function of the jet p_T. The scaling factor SF_c for jets with a p_T larger than 25 GeV is found to be $0.88 \pm 0.03\,(\text{stat}) \pm 0.02\,(\text{syst})$ ($0.92 \pm 0.03\,(\text{stat}) \pm 0.02\,(\text{syst})$) for SSV secondary vertices (IVF vertices). The systematic component accounts for uncertainties in pileup description, jet energy scale and resolution, lepton efficiencies, background subtraction and modelling of charm production and decay fractions in the simulation.

Detailed studies of the behaviour of the b-tagging methods developed in CMS and scaling factors SF_b that correct for differences between data and MC simulation performance are available in Ref. [47]. The level of data-to-MC agreement is tested with the $e\mu$-$t\bar{t}$ sample and b-tagging efficiencies in data and simulated events are computed as the fraction of $e\mu$-$t\bar{t}$ events with a *muon-inside-a-jet* participating in a secondary vertex with respect to the number of events when the secondary vertex condition is released. The scaling factor $SF_b = \epsilon_b^{\text{data}} / \epsilon_b^{\text{MC}}$ is measured to be 0.96 ± 0.03 for both SSV and IVF secondary vertices, where the uncertainty includes statistical and systematic effects due to jet energy scale and resolution and pileup description.

5 Analysis strategy

The extraction of $Z + c$ and $Z + b$ event yields is based on fits to distributions that have a different shape, depending on the jet flavour. In the semileptonic channel we examine the invariant mass distribution of the charged particles constituting the secondary vertex, the *muon-inside-a-jet* being one of them. All charged particles are given the mass of the pion, except for the identified muon. A correction is included to account for additional particles, either charged or neutral ones, that may have been produced in the semileptonic decay but were not reconstructed [48],

$$M_{\text{vertex}}^{\text{corr}} = \sqrt{M_{\text{vertex}}^2 + p_{\text{vertex}}^2 \sin^2\theta} + p_{\text{vertex}} \sin\theta,$$

where $M_{\text{vertex}}^{\text{corr}}$ denotes the corrected secondary vertex mass, M_{vertex} and p_{vertex} are the invariant mass and modulus of the momentum-sum of all reconstructed particles that form the secondary vertex and θ is the angle between the momentum vector sum and the direction of flight of the secondary vertex.

In the $D^{\pm}$ and $D^{*\pm}(2010)$ exclusive decay channels a likelihood estimate of the probability that the jet tracks come from the primary vertex, Jet Probability (JP) discriminant [37], is used.

The shapes of the $Z + c$ discriminant distributions are modelled using OS $W + c$ events, after subtraction of the SS $W + c$ distributions. This procedure is validated with simulated samples. Main features of the jets, such as p_T, η, jet charged multiplicity and the number of secondary vertices are found to be consistent between $Z + c$ and $W + c$ simulated samples and are in agreement with the observed distributions in the experimental sample of $W + c$ events. Figure 1 (left) shows the simulated $W + c$ and $Z + c$ p_T^{jet} distributions compared to $W + c$ data. The number of secondary vertices, identified with the IVF algorithm, is shown in Fig. 1 (right).

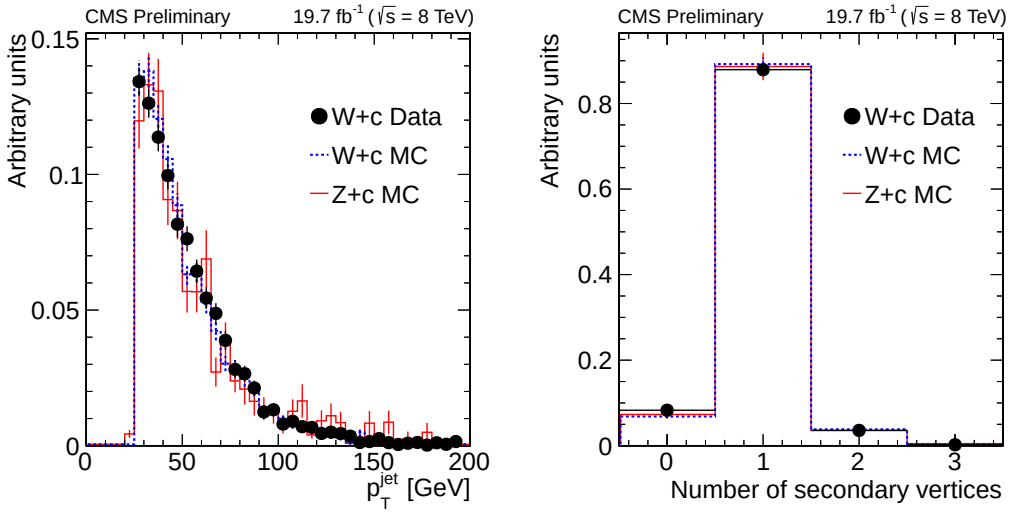

Figure 1: Transverse momentum distribution of the c-tagged jet (left) and number of reconstructed secondary vertices (right) in simulated $W + c$ and $Z + c$ samples, and in $W + c$ data events. Events with no reconstructed IVF vertices have at least one reconstructed vertex with the SSV vertex algorithm. The $W + c$ distributions are presented after OS $-$ SS subtraction.

The corrected secondary vertex mass and JP discriminant distributions, normalized to unity are presented in Fig. 2 for the three analysis categories. The simulated $W + c$ and $Z + c$ distributions are compared to data. The simulated $W + c$ and $Z + c$ distributions agree in all categories. Some discrepancy is observed between the simulated and experimental distributions of the corrected secondary vertex mass in $W + c$ events. It has been checked that the expected fraction of events with two-track and three-track vertices from $W + c$ and $Z + c$ agree in the simulation, but it is different in the $W + c$ data sample. Therefore we expect that the $W + c$ data distribution still properly describes the $Z + c$ shape in data even if the simulation does not completely describe the data. The distributions obtained in the electron and muon decay channels are consistent and are averaged to obtain the final templates, thus decreasing the associated statistical uncertainty.

The shape of the discriminant variables for $Z + b$ events is modelled with the simulated samples. The simulated distribution of the corrected secondary vertex mass is validated with the

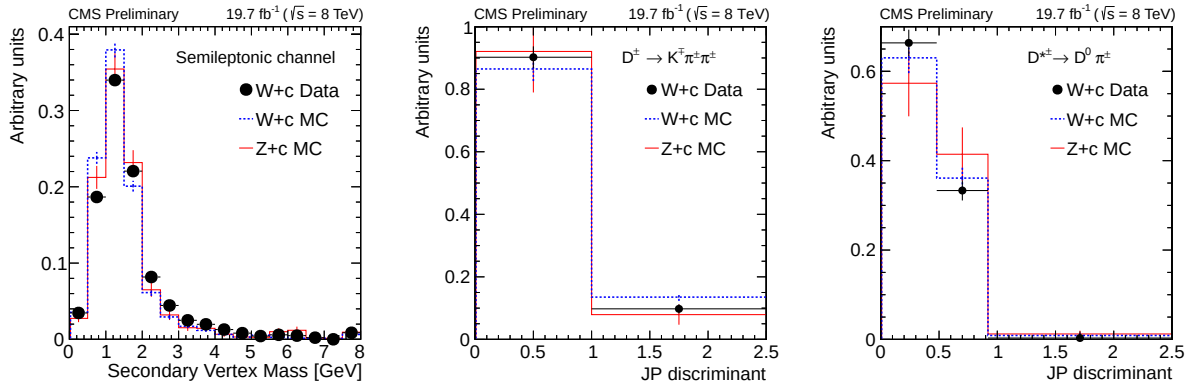

Figure 2: Distributions of the corrected secondary vertex mass (left plot) and JP discriminant ($D^{\pm}$ mode in the middle plot and $D^{*\pm}(2010)$ mode in the right plot), normalized to unity, in simulated $W + c$ and $Z + c$, and in $W + c$ data events. The $W + c$ distributions are presented after OS $-$ SS subtraction.

sample of eμ-$t\bar{t}$ events as shown in Fig. 3. The simulation describes the data well, apart from the mass region between 3–4 GeV, and above 7.5 GeV. The observed differences, $\sim$ 13% in the 3–4 GeV mass region and $\sim$ 50% above 7.5 GeV are used to correct the simulated $Z + b$ distribution. The number of selected events in eμ-$t\bar{t}$ sample does not allow for a meaningful validation of the shape of JP discriminant distributions for $Z + b$ events in the exclusive channels.

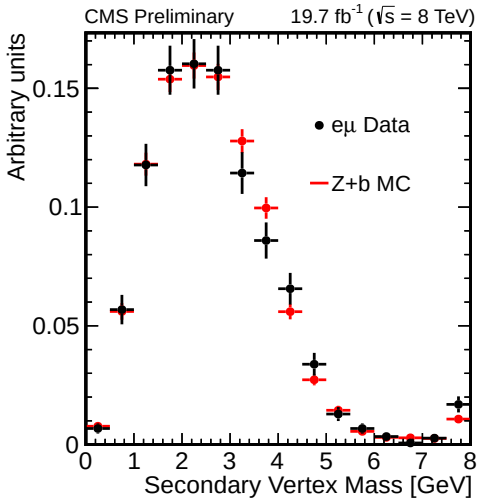

Figure 3: Distributions of the corrected secondary vertex mass normalized to unity from simulated $Z + b$ and data eμ-$t\bar{t}$ events.

The distributions obtained in data are corrected for background sources:

- The contribution from $t\bar{t}$ production is evaluated with the eμ-$t\bar{t}$ sample. The normalization to the signal region, where the two leptons are of the same flavour, $N_{ee}^{t\bar{t}}/N_{e\mu}^{t\bar{t}}$ ($N_{\mu\mu}^{t\bar{t}}/N_{e\mu}^{t\bar{t}}$) is determined in the region with large E_T^{miss}, $E_T^{miss} > 80$ GeV. This normalization factor is propagated to the signal region with $E_T^{miss} < 40$ GeV.
- The $Z +$ light-quark background in the semileptonic channel is evaluated with the

simulated samples. Discrepancies in the mistagging rate between data and simulation are corrected in the MC through the appropriate scaling factors. No background from Z + light-quark process is expected in the exclusive channels.

- Remaining background from dibosons production is taken from the simulation.

Figure 4 shows the distributions of the corrected secondary vertex mass for the Z + jets events with a *muon-inside-a-jet* associated with a secondary vertex. The corrected vertex mass tends to be larger for Z + b events than for Z + c events because of the larger mass of the b-quark giving rise to heavier hadrons ($m_{\text{b-hadrons}} \sim 5\,\text{GeV}$, $m_{\text{c-hadrons}} \sim 2\,\text{GeV}$).

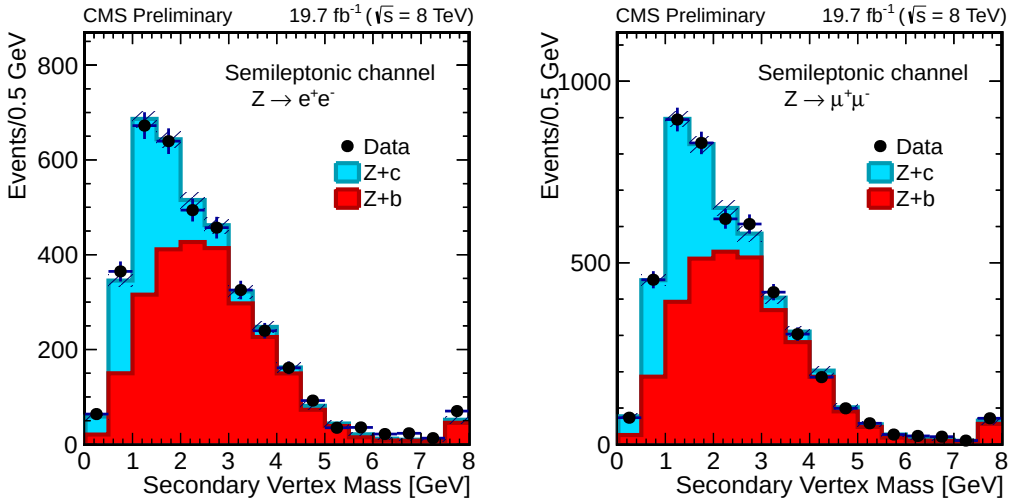

Figure 4: Corrected secondary vertex mass distributions in the dielectron (left) and dimuon (right) channels. The shape of the Z + c and Z + b contributions is estimated as explained in the text. Their normalization is adjusted to the result of the signal extraction fit.

The JP discriminant takes lower values for Z + c events than for Z + b events. The $D^{\pm}$ or $D^{*\pm}(2010)$ mesons in Z + b events are mostly "secondary" particles, i.e. they do not originate from the hadronization of a c-quark produced at the primary vertex but are decay products of previous B-hadron decays, that happen in unobserved secondary vertices. Figure 5 shows the distribution of the JP discriminant for the Z + jets events with a $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ candidate and Fig. 6 presents the distribution of the JP discriminant for the Z + jets events with a $D^{*\pm}(2010)$ candidate. In this latter channel the particle identified as the *soft pion* in the $D^{*\pm}(2010) \to D^0\pi_s^{\pm}$ decay is a true primary particle in the case of Z + c events whereas it arises in a secondary decay (B-hadron $\to D^{*\pm}(2010) + X \to D^0\pi_s^{\pm} + X$) for Z + b events.

We measure the fraction of Z + c and Z + b events in the selected sample by performing a χ^2-test fit to the observed data distributions. The parameters to fit are scaling factors SF_{Z+c} and SF_{Z+b} with respect to the initial normalization predicted by the simulation. They are used to measure the event yields N^{signal}, that are summarized in Table 1. The typical scaling factors are found to be in the range 0.95–1.05 and are applied in the data-to-predictions comparisons of Figs. 4, 5 and 6.

The sensitivity to the Z + b component in the $D^{\pm}$ and $D^{*\pm}(2010)$ exclusive channels is very reduced and the fitted SF_{Z+b} scaling factors in these channels are used for an effective subtraction of the remaining Z + b contribution only and no Z + b cross section is measured.

Semileptonic mode

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ (pb)
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	1066 ± 95	0.63 ± 0.03	$8.5 \pm 0.7 \pm 1.0$
$\mathrm{Z} \to \mu^+\mu^-$	1449 ± 143	0.81 ± 0.03	$9.0 \pm 0.7 \pm 1.0$
Channel	$N_{\mathrm{Z+b}}^{\mathrm{signal}}$	$\mathcal{C}_{\mathrm{Z+b}}$ (%)	$\sigma(\mathrm{Z+c})/\sigma(\mathrm{Z+b})$
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	2606 ± 114	2.90 ± 0.08	$1.9 \pm 0.2 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	3241 ± 147	3.93 ± 0.10	$2.2 \pm 0.3 \pm 0.2$

$\mathrm{D}^{\pm}$ mode

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ (pb)
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	275 ± 55	0.13 ± 0.02	$11.0 \pm 2.1 \pm 0.9$
$\mathrm{Z} \to \mu^+\mu^-$	315 ± 75	0.18 ± 0.02	$9.0 \pm 2.1 \pm 0.8$

$\mathrm{D}^{*\pm}(2010)$ mode

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ (pb)
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	151 ± 31	0.11 ± 0.01	$7.3 \pm 1.5 \pm 0.6$
$\mathrm{Z} \to \mu^+\mu^-$	228 ± 30	0.14 ± 0.01	$8.6 \pm 1.1 \pm 0.6$

Table 1: Cross section $\sigma(\mathrm{Z+c})$, and cross sections ratio $\sigma(\mathrm{Z+c})/\sigma(\mathrm{Z+b})$ in the three categories of this analysis and in the two Z-boson decay channels. $N_{\mathrm{Z+c}}^{\mathrm{signal}}$ and $N_{\mathrm{Z+b}}^{\mathrm{signal}}$ are the yields of $\mathrm{Z+c}$ and $\mathrm{Z+b}$ events extracted from the fit to the corrected secondary vertex mass (semileptonic mode) or JP discriminant ($\mathrm{D}^{\pm}$ and $\mathrm{D}^{*\pm}(2010)$ modes) distributions. The factors $\mathcal{C}$ that correct the selection inefficiencies are also given. They include the relevant branching fraction for the corresponding channel. All uncertainties quoted in the table are statistical except for the measured cross sections and cross sections ratio where the first uncertainty is statistical and the second one is the estimated systematic uncertainty from the sources discussed in the text.

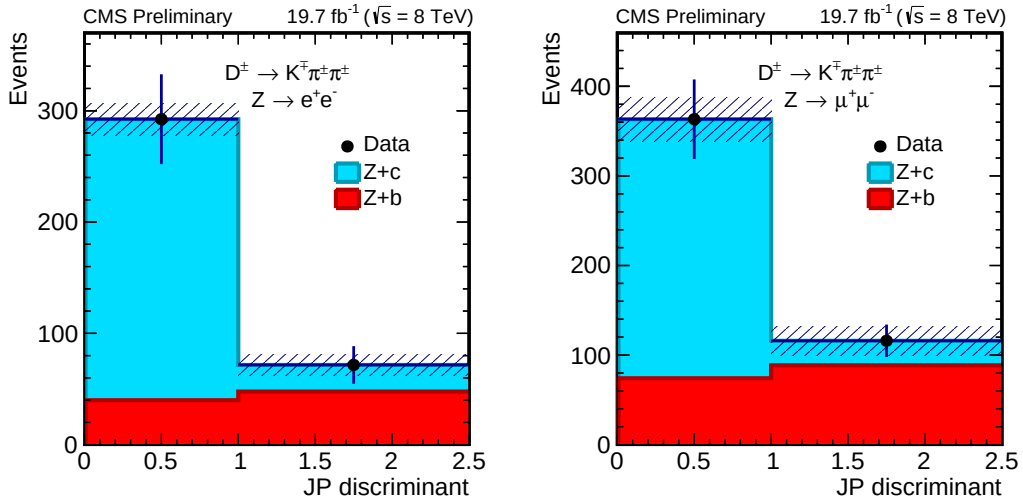

Figure 5: Distributions of the JP discriminant in the dielectron (left) and dimuon (right) channels for Z + jets events with a $D^{\pm} \to D^0\pi_s^{\pm} \to K^{\mp}\pi^{\pm}\pi_s^{\pm}$ candidate. The shape of the Z + c and Z + b contributions is estimated as explained in the text. Their normalization is adjusted to the result of the signal extraction fit.

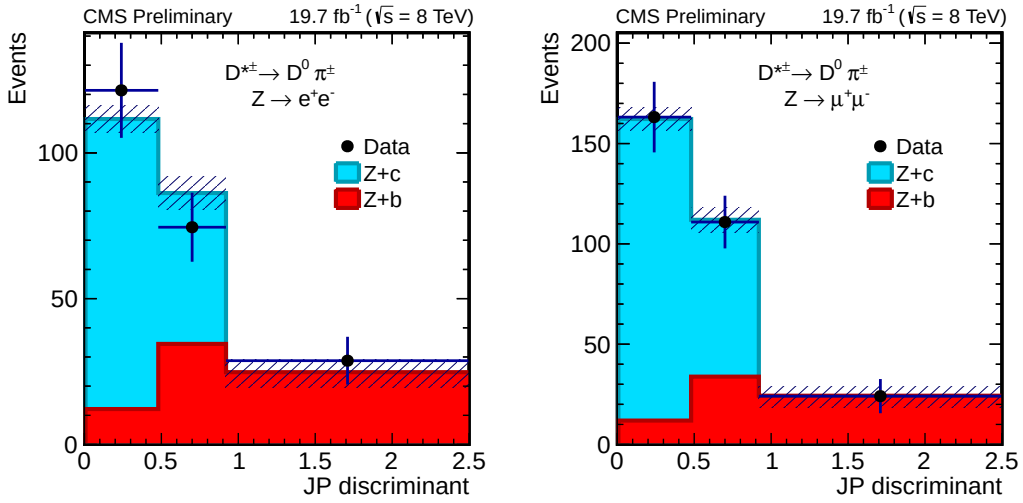

Figure 6: Distributions of the JP discriminant in the dielectron (left) and dimuon (right) channels for Z + jets events with a $D^{*\pm}(2010) \to D^0\pi_s^{\pm} \to K^{\mp}\pi^{\pm}\pi_s^{\pm}$ candidate. The shape of the Z + c and Z + b contributions is estimated as explained in the text. Their normalization is adjusted to the result of the signal extraction fit.

6 Systematic uncertainties

The major sources of systematic uncertainties are shown in Fig. 7. The contributions from the several sources are combined into fewer categories for presentation in the figure.

The main uncertainty is related with the charm fractions for the production and decay of c-hadrons in the simulated samples, and with the determination of the c-tagging efficiency. The average of the inclusive charm semileptonic branching fractions is $\mathcal{B}(c \to \ell) = 0.096 \pm 0.004$ [40], and of the exclusive sum of the individual contributions from all weakly decaying

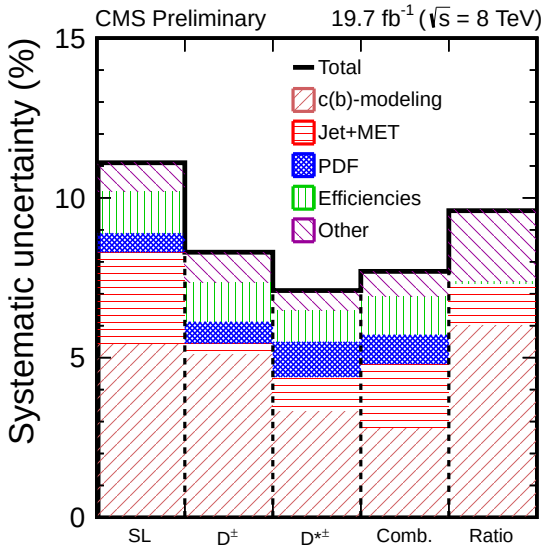

Figure 7: Contributions to the systematic uncertainty in the measured $Z + c$ cross section and in the $Z + c/Z + b$ cross sections ratio. The first three bins in the graphic show the uncertainties in the $Z + c$ cross section in the three decay modes, semileptonic , $D^{\pm}$, and $D^{*\pm}(2010)$, calculated from the combination in the dimuon and dielectron Z-boson decay channels. The fourth bin shows the systematic uncertainties in the combined $Z + c$ cross section. The last bin presents the systematic uncertainty in the $Z + c/Z + b$ cross sections ratio measured in the semileptonic mode. The uncertainty from every source is added on top of the already displayed ones according to its contribution to the total uncertainty.

charm hadrons is 0.086 ± 0.004 [40, 41]. The average of these two values, $\mathcal{B}(c \to \ell) = 0.091 \pm 0.003$, is consistent with the PYTHIA value present in our simulations (9.3%). We assign a 5% uncertainty in order to cover both central values within one standard deviation. The average of the inclusive bottom semileptonic branching fractions is $\mathcal{B}(b \to \ell) = 0.1069 \pm 0.0022$ [40], that is consistent with the PYTHIA value used in our simulations (10.5%). The corresponding uncertainty of 2% is propagated. Since the simulation in the $D^{\pm}$ and $D^{*\pm}(2010)$ modes is reweighted to match the experimental values [41], the uncertainty in the reweighting factors (5% for $D^{\pm}$ and 3.2% for $D^{*\pm}(2010)$) is propagated to the cross section.

The effect of a possible mismodelling of the gluon splitting process into $c\bar{c}$ ($b\bar{b}$) pairs in our reference $Z +$ jets simulated sample is evaluated independently by increasing by a 50% the weight of the events with at least two c-hadrons (b-hadrons) at generator level close to the reconstructed jet ($\Delta R(\text{jet}, c(b)) < 0.5$).

The uncertainty in the c-tagging scale factors is in the range 3.5-4% and it is around 2.5% for the b-tagging efficiency. In the $D^{*\pm}(2010)$ mode, the candidate building procedure is repeated changing independently by one sigma, in terms of the p_T resolution, the different p_T-thresholds imposed and the decay length significance requirement. We take as uncertainty the quadratic sum of the respective differences between MC and data in the change of the number of $D^{*\pm}(2010)$ candidates (2.8%).

The uncertainty from the lepton-efficiency correction factors are 2% in the $Z \to \mu^+\mu^-$ and 4% in the $Z \to e^+e^-$ channels. The uncertainty in the efficiency for the identification of muons inside jets is taken to be 3% according to dedicated studies in multijet events.

The effects due to the uncertainty in the knowledge of the jet energy scale and jet energy resolution are assessed varying the corresponding jet energy scale (jet energy resolution) correction factors within their uncertainties according to the results of CMS dedicated studies. The uncertainty from a mismeasurement of the missing transverse energy of the event is estimated by adding to the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ a 10% of the energy unassociated with reconstructed particle-flow objects.

An additional systematic uncertainty is assigned to account for a possible mismodelling of the subtracted backgrounds. For the $\mathrm{t\bar{t}}$ background the uncertainty is taken as the difference between the estimate based on data, as described in Section 5, and on simulation. For Z + light-quark-jet, the systematic uncertainty is evaluated by using different tagging working points. Finally, the dibosons contribution is varied by the uncertainties in their theoretical cross sections ($\sim 6\%$).

The reference signal simulated sample is generated with MADGRAPH +PYTHIA6 using the PDF CTEQ6L1 and reweighted to NNLO PDF set MSTW2008NNLO. The effect of using other NNLO PDF sets, NNPDF2.3 [49] and CT10 is evaluated and the largest difference in the final cross section, obtained using the NNPDF2.3 PDF set, is taken as the systematic uncertainty due to an imperfect knowledge of the PDFs.

The shapes of the discriminant distributions obtained from the W + c event sample are observed to be very stable. Changes in the jet energy scale and variations in the p_{T}-threshold imposed to select W-boson candidates do not affect the shape of the templates. The correction factors applied in certain regions to the corrected secondary vertex mass template for Z + b events are varied within their uncertainties.

Uncertainties due to the pileup modeling are calculated using a modified pileup profile obtained with a minimum bias cross section changed by its estimated uncertainty, 6%. The uncertainty in the determination of the integrated luminosity of the data sample is 2.6% [50].

Systematic uncertainties in the differential Z + c cross section and in the Z + c/Z + b cross sections ratio are in the range $11 - 15\%$. The main sources of systematic uncertainty in the differential distributions are due to the jet energy scale determination, the charm fractions for c-hadron production and decay in the MC and the efficiencies of heavy-flavour tagging. The uncertainty in the binned c-tagging efficiency scaling factors are $\sim$ 7–8%. Uncertainties in the b-tagging efficiencies are also higher. An additional source of systematic uncertainty in the differential measurement as a function of the transverse momentum of the jet arises from the statistical uncertainty in the determination of the response matrix. Its impact is evaluated repeating the unfolding procedure using a large number of unfolding matrices, each of them built from the nominal reponse matrix, and varying their components according to their statistical uncertainties. The effect is in the range 4–6% for the Z + c cross section and 4.5–7% for the Z + c/Z + b cross sections ratio.

7 Inclusive Z + c cross section and Z + c/Z + b cross sections ratio

For all channels under study, the Z + c cross section is determined in the fiducial region $p_{\mathrm{T}}^{\ell} > 20\,\mathrm{GeV}$, $|\eta^{\ell}| < 2.1$, $71 < m_{\ell\ell} < 111$ GeV, $p_{\mathrm{T}}^{\mathrm{jet}} > 25\,\mathrm{GeV}$, $|\eta^{\mathrm{jet}}| < 2.5$, and $\Delta R(\mathrm{jet}, \ell) > 0.5$, using the following expression:

$$\sigma(\mathrm{Z} + \mathrm{c}) = \frac{N_{\mathrm{Z+c}}^{\mathrm{signal}}}{\mathcal{C} \times \mathcal{L}}, \tag{1}$$

where $N^{\text{signal}}(\mathrm{Z}+\mathrm{c})$, is the fitted yield of $\mathrm{Z}+\mathrm{c}$ events and $\mathcal{L}$ is the integrated luminosity. The factor $\mathcal{C}$ corrects for events losses in the selection process and is estimated using simulated events.

Similarly, the ratio of cross sections $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ is calculated in the same fiducial region applying the previous expression also for the $\mathrm{Z}+\mathrm{b}$ contribution:

$$\frac{\sigma(\mathrm{Z}+\mathrm{c})}{\sigma(\mathrm{Z}+\mathrm{b})} = \frac{N^{\text{signal}}_{\mathrm{Z}+\mathrm{c}}}{N^{\text{signal}}_{\mathrm{Z}+\mathrm{b}}} \times \frac{\mathcal{C}(\mathrm{Z}+\mathrm{b})}{\mathcal{C}(\mathrm{Z}+\mathrm{c})}, \tag{2}$$

Table 1 presents the $\mathrm{Z}+\mathrm{c}$ production cross section and the $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ cross sections ratio (semileptonic sample only) obtained with the three decay modes.

The $\mathrm{Z}+\mathrm{c}$ cross sections obtained in the three categories of this analysis are consistent and can be combined. Statistical and systematic uncertainties of the individual measurements are taken into account in the combination process. Systematic uncertainties arising from a common source and affecting several measurements are considered as fully correlated. The combined values are:

$$\begin{aligned}
\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})\times\mathcal{B}(\mathrm{Z}\to\mathrm{e}^+\mathrm{e}^-) &= 8.5\pm0.7\,(\text{stat.})\pm0.7\,(\text{syst.})\,\text{pb},\\
\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})\times\mathcal{B}(\mathrm{Z}\to\mu^+\mu^-) &= 8.8\pm0.6\,(\text{stat.})\pm0.7\,(\text{syst.})\,\text{pb},\\
\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})\times\mathcal{B}(\mathrm{Z}\to\ell^+\ell^-) &= 8.6\pm0.5\,(\text{stat.})\pm0.7\,(\text{syst.})\,\text{pb}.
\end{aligned}$$

In both Z-boson decay channels the combination is dominated by the result in the semileptonic channel ($\sim$ 50%). Despite the limited size of the samples selected in the $\mathrm{D}^{*\pm}(2010)$ channel, their contribution to the respective averages is significant ($\geq$ 35%). The contribution of the $\mathrm{D}^{\pm}$ channel to the cross section combination is $\sim$ 10%.

The cross sections ratio $\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})/\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{b}+\mathrm{X})$ has been measured in the semileptonic mode, in the two Z-boson decay modes, and the results among them are consistent. Both cross sections ratio are combined taking into account the statistical and systematic uncertainties of the two channels and the correlations among them,

$$\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})/\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{b}+\mathrm{X}) = 2.0\pm0.2\,(\text{stat.})\pm0.2\,(\text{syst.}).$$

The measured $\mathrm{Z}+\mathrm{c}$ cross section and $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ cross sections ratio are compared to predictions from two MC event generators and to pQCD calculations at NLO using the MCFM program. The predicted $\mathrm{Z}+\mathrm{c}$ cross section from MADGRAPH is $\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})\times\mathcal{B}(\mathrm{Z}\to\ell^+\ell^-) = 8.14\pm0.03\,(\text{stat})\pm0.25\,(\text{PDF})$ pb is in agreement with the measured value. The quoted PDF uncertainty corresponds to the largest difference of the prediction using the central value of two different PDF sets (MSTW2008 vs NNPDF2.3), the uncertainty from the several members within each individual PDF set is about half this value.

We have also compared with the predictions from MADGRAPH5_AMC@NLO [51] (hereafter denoted as MG5_AMC) interfaced with PYTHIA v8.1 [52] using the CUETP8M1 tune [53] for parton showering and hadronization. The matrix element calculation includes the Z-boson production processes with 0, 1 and 2 partons at NLO. The FxFx [54] merging scheme is used with a merging scale parameter set to 20 GeV. The NNPDF3.0 PDF set [55] is used for the matrix element calculation while the NNPDF2.3 LO is used for the showering and hadronization. The $\mathrm{Z}+\mathrm{c}$ cross section predicted by MG5_AMC is slightly higher $\sigma(\mathrm{pp}\to\mathrm{Z}+\mathrm{c}+\mathrm{X})\times\mathcal{B}(\mathrm{Z}\to\ell^+\ell^-) = 9.47\pm0.04\,(\text{stat})\pm0.15\,(\text{PDF})\pm0.50\,(\text{scales})$ also in agreement with the measurement. Uncertainties in the prediction are evaluated making use of the reweighting features implemented in

the generator [56]. The PDF uncertainty quoted in this case corresponds to the variation from the one hundred PDF replicas in the set. The scale uncertainty is the envelope of the predictions when the factorization and renormalization scales are varied in a factor of two or one half independently, keeping always the ratio between them smaller than or equal to two.

Theoretical predictions in pQCD at NLO for the associated production of a Z boson and a HF-quark are obtained with the MCFM 7.0 program [57, 58] using the NLO PDF sets MSTW08, CT10, NNPDF3IC and NNPDF3nIC [10], accessed through the LHAPDF6 [59] library interface. Kinematic selection follows the experimental requirements: the two leptons from the Z-boson decay with $p_{\mathrm{T}}^{\ell} > 20\,\mathrm{GeV}$, $|\eta^{\ell}| < 2.1$, and $71 < m_{\ell\ell} < 111$ GeV, and a parton-jet with $p_{\mathrm{T}}^{\mathrm{parton-jet}} > 25\,\mathrm{GeV}$, $|\eta^{\mathrm{parton-jet}}| < 2.5$, and separated from the leptons by $\Delta R(\mathrm{jet}, \ell) > 0.5$. Partons are joined using an anti-k_t algorithm with a size parameter of 0.5. The renormalization and factorization scales are set to the mass of the Z boson. Corrections are applied to account for non-perturbative effects, they are on average of the order of 10% for $\mathrm{Z} + \mathrm{c}$ processes as estimated with our reference MC of MADGRAPH. No significant differences in the predictions were obtained using either NNPDF3IC or NNPDF3nIC PDF sets. Differences among them start to be sizeable when the transverse momentum of the Z boson is $\gtrsim 200\,\mathrm{GeV}$ [10], that is at the border of the sensitivity region of the present measurements. Only the predictions using NNPDF3IC set will be discussed in the following.

All pQCD predictions are smaller than the measured values. The largest prediction is obtained using MSTW08 PDF set, $\sigma(\mathrm{pp} \to \mathrm{Z} + \mathrm{c} + \mathrm{X}) \times \mathcal{B}(\mathrm{Z} \to \ell^+\ell^-) = 5.32 \pm 0.01\ ^{+0.12}_{-0.06}\ (\mathrm{PDF})$ pb. Predictions obtained using CT10 and NNPDF3IC are about 5% smaller, with uncertainties of the same order than MSTW08.

In what the $\mathrm{Z} + \mathrm{c}/\mathrm{Z} + \mathrm{b}$ cross sections ratio is concerned, the prediction from MADGRAPH is $1.805 \pm 0.006\,(\mathrm{stat}) \pm 0.004\ (\mathrm{PDF})$, where the PDF uncertainty corresponds to the largest variation in the predicted ratio using the several PDF sets. Uncertainties from the several members within one set are negligible in the cross sections ratio prediction. The expectation from MG5_AMC is $1.87 \pm 0.07\,(\mathrm{stat}) \pm 0.50\ (\mathrm{scales})$, the uncertainties from PDF essentially vanish in the ratio. Both predictions agree with the measured ratio.

A prediction for the cross sections ratio is also obtained with MCFM, as the ratio of the predictions for $\sigma(\mathrm{Z} + \mathrm{c})$ and $\sigma(\mathrm{Z} + \mathrm{b})$, using for both processes the same parameters emulating the experimental scenario. The highest predicted value is $\sigma(\mathrm{pp} \to \mathrm{Z} + \mathrm{c} + \mathrm{X})/\sigma(\mathrm{pp} \to \mathrm{Z} + \mathrm{b} + \mathrm{X}) = 1.58 \pm 0.01\ (\mathrm{stat.} + \mathrm{syst.})$ obtained when the CT10 PDF set is used. The prediction from NNPDF3IC is about 10% lower, mainly because the predicted $\mathrm{Z} + \mathrm{b}$ cross section using this PDF is the highest one. All predictions from MCFM are lower than the experimental cross sections ratio.

8 Differential $\mathrm{Z} + \mathrm{c}$ cross section and $\mathrm{Z} + \mathrm{c}/\mathrm{Z} + \mathrm{b}$ cross sections ratio

The $\mathrm{Z} + \mathrm{c}$ production cross section and the $\mathrm{Z} + \mathrm{c}/\mathrm{Z} + \mathrm{b}$ cross sections ratio are measured differentially as a function of the transverse momentum of the Z boson and of the transverse momentum of the HF-jet with the sample selected in the semileptonic mode. The transverse momentum of the Z boson is reconstructed from the momenta of the two selected leptons. The sample is divided in three different subsamples according to the value of the variable of interest, $p_{\mathrm{T}}^{\mathrm{Z}}$ or $p_{\mathrm{T}}^{\mathrm{jet}}$, and the fit procedure is performed independently for each of them, and for the two lepton-flavour Z-decay modes. The number and size of the bins is chosen such that the

corrected secondary vertex mass distribution for each bin is sufficiently populated to perform efficiently the signal extraction fit.

Potential effects of events migration between neighbouring bins and inside/outside the acceptance due to the detector resolution has been studied using simulated samples.

Migration effects in $p_{\mathrm{T}}^{\mathrm{Z}}$ are found to be negligible and no corrections are applied. This has an impact of less than $\sim$ 1% in the measured cross section and is included as a systematic uncertainty in the results.

Some migration of events in $p_{\mathrm{T}}^{\mathrm{jet}}$ between neighbouring bins is expected due to the experimental jet energy resolution, mainly between the first and second bins ($<$ 30%), migrations between the second and third bins are smaller than 10%. Migration effects are expected to be the same in the two Z-boson decay modes. The response matrix is used to unfold the fitted signal yields to signal yields at particle level. Acceptance losses in the border of the kinematical region due to resolution effects and reconstruction inefficiencies are also corrected. The unfolding is performed with an analytical inversion of the matrix defining the bin migrations. Statistical and systematic uncertainties are propagated through the unfolding procedure.

Tables 2 and 3 summarize the fitted $\mathrm{Z}+\mathrm{c}$ and $\mathrm{Z}+\mathrm{b}$ signal yields, the $\mathrm{Z}+\mathrm{c}$ cross section and the $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ cross sections ratio in the three $p_{\mathrm{T}}^{\mathrm{Z}}$ and $p_{\mathrm{T}}^{\mathrm{jet}}$ bins and in the two Z-boson decay channels. The differential cross section and cross sections ratio measured in the two Z-boson

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}}}$ (pb)	$N_{\mathrm{Z+b}}^{\mathrm{signal}}$	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}}} / \frac{\mathrm{d}\sigma(\mathrm{Z+b})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}}}$
$0 < p_{\mathrm{T}}^{\mathrm{Z}} < 30\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	212 ± 44	$0.066 \pm 0.014 \pm 0.010$	578 ± 52	$1.5 \pm 0.4 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	380 ± 61	$0.103 \pm 0.017 \pm 0.018$	693 ± 68	$2.7 \pm 0.6 \pm 0.4$
$30 < p_{\mathrm{T}}^{\mathrm{Z}} < 60\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	501 ± 60	$0.144 \pm 0.017 \pm 0.019$	1035 ± 66	$2.4 \pm 0.4 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	586 ± 92	$0.123 \pm 0.019 \pm 0.018$	1422 ± 87	$1.9 \pm 0.4 \pm 0.3$
$60 < p_{\mathrm{T}}^{\mathrm{Z}} < 200\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	363 ± 53	$0.017 \pm 0.002 \pm 0.002$	913 ± 67	$1.7 \pm 0.3 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	474 ± 73	$0.017 \pm 0.003 \pm 0.002$	1056 ± 81	$2.1 \pm 0.4 \pm 0.3$

Table 2: Differential cross section $d\sigma(\mathrm{Z}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}}$, and cross section ratio $(d\sigma(\mathrm{Z}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}})/(d\sigma(\mathrm{Z}+\mathrm{b})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{Z}})$ in the semileptonic mode and in the two Z-boson decay channels. $N_{\mathrm{Z+c}}^{\mathrm{signal}}$ and $N_{\mathrm{Z+b}}^{\mathrm{signal}}$ are the yields of $\mathrm{Z}+\mathrm{c}$ and $\mathrm{Z}+\mathrm{b}$ events extracted from the fit. All uncertainties quoted in the table are statistical except for the measured cross sections and cross sections ratio where the first uncertainty is statistical and the second one is the estimated systematic uncertainty from the sources discussed in the text.

decay channels are consistent and are combined to obtain the final results taking into account the statistical and systematic uncertainties of the two channels and the correlations among them. The combined cross section and cross sections ratio are presented in Table 4. They are also shown graphically in Fig. 8 in bins of $p_{\mathrm{T}}^{\mathrm{Z}}$ (top) and $p_{\mathrm{T}}^{\mathrm{jet}}$ (bottom).

Theoretical predictions for the differential cross section and cross sections ratio are also obtained with the two generator programs and with MCFM. They are shown in Fig. 8 for comparison with the measured values. Error bars in the MADGRAPH predictions include the statistical and PDF uncertainties. Scale variations are also included in the uncertainties from MG5_AMC. Predictions from MG5_AMC are higher than the predictions from MADGRAPH in the three bins

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}}$ (pb)	$N_{\mathrm{Z+b}}^{\mathrm{signal}}$	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}} / \frac{\mathrm{d}\sigma(\mathrm{Z+b})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}}$
$25 < p_{\mathrm{T}}^{\mathrm{jet}} < 40\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	476 ± 58	$0.342 \pm 0.048 \pm 0.041$	1022 ± 67	$2.3 \pm 0.6 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	583 ± 91	$0.337 \pm 0.059 \pm 0.055$	1393 ± 90	$2.4 \pm 0.5 \pm 0.3$
$40 < p_{\mathrm{T}}^{\mathrm{jet}} < 60\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	289 ± 47	$0.090 \pm 0.027 \pm 0.018$	843 ± 59	$1.3 \pm 0.6 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	456 ± 66	$0.103 \pm 0.027 \pm 0.014$	1044 ± 75	$1.9 \pm 0.5 \pm 0.3$
$60 < p_{\mathrm{T}}^{\mathrm{jet}} < 200\,\mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	310 ± 56	$0.012 \pm 0.003 \pm 0.008$	686 ± 64	$1.7 \pm 0.5 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	369 ± 63	$0.013 \pm 0.003 \pm 0.007$	800 ± 75	$1.9 \pm 0.5 \pm 0.3$

Table 3: Differential cross section $d\sigma(\mathrm{Z} + \mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}$, and cross section ratio $(d\sigma(\mathrm{Z} + \mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}})/(d\sigma(\mathrm{Z} + \mathrm{b})/\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}})$ in the semileptonic mode and in the two Z-boson decay channels. $N_{\mathrm{Z+c}}^{\mathrm{signal}}$ and $N_{\mathrm{Z+b}}^{\mathrm{signal}}$ are the yields of $\mathrm{Z} + \mathrm{c}$ and $\mathrm{Z} + \mathrm{b}$ events extracted from the fit. All uncertainties quoted in the table are statistical except for the measured cross sections and cross sections ratio where the first uncertainty is statistical and the second one is the estimated systematic uncertainty from the sources discussed in the text.

$[p_{\mathrm{T}}^{Z}{}_{\mathrm{min}}, p_{\mathrm{T}}^{Z}{}_{\mathrm{max}}]$ [GeV]	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{Z}}$ [pb]	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{Z}} / \frac{\mathrm{d}\sigma(\mathrm{Z+b})}{\mathrm{d}p_{\mathrm{T}}^{Z}}$
$[0, 30]$	$0.075 \pm 0.011 \pm 0.012$	$1.8 \pm 0.3 \pm 0.2$
$[30, 60]$	$0.133 \pm 0.013 \pm 0.018$	$2.1 \pm 0.3 \pm 0.3$
$[60, 200]$	$0.017 \pm 0.002 \pm 0.002$	$1.9 \pm 0.3 \pm 0.2$
$[p_{\mathrm{T}}^{\mathrm{jet}}{}_{\mathrm{min}}, p_{\mathrm{T}}^{\mathrm{jet}}{}_{\mathrm{max}}]$ [GeV]	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}}$ [pb]	$\frac{\mathrm{d}\sigma(\mathrm{Z+c})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}} / \frac{\mathrm{d}\sigma(\mathrm{Z+b})}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}}$
$[25, 40]$	$0.341 \pm 0.037 \pm 0.042$	$2.5 \pm 0.4 \pm 0.3$
$[40, 60]$	$0.097 \pm 0.019 \pm 0.012$	$1.5 \pm 0.3 \pm 0.2$
$[60, 200]$	$0.013 \pm 0.002 \pm 0.002$	$1.8 \pm 0.4 \pm 0.3$

Table 4: Differential $\mathrm{Z} + \mathrm{c}$ cross section and $\mathrm{Z} + \mathrm{c}/\mathrm{Z} + \mathrm{b}$ cross sections ratio. The first block presents the differential measurements as a function of the transverse momentum of the Z boson. The second block shows the cross section and ratio as a function of the transverse momentum of the jet with heavy flavour content. The first uncertainty is the statistical and the second one is the systematic uncertainty arising from the sources discussed in the text.

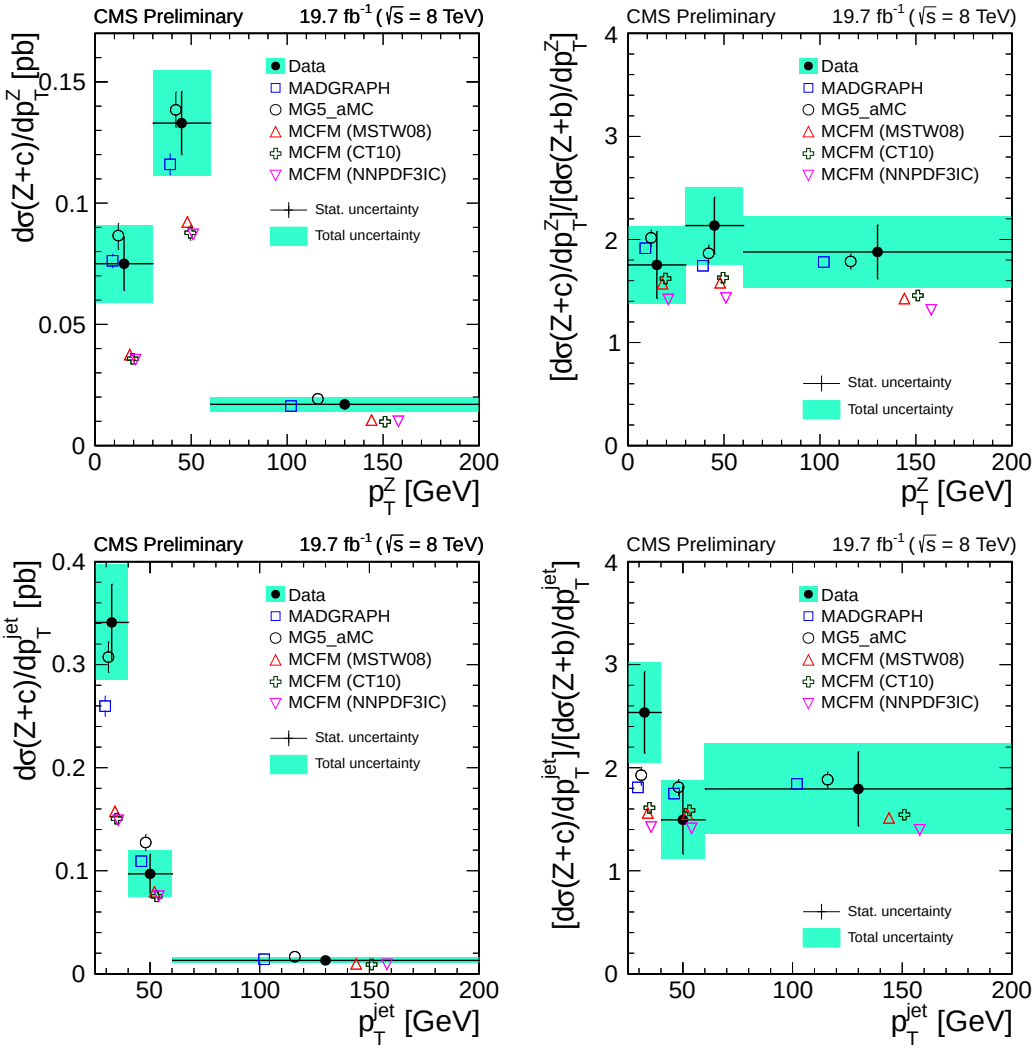

Figure 8: Differential $Z + c$ cross section and $Z + c/Z + b$ cross sections ratio as a function of the transverse momentum of the Z boson (top) and the transverse momentum of the jet (bottom). The combination of the results in the dielectron and dimuon channels are shown. The $Z + c$ differential cross section is shown in the left and $Z + c/Z + b$ cross sections ratio is shown in the right. Statistical uncertainties in the data are shown as error bars. The solid rectangles indicate the total (statistical plus systematic) experimental uncertainty. Statistical and systematic uncertainties in the theoretical predictions are shown added in quadrature.

of the $Z + c$ differential distributions. A higher $Z + c/Z + b$ cross sections ratio is also predicted, although consistent within uncertainties.

The predictions from MADGRAPH and MG5_AMC successfully reproduce the measurements, the level of agreement is similar in terms of the $Z + c$ cross section and the $Z + c/Z + b$ cross sections ratio.

Concerning the theoretical predictions calculated with MCFM and the various PDF sets, the same ordering than discussed above for the inclusive cross section follows in the differential cross sections. The highest $Z + c$ cross section is predicted using MSTW08 PDF set, the largest differential $Z + c/Z + b$ cross sections ratio in the two variables is obtained with CT10 PDF set.

All MCFM predictions are lower than the differential cross section measurements in the first and second bins in $p_{\mathrm{T}}^{\mathrm{Z}}$. This discrepancy concentrates in the first bin in $p_{\mathrm{T}}^{\mathrm{jet}}$, the MCFM predictions reproducing the measurements for $p_{\mathrm{T}}^{\mathrm{jet}} > 40\,\mathrm{GeV}$. Differences between predictions and data are reduced in the $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ cross sections ratio comparison.

9 Summary

The cross section of the production of a Z boson associated with at least one jet originated by a c-quark in pp collisions at a center-of-mass energy of 8 TeV is measured with a data sample corresponding to an integrated luminosity of $19.7 \pm 0.5\,\mathrm{fb}^{-1}$. The cross sections ratio of the production of a Z boson and at least one c- or b-quark jet is also determined. The measurements are performed in the kinematic region with two leptons with $p_{\mathrm{T}}^{\ell} > 20$ GeV, pseudorapidity $|\eta^{\ell}| < 2.1$, and dilepton invariant mass $71 < m_{\ell\ell} < 111$ GeV, and a HF-quark jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 25\,\mathrm{GeV}$, $|\eta^{\mathrm{jet}}| < 2.5$, and separated from the leptons from the Z-boson candidate by a distance $\Delta R(\mathrm{jet}, \ell) > 0.5$.

The measured $\mathrm{Z}+\mathrm{c}$ production cross section is $\sigma(\mathrm{pp} \to \mathrm{Z}+\mathrm{c}+\mathrm{X}) = 8.6 \pm 0.5\ (\mathrm{stat.}) \pm 0.7\ (\mathrm{syst.})$ pb, and the cross sections ratio is $\sigma(\mathrm{pp} \to \mathrm{Z}+\mathrm{c}+\mathrm{X})/\sigma(\mathrm{pp} \to \mathrm{Z}+\mathrm{b}+\mathrm{X}) = 2.0 \pm 0.2\ (\mathrm{stat.}) \pm 0.2\ (\mathrm{syst.})$. Both, the $\mathrm{Z}+\mathrm{c}$ production cross section and the cross sections ratio are measured inclusively and differentially as a function of transverse momentum of the Z boson and of the heavy flavour jet. The measurements are in agreement with the LO predictions from MADGRAPH and NLO predictions from MG5_AMC. Predictions from the MCFM program are lower than the measured $\mathrm{Z}+\mathrm{c}$ cross section, both inclusive and differentially. A better description is reached in terms of the $\mathrm{Z}+\mathrm{c}/\mathrm{Z}+\mathrm{b}$ cross sections ratio.

References

[1] CDF Collaboration, "Search for the Flavor Changing Neutral Current Decay $t \to Zq$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **101** (2008) 192002, doi:10.1103/PhysRevLett.101.192002, arXiv:0805.2109.

[2] D0 Collaboration, "Search for flavor changing neutral currents in decays of top quarks", *Phys. Lett. B* **701** (2011) 313, doi:10.1016/j.physletb.2011.06.014, arXiv:1103.4574.

[3] ATLAS Collaboration, "Search for pair-produced third-generation squarks decaying via charm quarks or in compressed supersymmetric scenarios in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Phys. Rev. D* **90** (2014), no. 5, 052008, doi:10.1103/PhysRevD.90.052008, arXiv:1407.0608.

[4] ATLAS Collaboration, "Search for Scalar Charm Quark Pair Production in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector", *Phys. Rev. Lett.* **114** (2015), no. 16, 161801, doi:10.1103/PhysRevLett.114.161801, arXiv:1501.01325.

[5] S. J. Brodsky et al., "A review of the intrinsic heavy quark content of the nucleon", *Adv. High Energy Phys.* **2015** (2015) 231547, doi:10.1155/2015/231547, arXiv:1504.06287.

[6] P.-H. Beauchemin, V. A. Bednyakov, G. I. Lykasov, and Yu. Yu. Stepanenko, "Search for intrinsic charm in vector boson production accompanied by heavy flavor jets", *Phys. Rev.*

D **92** (2015), no. 3, 034014, doi:10.1103/PhysRevD.92.034014, arXiv:1410.2616.

[7] G. Bailas and V. P. Goncalves, "Phenomenological implications of the intrinsic charm in the Z boson production at the LHC", *Eur. Phys. J. C* **76** (2016) 105, doi:10.1140/epjc/s10052-016-3941-z, arXiv:1512.06007.

[8] A. V. Lipatov, G. I. Lykasov, Yu. Yu. Stepanenko, and V. A. Bednyakov, "Probing proton intrinsic charm in photon or Z boson production accompanied by heavy jets at LHC", arXiv:1606.04882.

[9] T. Boettcher, P. Ilten, and M. Williams, "Direct probe of the intrinsic charm content of the proton", *Phys. Rev. D* **93** (2016), no. 7, 074008, doi:10.1103/PhysRevD.93.074008, arXiv:1512.06666.

[10] NNPDF Collaboration, "A Determination of the Charm Content of the Proton", arXiv:1605.06515.

[11] D0 Collaboration, "Measurement of associated production of Z bosons with charm quark jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **112** (2014), no. 4, 042001, doi:10.1103/PhysRevLett.112.042001, arXiv:1308.4384.

[12] CDF Collaboration, "Measurement of vector boson plus $D^*(2010)^+$ meson production in $\bar{p}p$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. D* **93** (2016), no. 5, 052012, doi:10.1103/PhysRevD.93.052012, arXiv:1508.06980.

[13] LHCb Collaboration, "Observation of associated production of a Z boson with a D meson in the forward region", *JHEP* **04** (2014) 091, doi:10.1007/JHEP04(2014)091, arXiv:1401.3245.

[14] 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.

[15] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[16] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[17] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi, and W events at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, CERN, 2010.

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

[19] 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.

[20] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", arXiv:1607.03663.

[21] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

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

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

[24] J. M. Campbell, R. K. Ellis, P. Nason, and E. Re, "Top-pair production and decay at NLO matched with parton showers", *JHEP* **04** (2015) 114, doi:10.1007/JHEP04(2015)114, arXiv:1412.1828.

[25] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[26] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[27] 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* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[28] J. Gao et al., "The CT10 NNLO Global Analysis of QCD", *Phys. Rev. D* **89** (2014) 033009, doi:10.1103/PhysRevD.89.033009, arXiv:1302.6246.

[29] R. Field, "Early LHC Underlying Event Data - Findings and Surprises", in *Hadron collider physics. Proceedings, 22nd Conference, HCP 2010, Toronto, Canada, August 23-27, 2010*. 2010. arXiv:1010.3558.

[30] A. Buckley et al., "Systematic event generator tuning for the LHC", *Eur. Phys. J. C* **65** (2010) 331, doi:10.1140/epjc/s10052-009-1196-7, arXiv:0907.2973.

[31] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. and Methods A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[32] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[33] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

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

[35] 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.

[36] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.

[37] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[38] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.

[39] CMS Collaboration, "Measurement of the cross section and angular correlations for associated production of a Z boson with b hadrons in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **12** (2013) 039, doi:10.1007/JHEP12(2013)039, arXiv:1310.1349.

[40] Particle Data Group, K.A. Olive et al., "Review of Particle Physics", *Chin. Phys. C* **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[41] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP", *Eur. Phys. J. C* **75** (2015) 19, doi:10.1140/epjc/s10052-014-3250-3, arXiv:1404.3888.

[42] OPAL Collaboration, "Measurement of $f(c \to D^* + X)$, $f(b \to D^* + X)$ and $\Gamma(c\bar{c})/\Gamma(had)$ using $D^{*\pm}$ mesons", *Eur. Phys. J. C* **1** (1998) 439, doi:10.1007/s100520050095, arXiv:hep-ex/9708021.

[43] ALEPH Collaboration, "Study of charm production in Z decays", *Eur. Phys. J. C* **16** (2000) 597, doi:10.1007/s100520000421, arXiv:hep-ex/9909032.

[44] DELPHI Collaboration, "Determination of $P(c \to D^{*+})$ and $BR(c \to \ell^+)$ at LEP 1", *Eur. Phys. J. C* **12** (2000) 209, doi:10.1007/s100529900227.

[45] R. Frühwirth, R. Kubinec, W. Mitaroff, and M. Regler, "Vertex reconstruction and track bundling at the LEP collider using robust algorithms", *Comp. Phys. Comm.* **96** (1996) 189.

[46] CMS Collaboration, "Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **02** (2014) 013, doi:10.1007/JHEP02(2014)013, arXiv:1310.1138.

[47] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CERN, 2013.

[48] LHCb Collaboration, "Identification of beauty and charm quark jets at LHCb", *JINST* **10** (2015), no. 06, P06013, doi:10.1088/1748-0221/10/06/P06013, arXiv:1504.07670.

[49] R. D. Ball et al., "Parton distributions with LHC data", *Nucl. Phys. B* **867** (2013) 244–289, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.

[50] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update", Technical Report CMS-PAS-SMP-12-008, CERN, Geneva, 2012.

[51] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[52] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[53] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[54] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[55] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[56] R. Frederix et al., "Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties", *JHEP* **02** (2012) 099, doi:10.1007/JHEP02(2012)099, arXiv:1110.4738.

[57] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, "Associated production of a Z Boson and a single heavy quark jet", *Phys. Rev. D* **69** (2004) 074021, doi:10.1103/PhysRevD.69.074021, arXiv:hep-ph/0312024.

[58] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, "Production of a Z boson and two jets with one heavy-quark tag", *Phys. Rev. D* **73** (2006) 054007, doi:10.1103/PhysRevD.77.019903, 10.1103/PhysRevD.73.054007, arXiv:hep-ph/0510362. [Erratum: Phys. Rev. D 77,019903(2008)].

[59] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.