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

Juan Pablo Fernández Ramos (CIEMAT)
on behalf of the CMS Collaboration

BTV Workshop @ Brussels
April 11, 2018

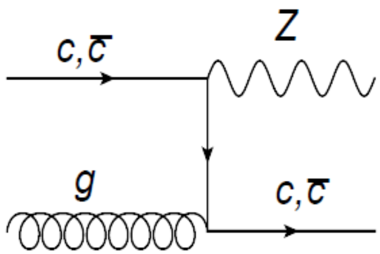
CMS-SMP-15-009 (https://arxiv.org/abs/1711.02143)

Introduction

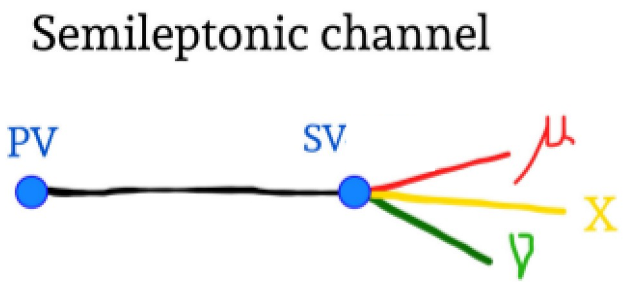
- Measurement of σ(pp→Z+c) and its relative production w.r.t. σ(pp→Z+b) both inclusively and differentially.
- Provides tests QCD predictions.
- Allows gluon (g→ $c\bar{c}$) PDF studies.
- Intrinsic Charm-quark component inside the proton enhances σ(Z+c) @ ↑pt(Z).
- Background in searches (relative contributions of the different flavors to the background is important since Z+c can be misidentified as Z+b jet events) and Higgs (Z+c is a background to ZH→ Z$c\bar{c}$).
- Study the performance of c-tagging algorithms (new in CMS since 2015) [CMS-BTV-16-001 & CMS-BTV-16-002]
  - c-tagging = identification of jets coming from charm quarks

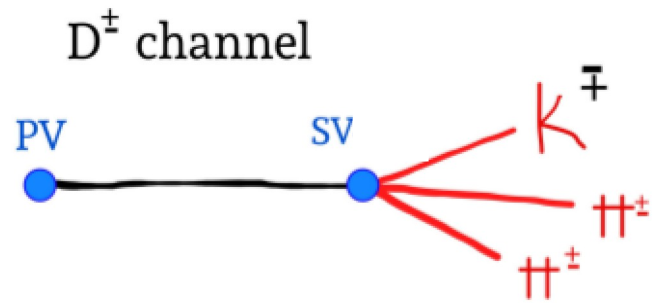
Analysis strategy:

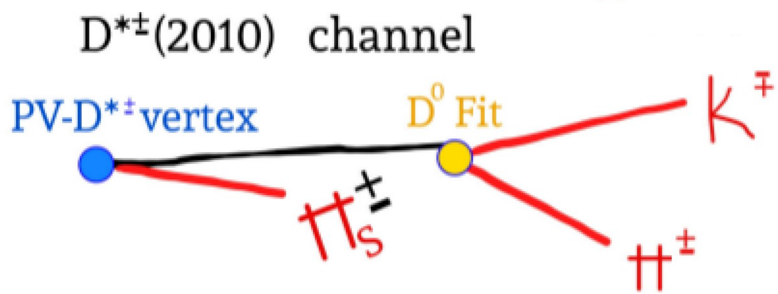
- Standard $Z \rightarrow l^{+}l^{-}$ selection
  - Isolated leptons with $p_{T}(l) > 20$ GeV and $|\eta(l)| < 2.1$
  - Dilepton invariant mass: [71,111] GeV
- anti-k_{T}(R=0.5) jet: $p_{T}^{jet} > 25$ GeV & $|\eta^{jet}| < 2.5$



Z+heavy flavor selection:







Identify jets from heavy flavor quarks through the decay of charm hadrons in 3 final states:

- Semileptonic decay of c/b hadrons: muon in a jet
- $D^{*\pm}$ and $D^{\pm}$ exclusive decays in jet

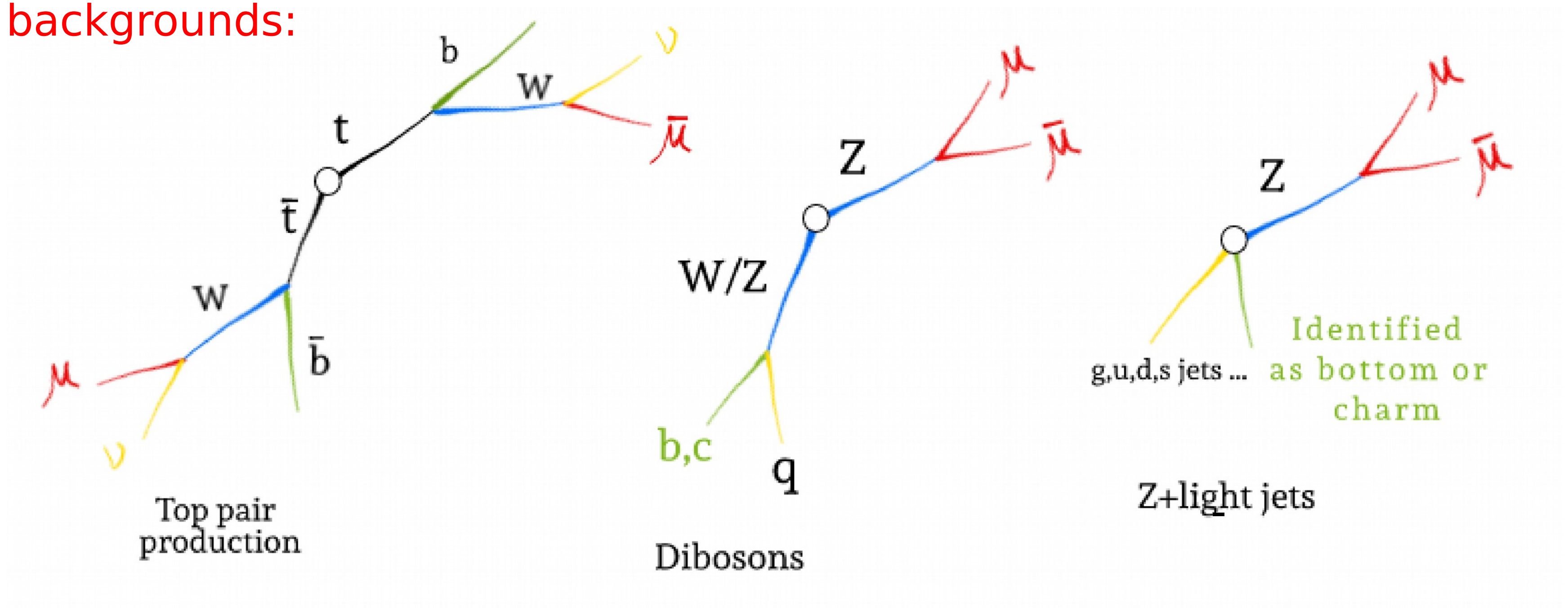
Samples

DATA: 2012 8 TeV (19.7 fb^{-1} ± 0.5)

Signal MC: DY+jets generated w. MADGRAPH5@LO+PYTHIA6(PDF set CTEQ6L)

σ(pp→Z+X) calculated at NNLO with FEWZ (PDF set MSTW2008NNLO)

Main backgrounds:

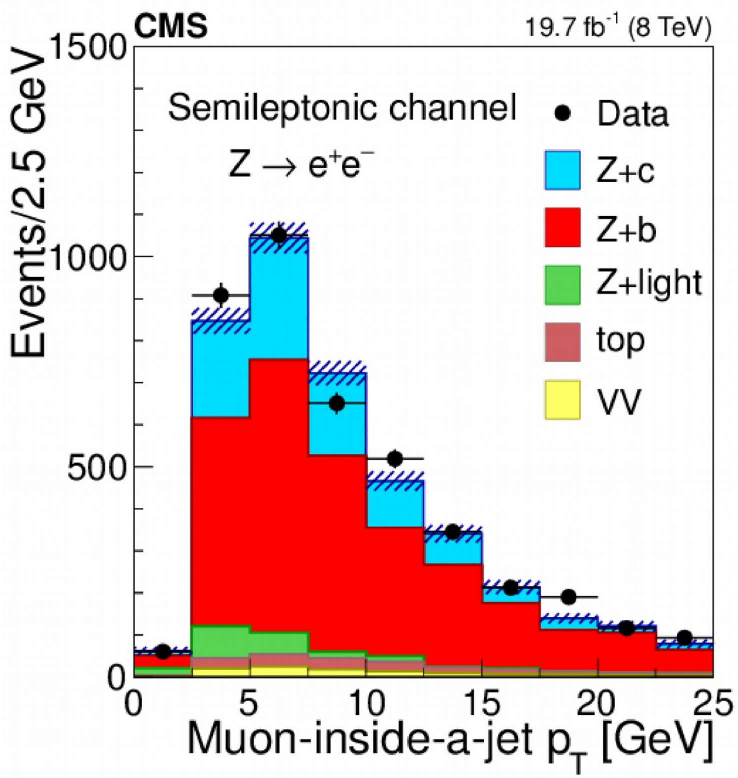


Contributions from ttbar, diboson, Z+light processes (from simulations except ttbar from data).

Missing transverse energy < 40 GeV (to reduce $t\bar{t}$ background).

Data-MC differences in lepton trigger, identification and isolation efficiencies corrected (tag & probe method). Pileup events included in the MC.

Semileptonic selection



- μ inside a jet and taking part of a secondary vertex (SSV or IVF). This reduces the light contribution more than standard b-tagging algorithms.

- $p_T^{\mu} < 25$ GeV, with $p_T^{\mu}/p_T^{jet} < 0.6$, $|\eta^{\mu}| < 2.5$

- non-isolated, $I_{comb}/p_T^{\mu} > 0.2$

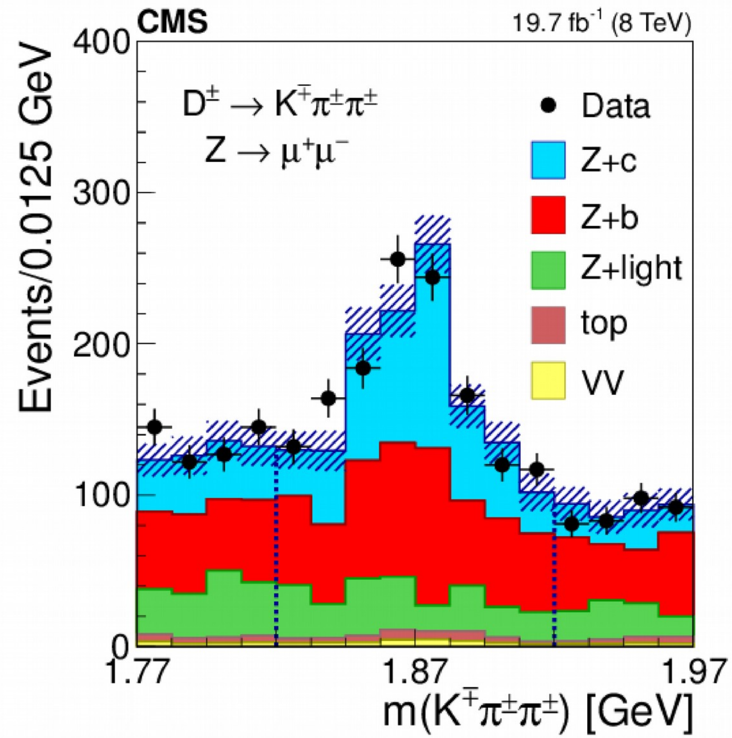
Semileptonic candidates:

4145 $Z \rightarrow e^+e^-$ 5258 $Z \rightarrow \mu^+\mu^-$

Relative contributions:

Z+c: ~25% Z+b: ~65% Z+light: ~5% Others:~5%

$D^{\pm}$ channel – Selection



Use jets with a 3 tracks secondary vertex (SSV, IVF) & search for $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ resonant peak.

Define signal region :
$|m(D^{\pm}) - 1.87| < 0.05$ GeV

Sideband region :
$0.05 < |m(D^{\pm}) - 1.87| < 0.10$ GeV

Non resonant background in the signal region subtracted from the neighboring sidebands

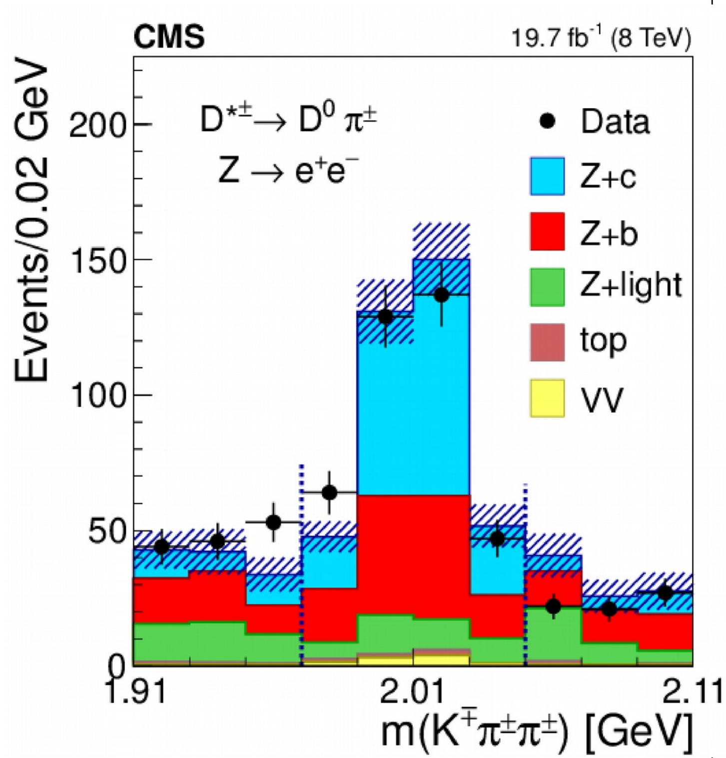
After sideband subtraction:

490 ± 48 $D^{\pm}$ ($Z \to \mu^{+}\mu^{-}$) 375 ± 44 $D^{\pm}$ ($Z \to e^{+}e^{-}$)

Z+c: ~60% Z+b: ~35% Z+light: <1% Others($t\bar{t}$+VV):<4%

- The simulation is reweighted to match the experimental values of $c \to D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ (PDG + L.Gladilin, Eur. Phys. J. C 75 (2015) 19)

D(2010)$^{*\pm}$ channel - Selection



- $D^{*\pm} \to D^0 \pi_s^\pm$ [$D^0 \to K^-\pi^+$(+c.c.)] decay chain.
- Loop over all tracks in the jet .
- Kaon: track with sign opposite to π_s

- D^0 vertex with $L_{xy}/\sigma(L_{xy})>3$, D^0 vertex prob.>0.05
- p_T (K) > 1.75, p_T (π) > 0.75, p_T (π_s) > 0.5 GeV
- |$\Delta R(D^*, \text{jet})$| < 0.5, |$\Delta R(D^0, \pi_s)$| < 0.1.
- |m(D^0)−1.865|<0.1 GeV, |Δm−145|< 5 MeV
- Signal region : 1.97 < m(D*) < 2.05 GeV
- Sidebands : 0.06 < |m($D^{*\pm}$)−2.01| < 0.12 GeV

Non resonant background in the signal region subtracted from the neighboring sidebands

After sideband subtraction:

309 ± 22 D*±(Z→ $\mu^+\mu^-$) 234 ± 22 D*±(Z→ e^+e^-)

Z+c:~65% Z+b:~30% Z+light:<1% Others($t\bar{t}$+VV) :<4%

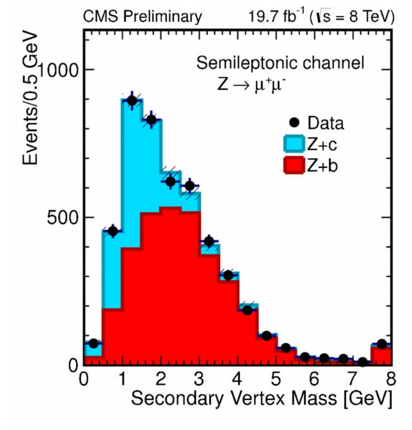
- The simulation is reweighted to match the experimental values of c→ $D^{*\pm} \to D^0 \pi_s^\pm$ [$D^0 \to K^- \pi^+$] (PDG + L.Gladilin, Eur. Phys. J. C 75 (2015) 19)

b/c separation (discriminants)

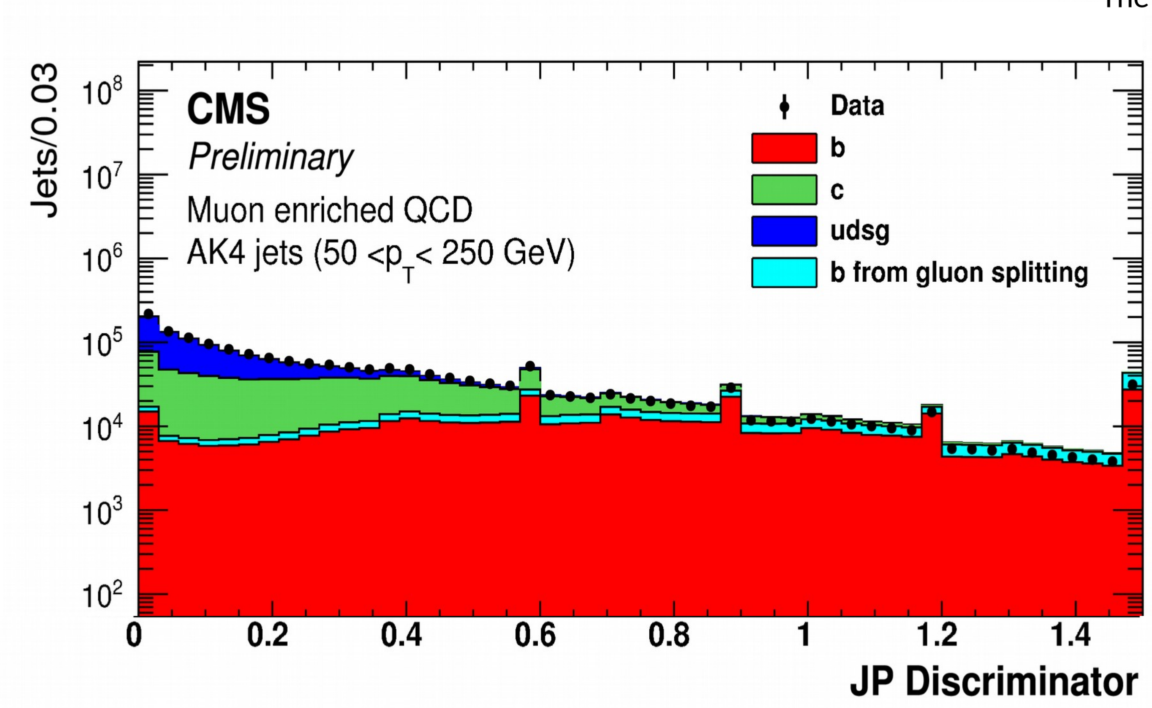
- Vertex mass (for semileptonic mode)

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

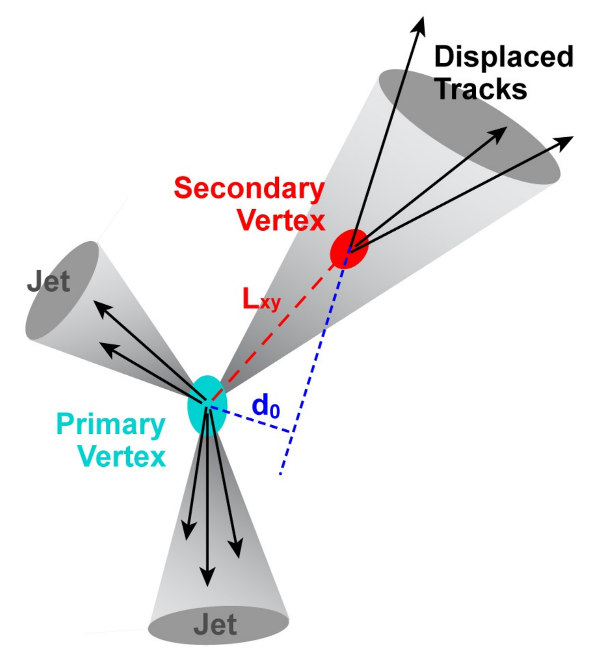
Correction included to account for unidentified neutral decay products



- JP (for D hadron modes): likelihood estimate of prob. of jet tracks to come from primary vertex



The larger the IP of a track the more inconsistent w.r.t. PV



Modeling strategy

Now that we have chosen the variables to separate the different contributions we need a way to model properly each of them

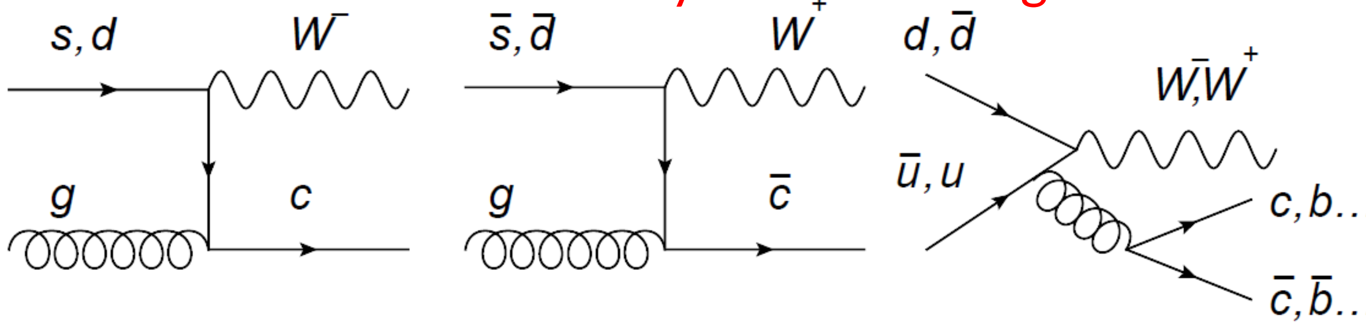
This is called template modeling and has two parts:

- Modeling properly the shape
- Accurate determination of tagging efficiency

- Z+c :
  - Shape : data driven (W+charm) [1st time]
  - Normalization taken from MC after applying vertex-efficiency corrections
- Z+b :
  - Shape : from MC but corrected with data (ttbar)
  - Normalization from MC after vertex-efficiency corrections
- Z+light and Dibosons: shape and normalization from MC
- ttbar: Data driven

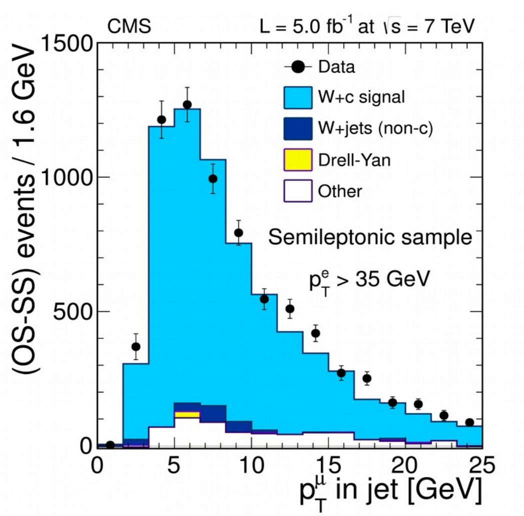
Selection of W+c sample

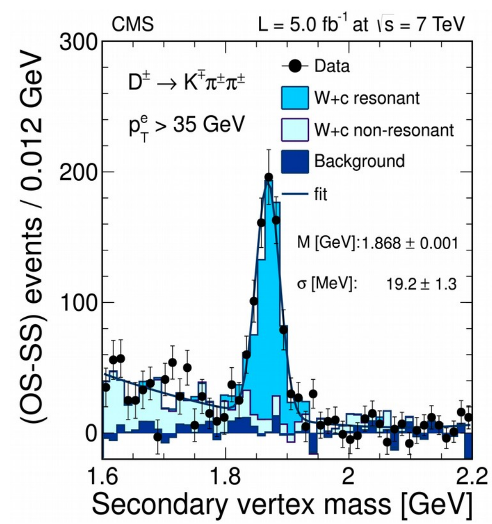
- W⟶ eν plus jets with similar selection to Z+HF
- Same identification of heavy flavor jet: **μ in jet or D-hadron exclusive decays**
- OS–SS subtraction to remove symmetric backgrounds

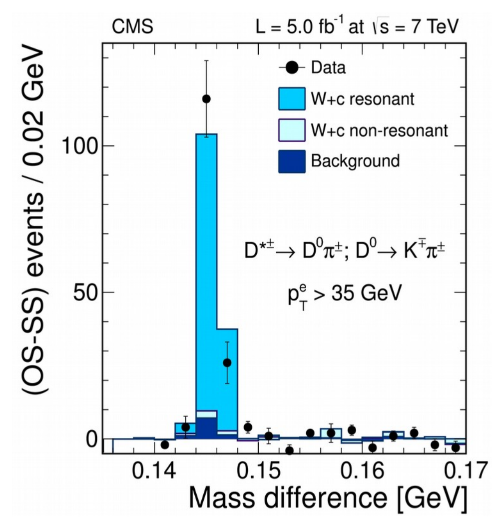


OS event *OS event* *Bckg: 50% OS, 50% SS*

After OS-SS subtraction the purity in W+c of the resulting sample is > 90% (semil. channel) and >98% (D-hadron chan.)



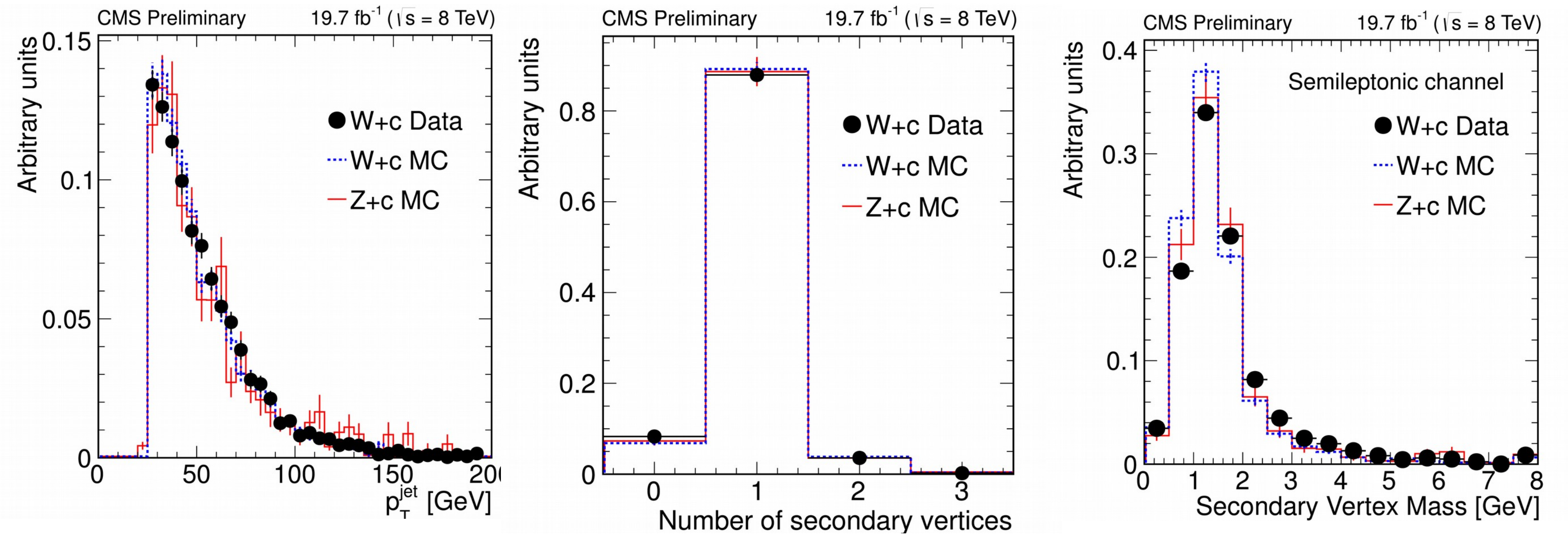




JHEP 02 013(W+c)

Template (shape) modeling for Z+c

Comparison of c-jets from Z+c and W+c processes
(data from W+c : after subtraction of remaining (little) background)



- *Agreement in general distributions (p_T^{jet} , N_{SV})*
- Discriminant distributions (SV-mass and JP) **W+c MC** and **Z+c MC** agree
- JP prob **W+c MC** and **W+c data** agree and validates the **Z+c MC** description
- SV-mass **W+c MC** and **W+c data** do not agree

The shape is not well modeled by the **W+c MC**. We take the shape of SV-mass from **W+c data** since there is agreement in the kinematic properties between **Z+c MC** and **W+c MC**

c-tagging efficiency

With the W+c sample we compute the secondary vertex efficiency & SF_c and apply it to the Z+c MC to have a proper description of the detector and algorithms of the c-tagging

$$\epsilon_c^{data} = \frac{N_{\mathrm{W+c}}^{tagged,\, data}}{N_{\mathrm{W+c}}^{data}}$$

Tagged : with the full selection of slide 5
Denominator : releasing the SV requirement
$\epsilon_c^{data} \sim 30$ % of muons take part of SV.

$$SF_c = \frac{\epsilon_c^{data}}{\epsilon_c^{sim}}$$

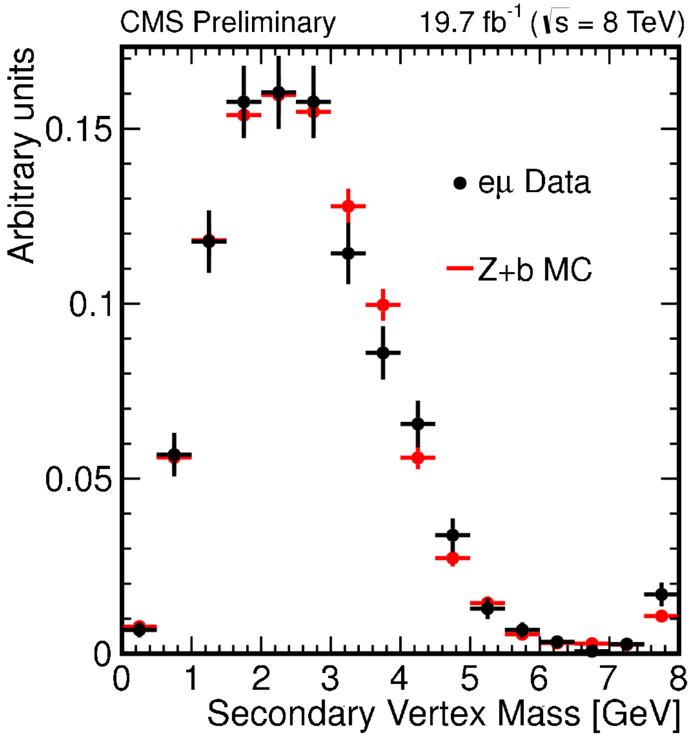
SFc (SSV μ-in-vertex) = $0.882 \pm 0.032 \pm 0.016$
SFc (IVF μ-in-vertex) = $0.918 \pm 0.026 \pm 0.018$

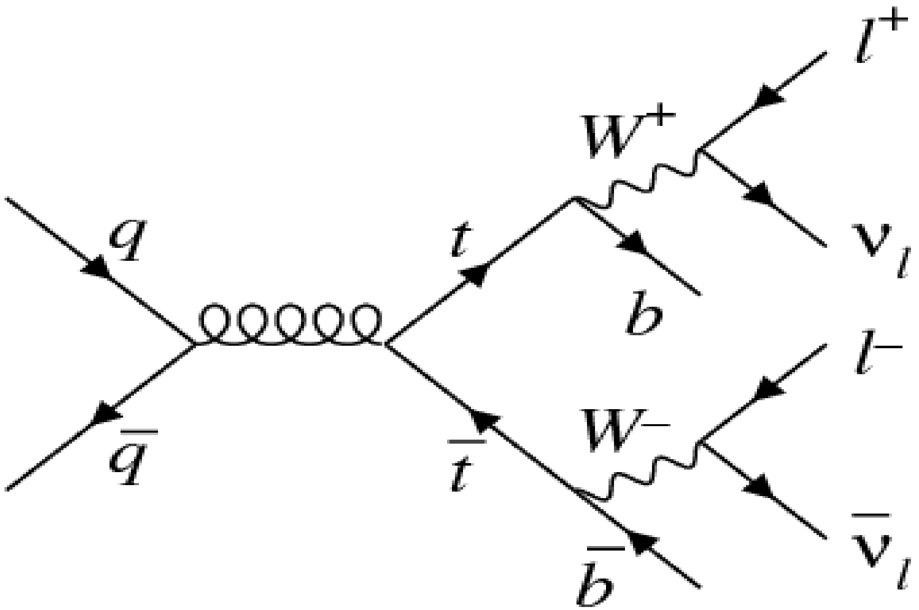
With these SF_c we correct for differences between data and MC simulation in the performance of our way to identify charm

In bins of p_T of the jet SF_c, compatible with 1 for high p_T jets (p_T > 40 GeV)

Template (shape) modeling for Z+b

Evaluated in a clean sample of b-jets from $t\bar{t}$ production where the two W bosons from the $t(\bar{t})$ quark decay leptonically into leptons of different flavor





Identification of heavy flavor: **muon inside the jet**

Correct Z+b MC in the bins where difference > 1 σ

Not enough statistics to validate the Z+b templates in the D-hadron exclusive channels

b-tagging efficiency

With the ttbar sample we compute the vertex efficiency & SF_b and apply it to the Z+b MC to have a proper description of the detector and algorithms of the b-tagging

$$\epsilon_b^{data} = \frac{N_{tt}^{tagged,\, data}}{N_{tt}^{data}}$$

70 % of muons take part of SV.

$$SF_b = \frac{\epsilon_b^{data}}{\epsilon_b^{sim}}$$

SF_b (IVF μ-in-vertex) = 0.96 ± 0.03 (same for SSV), where the uncertainty includes statistical and systematic effects due to jet energy scale and resolution and pileup description.

Now that we have a good description of the shape of the discriminant distributions and of the tagging efficiency for b and c jets we proceed ...

Signal extraction

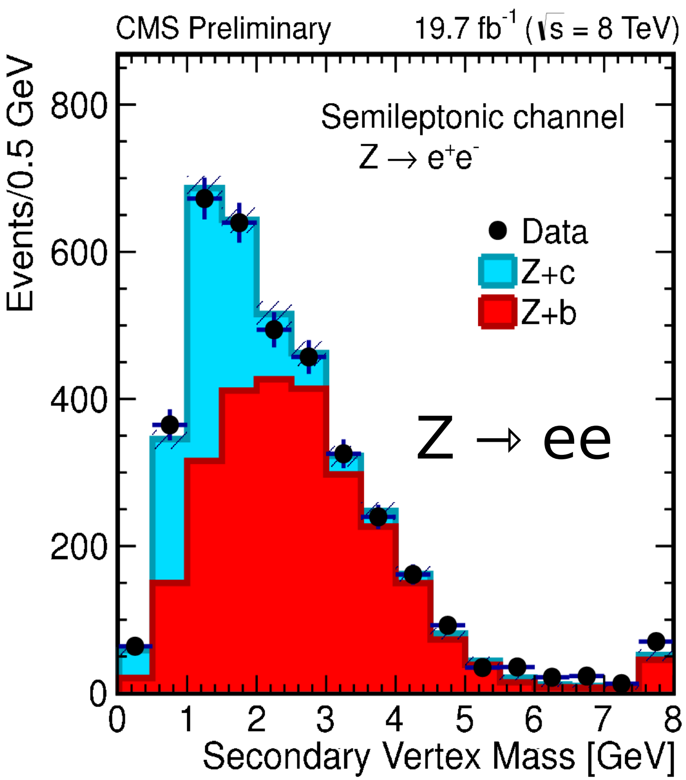
Total # of observed Z+c/Z+b extracted from a χ^2 minimization fit of the Z+c/Z+b templates to the experimental distributions of vertex mass and JP discriminants

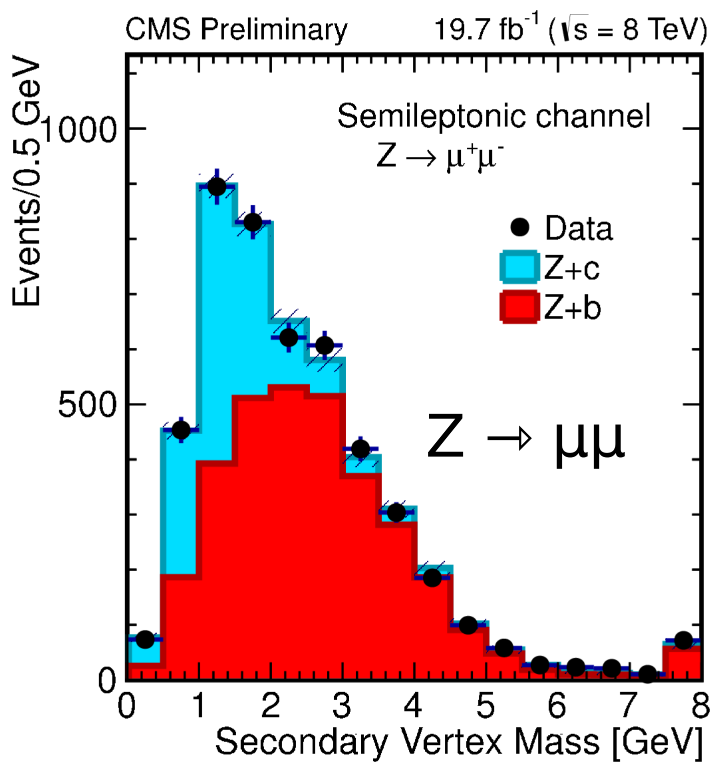
$$\chi^2 = \sum_i \frac{(n_i^{data} - \mu_{\mathrm{Z+c}} N_i^{\mathrm{Z+c}} - \mu_{\mathrm{Z+b}} N_i^{\mathrm{Z+b}})}{(\sigma_i^d)^2 + (\sigma_i^{\mathrm{MC}})^2}$$

n_i = Number of events in data (after subtraction of remaining background from Z+light, $t\bar{t}$ and VV)

N_i^{Z+c}, N_i^{Z+b} = Number of Z+c, Z+b

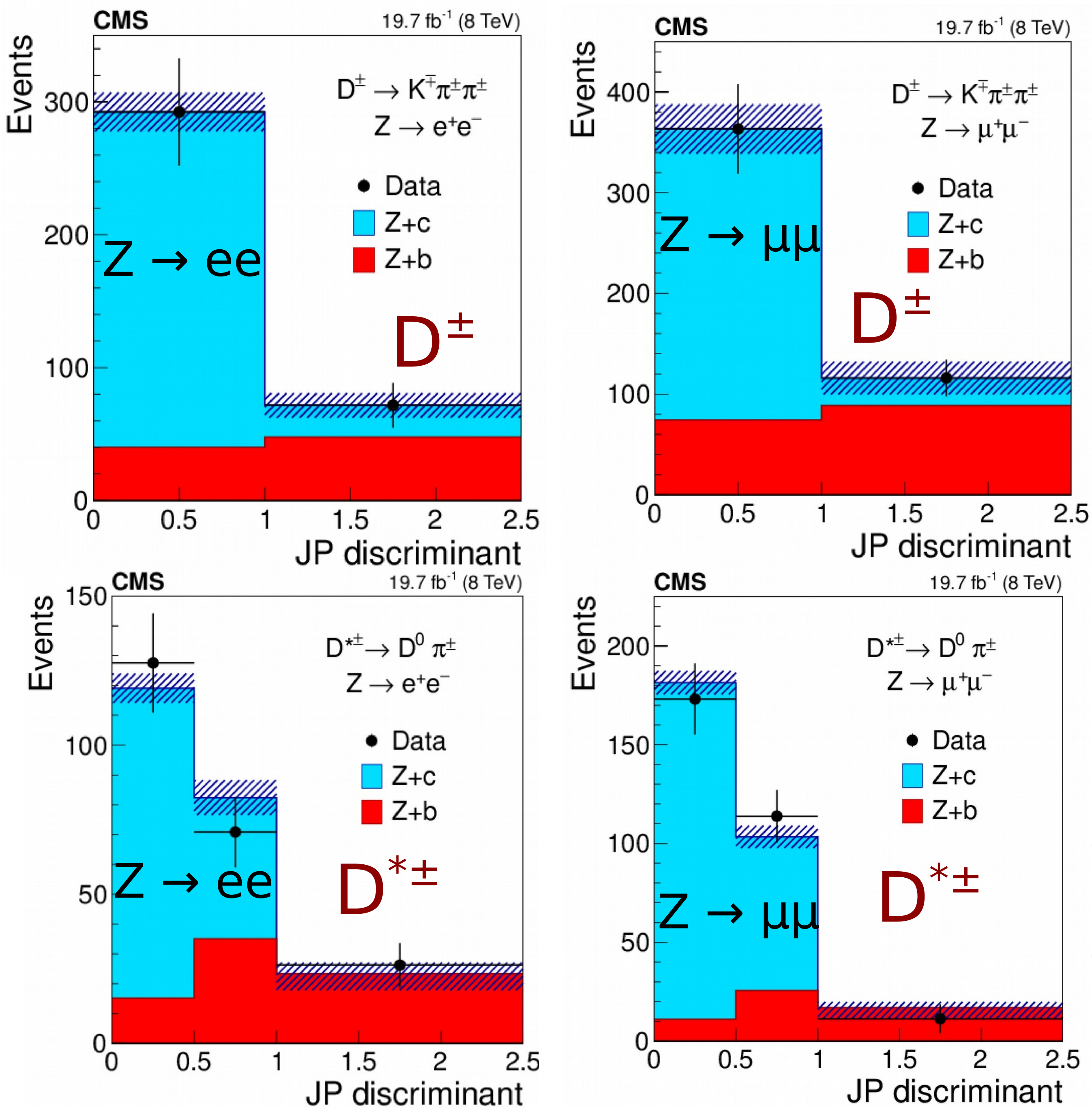
Parameters to fit: μ_{z+c} & μ_{z+b}





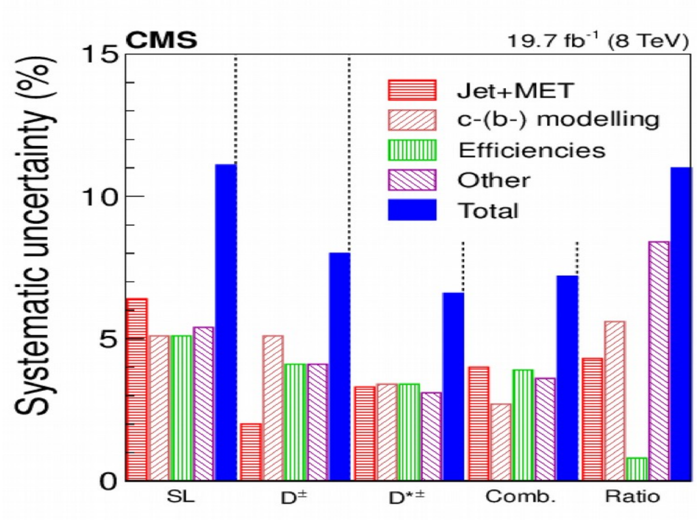
μ_{z+c} & μ_{z+b} in the 0.9-1.1 range

Signal extraction



c/b separation clearer in the D* mode (the soft pion comes from the PV for c→ D* and not for b→ B→D*)

Systematic uncertainties



- **Charm fraction (production and decay)** :
  - SL :take the difference between the BR from inclusive or exclusive individual contributions (5%)
  - D-hadrons: uncertainty on the reweighting factors (5% for $D^{+/-}$, 3% for D*)
- **c(b)gluon-splitting**: increase by three times the experimental uncertainty in the g→ $c\bar{c}$ (g→ $b\bar{b}$) rate, the weight on events with 2c(b) with ΔR (jet,c(b)) < 0.5 (<1%)

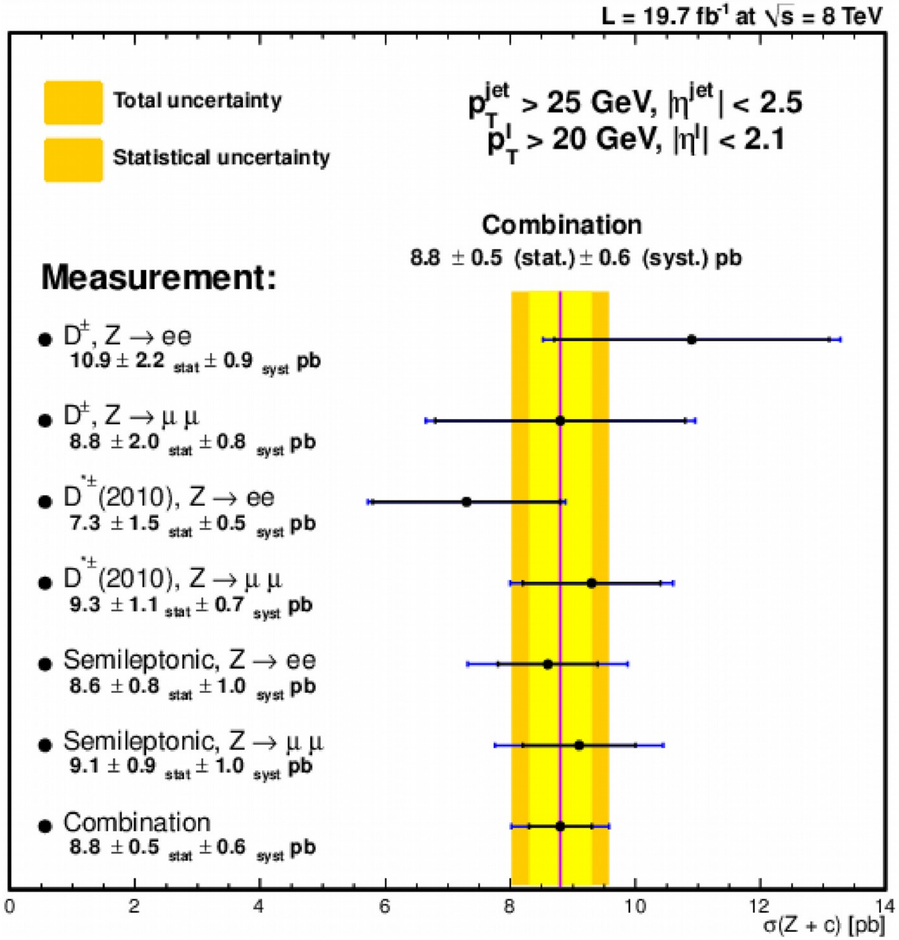
- **Jet Energy scale and Resolution :** change scale and resolution correction factors by their uncertainties (2-5%)
- **Missing-et :** Misestimations on the missing transverse energy: modify the missing E_T by 10% of the unclustered missing E_T (1-2%).

- **c-(b-) tagging efficiency**. Use uncertainty of c(b)-tagging efficiencies (2.5-4%)
- **Lepton efficiencies** :change efficiencies by their errors (4% for electrons, 2% for muons)

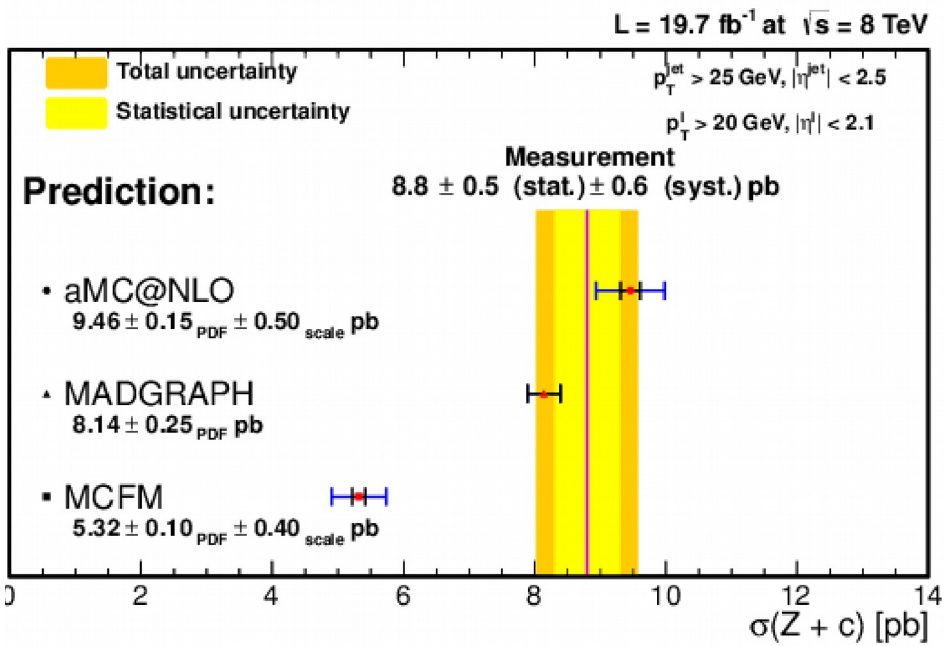
- **PDFs** : difference resulting from using other NNLO PDF sets (<1%)
- **Shape** (semileptonic mode) : change Z+b template correction factor by its error (4-7%)
- **Pileup** profile : assuming a different inelastic cross section (1-3%)
- **Luminosity** : 2.6 %

Cross section determination

σ (Z+c)



$$\sigma(\mathrm{Z+c}) = \frac{N^{fitted}_{\mathrm{Z+c}}}{\epsilon_c^{\mathrm{Z+c}}\mathcal{L}}$$

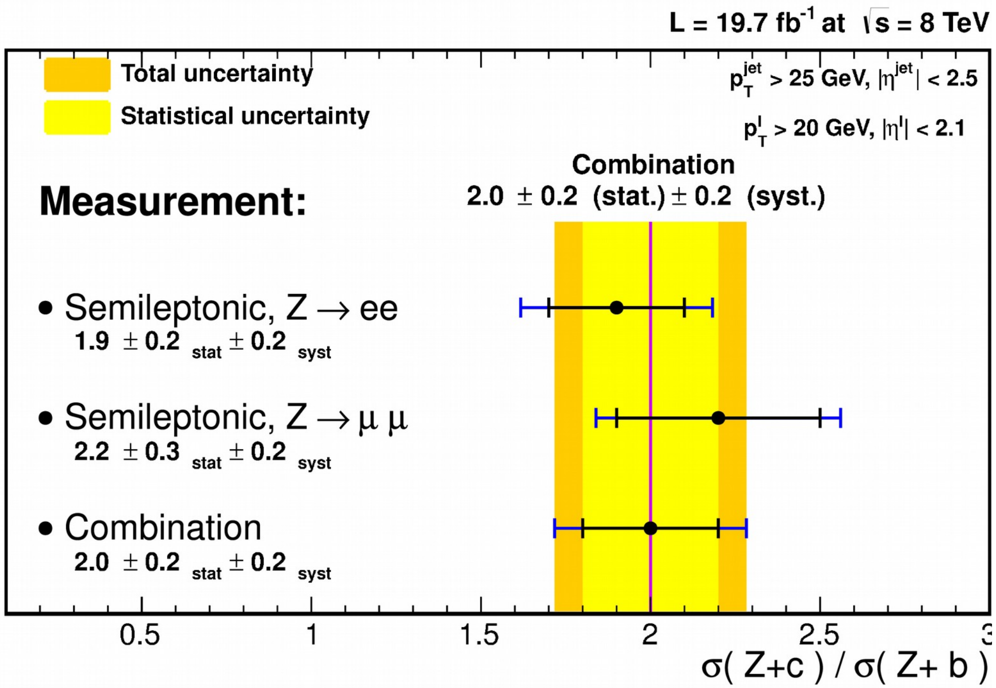


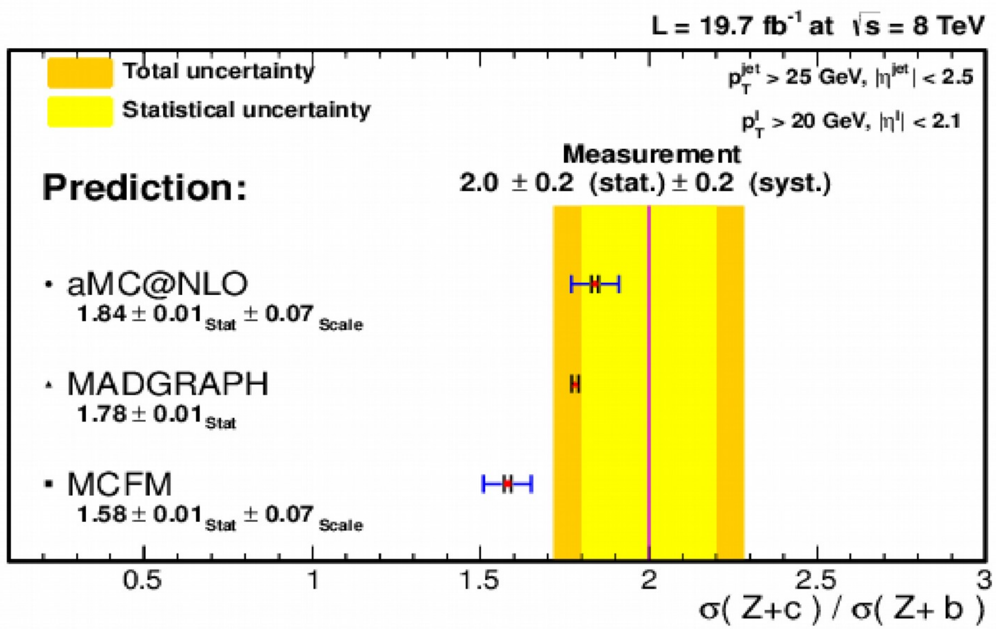
Cross section ratio

$$\frac{\sigma(\mathrm{Z+c})}{\sigma(\mathrm{Z+b})} = $$

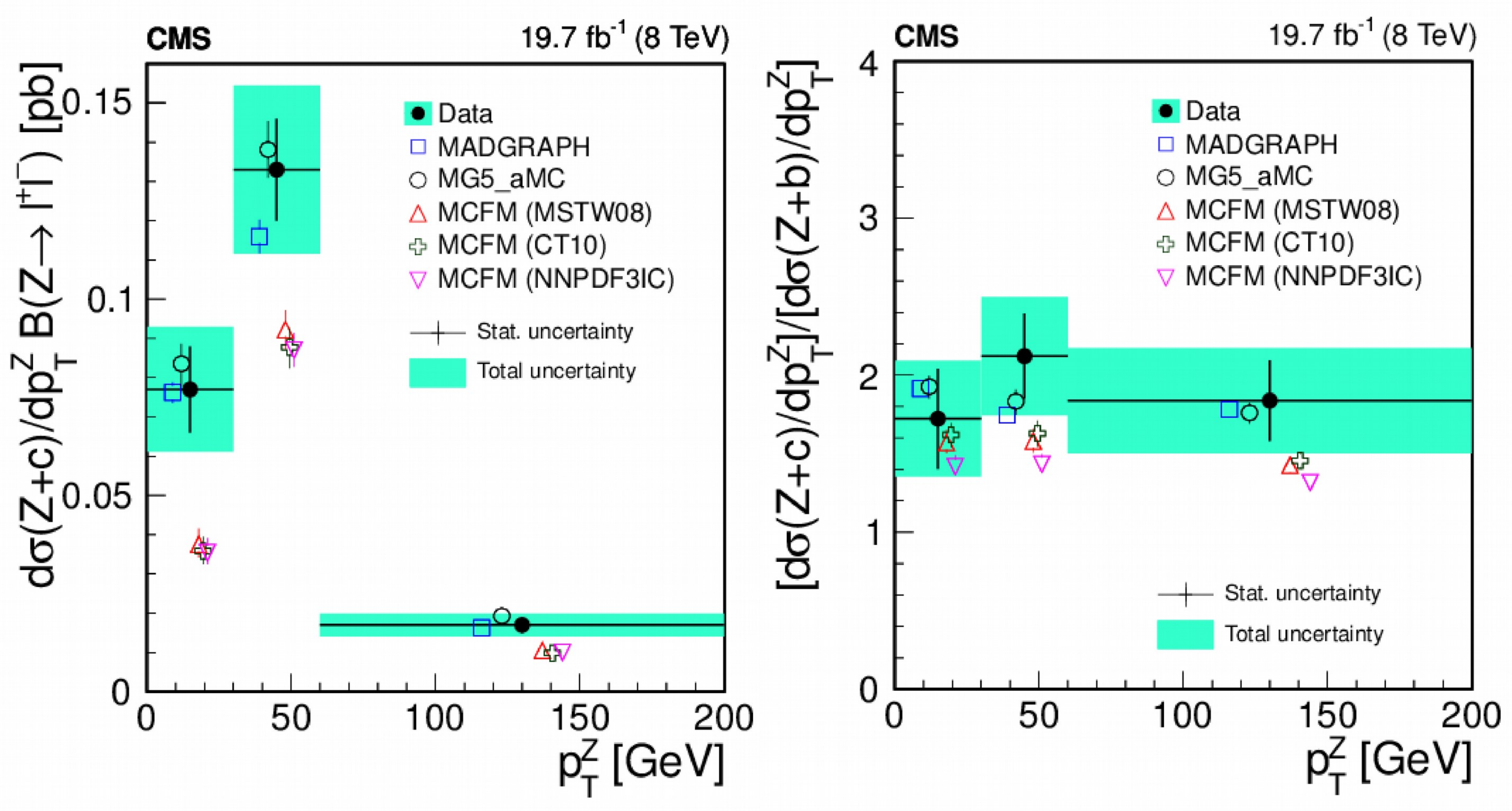
$$= \frac{N^{fitted}_{\mathrm{Z+c}}}{N^{fitted}_{\mathrm{Z+b}}} \frac{\epsilon_c^{\mathrm{Z+b}}}{\epsilon_c^{\mathrm{Z+c}}}$$

Only the semileptonic channel is used





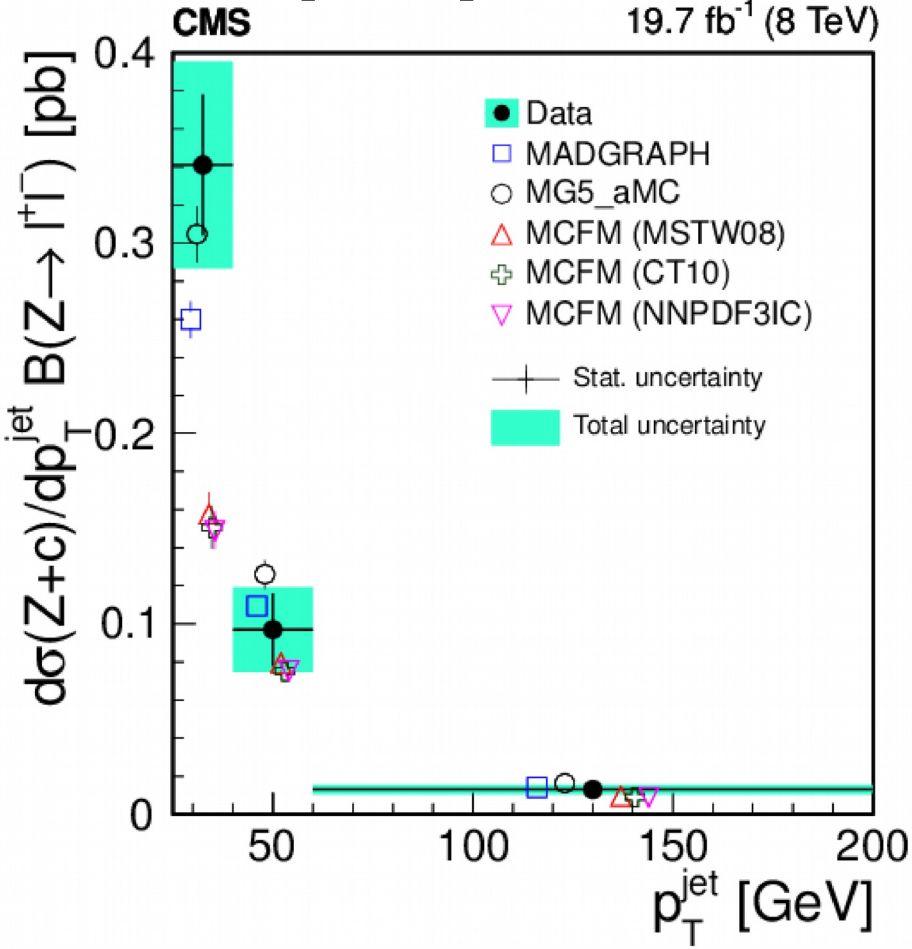
Differential cross sections as a function of p_T^Z

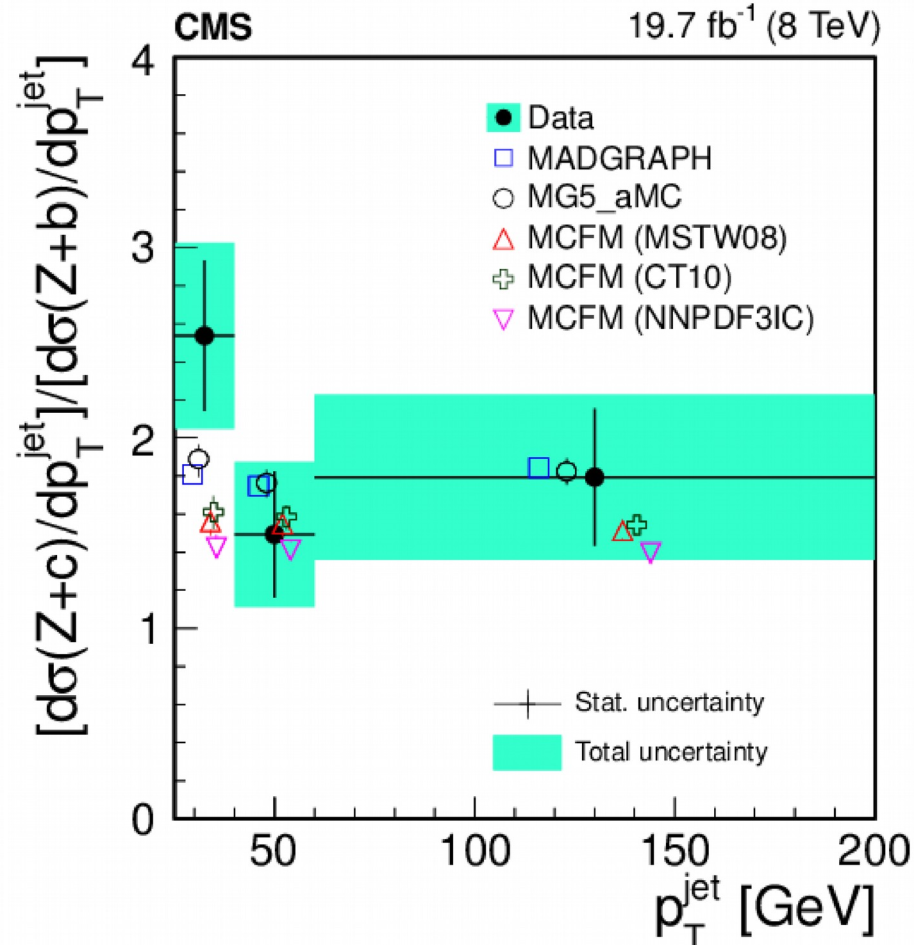


Bins[GeV] : 0-30, 30-60 and 60-200

Differential cross sections as a function of p_T^{jet}

Bins [GeV] : 25-40, 40-60 and 60-200





If the proton momentum fraction taken by the charm quark component (intrinsic + perturbative) is of ~2%,an increase in the production of Z+c events with a $p_T^Z \approx 100$ GeV of at least 20–25% would be expected. No increase in the production rate in the highest p_T^{jet} bin is observed (in agreement with current upper limits on IC component)

Conclusions

Measured **σ(Z+c)** and ratio **σ(Z+c)/σ(Z+b)** inclusive and differential for two opposite sign leptons from the Z with p_T^{lepton} > 20 GeV and $|\eta^{lepton}|$ < 2.1 and p_T^{jet} >25 GeV & $|\eta^{jet}|$<2.5

σ(Z+c) = 8.8 ± 0.5 ± 0.6 pb

σ(Z+c)/σ(Z+b) = 2.0 ± 0.2 ± 0.2

In agreement with predictions from MadGraph5 amc@nlo and Madgraph renormalized to a FEWZ calculation.

https://arxiv.org/abs/1711.02143, CMS-SMP-15-009 , accepted by EPJC.

Back up

Cross section determination

$$\sigma(\mathrm{Z}+\mathrm{c}) = \frac{N^{fitted}_{\mathrm{Z+c}}}{\epsilon_c^{\mathrm{Z+c}}\mathcal{L}}$$

Semileptonic mode			
Channel	$N^{\mathrm{signal}}_{\mathrm{Z+c}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ [pb]
$\mathrm{Z} \rightarrow \mathrm{e^+e^-}$	1066 ± 95	0.63 ± 0.03	$8.6 \pm 0.8 \pm 1.0$
$\mathrm{Z} \rightarrow \mu^+\mu^-$	1449 ± 144	0.81 ± 0.03	$9.1 \pm 0.9 \pm 1.0$
$\mathrm{Z} \rightarrow \ell^+\ell^-$	$\sigma(\mathrm{Z+c}) = 8.8 \pm 0.6\,(\mathrm{stat}) \pm 1.0\,(\mathrm{syst})\,\mathrm{pb}$		
Channel	$N^{\mathrm{signal}}_{\mathrm{Z+b}}$	$\mathcal{C}_{\mathrm{Z+b}}$ (%)	$\sigma(\mathrm{Z+c})/\sigma(\mathrm{Z+b})$
$\mathrm{Z} \rightarrow \mathrm{e^+e^-}$	2606 ± 114	2.90 ± 0.08	$1.9 \pm 0.2 \pm 0.2$
$\mathrm{Z} \rightarrow \mu^+\mu^-$	3240 ± 147	3.93 ± 0.10	$2.2 \pm 0.3 \pm 0.2$
$\mathrm{Z} \rightarrow \ell^+\ell^-$	$\sigma(\mathrm{Z+c})/\sigma(\mathrm{Z+b}) = 2.0 \pm 0.2\,(\mathrm{stat}) \pm 0.2\,(\mathrm{syst})$		
$\mathrm{D}^{\pm}$ mode			
Channel	$N^{\mathrm{signal}}_{\mathrm{Z+c}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ [pb]
$\mathrm{Z} \rightarrow \mathrm{e^+e^-}$	276 ± 55	0.13 ± 0.02	$10.9 \pm 2.2 \pm 0.9$
$\mathrm{Z} \rightarrow \mu^+\mu^-$	316 ± 75	0.18 ± 0.02	$8.8 \pm 2.0 \pm 0.8$
$\mathrm{Z} \rightarrow \ell^+\ell^-$	$\sigma(\mathrm{Z+c}) = 9.7 \pm 1.5\,(\mathrm{stat}) \pm 0.8\,(\mathrm{syst})\,\mathrm{pb}$		
$\mathrm{D}^{*\pm}(2010)$ mode			
Channel	$N^{\mathrm{signal}}_{\mathrm{Z+c}}$	$\mathcal{C}_{\mathrm{Z+c}}$ (%)	$\sigma(\mathrm{Z+c})$ [pb]
$\mathrm{Z} \rightarrow \mathrm{e^+e^-}$	151 ± 31	0.11 ± 0.01	$7.3 \pm 1.5 \pm 0.5$
$\mathrm{Z} \rightarrow \mu^+\mu^-$	247 ± 28	0.14 ± 0.01	$9.3 \pm 1.1 \pm 0.7$
$\mathrm{Z} \rightarrow \ell^+\ell^-$	$\sigma(\mathrm{Z+c}) = 8.5 \pm 0.9\,(\mathrm{stat}) \pm 0.6\,(\mathrm{syst})\,\mathrm{pb}$		
Combination			
$\mathrm{Z} \rightarrow \ell^+\ell^-$	$\sigma(\mathrm{Z+c}) = 8.8 \pm 0.5\,(\mathrm{stat}) \pm 0.6\,(\mathrm{syst})\,\mathrm{pb}$		