

CMS Draft Analysis Note

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

2020/01/17

Archive Hash: 2198964-D

Archive Date: 2020/01/15

Measurement of the b-quark fragmentation function using charmed mesons in pp collisions at $\sqrt{s} = 13$ TeV

Brent R. Yates, Mirena Paneva, Robert Clare, Stephen Wimpenny
University of California, Riverside

Abstract

In this analysis we present the first results of the measurement of the shape parameter r_b in the Lund–Bowler fragmentation function for the b-quark in a $t\bar{t}$ environment. The analysis uses charmed mesons produced in the leptonic decays of $t\bar{t}$ at $\sqrt{s} = 13$ TeV in the CMS detector using the full 2016 dataset with an integrated luminosity of 35.9 fb^{-1} . The tracker only decays of $J/\psi \rightarrow \mu^+\mu^-$ and $D^0 \rightarrow K^\pm\mu^\mp$ are used as proxies for the parent b quark via the ratio x_B (the p_T of the charmed meson over the $\sum p_T^{\text{ch}}$ of all charged particles in the jet) to fit the shape parameter r_b . The final fit result is $r_b = 0.855 \pm 0.030 \text{ (stat)} \pm 0.014 \text{ (syst)}_{+0.092}^{-0.076} \text{ (FSR)}$.

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

PDFAuthor:	Brent R. Yates, Mirena Paneva, Robert Clare, Stephen Wimpenny
PDFTitle:	Measurement of the b-quark fragmentation function using charmed mesons in pp collisions at $\sqrt{s} = 13$ TeV
PDFSubject:	CMS
PDFKeywords:	CMS, physics, software, computing

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

1 Intro

The fragmentation function is an essential component in modeling physical processes. The decay of parton level objects to final state, or detectable, particles requires an empirically measured fragmentation function. Current measurements show the Lund–Bowler function [1] best describes this function for light (u, d, s) and heavy (c, b) quarks [2]. This function contains quark independent parameters, which have been determined via light quark fragmentation, and quark specific parameters, such as the shape parameter r_b , which must still be pinned down for the b quark. This parameter has been measured [3] using e^+e^- data from The Large Electron Positron collider (LEP) [4–6] and The SLAC Linear Collider (SLC) [7], but this is a very different from the color rich environment from the proton-proton collisions of The Large Hadron Collider (LHC) [8]. To measure r_b this analysis uses the $t\bar{t}$ environment produced at the LHC using the Compact Muon Solenoid (CMS) detector [9]. Since the top quark only decays via the weak force to a bottom quark ($t \rightarrow b + W$) it can be used to study the bottom quark fragmentation. The b -fragmentation mismodeling was a significant uncertainty in top mass measurements in Run I. The fragmentation function has also never been measured quantitatively in a $t\bar{t}$ environment.

In this analysis we follow well established method [10, 11] of partially reconstructing the top quark in leptonic final states using charmed mesons, specifically the J/ψ and D^0 mesons. This method looks at $t\bar{t}$ events with at least one top quark decaying leptonically. The charmed mesons are produced in the fragmentation of a bottom quark. This will allow us to experimentally measure the b -fragmentation model. The branching ratio for these decays are quite small, hence the full 35.9 fb^{-1} from the 2016 proton-proton collisions in the center of mass-energy $\sqrt{s} = 13 \text{ TeV}$ at the LHC will provide much needed statistics for these events.

The branching ratios for these events are $BR(W \rightarrow \ell\nu) \cdot BR(b \rightarrow J/\psi + X) \sim 32.57 \times 10^{-2} \cdot 2.5 \times 10^{-2} \sim 0.81\%$ and $BR(B^\pm \rightarrow D^0 + X) \cdot BR(D^0 \rightarrow K^\pm \pi^\mp) \sim 0.79 \cdot 0.039 \sim 3.1\%$. As the $J/\psi \rightarrow e^+e^-$ and $W \rightarrow \tau\nu_\tau$ are harder to reconstruct, they are not used in this analysis. This reduces the J/ψ branching ratio by $5.93 \times 10^{-2} \cdot 65.5 \times 10^{-2}$ and the D meson branching ratios by 65.5×10^{-2} . The final branching ratio for J/ψ is then $\sim 3.2 \times 10^{-4}$, while the final branching ratios for D^0 is 2×10^{-2} .

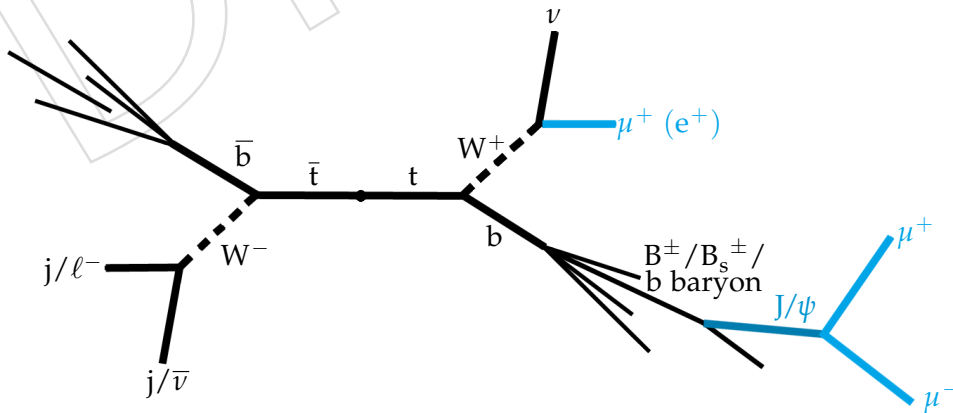


Figure 1: Pictorial view of an exclusive J/ψ production in a $t\bar{t}$ system.

2 Data and Simulation Samples

2.1 Datasets

The triggers used for Single Muon (μ), Single Electron (e), Double Muon ($\mu\mu$), Double Electron (ee), and Electron Muon ($e\mu$) samples can be found in Table 1. Triggers with a DZ in the title are only used in epoch H as some triggers were prescaled during these runs.

The complete 2016 dataset, comprising of 23SepReReco for epochs B-G and PromptReco for epoch H, is used in this analysis. They have a total integrated luminosity of 35.9 fb^{-1} . An exhaustive list these samples can be found in Table 2.

Table 1: HLT Paths

Channel	HLT	Epoch
μ	HLT_IsoMu24_v	BCDEFGH
	HLT_IsoTkMu24_v	
e	HLT_Ele32_eta2p1_WPTight_Gsf_v	BCDEFGH
$\mu\mu$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v	BCDEFG
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v	
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v	H
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v	
ee	HLT_DoubleEle24_22_eta2p1_WPLoose_Gsf_v	BCDEFGH
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	
$e\mu$	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v	BCDEFG
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	H
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	

Table 2: Datasets and integrated luminosity

Dataset	Integrated lumi (fb^{-1})				
	e	μ	ee	$e\mu$	$\mu\mu$
/Run2016B-23Sep2016-v3/MINIAOD	5.78	5.78	5.78	5.78	5.78
/Run2016C-23Sep2016-v1/MINIAOD	2.57	2.57	2.57	2.57	2.57
/Run2016D-23Sep2016-v1/MINIAOD	4.25	4.24	4.25	4.24	4.25
/Run2016E-23Sep2016-v1/MINIAOD	4.01	4.01	4.01	4.01	4.01
/Run2016F-23Sep2016-v1/MINIAOD	3.10	3.10	3.10	3.10	3.10
/Run2016G-23Sep2016-v1/MINIAOD	7.54	7.54	7.54	7.54	7.54
/Run2016H-PromptReco-v2/MINIAOD	8.39	8.39	8.39	8.39	8.39
/Run2016H-PromptReco-v3/MINIAOD	0.22	0.22	0.22	0.22	0.22
	35.86	35.86	35.86	35.86	35.86

2.2 Monte Carlo Simulations

Official Monte Carlo (MC) simulations from Moriond17 TrancheIV are used to determine the signal and background (Table 3). Signal for the $t\bar{t}$ samples is modeled using POWHEG [12–

[15] at next-to-leading order (NLO) with a top quark mass of 172.5 GeV and a next-to-next-to leading order (NNLO) cross section of 832^{+40}_{-46} pb [16, 17]. The events are then processed in PYTHIA 8 (v. 8.219) [18, 19] using tunes CUETP8M1 [20] for the underlying event (UE) and tune CUETP8M2T4 [21] for parton shower (PS) modeling. The parton distribution functions (PDF) used in POWHEG are the NLO set NNPDF30 [22]. The CMS detector is modeled using GEANT4 v9.4 [23].

The background processes are modeled using POWHEG and MADGRAPH5_aMC@NLO [24] for matrix element (ME) calculations, while parton shower and fragmentation are modeled using PYTHIA 8. The processes for W boson and Drell–Yan (DY) production are produced using MADGRAPH5_aMC@NLO and PYTHIA using the MLM [25] matching scheme. The simulation of single top quarks using the NLO cross section from HATHOR v2.1 [26] is modeled through the t-channel production using MADSPIN [27] and tW production is modeled using POWHEG. Multi-boson events are modeled using the NLO cross sections given by MCFM [24] where ZZ and WZ events are simulated using PYTHIA and WW events are simulated using POWHEG.

The lepton selection efficiency and the trigger efficiency in MC are corrected using scale factors to better reproduce the Data. This includes μ and e ID, μ isolation, and e reconstruction. The statistical uncertainties are centrally provided. The recommendation for systematic uncertainties is an additional 1% uncertainty for the μ ID, 0.5% uncertainty for the μ isolation, 0.5% for the single μ triggers, and 1% for the e reconstruction (only for e $p_T < 20$ GeV and e $p_T > 80$ GeV).

Table 3: MC Samples

Process	Dataset	$\sigma(\text{pb})$
$t\bar{t}$	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	832
	/TTWjetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	/TTWjetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
	/TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
Di-Boson	/ZZ_TuneCUETP8M1_13TeV-pythia8	16.523
	/WWToLNuQQ_13TeV-powheg	49.997
	/WWToL2Nu_13TeV-powheg	12.178
	/WZ_TuneCUETP8M1_13TeV-pythia8	47.13
	/W1jetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9493.0
$W \rightarrow l\nu$	/W2jetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3120.0
	/W3jetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	942.3
	/W4jetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	524.2
	/DYjetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	16270.0
$Z \rightarrow ll$	/DYjetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4963.0
	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	44.33
	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	26.38
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85
	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85

3 Object Selection

A clean $t\bar{t}$ environment is essential to this analysis. Combinatorial background can be significantly reduced by requiring each event to have a certain number of leptons and jets. This analysis is split into 5 channels. Single muon with jets (μ + jets), single electron with jets (e + jets), double muon ($\mu\mu$), double electron (ee), and electron muon (e μ). A Kalman Vertex Filter [28] is used to reconstruct charmed mesons (see section: Charmed Mesons), and flag all jets containing at least one charmed meson. This allows for more relaxed jet criteria than recommended by the Top POG. The momentum scale and resolution of all isolated muons is also correct with the Rochester corrections [29] to remove any bias from detector misalignment or any error in the magnetic field. All particles are reconstructed using the Particle-flow (PF) algorithm [30] and jets are clustered using the anti- k_T algorithm [31, 32]. All PF tracks are limited to either

tracks used in the primary vertex (PV) fit, or the track that is the closest in the z-axis to the PV (PVTight ID). This helps remove pile-up (PU) tracks.

3.1 Isolated Muons (mu+jets)

- $p_T > 26 \text{ GeV}$
 - $|\eta| < 2.4$
 - PF-based Relative Isolation
- $$\frac{\Sigma p_T(\text{ch. had. from PV}) + \max[0, \Sigma E_T(\text{neutral hadrons}) + \Sigma E_T(\gamma) - 0.5 * \Sigma p_T(\text{ch. had. from PU})]}{p_T}$$

in a $\Delta R=0.4$ cone around the muon candidate must be less than 0.15

- At least 1 jet with $p_T > 30 \text{ GeV}$
- At least 1 more jet flagged by the Kalman filter

Passing the tight muon ID

- Is a Global Muon
- Is a PF Muon
- Global track normalized $\chi^2 < 10$
- At least 1 muon chamber hit
- Muon segments in at least 2 muon stations
- Tracker transverse impact parameter $d_{xy} < 2 \text{ mm}$ w.r.t the PV
- Tracker longitudinal impact parameter $d_z < 5 \text{ mm}$ w.r.t the PV
- At least 1 pixel hit
- At least 6 tracker layers

While vetoing on additional muons with

- $p_T > 15 \text{ GeV}$ (passing loose ID) (is a PF Muon and is a Global or Tracker Muon)
- $|\eta| < 2.4$
- Relative Isolation less than 0.24

3.2 Isolated Electrons (e+jets)

- $p_T > 35 \text{ GeV}$
- $|\eta| < 2.1$
- At least 1 jet with $p_T > 30 \text{ GeV}$
- At least 1 more jet flagged by the Kalman filter

Passing the tight electron ID (Table 4)

While vetoing on additional electrons with the Veto ID (Table 5)

3.3 Di-leptons (mumu, emu, and ee)

- Muon $p_T > 20 \text{ GeV}$ passing the tight muon ID
- Electron $p_T > 30 \text{ GeV}$ passing the tight electron ID

Table 4: Electron Tight ID

	$ \eta \leq 1.479$	$ \eta > 1.479$
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.00998	< 0.0292
$\Delta\eta_{\text{seed}}$	< 0.00308	< 0.00605
$\Delta\phi_{\text{in}}$	< 0.0816	< 0.0394
H/E	< 0.0414	< 0.0641
Rel. comb. PF iso	< 0.0588	< 0.0571
$ 1/E - 1/p $	< 0.0129	< 0.0129

At most 1 missing hit
Conv. veto

Table 5: Electron Veto ID

	$ \eta \leq 1.479$	$ \eta > 1.479$
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0115	< 0.037
$\Delta\eta_{\text{seed}}$	< 0.00749	< 0.00895
$\Delta\phi_{\text{in}}$	< 0.228	< 0.213
H/E	< 0.356	< 0.211
Rel. comb. PF iso	< 0.175	< 0.159
$ 1/E - 1/p $	< 0.299	< 0.15

At most 2 missing hit At most 3 missing hit
Conv. veto

- Muon $|\eta| < 2.4$
- Electron $|\eta| < 2.1$
- Relative Isolation less than 0.15
- At least 1 jet flagged by the Kalman filter
- Z-mass resonance suppression $|M_Z - M_{\text{ll}}| > 15 \text{ GeV}$ for same flavor di-leptons
- Drell-Yan suppression $E_{\text{T}}^{\text{miss}} > 40 \text{ GeV}$ for same flavor di-leptons
- QCD suppression $M_{\text{ll}} > 20 \text{ GeV}$

3.4 Jets

All jets must be defined by the anti- k_{T} algorithm [31, 32] with $p_{\text{T}} > 30 \text{ GeV}$ and $|\eta| < 2.4$. Jets must pass the Loose ID with

- Neutral Hadron Fraction < 0.99
- Neutral EM Fraction < 0.99
- Number of Constituents > 1
- Charged Hadron Fraction > 0
- Charged Multiplicity > 0
- Charged EM Fraction < 0.99

Jets are also cleaned so none overlap with the isolated leptons within a cone of $\Delta R = 0.4$.

Some control plots for the combined channels can be seen in Figure 2.

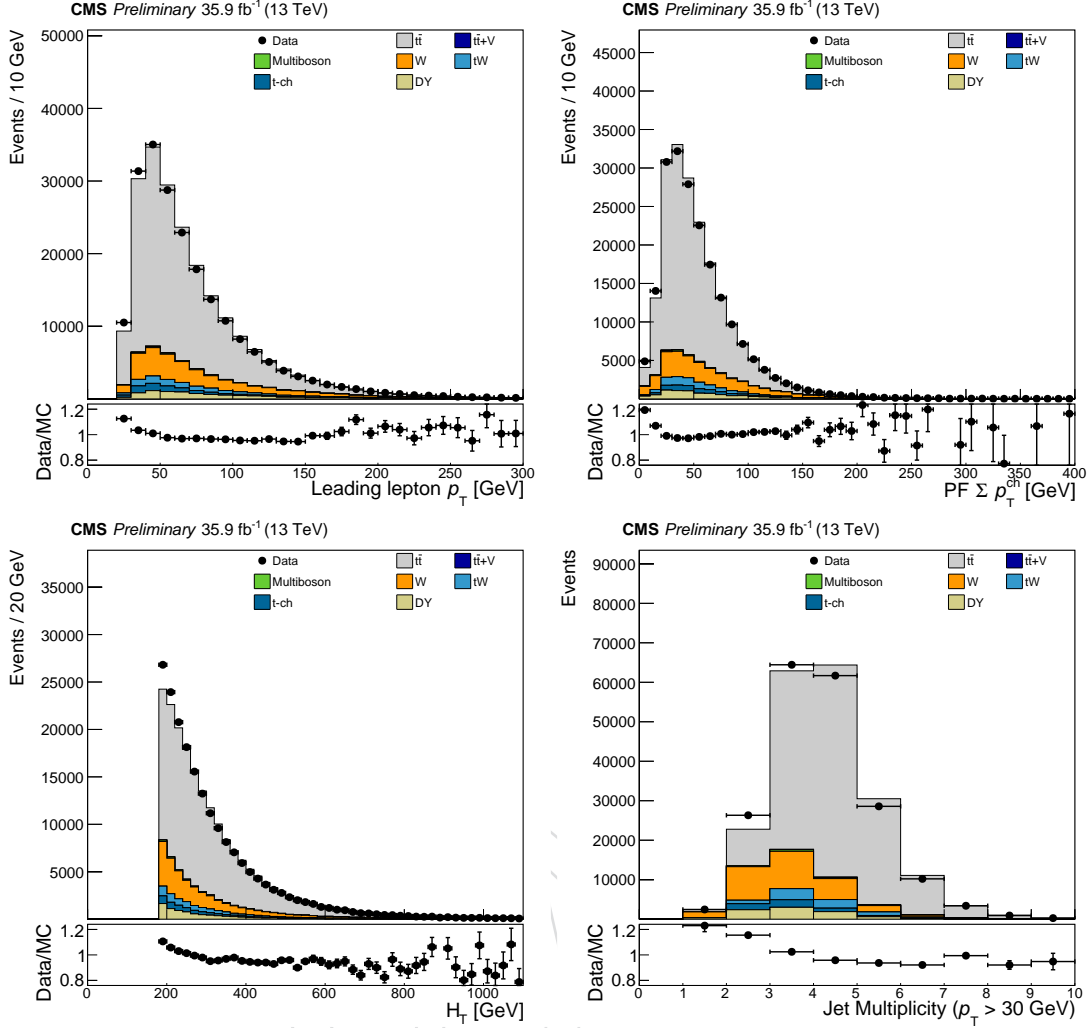


Figure 2: (D^0 selection) All channels isolated lepton p_T (top left) all channels jet $\sum p_T^{\text{ch}}$ (top right) all channels H_T (bottom left) and all channels jet multiplicity (bottom right)

4 Charmed Mesons

While the branching ratio for $b \rightarrow c$ may be small, charmed mesons produced in b quark fragmentation provide a relatively clean sample. A Kalman Vertex Filter [28] is used to fit the vertices of the J/ψ and D^0 candidates to help reduce the combinatorial background. This also allowed us to relax some constraints such as b -tagging because these mesons are ensured to be produced via b quark decay. All tracks are required to be high purity tracks, and all muons are required to be either global or tracker muon. The number of events for each channel is in Table 6. Additional criteria for the fitted vertices are as follows.

- Jet $p_T > 30 \text{ GeV}$
- Vertex fit $\chi^2 < 5$ to remove combinatorial background
- Meson $c\tau$ significance > 10 to remove prompt mesons (similar to [11])
- Meson $\sigma_{c\tau} > 2 \times 10^{-4}$ to remove residual prompt background and miss-reconstructed events

Table 6: Number of Events

Sample	J/ ψ Events	D ⁰ Events	D ⁰ _{μ} Events
Before Background Subtraction			
Data	2936	198859	6424
MC	2930	196151	6786
Data/MC (%)	100	101	95
After Background Subtraction			
Data	1960	11007	1201
MC	1932	10459	1376
Data/MC (%)	100	105	87

J/ ψ Reconstruction To help isolate J/ ψ decays all events must have $H_T > 80$ GeV. All soft muons used to reconstruct a J/ ψ must have $p_T > 3$ GeV as can be seen in Figure 3. High p_T isolated muon cannot be used in any events triggered with a muon (μ +jets, $\mu\mu$, and $e\mu$). The invariant mass of the J/ ψ candidates is in Figure 4 left. This plot shows how clean the J/ ψ channel is in spite of the low statistics. Any disagreement between Data and MC is due to statistical fluctuations in the background samples.

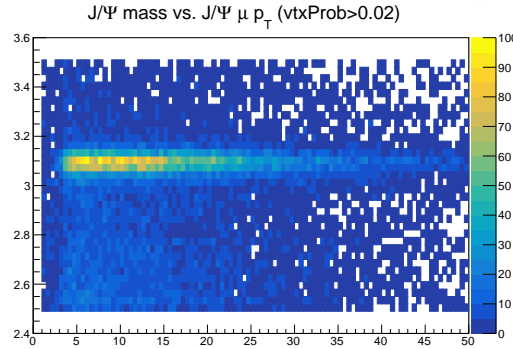
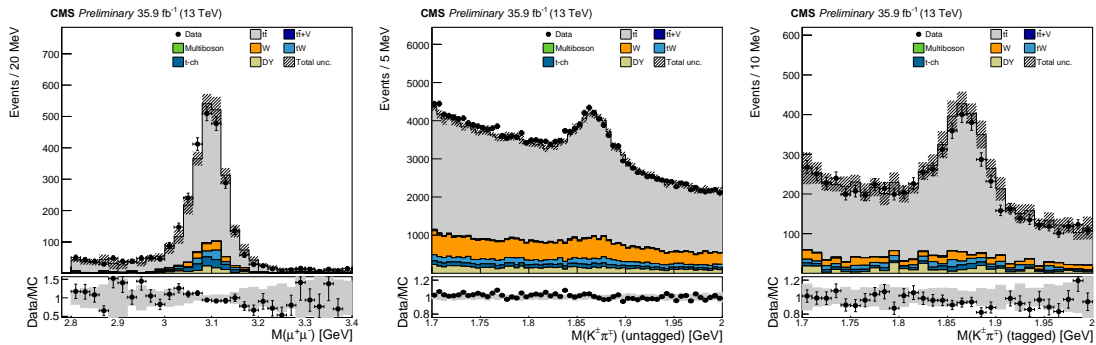
Figure 3: J/ ψ mass vs. μp_T 

Figure 4: J/ ψ mass (left), D⁰ mass (middle), and D⁰ _{μ} mass (right) where the bands are the quadrature sum of the MC statistical and systematic uncertainties

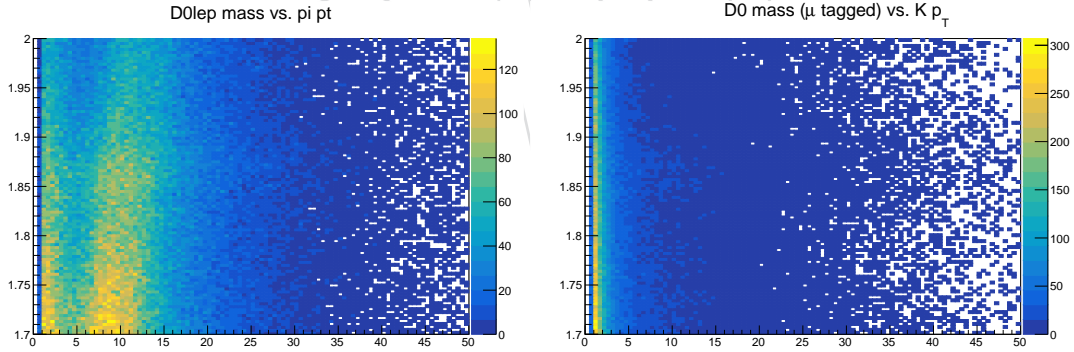
D⁰ Reconstruction To help isolate D⁰ decays all events must have $H_T > 180$ GeV. All π and K used to reconstruct a D⁰ must have $p_T > 5$ GeV and $p_T > 1$ GeV respectively as can be seen in Figure 5. Due to the high combinatorics in this sample a single mis modeled track

can create a large discrepancy between Data and MC. The π and K for the un-tagged D^0 must also have $|\eta| < 1.5$ as the large η particles were not modeled well in MC. The CMS detector cannot distinguish between a π and a K, so the mass hypothesis must be tested. This involves assigning one reconstructed particle as a π and the other as a K, and only saving the pair if their invariant mass is between 1.7 GeV and 2.0 GeV. Then the particle assignment is flipped and re-tested. Most events only contain one pair within this loose D^0 mass window. However, when both versions are valid, each copy is saved and given a weight of $\frac{1}{2}$ to preserve the total number of D^0 candidates. The invariant mass of the D^0 candidates is in Figure 4 middle.

Flavor-tagging In addition to reconstructing the D^0 a soft muon can also be required. This allows for flavor tagging, in which the charge of the parent meson is followed to ensure the correct mass assumptions for the π and K (see Table 7). The soft muon must have a minimum $p_T > 1$ GeV to reach the outer muon systems. The flavor tagged D^0 is denoted as D^0_μ . Flavor tagging significantly reduces the combinatorial background in D^0 events as can be seen in Figure 4 right. Note that this reduction allows the $\sum p_T^{\text{ch}}$ and decay product η cuts to be removed.

Table 7: Flavor tagging

Decay chain				Final Charged Products	
B^\pm	\rightarrow	D^0	\rightarrow	K^\pm	$K^\pm \pi^\mp + \mu^\pm$
u/\bar{u}	\rightarrow	u/\bar{u}	\rightarrow	u/\bar{u}	K^\pm
\bar{b}/b	\rightarrow	\bar{c}/c	\rightarrow	\bar{s}/s	π^\mp
	\downarrow		W^\mp	\rightarrow	μ^\pm
	W^\pm	\rightarrow	$\mu^\pm + \nu_\mu$		

Figure 5: πp_T (left) and K p_T (right)

5 Study of b quark Fragmentation Function

The b-fragmentation function is modeled in MC by a Lund–Bowler function [1]

$$\frac{1}{z^{1+r_b * b * m_b^2}} * (1-z)^a * \exp\left(-\frac{b * m_T^2}{z}\right) \quad (1)$$

where z is the fraction of momentum the particle receives from the quark, a and b are general fit parameters, and r_b and m_b are specific to the fragmenting b quark. By default $a = 0.68$,

171 $b = 0.98 \text{ GeV}^{-2}$, $r_b = 0.855$, $m_b = 4.78 \text{ GeV}$ (the b quark pole mass), and $m_T^2 = m_{\text{had}}^2 + p_{T,\text{had}}^2$
 172 [2]. In previous analyses the b-quark fragmentation function was fit by fully reconstructing
 173 B mesons using e^+e^- data at the Z pole ($Z^0 \rightarrow b\bar{b}$) [4–7] to obtain a value of $r_b = 0.8949^{+0.1841}_{-0.1968}$
 174 in PYTHIA 8. In order to measure the b-fragmentation in this analysis the charmed mesons are
 175 used as a proxy for the parent b quark. The ratio of the p_T the charmed meson divided by the
 176 sum of the p_T of all the charged tracks in the jet

$$\frac{\text{charged meson}}{\sum p_T^{\text{ch}}} \quad (2)$$

177 is used to measure the fragmentation of the charmed mesons. This is done by using the PF
 178 tracks described in Section 3, with an additional cuts that charged tracks have a $p_T > 0.5 \text{ GeV}$
 179 and an $|\eta| < 2.4$. Restricting the analysis to charged particles makes use of inner tracker and
 180 bypasses many of the issues with the jet energy scale (JES) and jet energy resolution (JER). It
 181 also provides increased statistics over fully reconstructing the B mesons. The samples split
 182 into two separate data epochs, B–F and GH, due to issues with the silicon tracker in 2016 (see
 183 Appendix A), and are recombined after background subtraction is performed (see Section 5).
 184 The flavor tagged D^0 combined with the muon used for tagging gives the closest kinematics to
 185 the b quark (Fig. 6 right).

$$\frac{p_T(D^0_\mu + \mu)}{\sum p_T^{\text{ch}}} \quad (3)$$

186 This is because only the ν_μ and soft tracks from the fragmentation are missing. The final ratio of
 187 Data over MC for all three samples can be found in Fig. 9 and 10. The un-tagged D^0 sample has
 188 much larger statistics, making it just as valuable as a proxy for the b-fragmentation function.
 189 The J/ψ sample has the lowest statistics, but the highest purity, making it an important sample
 190 as a cross check.

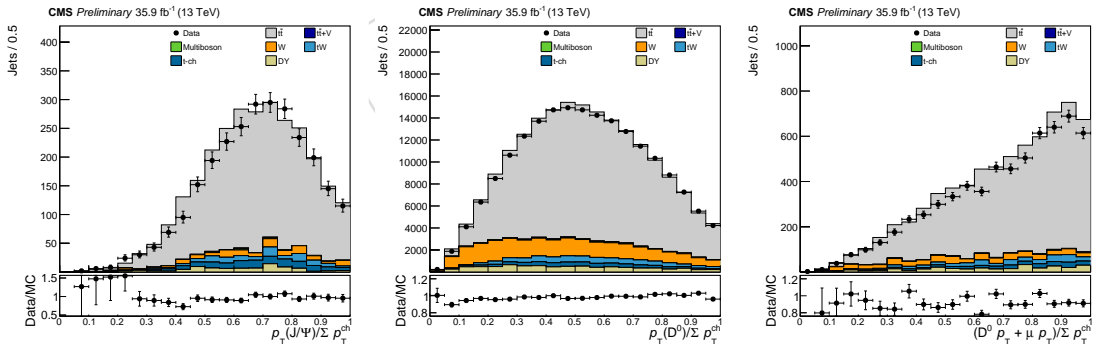


Figure 6: $J/\psi p_T$ over $\sum p_T^{\text{ch}}$, $D^0 p_T$ over $\sum p_T^{\text{ch}}$, and $(D^0_\mu p_T + \mu p_T)$ over $\sum p_T^{\text{ch}}$

191 **Background Subtraction** In order to get a clean sample, we must remove as much back-
 192 ground as possible. The Data and Monte Carlo are background subtracted using sPlot [33].
 193 This method uses the maximum likelihood fit of the invariant mass to produce event weights,
 194 which separate the signal from the background. The fit for the D^0 samples are obtained using a
 195 Gaussian for the peak, and an exponential for the background (Fig. 7 and Fig. 8). The J/ψ peak
 196 is fit with a Crystal Ball function for the peak, and an exponential for the background.

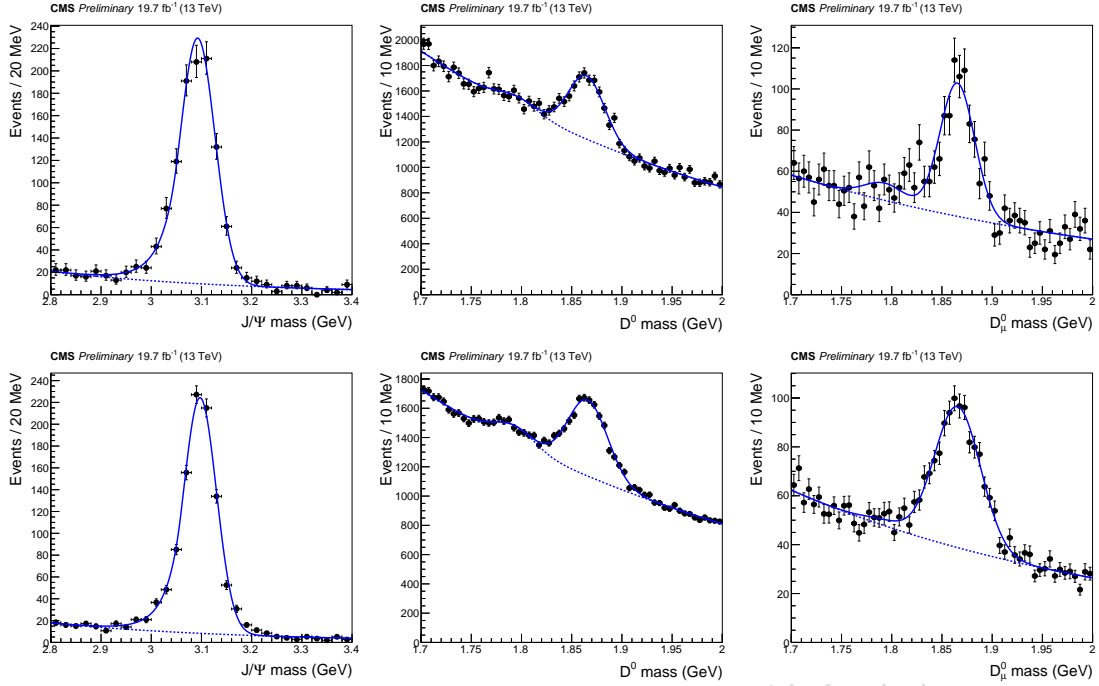


Figure 7: Mass fit for the J/ψ (left) D^0 (middle) and D^0_μ samples (right) for epochs B–F in the Data (top) and MC (bottom)

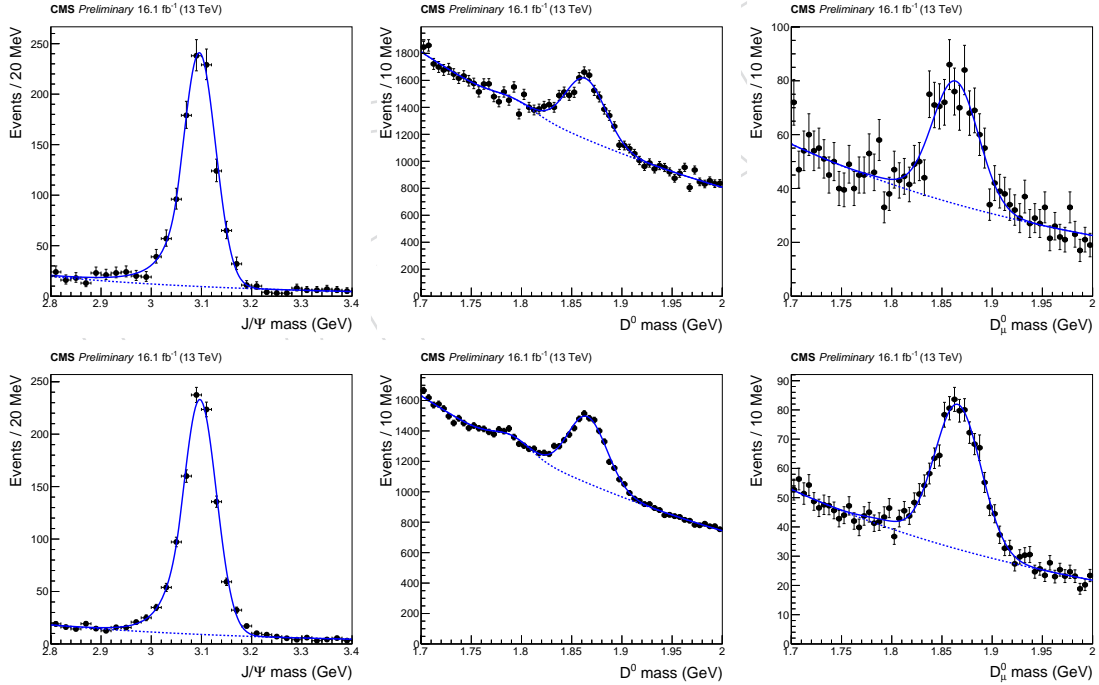


Figure 8: Mass fit for the J/ψ (left) D^0 (middle) and D^0_μ samples (right) for epochs GH in the Data (top) and MC (bottom)

198 The untagged D^0 sample contains a small contamination from the Cabibbo suppressed decay
 199 of $D^0 \rightarrow K^+ K^-$. In these decays, one K is misidentified as a π , resulting in a mis-reconstructed
 200 D^0 mass. This mis-reconstructed mass is clear in the MC (bottom of Figs. 7–8), and is fit

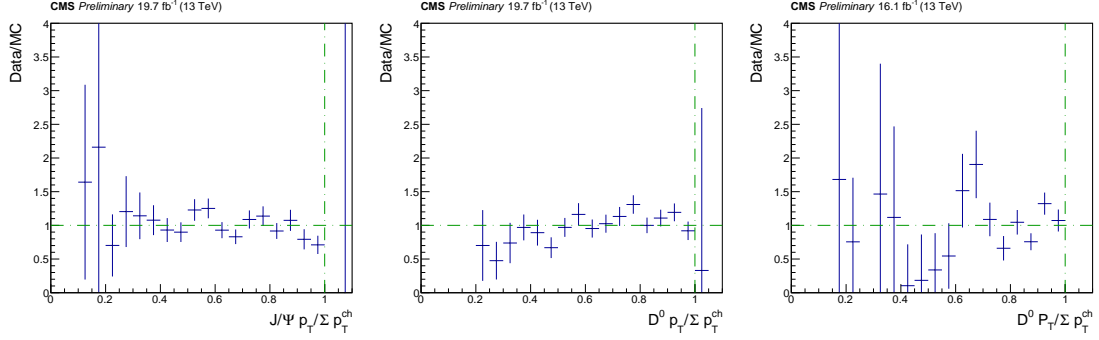


Figure 9: Ratio of $J/\psi p_T$ over Σp_T^{ch} (left), $D^0 p_T$ over Σp_T^{ch} (middle), and ratio of $(D^0 p_T + \mu p_T)$ over Σp_T^{ch} (right) for Data epochs B–F

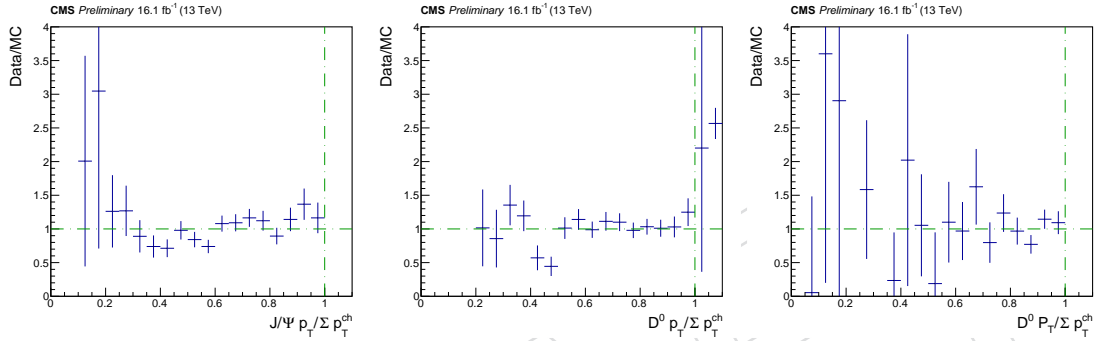


Figure 10: Ratio of $J/\psi p_T$ over Σp_T^{ch} (left), $D^0 p_T$ over Σp_T^{ch} (middle), and ratio of $(D^0 p_T + \mu p_T)$ over Σp_T^{ch} (right) for Data epochs GH

201 using a second Gaussian. The Gaussian is then applied to the data, with the fit starting at the
 202 values found by in the MC. This fit is then treated as background in both the data and MC,
 203 and is properly subtracted. To properly capture the Cabibbo suppressed decays in the D^0_{μ} ,
 204 the number of $D^0 \rightarrow K^+K^-$ was set constant, and scaled by the nubmer of D^0_{μ} events in the
 205 signal peak divided by the number of D^0 events in the signal peak. The response in the r_b
 206 measurment to this subtraction is negligible.

Fragmentation Templates Performing a fit of the fragmentation function requires templates generated for different tune parameters. MC samples are produced at generator level (GEN) for various values of r_b : 0.655, 0.700, 0.725, 0.755, 0.775, 0.800, 0.825, 0.875, 0.900, 0.925, 0.955, 0.975, 1.00, and 1.055. The jet ratio

$$x_B = \frac{B \text{ meson } p_T}{B \text{ jet } p_T} \quad (4)$$

207 for each tune is computed, and an event weight is produced by dividing this value by the
 208 nominal x_B (Fig. 11). These weights are then used to alter the shape of the MC at reconstruction
 209 (RECO) level. Some of the re-weighted proxies after background subtraction are in Fig. 12. The
 210 weights are ignored in the region of $x_B < 0.2$ as the sensitivity is very low, and in the region of
 211 $x_B > 0.975$ as these are solely due to resolution effects. If the sample is split by x_B , templates
 212 are generated for each x_B region, and are then combined before the final fit is performed.

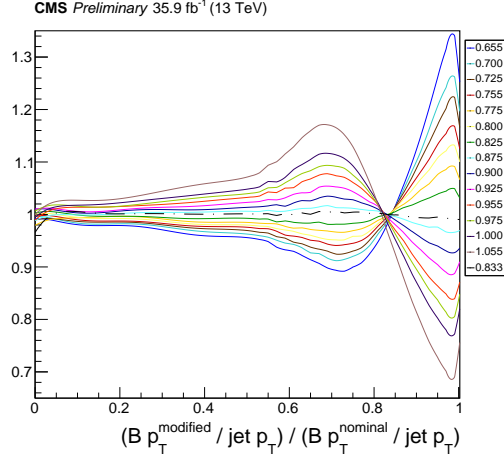


Figure 11: Event weights for various r_b values (combined fit value in dash-dotted line)

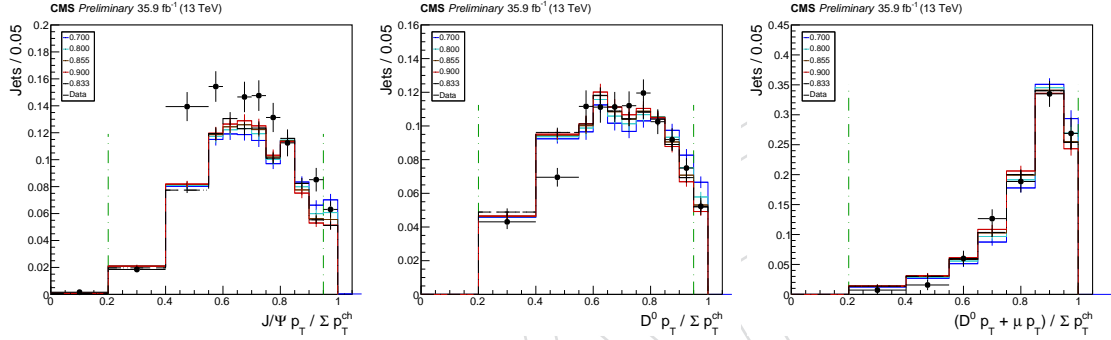


Figure 12: J/ψ , D^0 , and D^0_μ proxies for $r_b = 0.700$, $r_b = 0.800$, $r_b = 0.855$, $r_b = 0.900$, and the combined fit value (dash-dotted line) compared with Data

χ^2 Fit To find the best fit value of r_b a χ^2 goodness of fit is performed on each of the generated templates with respect to the data. The values of the χ^2 test are plotted versus the fragmentation parameter r_b used for the template. This plot is then fit with a 3rd order polynomial, and the minimum is extracted (Fig. 13). A 3rd order polynomial is needed due to the asymmetry of the χ^2 curves. The minima are $r_b = 0.883 \pm 0.049$ (stat), $r_b = 0.834 \pm 0.036$ (stat), and $r_b = 0.842 \pm 0.076$ (stat) for the J/ψ , D^0 , and D^0_μ samples respectively. The statistical error is the difference between the minimum r_b value and the r_b value at $\chi^2_{\min} + 1$ for each fit.

5.1 Kinematics cross check

The kinematic plots for the fitted r_b values are compared to the nominal samples for each charmed meson (Figures 14 – 16).

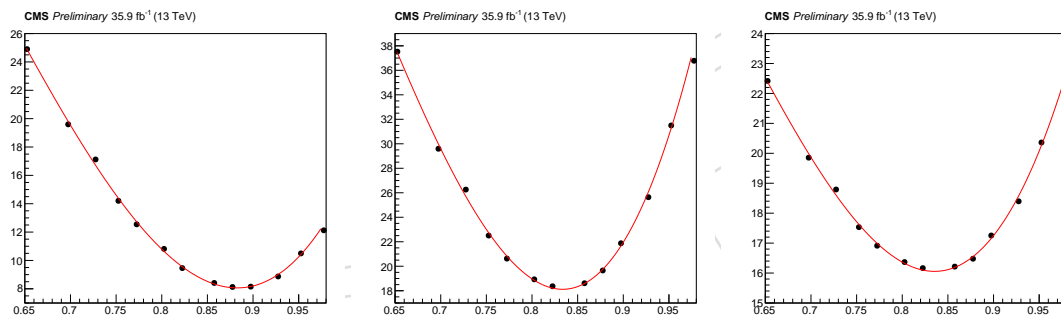
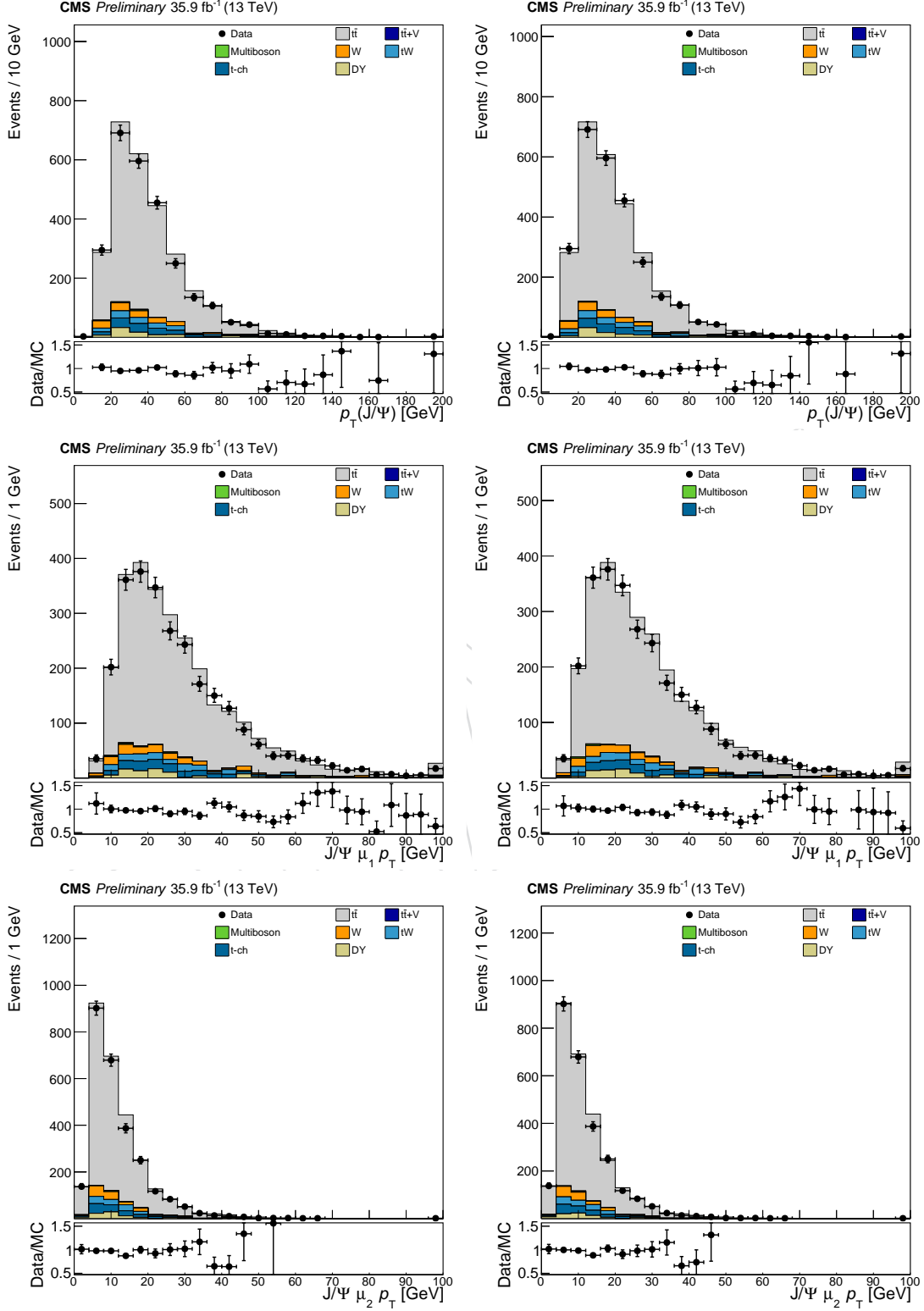
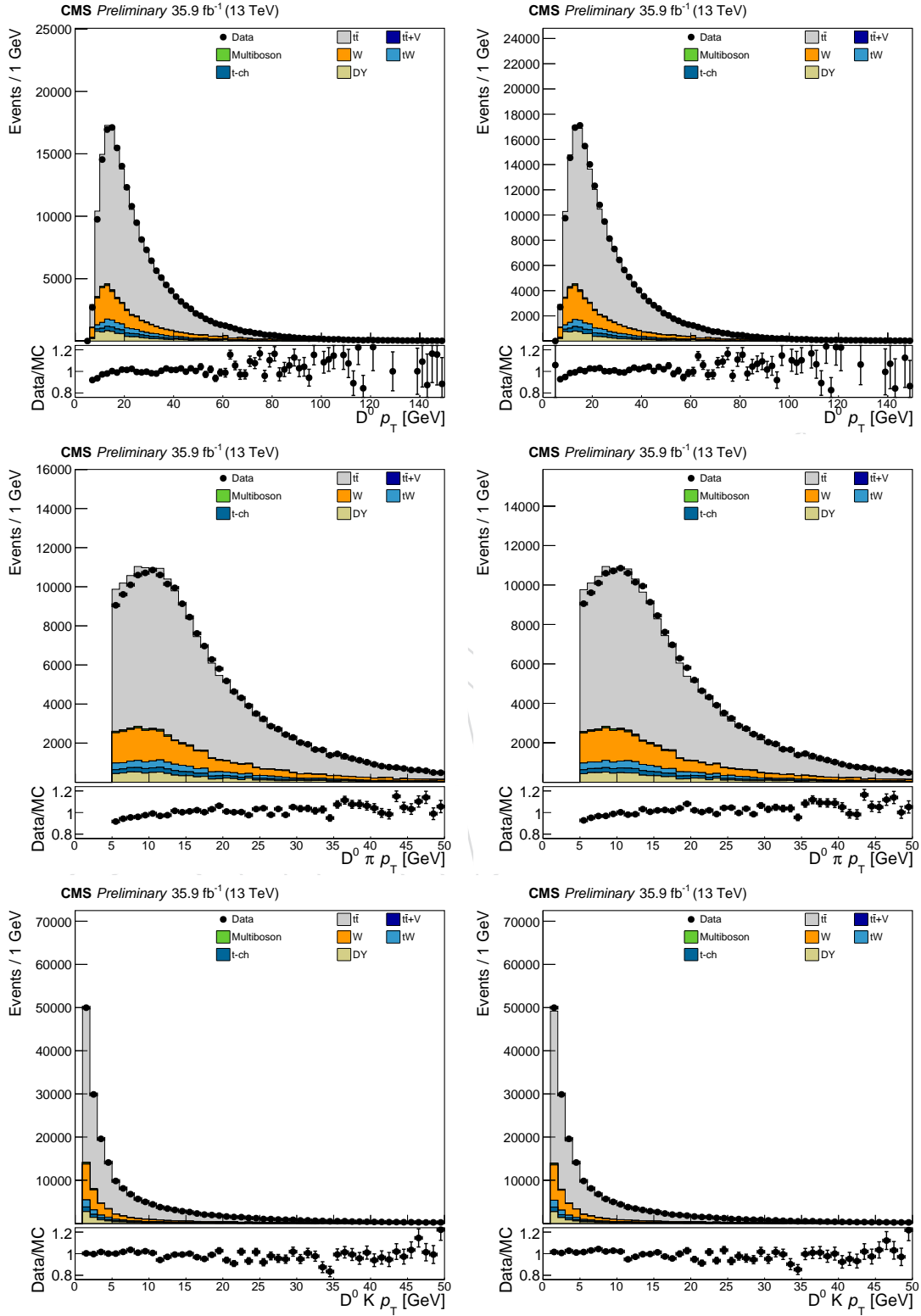
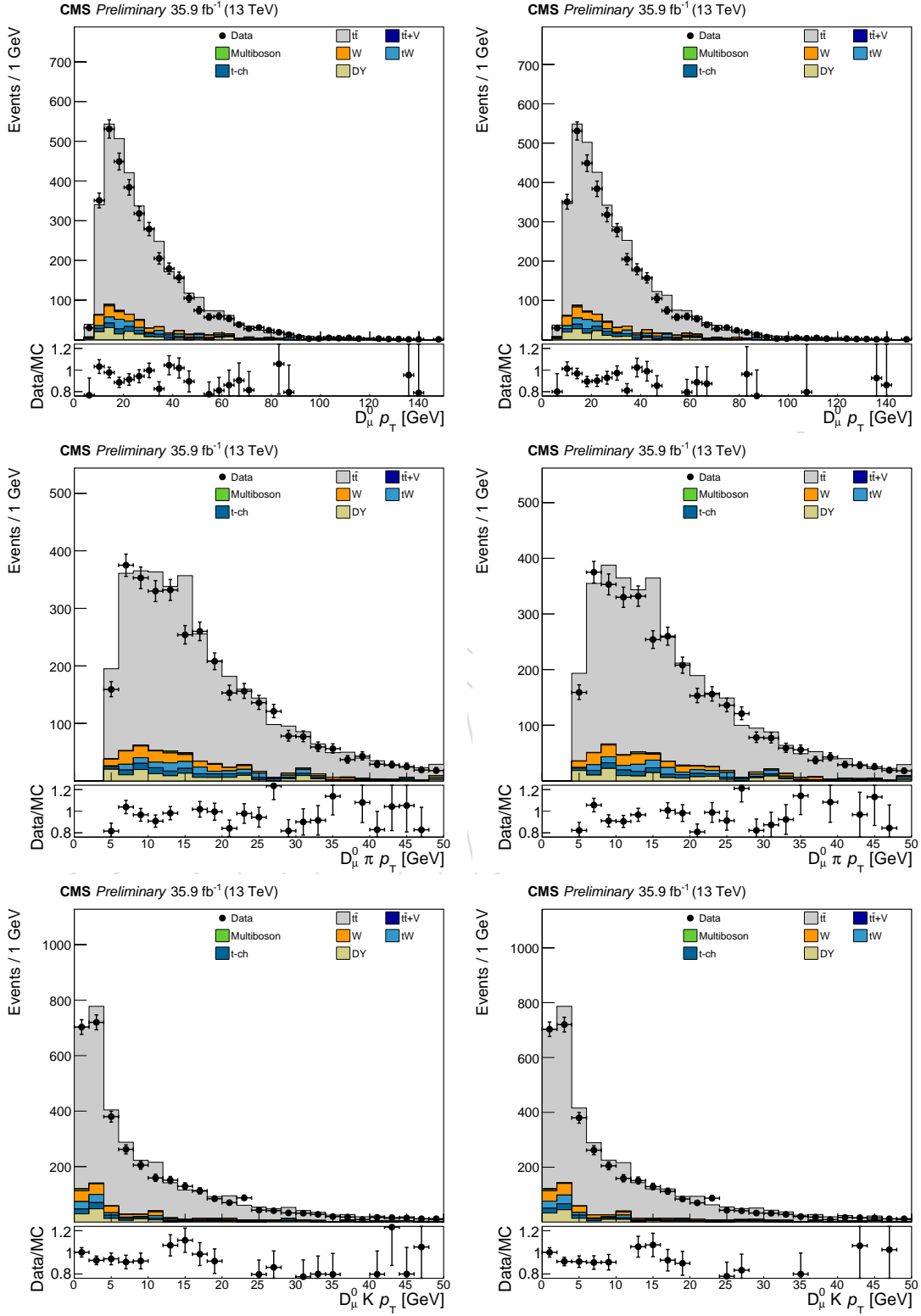


Figure 13: χ^2 goodness of fit vs. r_b for J/ψ (left), D⁰ (middle), and D⁰_μ (right)

Figure 14: J/ψ nominal (left) and fitted (right)

Figure 15: D^0 nominal (left) and fitted (right)

Figure 16: D_μ^0 nominal (left) and fitted (right)

6 Systematic uncertainties

Systematic uncertainties are evaluated by performing the χ^2 fit on each systematic sample, and subtracting the nominal fit value from the result. Statistical uncertainties on these measurements are extracted only when the systematic sample has independent statistics from the nominal, and are computed as the quadrature sum of the statistical uncertainty of the systematic fit ($\chi^2_{\min} + 1$) and the statistical uncertainty of the nominal fit. If the quadrature sum is larger than the systematic uncertainty, this number is used instead. A complete list of systematic uncertainties can be found in Table 8. Most of the systematic uncertainties are compatible with the statistical uncertainty of the measurements and are therefore symmetrized. A complete list of all the alternate samples used to evaluate the systematics can be found in Table 9.

Pseudo-experiments are performed on the nominal MC sample to cross-check the statistical uncertainty of the data. A random subset of the MC is picked corresponding to the total number of events in the data for each pseudo-experiment, and this subset is then treated as the toy data. The χ^2 fit procedure is then performed. The mean of the fit errors for the pseudo-experiments is extracted as the statistical uncertainty on the nominal MC sample.

Final state radiation The final state radiation (FSR) relies on a shape tune from LEP data for light-flavor quarks (u, d, s, c) [2]. The value of the renormalization scale (μ_R) is varied in PYTHIA by a factor of $\sqrt{2}$ up and $1/\sqrt{2}$ down to obtain the systematic uncertainties. It is worth noting that the parton-shower renormalization scale in PYTHIA 8 is defined as $\mu_R = z(1-z)Q^2$ [34] and is heavily dependent on previous fragmentation measurements for light (massless) quarks. The FSR systematic is evaluated using dedicated samples (see Table 9).

Matrix element parton shower matching The matching of ME from POWHEG to PS in PYTHIA is controlled by the model parameter h_{damp} [35] in PYTHIA 8. The ME/PS systematic uncertainty is obtained by varying the h_{damp} parameter up and down by 1σ .

Initial state radiation The initial state radiation (ISR), similar to FSR, relies on a shape tune from LEP data for light-flavor quarks (u, d, s, c) [2]. The value of the renormalization scale (μ_R) is varied in PYTHIA by a factor of 2 up and $1/2$ down to obtain the systematic uncertainties.

Color reconnection The color reconnection (CR) affecting the resonance decays is compared to alternative models to obtain the systematic uncertainties. There are three models available: early resonance decay [36] is enabled, a QCD inspired model [37] where the QCD color rules are also taken into account, and a Gluon move model [38] where gluons can be moved to another string. The model yielding the largest deviation in r_b from the nominal value is used as the systematic uncertainty for each sample.

Underlying event Remnants after initial parton scattering and multiple particle interactions (MPI) are known as the underlying events (UE). The UE is varied up and down by the uncertainties in PYTHIA [39].

Pile-up The inelastic proton-proton cross section (known as the minimum-bias cross section) is varied by $\pm 5\%$ to estimate the pile-up in the data [40]. The minimum-bias cross section is measured to be 69.2 mb.

Lepton selection efficiency Scale factors are used to correct the lepton selection efficiency in MC. This includes μ and e identification, μ isolation, and e reconstruction. The statistical uncertainties on these efficiencies are provided by the respective Physics Object Groups. The

recommendation for systematic uncertainties on these efficiencies is an additional 1% uncertainty for the μ identification, 0.5% uncertainty for the μ isolation, and 1% for the e reconstruction (only for e $p_T < 20$ GeV and e $p_T > 80$ GeV). The total uncertainties on these scale factors are used to shift the efficiencies up and down to obtain the systematic uncertainties.

Tracker efficiency Additional scale factors are derived for this analysis to account for issues seen in the silicon tracker during the 2016 data taking period (see sec:corr). The scale factors are used as a probability to drop reconstruction level particles in MC to help reproduce the Data. These scale factors are shifted up and down by 1σ to obtain the systematic uncertainties. The systematic uncertainty due to these scale factors is expected to be anti-correlated; shifting the scale factors up or down results in smaller or larger probability of losing tracks respectively.

W peak subtraction Figure 4 does show a small contribution to the D^0 mass peak from $b\bar{b} + W$ events. This peak is fit with a Gaussian using MC, then the fit parameters were frozen, and applied to the full MC and data. The number of events in the peak is also varied up and down by 25% to quantify the response in r_b . The D^0_μ sample is not sensitive to this peak due to the low statistics.

Top quark p_T re-weighting It was shown in [41, 42] that the p_T of the t quark in the Data was significantly softer than the MC. Scale factors are derived to correct this shape mismatch

$$SF(p_T) = e^{0.0615 - 0.0005 \cdot p_T} \quad (5)$$

$$w = \sqrt{SF(t) SF(\bar{t})} \quad (6)$$

where SF is the scale factor and w is an event weight. The uncertainty is calculated by taking the difference of the r_b fit value with the re-weighting turned on and off.

Fit procedure A large portion of the systematic shifts are well within the statistical uncertainty of their respective samples. An uncertainty on the fitting procedure has been adopted to encompass this effect. The uncertainty is obtained by performing pseudo-experiments (see MC statistics for details) only on the MC samples which are obtained from re-weighting the nominal sample (such as ME-PS and UE). These samples by definition contain the same number of events as the nominal MC, so any differences in the up and down variations are attributed to the error of the fit procedure. The sign of these shifts are dropped assuming the shift is a statistical effect. The up and down shifts are then averaged, resulting in the fit procedure uncertainty.

MC statistics The statistical limitations on the MC samples are tested using pseudo-experiments. The experiments involve treating the nominal r_b template as toy data, and performing the χ^2 fitting procedure as normal. For each pseudo-experiment, each bin in the toy data is shifted up or down by a random Gaussian number with a width corresponding the uncertainty on each bin. The results for r_b for all pseudo-experiment are fit with a Gaussian function, and the width of the Gaussian is extracted. This width is a convolution of the MC statistical uncertainty and the fit procedure uncertainty. The quadrature difference of the width and the fit procedure uncertainty is computed, and the result is taken as the MC statistical uncertainty.

Mass fit shape The shape of the PDFs used to model the signal and background before the sPlot method is performed is another possible source of uncertainty. The background shape for both D^0 samples is modified from an exponential PDF to the product of an exponential and a Gaussian (with $\sigma = 1$). The signal for the J/ψ is modified from the sum of two Gaussians to the sum of three Gaussians. In all cases the visual change is negligible. The fitting procedure is then repeated as normal, starting with generating the r_b templates and performing the χ^2 scan.

Top quark mass The current uncertainties on the top quark mass measurements due to b quark fragmentation are restricted by the r_b fit from e^+e^- data. As a result, the r_b response as a function of m_t should be checked. The sensitivity to the top quark mass is measured by varying m_t in PYTHIA 8 between 166.5 GeV and 178.5 GeV in steps of 1 GeV. The response in r_b is shown to be linear, and the shift from varying m_t by ± 0.5 GeV is interpolated. This corresponds to the current sensitivity of the top quark mass from CMS [43]. The response in r_b is essentially zero at the level of precision available in this analysis, and provides no additional uncertainty. This is because the measurement of r_b presented is compatible with the nominal value which entered into the m_t measurements. The previously measured uncertainties on r_b , which were used to measure systematic uncertainties for m_t , are much larger than the measurement presented.

Trigger Efficiency The efficiency of the triggers in MC are also corrected using scale factors. The uncertainties on these scale factors are used to shift the efficiencies up and down. The p_T of the isolated μ and e are selected beyond the trigger turn on curves. This results in the trigger efficiency only affecting the overall normalization, and not the shape of the fragmentation function. All distributions are normalized to unity before the r_b fit is performed, so the trigger efficiency provides no additional uncertainty.

Jet energy resolution The jet energy resolution (JER) can be corrected to better reproduce the data [44]. This involves scaling the resolution of MC particles matched to generator level particles, or a stochastic smearing otherwise. The systematic uncertainty is obtained by varying these corrections by their statistical and systematic uncertainties. Most of the uncertainty in the JER is due to reconstructing neutral particles in the ECAL and HCAL. As the analysis only uses charged tracks, it is not sensitive to these corrections. The systematic uncertainty due to shifting the JER was checked, and the response in r_b is indeed negligible.

Table 8: Sources of systematic uncertainty. The uncertainties quoted for the dominant FSR systematic correspond to the statistical uncertainty on the values.

Source	J/ψ	D^0	D^0_μ	Combined
Fit procedure	± 0.021	± 0.021	± 0.036	± 0.011
MC stat	± 0.014	± 0.018	± 0.027	± 0.008
Functional form	± 0.037	± 0.009	± 0.005	± 0.001
W subtraction	N/A	± 0.005	N/A	± 0.003
Total systematics	± 0.045	± 0.030	± 0.045	± 0.014
FSR up	-0.145 ± 0.078	-0.043 ± 0.059	-0.104 ± 0.113	-0.076 ± 0.043
FSR down	$+0.085 \pm 0.073$	$+0.032 \pm 0.051$	$+0.183 \pm 0.090$	0.092 ± 0.038

Table 9: MC Systematic Samples

UE up	/TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
UE down	/TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
	/TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
CR	/TT_TuneCUETP8M2T4_erdON_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_erdON_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
	/TT_TuneCUETP8M2T4_GluonMoveCRTune_erdON_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_QCDbasedCRTune_erdON_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_QCDbasedCRTune_erdON_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
h_{damp} up	/TT_HdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
h_{damp} down	/TT_HdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_HdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
FSR up	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
FSR down	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
ISR up	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
ISR down	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
	/TT_TuneCUETP8M2T4_13TeV-powheg-isrpythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-ext1-v1
$m(t) = 166.5 \text{ GeV}$	/TT_TuneCUETP8M2T4_mtop1665_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
$m(t) = 171.5 \text{ GeV}$	/TT_TuneCUETP8M2T4_mtop1715_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
$m(t) = 173.5 \text{ GeV}$	/TT_TuneCUETP8M2T4_mtop1735_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
$m(t) = 175.5 \text{ GeV}$	/TT_TuneCUETP8M2T4_mtop1755_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1
$m(t) = 178.5 \text{ GeV}$	/TT_TuneCUETP8M2T4_mtop1785_13TeV-powheg-pythia8/RuntISummer16MiniAODv2-PUMoriond17_80X_mcRun2.asymptotic.2016.TracheIV_v6-v1

330

7 Results

The final fits for the separate channels are

$$r_b = 0.883 \pm 0.049 (\text{stat}) \pm 0.045 (\text{syst})_{+0.085}^{-0.145} (\text{FSR}), \quad (7)$$

$$r_b = 0.834 \pm 0.036 (\text{stat}) \pm 0.030 (\text{syst})_{+0.032}^{-0.043} (\text{FSR}), \quad (8)$$

and

$$r_b = 0.842 \pm 0.076 (\text{stat}) \pm 0.045 (\text{syst})_{+0.183}^{-0.104} (\text{FSR}) \quad (9)$$

for the J/ψ , D^0 , and D^0_μ samples respectively. The comparison with the data can be seen in Figure 17. A combined fit is performed by fitting each sample and summing the χ^2 values for each value of r_b . The χ^2 scan is then fit using the same procedure as the separate channels, producing a final fit of

$$r_b = 0.855 \pm 0.030 (\text{stat}) \pm 0.014 (\text{syst})_{+0.092}^{-0.076} (\text{FSR}). \quad (10)$$

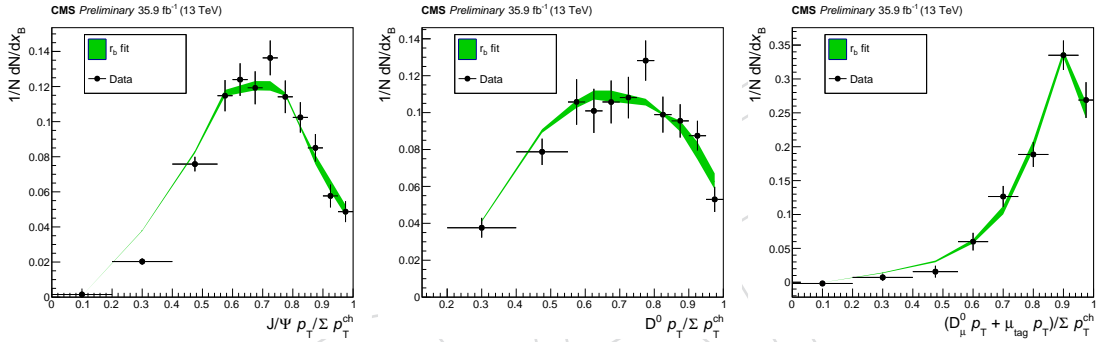


Figure 17: $J/\psi p_T$ over Σp_T^{ch} , $D^0 p_T$ over Σp_T^{ch} , and $(D^0_\mu p_T + \mu p_T)$ over Σp_T^{ch} (bands are the total fit uncertainties for the combined fit)

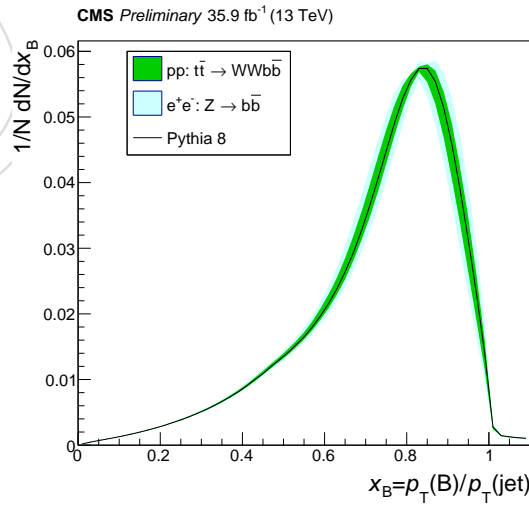


Figure 18: Fragmentation function (bands are the quadrature sum of the MC statistical and total fit uncertainties)

This analysis gives about a factor of three better precision than the analysis using e^+e^- Z pole data [3] ($r_b = 0.894^{+0.184}_{-0.197}$) while no disagreement between the measurements is observed. A comparison between the GEN level fragmentation function for this analysis, the e^+e^- tune, and the default PYTHIA tune is in Figure 18. The default PYTHIA 8 tune, $r_b = 0.855$ based on the e^+e^- data [2], is in good agreement with this analysis. There is no experimental evidence for an environmental dependence.

8 Conclusion

A fit for the b-quark fragmentation function has been performed in a $t\bar{t}$ environment using the 2016 dataset consisting of 35.9 fb^{-1} in a $t\bar{t}$ environment. This fit consists of three separate channels: J/ψ , D^0 , and D^0_μ for semi-leptonic (μ +jets and e+jets) and fully leptonic ($\mu\mu$, μe , and ee) events. The fit for the separate channels are

$$r_b = 0.883 \pm 0.049 (\text{stat}) \pm 0.045 (\text{syst})^{+0.145}_{-0.085} (\text{FSR}), \quad (11)$$

$$r_b = 0.834 \pm 0.036 (\text{stat}) \pm 0.030 (\text{syst})^{+0.043}_{-0.032} (\text{FSR}), \quad (12)$$

and

$$r_b = 0.842 \pm 0.076 (\text{stat}) \pm 0.045 (\text{syst})^{+0.104}_{-0.183} (\text{FSR}) \quad (13)$$

for the J/ψ , D^0 , and D^0_μ samples respectively. A combined fit is performed producing a fit for the parameter

$$r_b = 0.855 \pm 0.030 (\text{stat}) \pm 0.014 (\text{syst})^{+0.076}_{-0.092} (\text{FSR}) \quad (14)$$

which is in consistent with the nominal value $r_b = 0.855$. This improves the statistical significance of the fit by roughly a factor of three over the e^+e^- Z pole measurements. These measurements are consistent, so no environmental dependence is observed.

References

- [1] M. G. Bowler, “ e^+e^- Production of Heavy Quarks in the String Model”, *Z. Phys.* **C11** (1981) 169, doi:10.1007/BF01574001.
- [2] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J.* **C74** (2014), no. 8, 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.
- [3] M. Seidel, “Precise measurement of the top-quark mass at the CMS experiment using the ideogram method”. PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2015. doi:10.3204/DESY-THESIS-2015-034.
- [4] DELPHI Collaboration, “A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole”, *Eur. Phys. J.* **C71** (2011) 1557, doi:10.1140/epjc/s10052-011-1557-x, arXiv:1102.4748.
- [5] ALEPH Collaboration, “Study of the fragmentation of b quarks into B mesons at the Z peak”, *Phys. Lett.* **B512** (2001) 30–48, doi:10.1016/S0370-2693(01)00690-6, arXiv:hep-ex/0106051.
- [6] OPAL Collaboration, “Inclusive analysis of the b quark fragmentation function in Z decays at LEP”, *Eur. Phys. J.* **C29** (2003) 463–478, doi:10.1140/epjc/s2003-01229-x, arXiv:hep-ex/0210031.

- [7] SLD Collaboration, “Measurement of the b quark fragmentation function in Z^0 decays”, *Phys. Rev. D* **65** (2002) 092006, doi:10.1103/PhysRevD.66.079905, 10.1103/PhysRevD.65.092006, arXiv:hep-ex/0202031. [Erratum: *Phys. Rev. D* 66, 079905 (2002)].
- [8] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [9] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [10] CMS Collaboration, “Measurement of the top quark mass using charged particles in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. D* **93** (2016), no. 9, 092006, doi:10.1103/PhysRevD.93.092006, arXiv:1603.06536.
- [11] CMS Collaboration, “Measurement of the mass of the top quark in decays with a J/ψ meson in pp collisions at 8 TeV”, *JHEP* **12** (2016) 123, doi:10.1007/JHEP12(2016)123, arXiv:1608.03560.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] S. Alioli, S.-O. Moch, and P. Uwer, “Hadronic top-quark pair-production with one jet and parton showering”, *JHEP* **01** (2012) 137, doi:10.1007/JHEP01(2012)137, arXiv:1110.5251.
- [16] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [17] M. Czakon, P. Fielder, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$ ”, *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.
- [18] 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.
- [19] 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.
- [20] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

- [21] CMS Collaboration Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV”, Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.
- [22] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching matrix elements and shower evolution for top-quark production in hadronic collisions”, *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.
- [23] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [24] 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.
- [25] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5.
- [26] P. Kant et al., “Hathor for single top quark production: updated predictions and uncertainty estimates for single top quark production in hadron collisions”, *Comp. Phys. Comm.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001.
- [27] R. Artoisenet, R. Frederix, O. Mattelas, and R. Rietkerk, “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”, *JHEP* **03** (2013) 015, doi:10.1007/JHEP03(2013)015.
- [28] T. Speer et al., “Vertex Fitting with the Kalman Filter Formalism in the ORCA Reconstruction Program”,.
- [29] A. Bodek et al., “Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments”, *Eur. Phys. J.* **C72** (2012) 2194, doi:10.1140/epjc/s10052-012-2194-8, arXiv:1208.3710.
- [30] CMS Collaboration, “Particle-flow reconstruction and global event description with the cms detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [31] 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.
- [32] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [33] M. Pivk and F. R. L. Diberder, “sPlot: a statistical tool to unfold data distributions”, doi:10.1016/j.nima.2005.08.106, arXiv:arXiv:physics/0402083.
- [34] T. Sjostrand and P. Z. Skands, “Transverse-momentum-ordered showers and interleaved multiple interactions”, *Eur. Phys. J.* **C39** (2005) 129–154, doi:10.1140/epjc/s2004-02084-y, arXiv:hep-ph/0408302.
- [35] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV”,.

- [36] T. Sjostrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions", *Phys. Rev. D* **36** (1987) 2019, doi:10.1103/PhysRevD.36.2019.
- [37] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP* **08** (2015) 003, doi:10.1007/JHEP08(2015)003, arXiv:1505.01681.
- [38] J. R. Christiansen and T. Sjostrand, "Color reconnection at future e^+e^- colliders", *Eur. Phys. J. C* **75** (2015), no. 9, 441, doi:10.1140/epjc/s10052-015-3674-4, arXiv:1506.09085.
- [39] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements",.
- [40] ATLAS Collaboration, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.* **117** (2016), no. 18, 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.
- [41] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Phys. Rev. D* **95** (2017), no. 9, 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.
- [42] CMS Collaboration, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons", *JHEP* **02** (2019) 149, doi:10.1007/JHEP02(2019)149, arXiv:1811.06625.
- [43] CMS Collaboration Collaboration, "Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 tev", *Phys. Rev. D* **93** (Apr, 2016) 072004, doi:10.1103/PhysRevD.93.072004.
- [44] 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.
- [45] CMS Collaboration, "Operation and Performance of the CMS outer tracker", *PoS Vertex* **2017** (2018) 013, doi:10.22323/1.309.0013.

A Corrections for the problems in the 2016 Data

During first half of the 2016 data taking period the CMS silicon tracker had known issues with the APV configuration [45]. The instantaneous luminosity from the LHC was steadily increased over this period; and the APV25 readout per-amplifier was saturated due to the amount of charge deposited by tracks in the silicon tracker. This was corrected halfway through 2016 by changing the drain speed, and we see that the second half of the Data is unaffected by this problem.

From comparison of the unaffected Data and the MC simulation a normalization correction factor of $11\% \pm 0.42\%$ was derived as shown in Fig. 20.. This includes the multiplicity mis-modeling inside jets using PYTHIA 8. The correction is needed for both halves for the Data. To correct for the kinematic dependent effects in the early Data, a data-driven method is used. After the sub-samples of the Data are normalized to their respective luminosity, the ratio of the Data for part 1 divided by part 2 is compared on an (η, p_T) grid. The results are shown in Table 10. These values are used as a probability to drop particles from the simulation.

Table 10

	η					
	-1.5	-0.8	-0.4	0	0.4	0.8
300	0.90 ± 0.19	1.01 ± 0.20	1.11 ± 0.24	0.99 ± 0.21	1.07 ± 0.23	1.41 ± 0.30
100	1.00 ± 0.06	0.98 ± 0.06	1.08 ± 0.07	0.97 ± 0.06	1.07 ± 0.07	1.02 ± 0.06
50	0.96 ± 0.03	0.91 ± 0.03	0.96 ± 0.03	0.94 ± 0.03	0.92 ± 0.03	1.03 ± 0.03
30	0.92 ± 0.02	0.92 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.93 ± 0.03	0.98 ± 0.03
20	0.94 ± 0.02	0.87 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.91 ± 0.02	0.95 ± 0.02
10	0.91 ± 0.02	0.82 ± 0.02	0.86 ± 0.02	0.85 ± 0.02	0.86 ± 0.02	0.93 ± 0.02
6	0.88 ± 0.03	0.83 ± 0.03	0.87 ± 0.03	0.85 ± 0.03	0.81 ± 0.03	0.98 ± 0.03
4	0.88 ± 0.02	0.81 ± 0.02	0.79 ± 0.02	0.83 ± 0.02	0.83 ± 0.02	0.91 ± 0.02
2	0.87 ± 0.02	0.84 ± 0.02	0.79 ± 0.02	0.86 ± 0.02	0.81 ± 0.02	0.90 ± 0.02

It was observed that the $\sum p_T^{\text{ch}}$ comparison between Data from both run periods show a p_T and η dependence beyond the 11% normalization correction and that the results in Table 10 are not quite sufficient to correct for this. A similar data to data comparison was performed for this quantity, resulting in the ratios observed in Table 11. After correcting the MC using these ratios, following the same method as above, good agreement between the Data and simulation is observed (see Fig. 21). We believe that the differences between the corrections the result of a bias in the track selection coming from the Kalman filter.

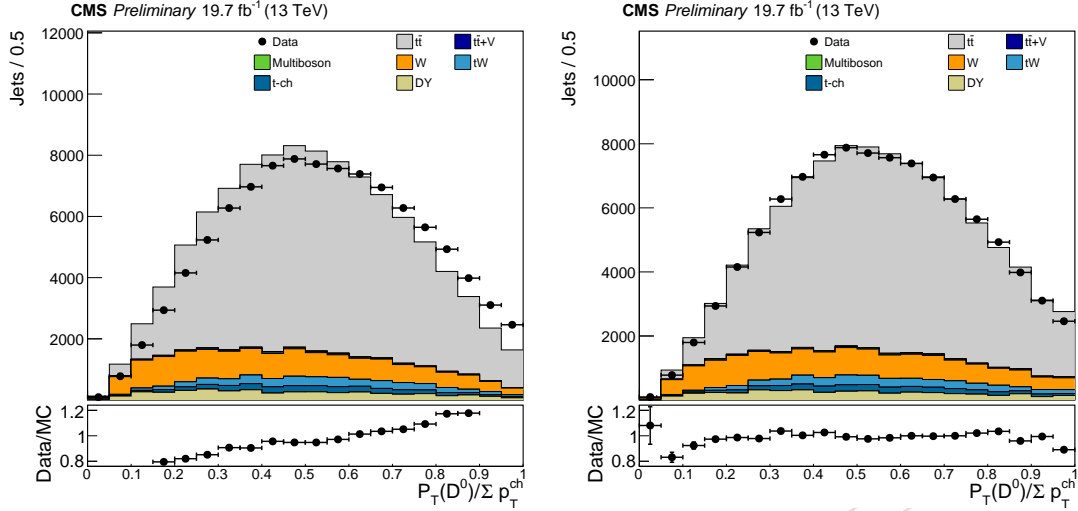


Figure 19: Affects of the tracker corrections (uncorrected left and corrected right) on the fragmentation proxies $D^0 p_T$ over $\sum p_T^{ch}$ for Data epochs B–F

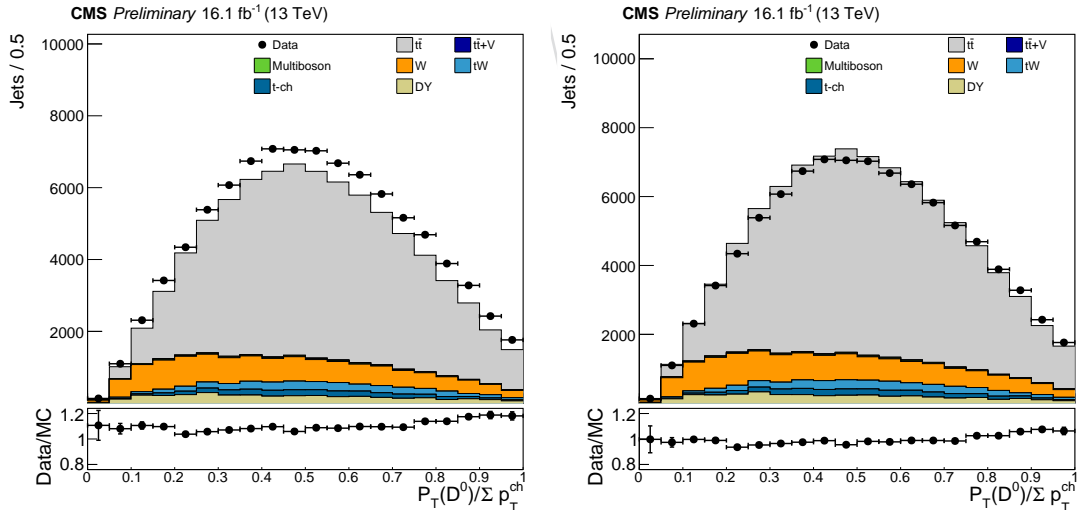


Figure 20: Affects of the normalization correction (uncorrected left and corrected right) on the fragmentation proxies $D^0 p_T$ over $\sum p_T^{ch}$ for Data epochs GH

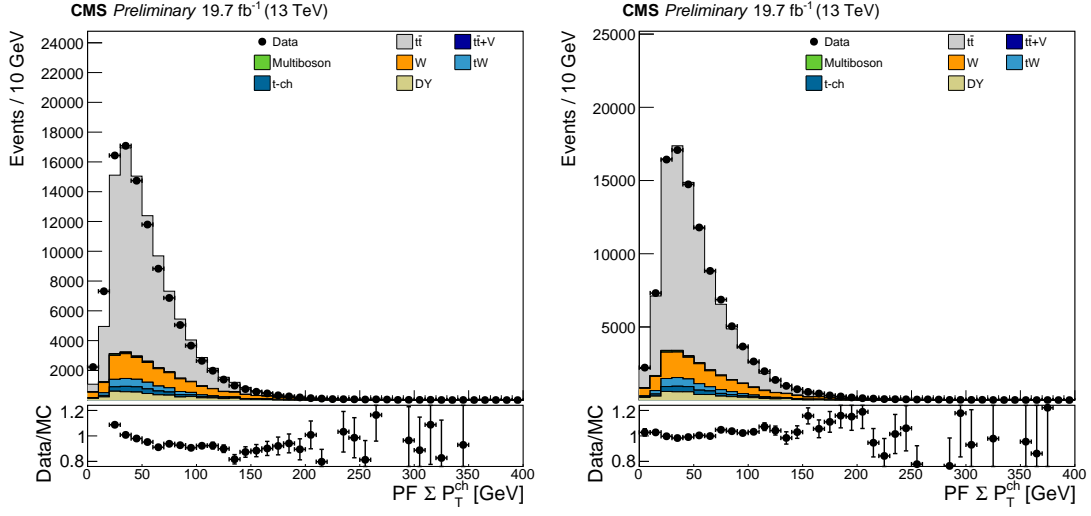


Figure 21: Affects of the tracker corrections (uncorrected left and corrected right) on the Σp_T^{ch} for the first half of the Data

Table 11

p_T (GeV)	-2.4	-1.5	-0.8	-0.4	0	0.4	0.8	1.5
300	1.17 ± 0.11	1.10 ± 0.08	0.76 ± 0.05	0.89 ± 0.05	0.80 ± 0.05	0.72 ± 0.05	1.03 ± 0.06	0.88 ± 0.04
100	0.88 ± 0.03	0.94 ± 0.02	0.93 ± 0.02	0.91 ± 0.02	0.86 ± 0.02	0.92 ± 0.02	0.96 ± 0.02	0.94 ± 0.02
50	0.87 ± 0.02	0.87 ± 0.01	0.91 ± 0.01	0.86 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.94 ± 0.01	0.94 ± 0.01
30	0.91 ± 0.02	0.89 ± 0.01	0.89 ± 0.01	0.86 ± 0.01	0.88 ± 0.01	0.88 ± 0.01	0.93 ± 0.01	0.85 ± 0.02
20	0.87 ± 0.01	0.91 ± 0.01	0.84 ± 0.01	0.83 ± 0.01	0.85 ± 0.01	0.87 ± 0.01	0.90 ± 0.01	0.90 ± 0.05
10	0.86 ± 0.01	0.87 ± 0.01	0.82 ± 0.01	0.80 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.02
6	0.85 ± 0.01	0.85 ± 0.01	0.83 ± 0.01	0.80 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.92 ± 0.01	0.84 ± 0.01
4	0.82 ± 0.01	0.86 ± 0.01	0.81 ± 0.01	0.79 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.90 ± 0.01	0.84 ± 0.02
2	0.86 ± 0.01	0.91 ± 0.01	0.87 ± 0.01	0.82 ± 0.01	0.93 ± 0.01	0.88 ± 0.01	0.96 ± 0.01	0.92 ± 0.06

B Cross-check of the result

B.1 Split epochs

Due to the differences in tracker efficiency between the B-F and GH data epochs, the r_b measurement should be checked in both epochs separately. The D^0 sample is the only one with enough statistics to allow for splitting between epochs. The proxies from the data compared with the nominal MC ($r_b = 0.855$) are in Figure 22. A statistical fluctuation around $x_B = 0.4$ in epoch GH causes a mismatch between the data and MC. The χ^2 scans for epochs B-F and GH are in Figure 23. The final values for the fragmentation shape parameter are $r_b = 0.852 \pm 0.056$ and $r_b = 0.807 \pm 0.077$ for the B-F and GH epochs respectively, and are consistent with both each other and the final result.

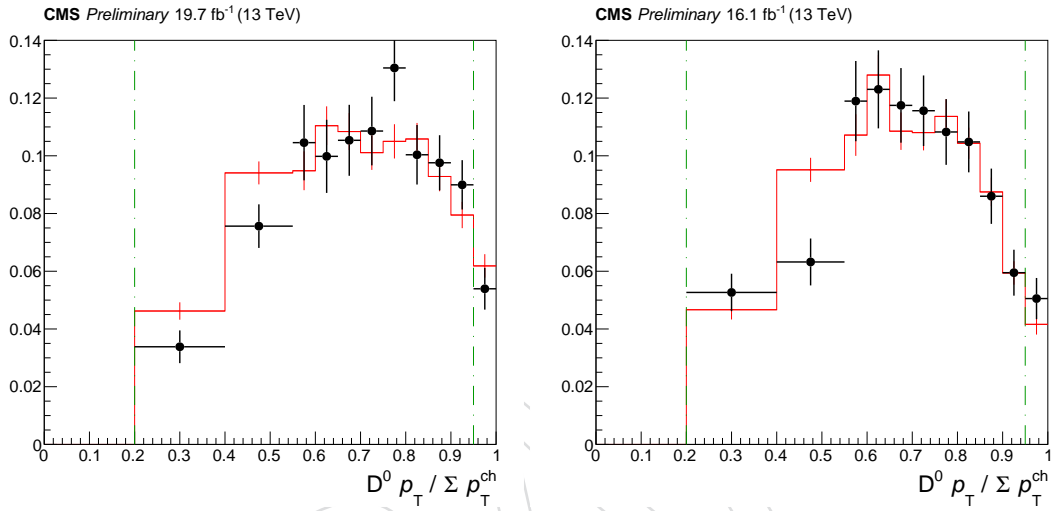


Figure 22: $D^0 p_T$ over $\sum p_T^{ch}$ for epochs B-F (left) and GH (right)

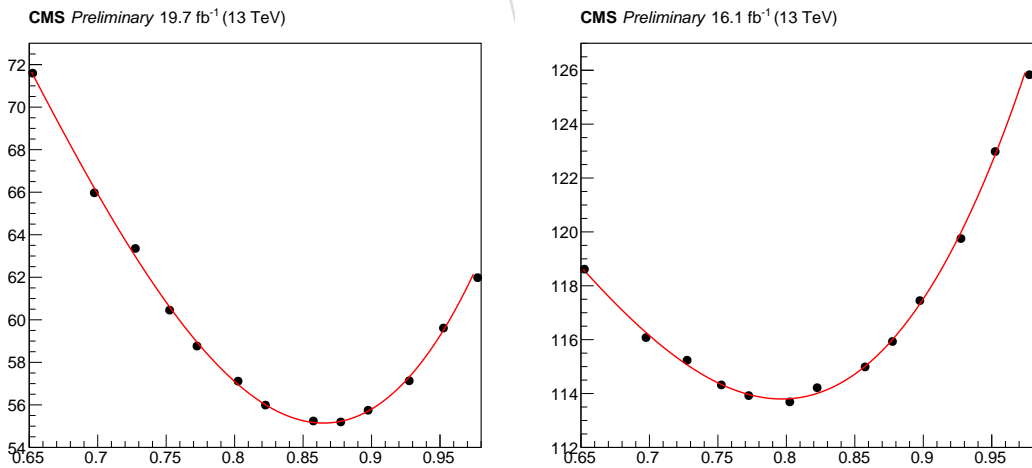


Figure 23: $D^0 \chi^2$ scans for epochs B-F (left) and GH (right)

B.2 Mass dependence on x_B

499 10.95

500 The D^0 invariant mass were checked for a dependence on the x_B binning. Figures 24–25 show
 501 the mean and width of the invariant mass has no noticeable dependence on x_B within the preci-
 502 sion available.

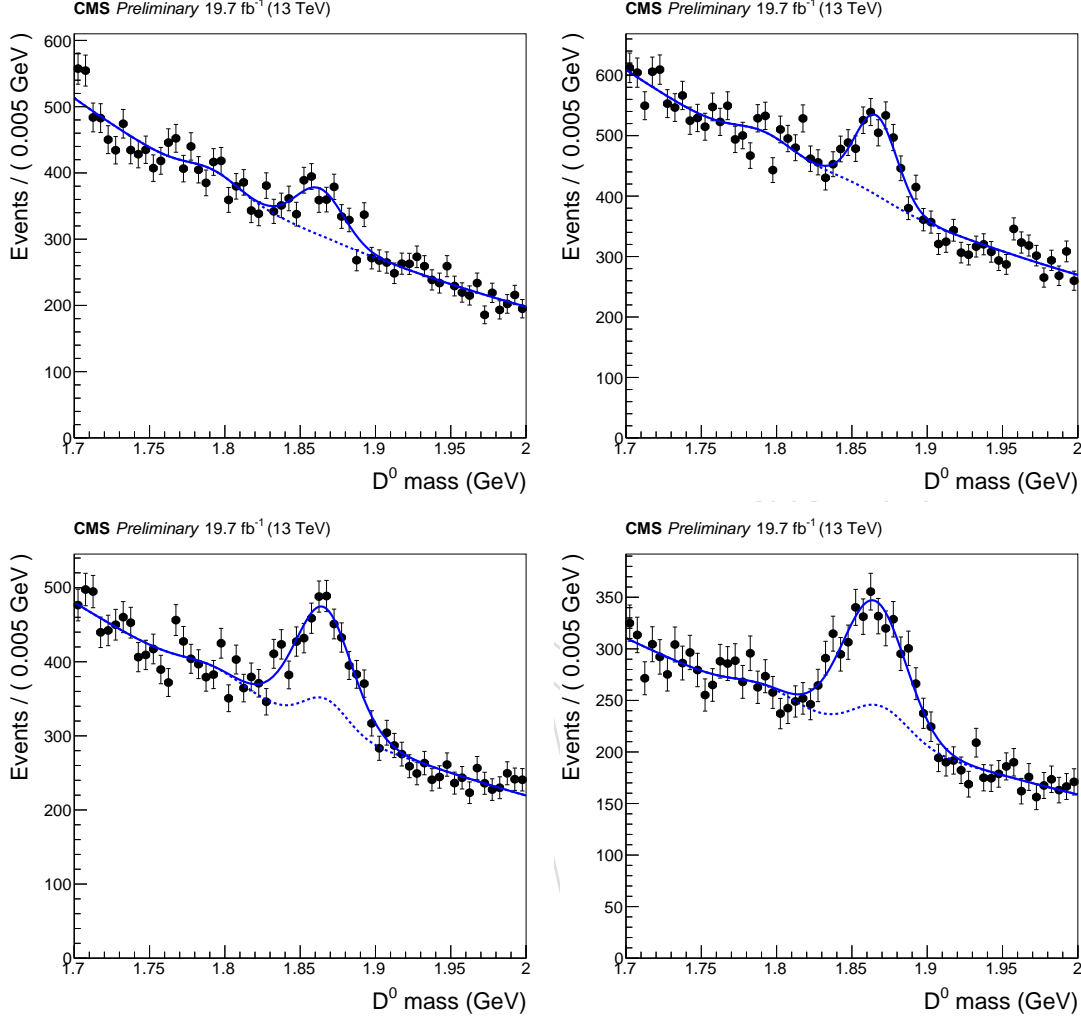


Figure 24: D^0 invariant mass for $x_B < 0.4$, $0.4 < x_B < 0.6$, $0.6 < x_B < 0.8$, and $x_B > 0.8$ for epochs B–F.

503 B.3 Full jet p_T

The result for the un-tagged D^0 fit was cross-checked using the full jet p_T . This turns the fragmentation proxy into

$$x_B = \frac{D^0 p_T}{\text{jet } p_T}. \quad (15)$$

504 The proxy for the D^0 sample is in Figure 26. The proxy peaks at a much lower value of x_B which
 505 this analysis is less sensitive to (see Fig. 11). As a result, the statistical and systematic uncer-
 506 tainties increase. By including the neutral particles in the jet from the ECAL and HCAL, the
 507 JSF and JER corrections must be included. A complete list of the full jet systematics are listed
 508 in Figure 13. The final fit for the D^0 sample is $r_b = 0.865 \pm 0.089$ (stat) ± 0.141 (syst) $^{+0.265}_{-0.005}$ (FSR).

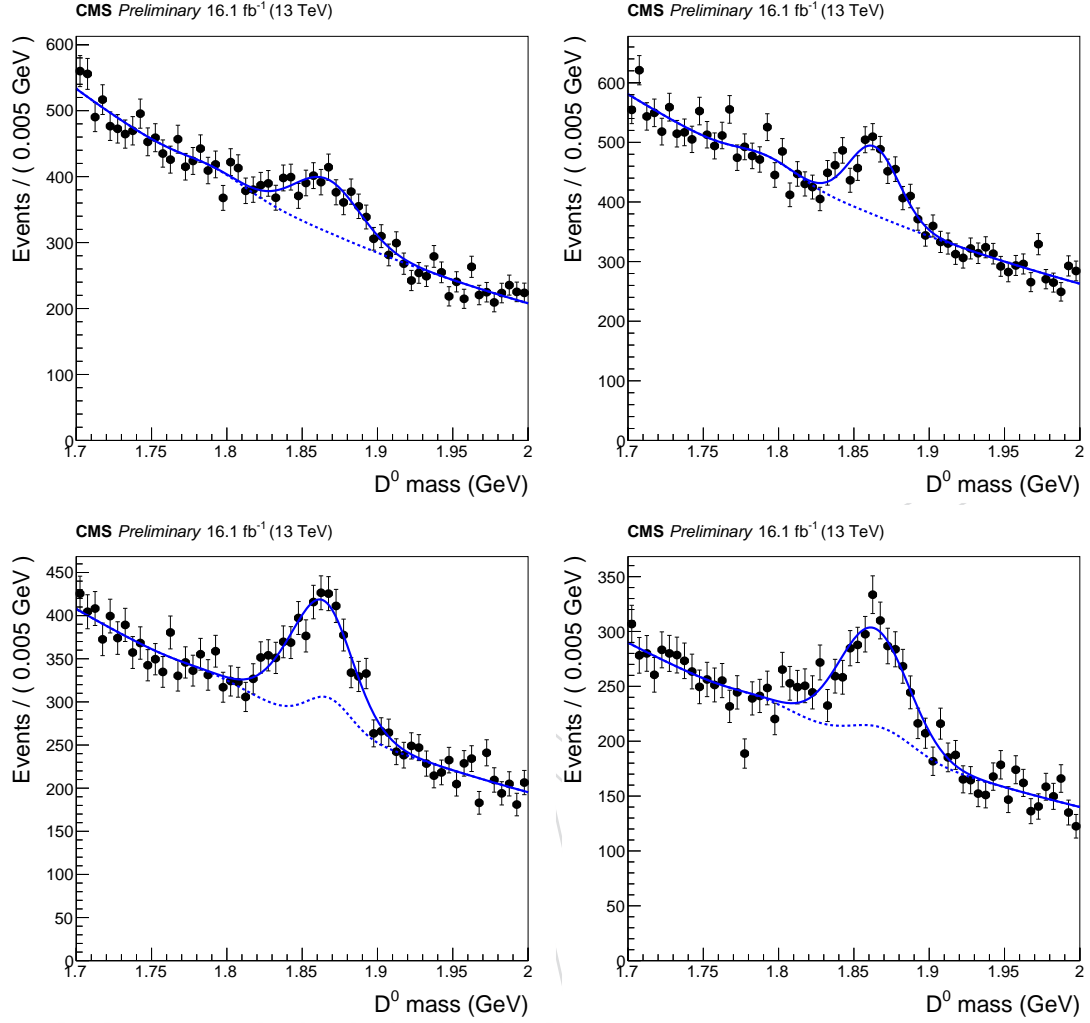
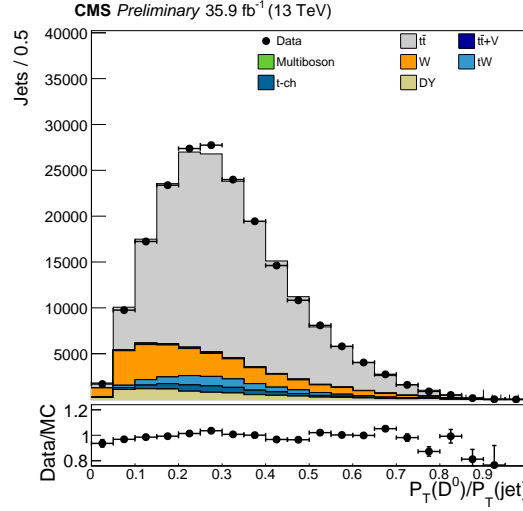


Figure 25: D^0 invariant mass for $x_B < 0.4$, $0.4 < x_B < 0.6$, $0.6 < x_B < 0.8$, and $x_B > 0.8$ for epochs GH.

Table 12: χ^2/ndf fits for the various mass x_B bins

$x_B < 0.4$	1.16212	0.802564
$0.4 < x_B < 0.6$	1.2698	1.09589
$0.6 < x_B < 0.8$	1.00772	0.769423
$x_B > 0.8$	0.72529	1.14939

Figure 26: D^0 p_T over full jet p_T Table 13: Sources of systematic uncertainty for D^0 full jet p_T proxy

ISR-up	-0.030
ISR-down	-0.045
FSR-up	-0.265 ± 0.099
FSR-down	0.005 ± 0.120
Underlying event up	-0.048
Underlying event down	0.028
Color reconnection	-0.028
Lepton selection up	-0.001
Lepton selection down	0.015
Pile-up up	-0.035
Pile-up down	-0.031
Trigger selection up	-0.022
Trigger selection down	0.069
JER up	0.015
JER down	-0.002
JSF up	-0.055
JSF down	-0.010
ME/PS up	-0.065
ME/PS down	-0.036

C b quark fragmentation templates

A complete set of r_b templates can be found here.

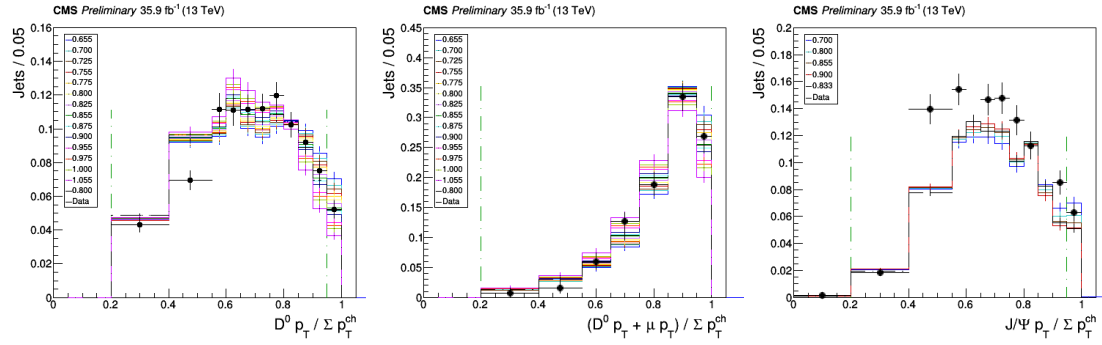


Figure 27: nominal $D^0, D^0_\mu, J/\psi$

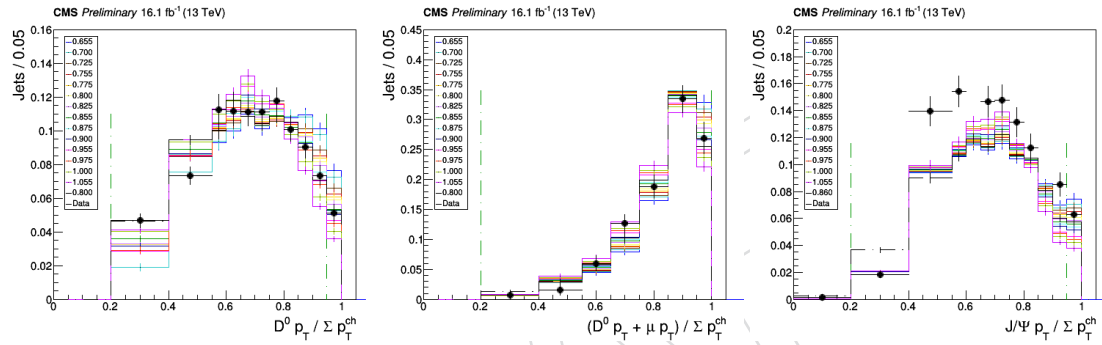


Figure 28: ISR-up $D^0, D^0_\mu, J/\psi$

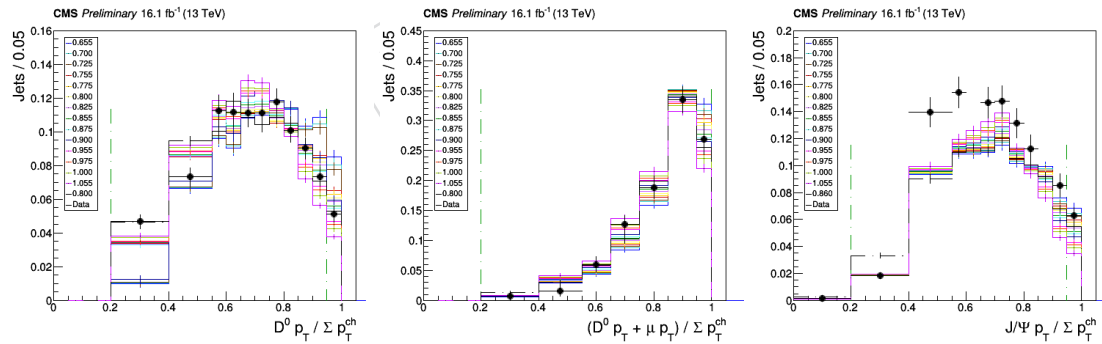
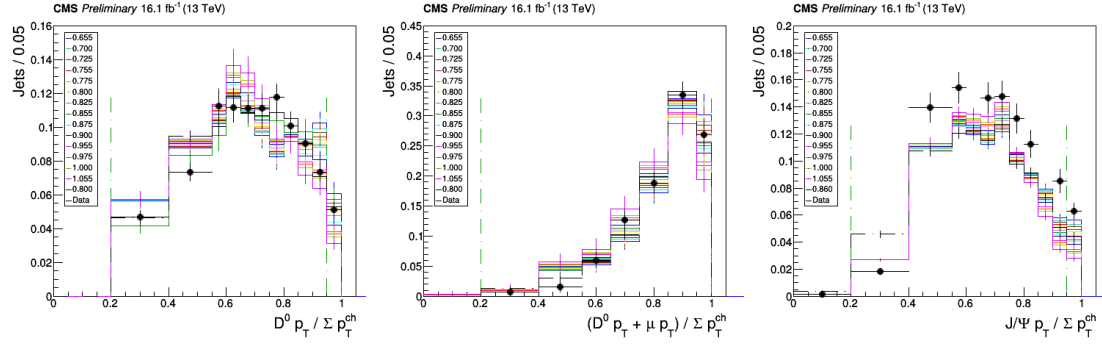
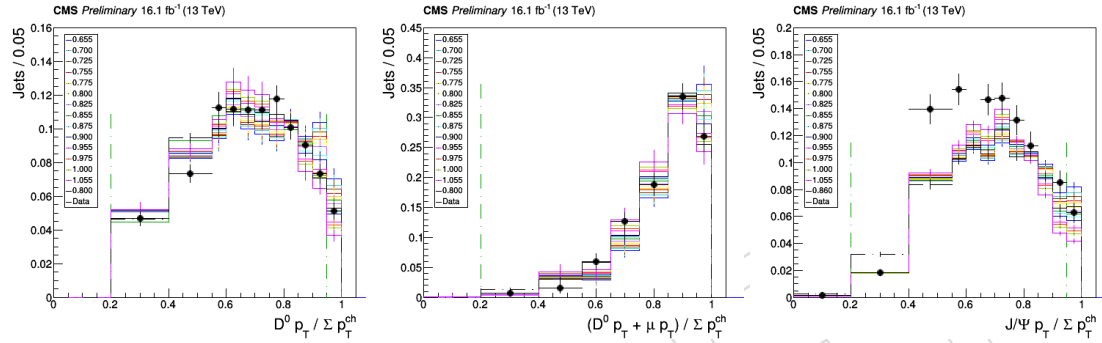
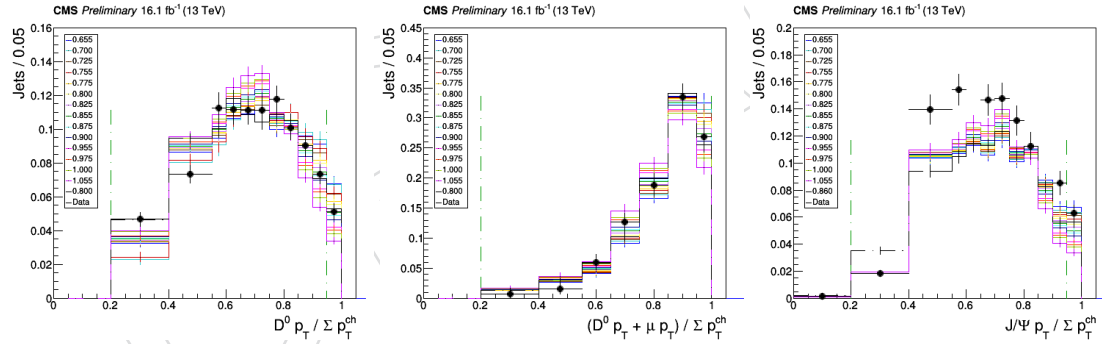
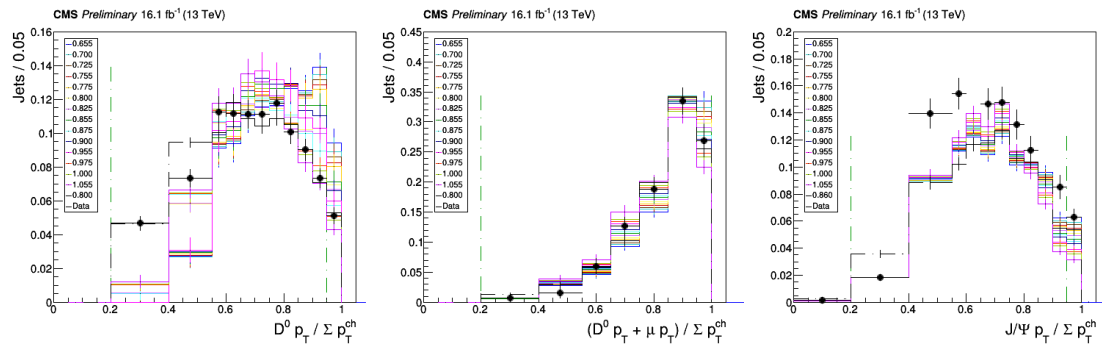
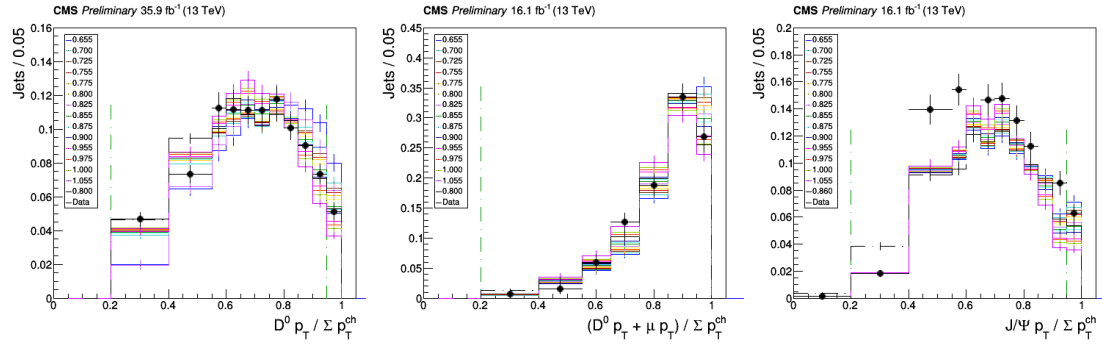
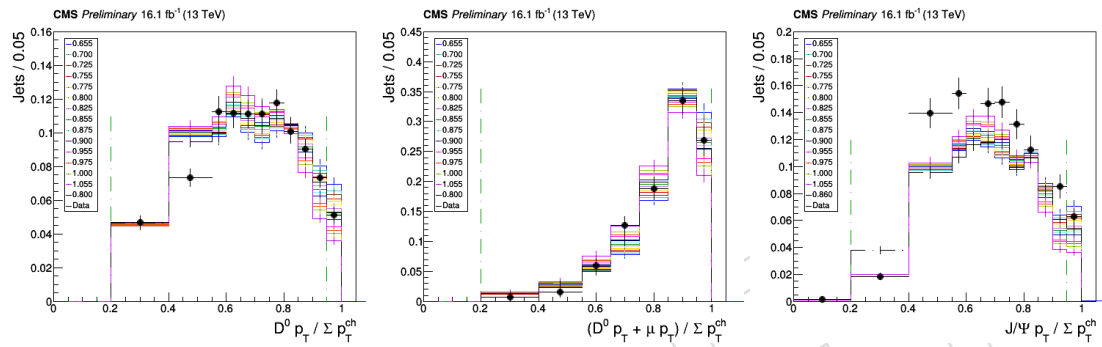
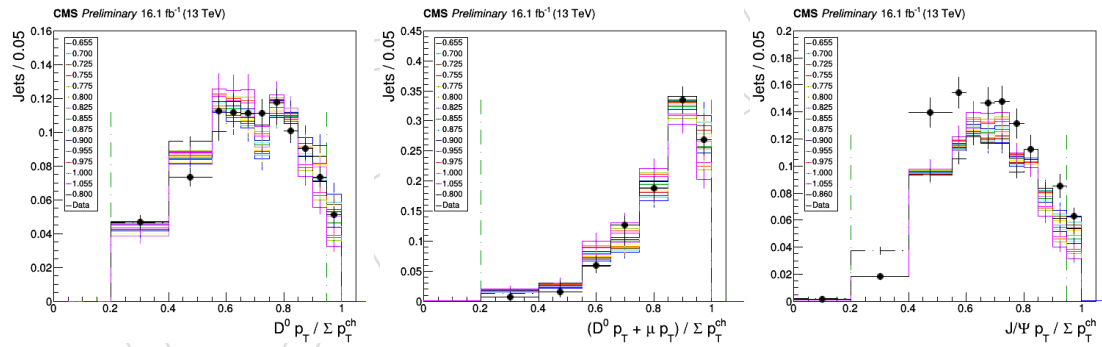
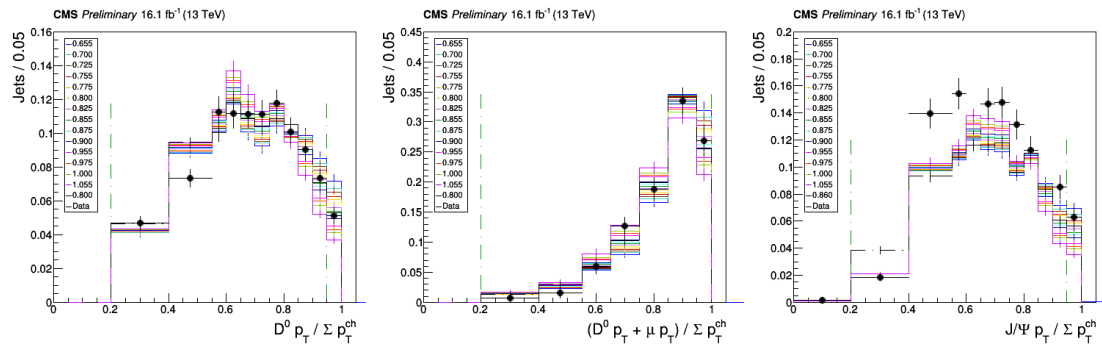
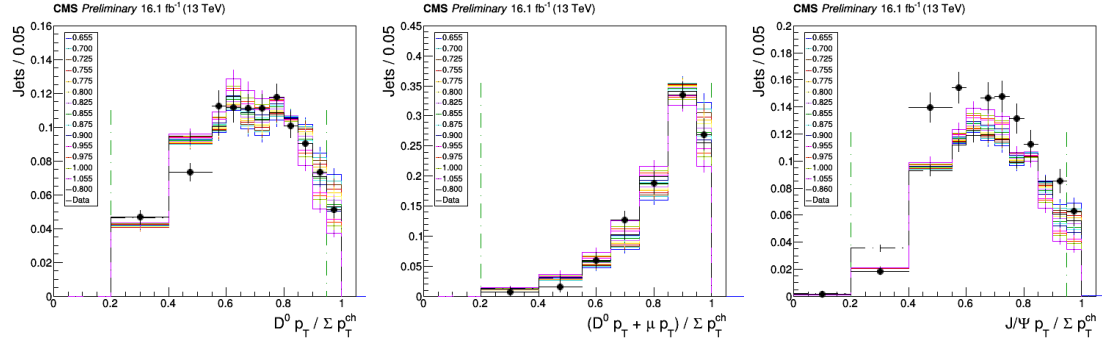
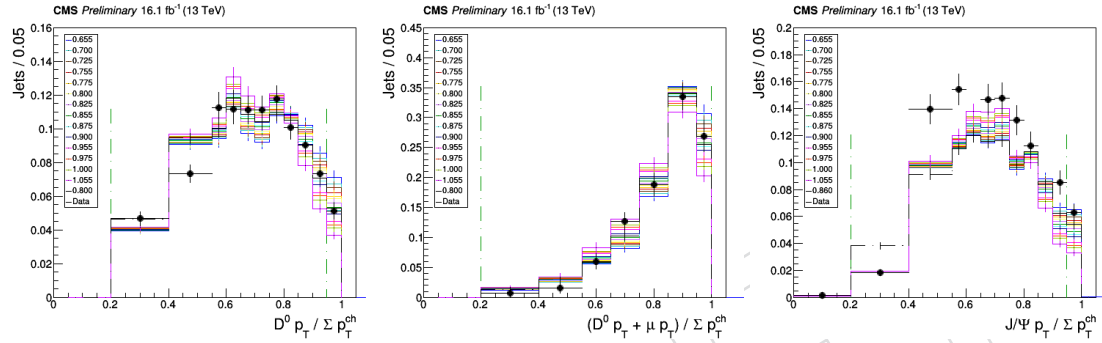
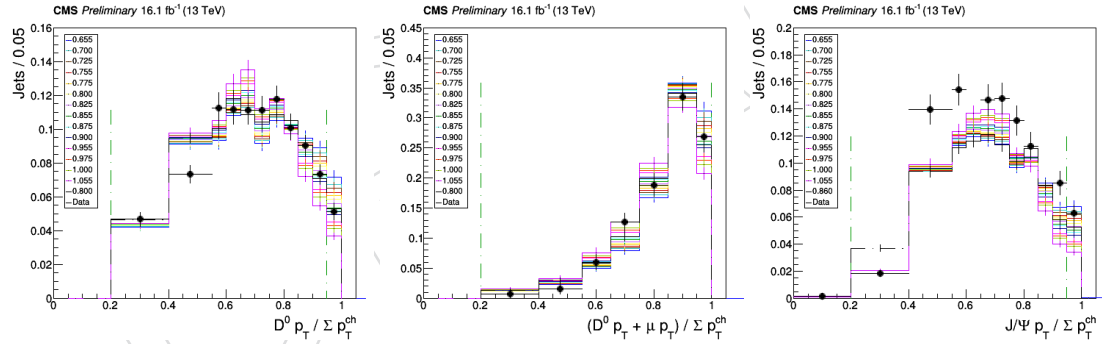
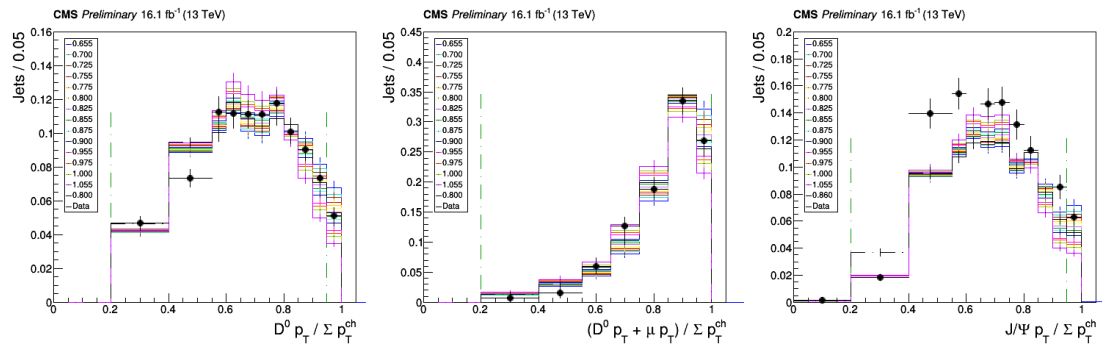
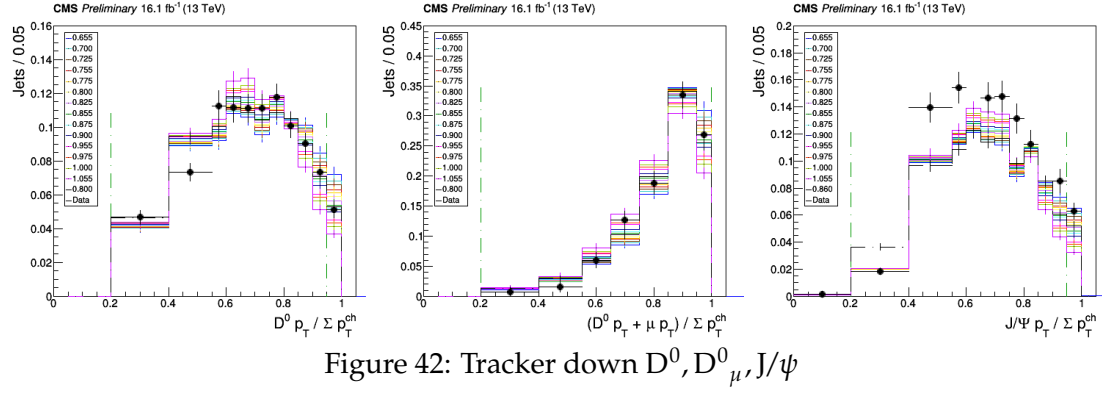
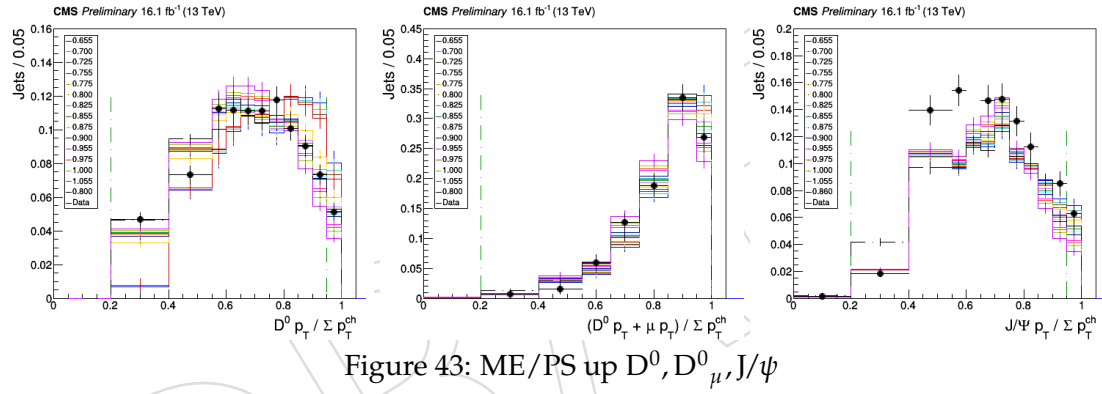
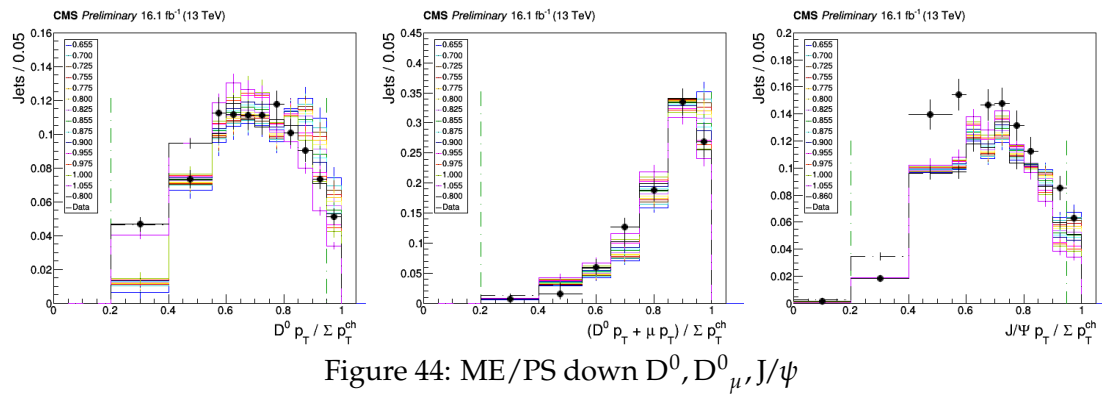


Figure 29: ISR-down $D^0, D^0_\mu, J/\psi$

Figure 30: FSR-up $D^0, D^0_\mu, J/\psi$ Figure 31: FSR-down $D^0, D^0_\mu, J/\psi$ Figure 32: UE up $D^0, D^0_\mu, J/\psi$ Figure 33: UE down $D^0, D^0_\mu, J/\psi$

Figure 34: CR $D^0, D^0_\mu, J/\psi$ Figure 35: TRIGGER $D^0, D^0_\mu, J/\psi$ Figure 36: TRIGGER $D^0, D^0_\mu, J/\psi$ Figure 37: LEP up $D^0, D^0_\mu, J/\psi$

Figure 38: LEP down $D^0, D^0_\mu, J/\psi$ Figure 39: PU up $D^0, D^0_\mu, J/\psi$ Figure 40: PU down $D^0, D^0_\mu, J/\psi$ Figure 41: Tracker up $D^0, D^0_\mu, J/\psi$

Figure 42: Tracker down $D^0, D^0_\mu, J/\psi$ Figure 43: ME/PS up $D^0, D^0_\mu, J/\psi$ Figure 44: ME/PS down $D^0, D^0_\mu, J/\psi$

D Systematic Fits

A complete set of χ^2 fits for the sytematic samples can be found [here](#).

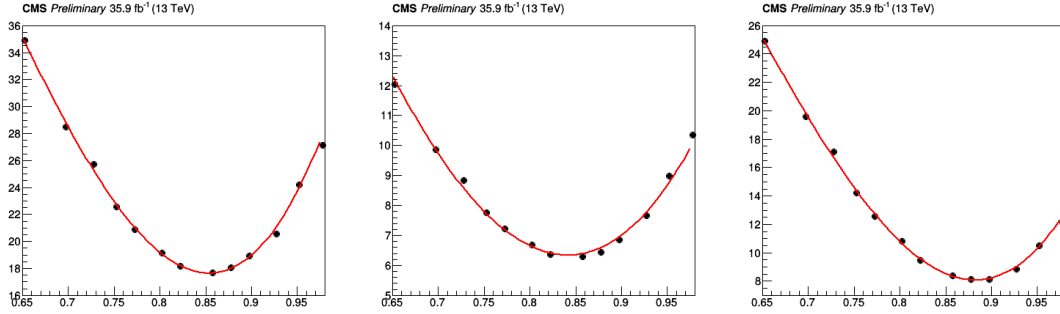


Figure 45: nominal $D^0, D^0_\mu, J/\psi$

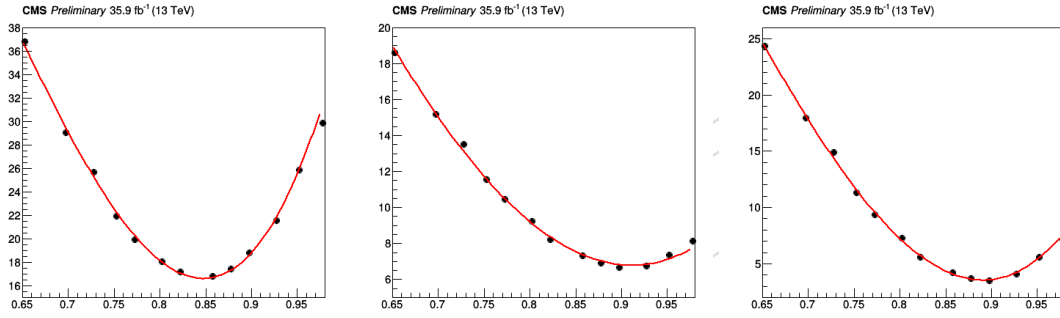


Figure 46: ISR-up $D^0, D^0_\mu, J/\psi$

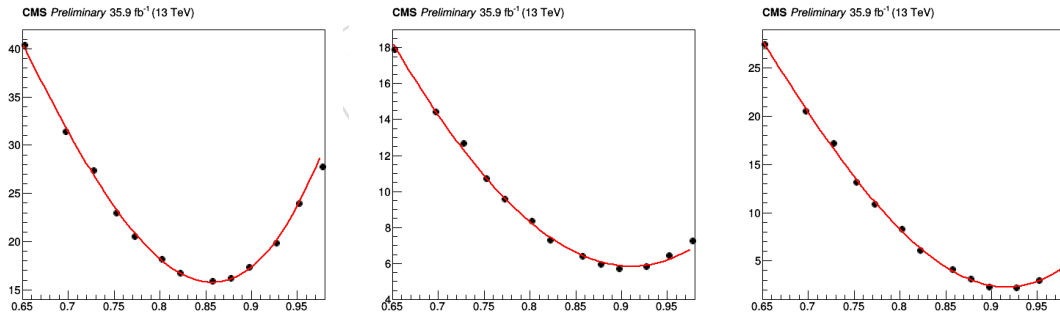
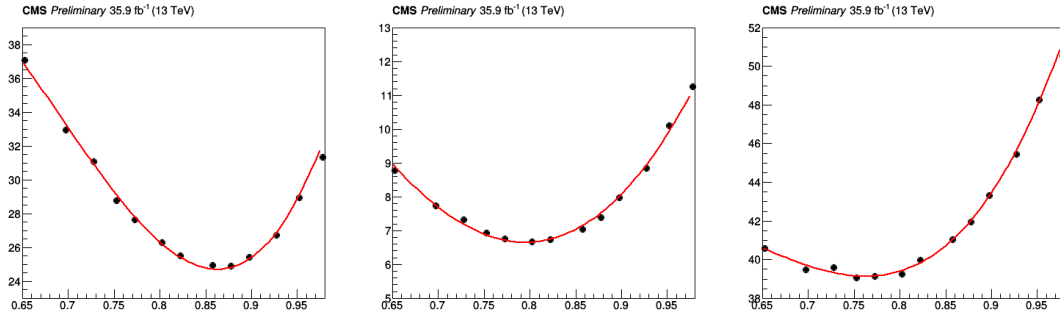
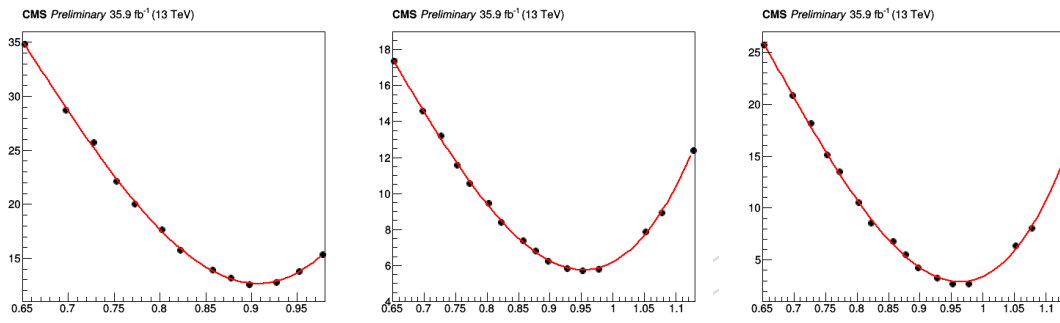
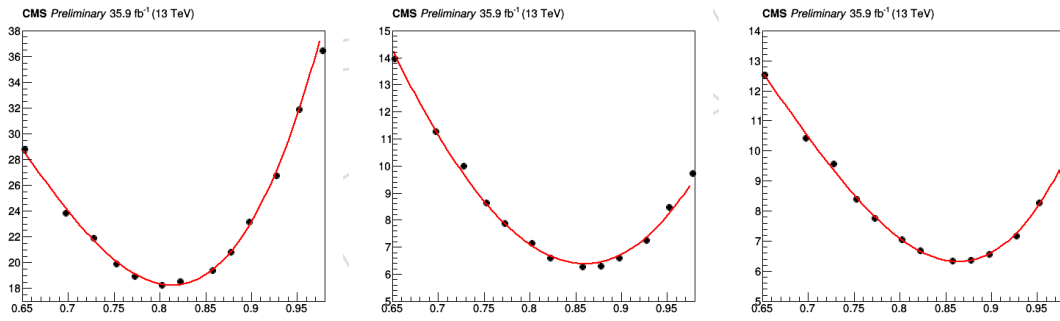
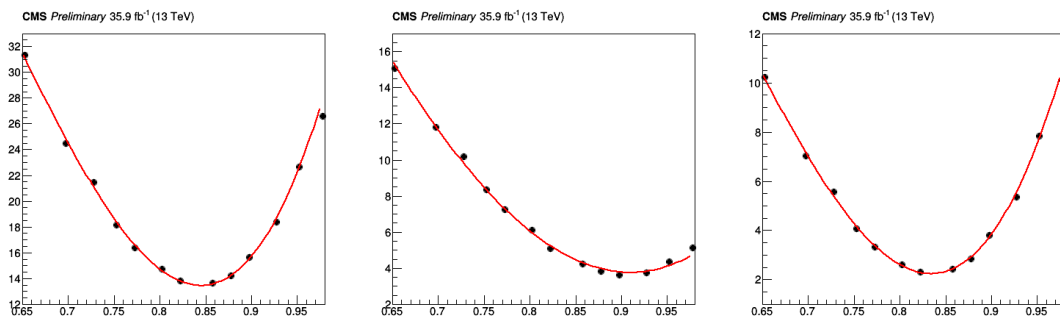
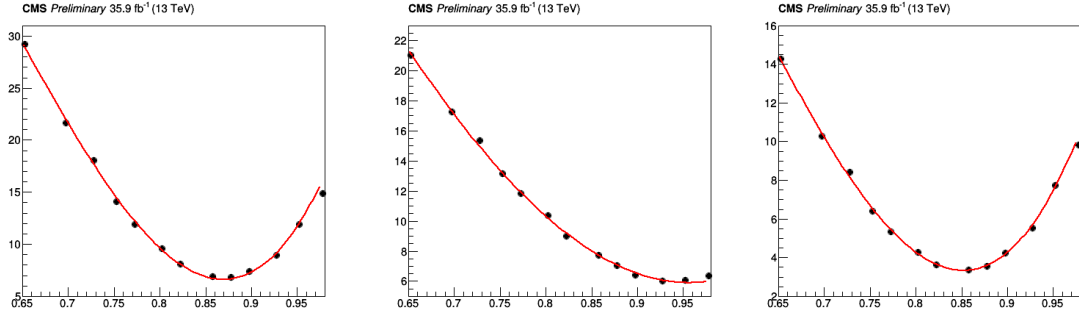
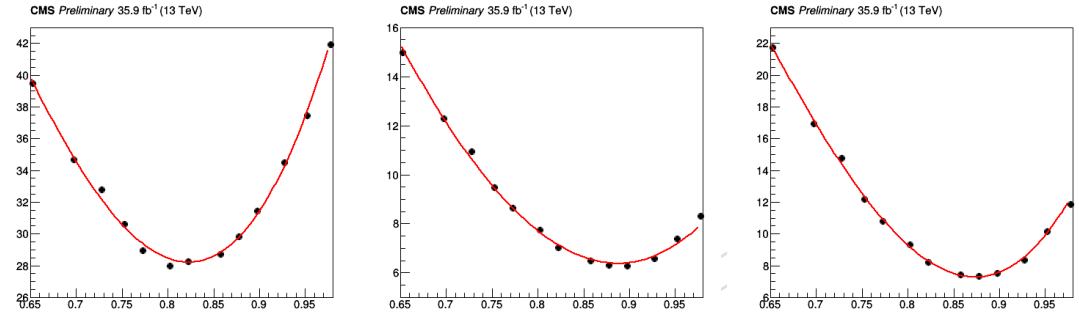
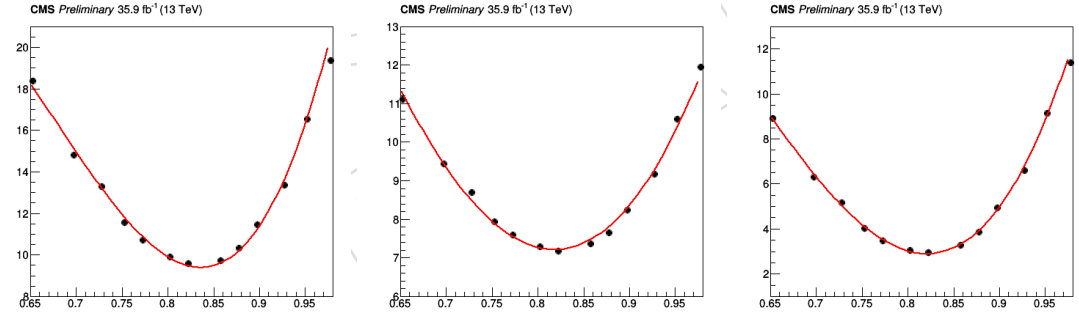
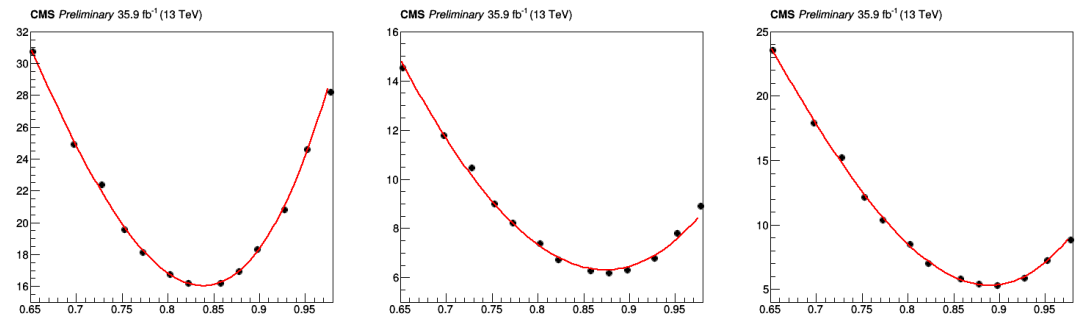


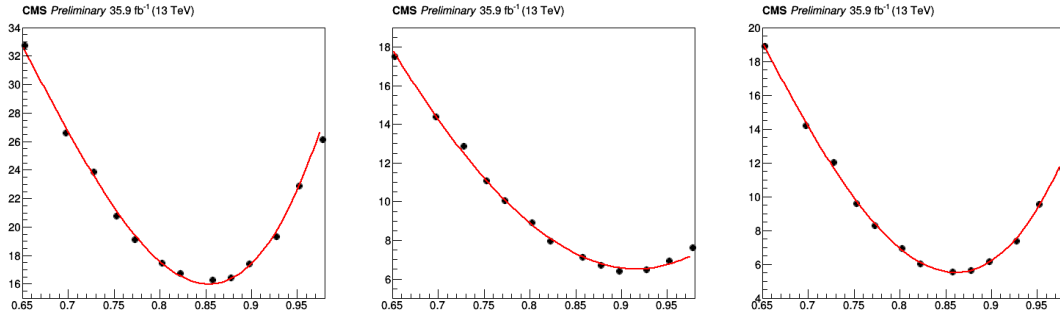
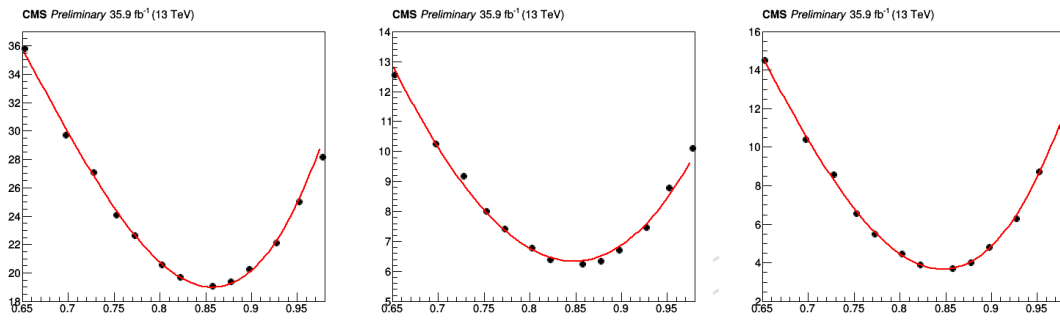
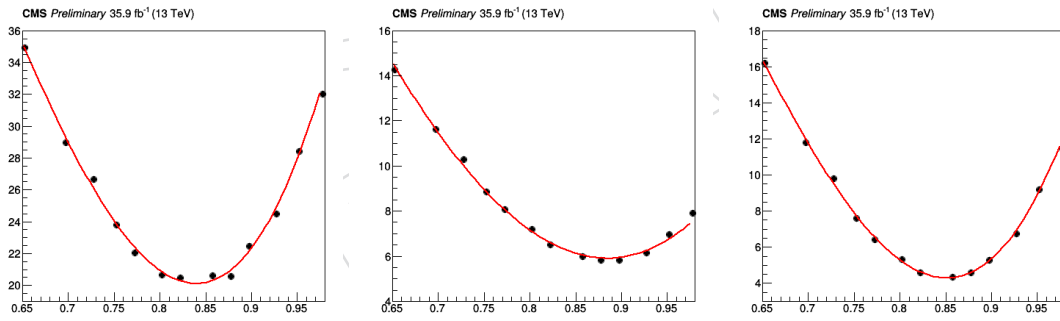
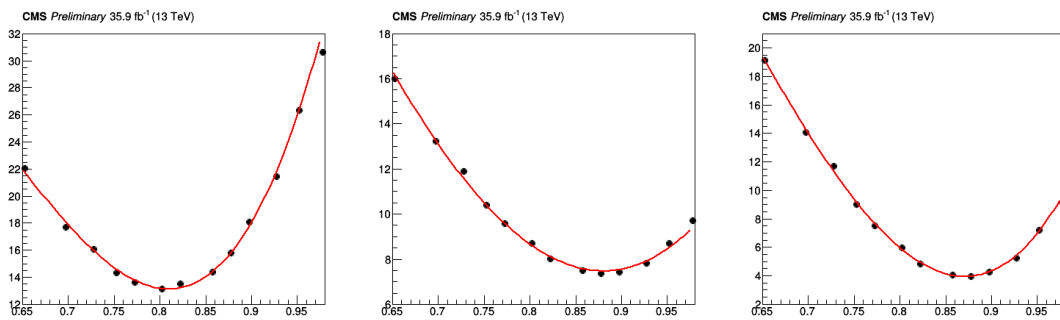
Figure 47: ISR-down $D^0, D^0_\mu, J/\psi$

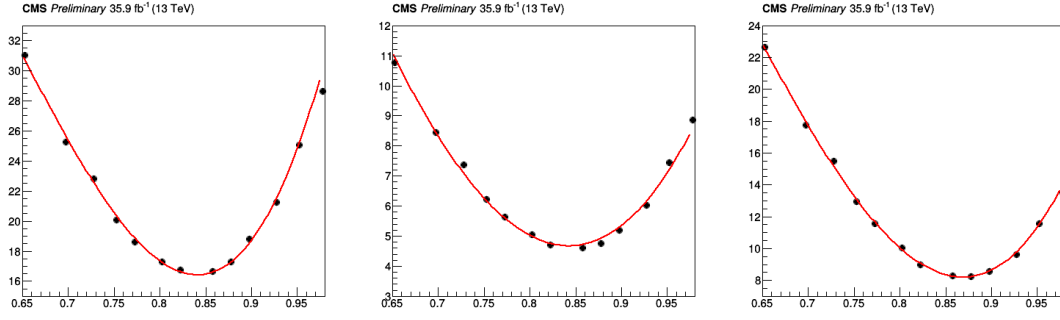
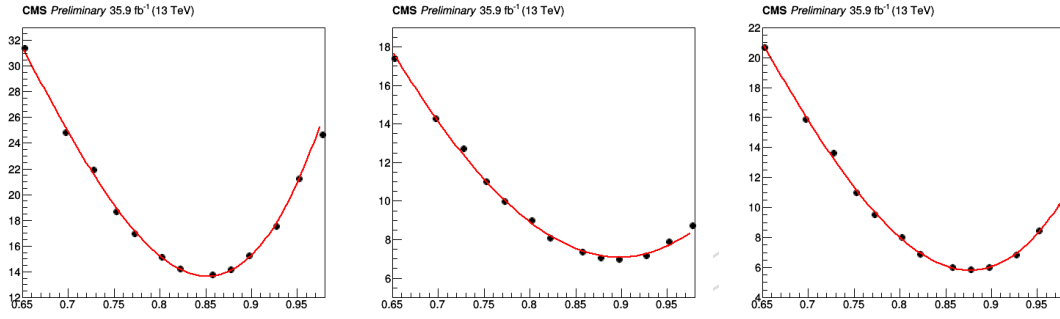
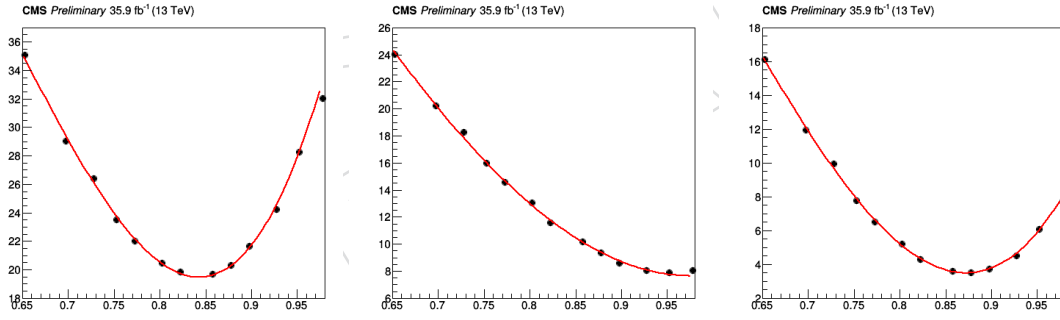
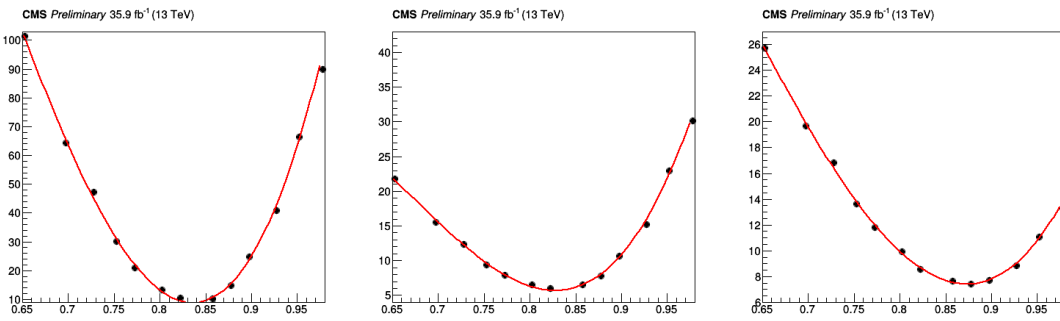
Figure 48: FSR-up $D^0, D^0_\mu, J/\psi$ Figure 49: FSR-down $D^0, D^0_\mu, J/\psi$ Figure 50: UE up $D^0, D^0_\mu, J/\psi$ Figure 51: UE down $D^0, D^0_\mu, J/\psi$



X

Figure 52: CR D⁰, D⁰_μ, J/ψFigure 53: TRIGGER up D⁰, D⁰_μ, J/ψFigure 54: TRIGGER down D⁰, D⁰_μ, J/ψFigure 55: LEP up D⁰, D⁰_μ, J/ψ

Figure 56: LEP down $D^0, D^0_\mu, J/\psi$ Figure 57: PU up $D^0, D^0_\mu, J/\psi$ Figure 58: PU down $D^0, D^0_\mu, J/\psi$ Figure 59: Tracker up $D^0, D^0_\mu, J/\psi$

Figure 60: Tracker down $D^0, D^0_\mu, J/\psi$ Figure 61: ME/PS up $D^0, D^0_\mu, J/\psi$ Figure 62: ME/PS down $D^0, D^0_\mu, J/\psi$ Figure 63: t quark p_T reweighting $D^0, D^0_\mu, J/\psi$

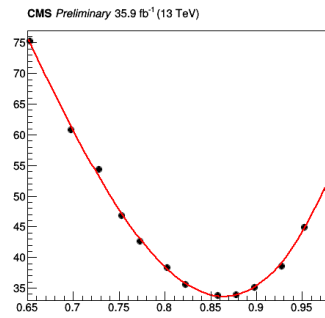


Figure 64: nominal combined

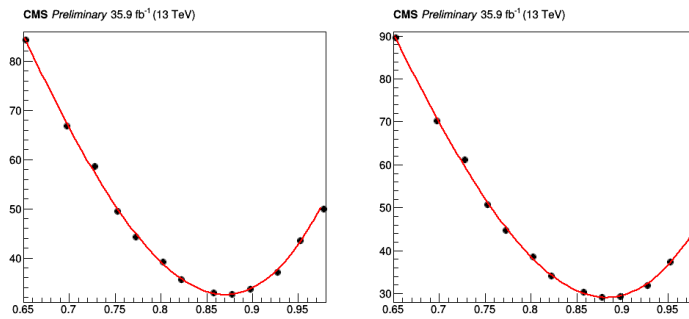


Figure 65: ISR-up combined, ISR-down combined

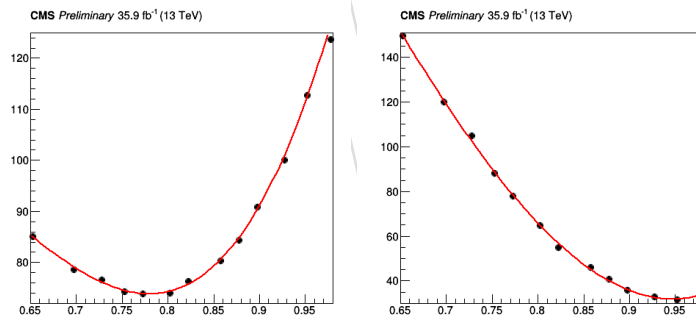


Figure 66: FSR-up combined, FSR-down combined

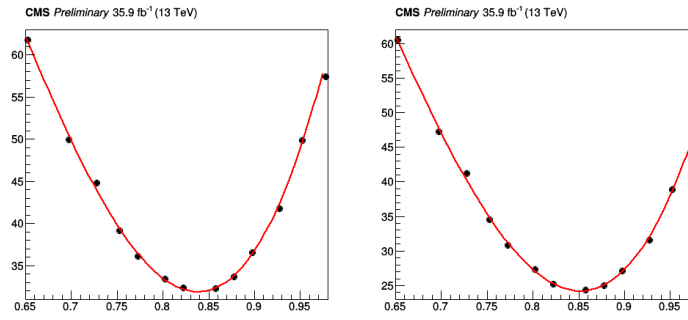


Figure 67: UE up combined, UE down combined

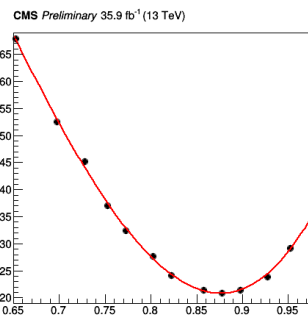


Figure 68: CR combined

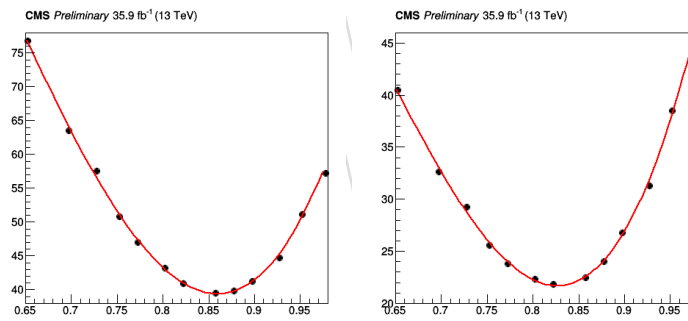


Figure 69: TRIGGER combined, TRIGGER combined

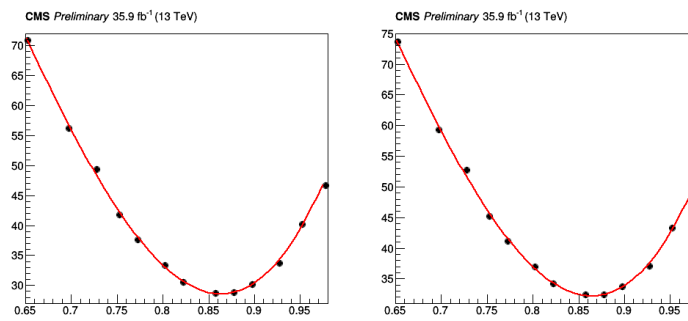


Figure 70: LEP up combined, LEP down combined

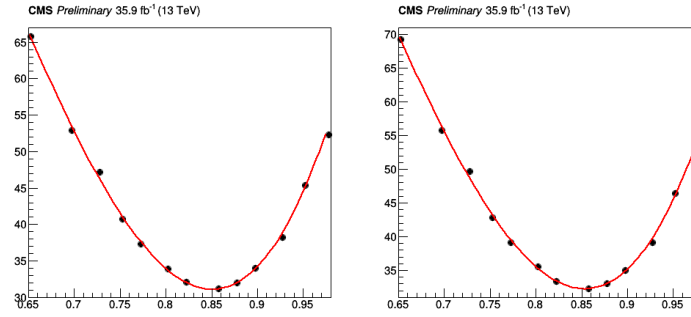


Figure 71: PU up combined, PU down combined

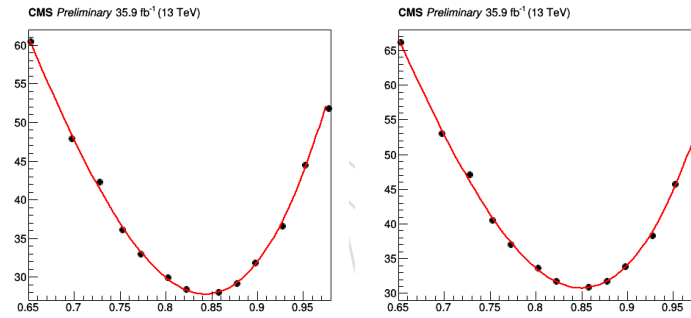


Figure 72: Tracker down combined, Tracker up combined

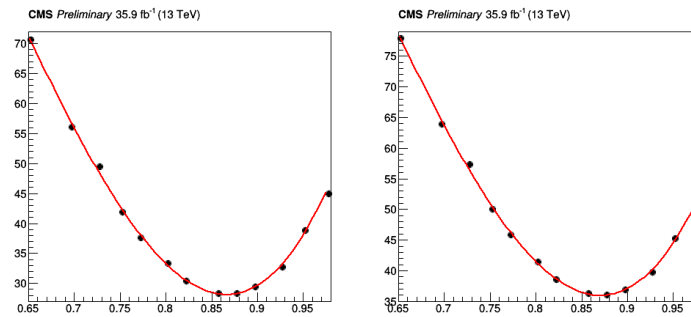


Figure 73: ME/PS up Combined, ME/PS down Combined