Measurement of Charm Production in CMS and Total Charm Cross Section with Non-Universal Charm Fragmentation

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Yewon Yang

Hamburg

2024

Gutachter/innen der Dissertation:	PD Dr. Achim Geiser Prof. Dr. Sven-Olaf Moch
Zusammensetzung der Prüfungskommission:	PD Dr. Achim Geiser Prof. Dr. Sven-Olaf Moch Prof. Dr. Daniela Pfannkuche Prof. Dr. Freya Blekman Dr. Katarzyna Wichmann
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Daniela Pfannkuche
Datum der Disputation:	27.06.2024
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Markus Drescher
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Hamburg, 16 May 2024

Yewon Yang 

Abstract

This thesis presents differential and total charm cross section measurements at various center-of-mass energies in pp collisions. For the purpose of measuring the total charm cross section, the D^{*+} differential cross sections are measured at $\sqrt{s}$ = 7, 0.9 and 13 TeV in the CMS detector by covering the largest possible phase space, resulting in the smallest extrapolation factor ever introduced at the LHC. The measured fiducial cross sections are extrapolated to the total cross section using a data-driven parametrization which is introduced for the first time to be applied to all the weakly-decaying ground states in pp collisions with non-universal charm fragmentation. Also adding other LHC measurements, the total charm-pair cross sections are measured at $\sqrt{s}$ = 0.9, 5, 7, and 13 TeV to be $1.83^{+0.35}_{-0.37}$, $8.43^{+1.05}_{-1.16}$, $9.39^{+1.35}_{-1.49}$, and $17.43^{+2.10}_{-2.57}$ mb, respectively. These measurements supersede all the earlier LHC measurements which were derived under the fragmentation universality assumption, and show consistency with NNLO QCD theory predictions. The measurement of the total charm cross section as a function of the center-of-mass energy can be used to constrain QCD parameters. The very first example is shown to constrain the charm mass and the parton distribution functions particularly in the low-x region.

Zusammenfassung

In dieser Arbeit werden Messungen differentieller und totaler Wirkungsquerschnitte für charm-Produktion in pp-Kollisionen bei verschiedenen Schwerpunktsenergien vorgestellt. Als Zwischenschritt zur Messung des totalen Wirkungsqurschnitts werden differentielle Wirkungsquerschnitte für die Produktion von D^{*+}-Mesonen gemessen bei $\sqrt{s} = 7$, 0.9 und 13 TeV im CMS-Detektor. Dabei wird der größtmögliche Phasenraum abgedeckt, so dass der finale Extrapolationsfaktor kleiner ist als jemals zuvor am LHC erzielt. Diese Extrapolation jenseits des gemessenen Phasenraums erfolgt mittels einer datengetriebenen Theorie-Parametrisierung, die zum ersten mal angewendet wird auf alle schwach zerfallenden Charm-Grundzustände unter Berücksichtigung der Nicht-Universalität der Charm-Fragmentation. Unter Hinzunahme anderer LHC-Messungen werden die totalen Wirkungsquerschnitte für Charm-Paarproduktion bei $\sqrt{s} = 0.9$, 5, 7, und 13 TeV bestimmt zu $1.83^{+0.35}_{-0.37}$, $8.43^{+1.05}_{-1.16}$, $9.39^{+1.35}_{-1.49}$, und $17.43^{+2.10}_{-2.57}$ mb. Diese Messungen ersetzen alle vorhergehenden LHC-Messungen, die unter der Annahme universeller Charm-Fragmentation erzielt worden sind. Sie bleiben konsistent mit NNLO-QCD Theorie-Vorhersagen. Die Darstellung dieser Messungen als Funktion der Schwerpunktenergie kann verwendet werden um QCD-Parameter zu beschränken. Die erste entsprechende Untersuchung liefert Einschränkungen auf die Masse des Charm-Quarks, und der Parton-Dichteverteilungen bei sehr kleinen Impulsanteilen x.

Contents

1 | Introduction

Elementary particle physics is to explain the ultimate constituents of matter and their interactions. Understanding the fundamental particles and interactions except gravity is embodied today in a unified description, the Standard Model (SM). The theory of Quantum ChromoDynamics (QCD) is a well established part of the SM, which describes the strong interactions. The strong interaction has the unique property of increasing strength with decreasing energy scale, such that quarks can never be measured as free particle in experiments. Thus it is inevitable to encounter non-perturbative QCD phenomena for any interactions involving quark states, which requires a strong interplay between experiment and theory. Heavy quark production in *pp* collisions is one of the best subjects to understand the strong interaction, since it is almost free from any other interactions and allows a comparison of theory to measurement with a broad energy coverage.

In particular, measurements of charm production in *pp* collisions at LHC center-of-mass-energies provide an important test of QCD in the transition region of the perturbative and non-perturbative regimes. Heavy quarks known as top, beauty and charm have mass scales which are larger than the QCD scale, Λ_{QCD}. Those mass scales make the strong coupling constant, α_s, to be small enough to calculate the production cross section perturbatively. However, the charm mass is very close to Λ_{QCD}, which challenges perturbative calculations with large theoretical uncertainty, and thus providing experimental constraints is of importance. Currently differential cross-section calculations are known for top [1] and beauty [2] up to Next-to-Next-to-Leading Order (NNLO, terms up to α_s^4), while for charm up to only Next-to-Leading Order (NLO, up to α_s^3) without [3] or with [4] Next-to-Leading Log (NLL) contributions. In the case of the total cross section, calculations are known up to NNLO [5] for all three heavy quarks. The total charm cross section, which is free of any fragmentation effects, can be used to constrain the charm mass and Parton Distribution Functions (PDF) in the low-x region where uncertainties are large.

For the total charm cross section measurement, differential cross sections measured in limited kinematic ranges and for a restricted set of final hadronic states need to be extrapolated to the total cross section, under certain theoretical assumptions. Several differential cross-section measurements have been performed on different inclusive charm final states in *pp* collisions by the LHC experiments so far: ALICE at $\sqrt{s} = 2.76$, 5, 7, 13 TeV [6–15], LHCb at $\sqrt{s} = 5$, 7, 13 TeV [16–18], ATLAS at $\sqrt{s} = 7$ TeV [19] and CMS at $\sqrt{s} = 5$, 13 TeV [20–23]. Some of these measurements were extrapolated to the full kinematic phase space with a large extrapolation factor in order to extract the total charm cross section [6, 12, 13, 24], with the assumption of charm fragmentation universality, i.e., that the fragmentation is independent of either collision systems and kinematics. The corresponding cross section values are thus strongly theory and model dependent.

Recent LHC measurements reported that the fragmentation significantly depends on collision systems and transverse momentum [10, 15, 22, 25–27], which indicates non-universality of charm fragmentation. This requires all the previous total charm cross section measurements to be updated. For the first time, based on the study in this thesis, the treatment of the non-universal fragmentation for the total charm cross section extraction was presented in [28] by introducing a novel phenomenological way not to assume any particular non-universal charm

fragmentation model; the non-universal fragmentation effects were extracted directly from the LHC measurements and combined with NLO+NLL perturbative theory. This was then applied into extrapolating differential cross section measurements at $\sqrt{s} = 5$ TeV. Meanwhile, [29] also presented the total charm cross section derived at $\sqrt{s} = 5$ TeV by using a PYTHIA [30, 31] tuning to deal with the non-universal fragmentation effects, but it appears that still not all of the final states can be well described by the tuning. No generally accepted model is available to describe the non-universal fragmentation at the moment of writing this thesis.

Measurements in the CMS detector result in the smallest extrapolation factor from a single experiment by covering the largest possible phase space up to about 2.5 in rapidity. Using LHCb measurements for the larger rapidity, the total charm cross section can be derived with the smallest extrapolation factor ever achieved in LHC measurements [32].

This thesis aims to eventually present the total charm-pair cross section measurements in pp collisions at various center-of-mass energies by applying the non-universal charm fragmentation with minimal theory and model dependence, which then supersede the earlier LHC measurements still provided based on the fragmentation universality assumption. With these measurements, the very first preliminary study to constrain the charm mass and the low-x PDFs also will be briefly introduced. For these goals, Chapter 2 provides theoretical predictions of charm quark pair production at the highest order available to date, which are taken as perturbative predictions for extrapolation and to constrain the QCD parameters in this thesis. All the non-perturbative inputs for the charm fragmentation are extracted from published measurements which will be introduced in Chapter 3. After introducing the CMS detector in Chapter 4, differential cross section measurements in CMS, for specifically D^{*+} and D^0, will be discussed in Chapter 5. Finally the total charm cross section measurements and the application of the results to constrain the QCD parameters will be presented in Chapter 6 starting from introducing the new extrapolation scheme to apply the non-universal charm fragmentation, which was presented in [28].

2 Theoretical Particle Physics and Heavy-Quark Production

The Standard Model (SM) of particle physics describes three of the four known fundamental forces in the Universe. To date, it is known that the four fundamental forces are gravitational, electromagnetic, weak and strong interaction. All massive particles feel the attractive force of gravity. However, the gravitational interaction has the weakest strength among the four, which results in no significant effect at the microscopic scale, while it is the most significant interaction at the macroscopic scale, governing the mechanics of the planets, stars and galaxies. Electromagnetism governs the interactions between electrically charged particles via electromagnetic fields. The weak interaction is responsible for particle decays. The coupling constant of the weak interaction is larger than that of electromagnetism, however, the interaction has smaller strength at low energies for decay processes due to the massive bosons, and thus is called "weak". The strong interaction is responsible for bindings of hadronic and nuclear states, and has the largest strength among the four. Except gravity, the three fundamental interactions are unified in the SM.

According to the SM, the elementary particles are classified as twelve *fermions* of half-integer spin and four kinds of *bosons* of integer spin. Fermions are classified again as six *quarks* and six *leptons*. There are three families composed of *up and down type* quarks: up (u) and down (d), charm (c) and strange (s), and top (t) and beauty (b). Similarly leptons have three families of *electron and neutrino type*: electron (e^-) and electron neutrino (ν_e), muon (μ^-) and muon neutrino (ν_μ), and tau (τ^-) and tau neutrino (ν_τ). Fermions have the corresponding antiparticles having opposite charge, which are called antiquark or antineutrino and denoted with a bar over the corresponding particle symbol. In the case of an electron type particle, the antiparticle is denoted by e^+, μ^+ and τ^+, and called positron, antimuon and antitau, respectively. On the other hand, it is known that there are four kinds of elementary bosons: the photon (γ), neutral and charged weak bosons (Z and $W^\pm$, respectively), eight gluons (g), and the Higgs boson (H). These bosons except the Higgs boson interact with fermions by carrying forces. The Higgs boson specifically contributes to the phenomenological mechanism of mass generation.

2.1 Standard model theory

An overview of the SM theory is shown in Figure 2.1. The underlying theory of the SM is quantum field theory [34]: a theoretical framework describing particle creation and annihilation by combining quantum mechanics and special relativity. In a field theory, the *Lagrangian* contains free-field terms and additional interaction terms. Combined with propagators for virtual or unstable particles, this gives information to compute scattering and decay amplitudes according to Feynman rules [35, 36]. Eventually, predictions are derived for quantities like the cross section and decay rate, which can be measured from experiment.

Let me first introduce the Lagrangian of fermions [37, p.34] in the absence of interactions:

$$\mathcal{L}_0 = \bar{\psi}(i\partial\!\!\!/ - m)\psi, \tag{2.1}$$

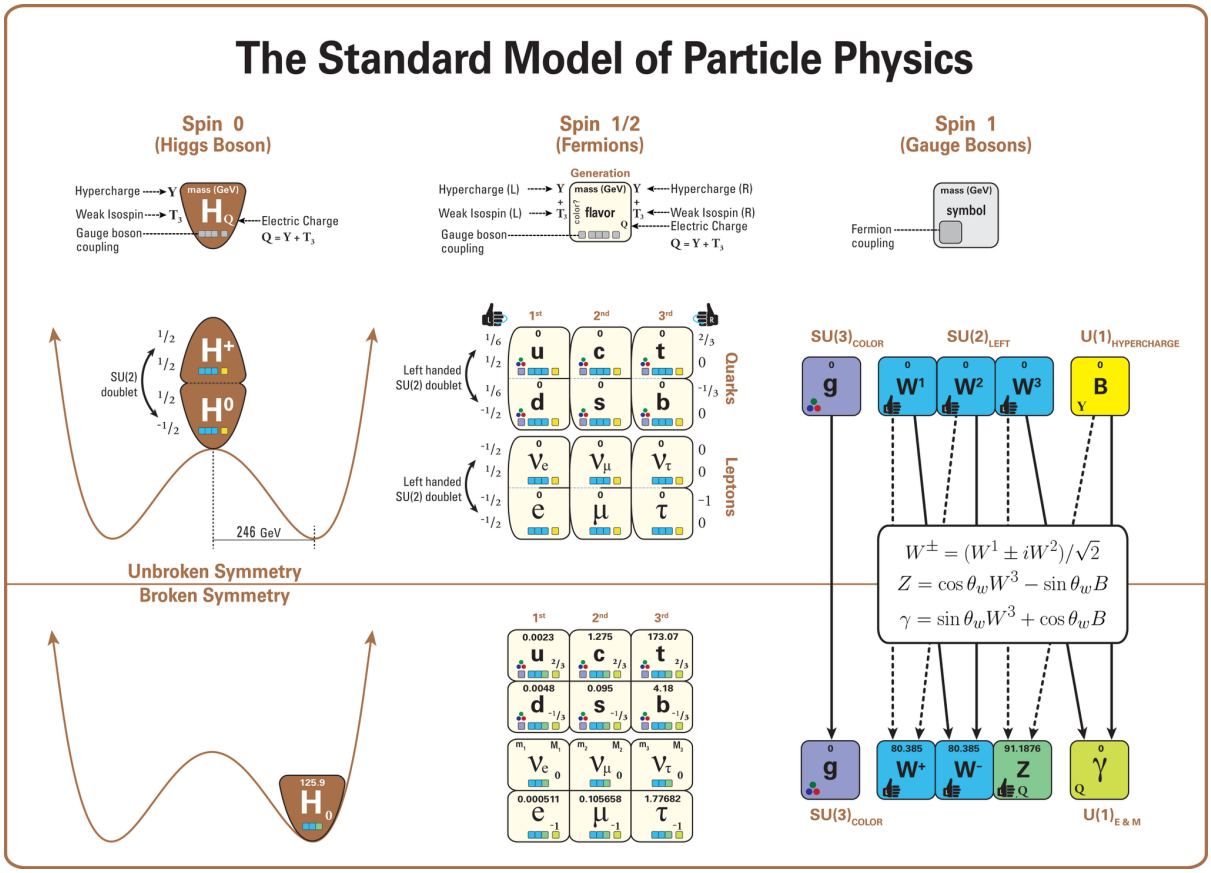


Figure 2.1: Overview of the Standard Model theory. The diagram shows the elementary particles of the Standard Model (the three families of quarks and leptons, and the gauge bosons), including their names, masses, spins, charges, chiralities, and interactions with the strong, weak and electromagnetic forces. It shows how the properties of the various particles differ in the (high-energy) symmetric phase (top) and the (low-energy) broken-symmetry phase (bottom). Figure is taken from [33].

where ψ is the Dirac field describing a four-component spin-$\frac{1}{2}$ particle with mass m and $\bar{\psi}$ is the Dirac adjoint. By minimizing the action

$$\mathcal{S} = \int d^4x \mathcal{L}_0, \tag{2.2}$$

the free-field Dirac equation is obtained as

$$(i\partial\!\!\!/ - m)\psi = 0. \tag{2.3}$$

Solutions of the Dirac equation can be given by a free-particle plane wave [38, p.100],

$$\psi = u(p)\exp(ip \cdot x) \tag{2.4}$$

for a relativistic spin-half particle, and

$$\psi = v(p)\exp(-ip \cdot x) \tag{2.5}$$

for its antiparticle. Here $u(p)$ and $v(p)$ are four-component Dirac spinors.

Conservation laws in physics are related mathematically to an invariance under a certain set of symmetries. The SM has the global Poincaré symmetry under which energy, momentum and angular momentum are conserved. The SM is also invariant under a local $\mathrm{SU(3)}\times\mathrm{SU(2)}\times\mathrm{U(1)_Y}$ gauge symmetry. The free-field Lagrangian of Eq.(2.1) alone is no longer invariant under the gauge symmetries and additional interaction terms should be introduced. Mathematically, this gauge symmetry is introduced with gauge fields and gauge covariant derivatives into the free-field Lagrangian, under which the color charge, weak isospin, weak hypercharge and electric charge are conserved.

2.1.1 QED Lagrangian

Quantum ElectroDynamics (QED) is an Abelian gauge theory describing the electromagnetic interaction, which is invariant under a local U(1) gauge symmetry. The Lagrangian of QED is described as

$$\mathcal{L} = \bar{\psi}(iD\!\!\!/ - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.6}$$

where a gauge covariant derivative [37, p.43],

$$D_\mu \equiv \partial_\mu - ieA_\mu, \tag{2.7}$$

is introduced instead of the partial derivative (∂_μ) in Eq.(2.1), and the field strength tensor [37, p.26] is

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.8}$$

Here a U(1) gauge field A_μ represents a vector boson, i.e., the photon and e is the coupling constant for an interaction vertex between fermions and photon. The free photon field A^μ, which is the electromagnetic vector potential, can be written in terms of a plane wave and a four-vector photon polarization, ϵ^μ_λ ($\lambda = 1, 2$) [38, p.121]:

$$A^\mu = \epsilon^\mu_\lambda \exp(ip \cdot x). \tag{2.9}$$

The Lagrangian, Eq.(2.6), is invariant under a gauge transformation where

$$A^\mu \mapsto A^\mu - \frac{1}{e}\partial^\mu \varkappa, \quad \psi \mapsto \exp(-i\varkappa)\psi \tag{2.10}$$

with a local infinitesimal variation, $\varkappa$. QED is invariant also under the discrete symmetries of charge conjugation [38, p.103] where

$$\psi \mapsto \hat{C}\psi \tag{2.11}$$

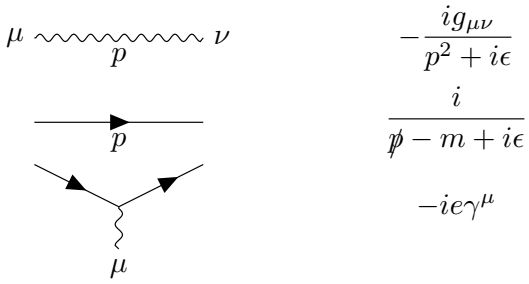


Figure 2.2: Feynman rules for QED.

and parity [38, p.110] where

$$\psi \mapsto \hat{P}\psi. \tag{2.12}$$

The charge conjugation operator $\hat{C}$ transforms a particle wavefunction into the corresponding antiparticle wavefunction and vice versa. The parity operator $\hat{P}$ results in spatial inversion through the origin.

Feynman rules introduced for QED are shown in Figure 2.2.

2.1.2 QCD Lagrangian

The theory of the strong interaction is described by a non-Abelian Yang-Mills gauge theory and called Quantum ChromoDynamics (QCD). The underlying symmetry associated with QCD is invariance under a local SU(3) transformation [38, p.244]:

$$G^C_\mu \mapsto G^C_\mu - \partial_\mu \varkappa_C - g_s f_{ABC} \varkappa_A G^B_\mu, \quad \psi \mapsto \exp(i g_s \varkappa^C \cdot t^C)\psi. \tag{2.13}$$

where G_μ is a SU(3) gauge field, i.e., the gluon field. f_{ABC} is the structure constant of the SU(3) colour group and t^C is the generator represented by the Gell-Mann matrices, λ^C [39, p.6]:

$$t^C = \frac{1}{2}\lambda^C. \tag{2.14}$$

Here the indices a and b run from 1 to 3, referring to the triplet representation of SU(3) for the (anti-)quark field $(\bar{q})q$, while $A, B, C = 1, 2, \cdots, 8$ are the colour degrees of freedom for G_μ. With a gauge covariant derivative [39, p.6] defined as

$$(D_\mu)_{ab} = \partial_\mu \delta_{ab} + i g_s (t^C G^C_\mu)_{ab}, \tag{2.15}$$

a Lagrangian density [39, p.5] for QCD can be given by

$$\mathcal{L} = \sum_{\text{flavours}} \bar{q}_a (i \not{D} - m)_{ab} q_b - \frac{1}{4} G^C_{\mu\nu} G^{\mu\nu}_C, \tag{2.16}$$

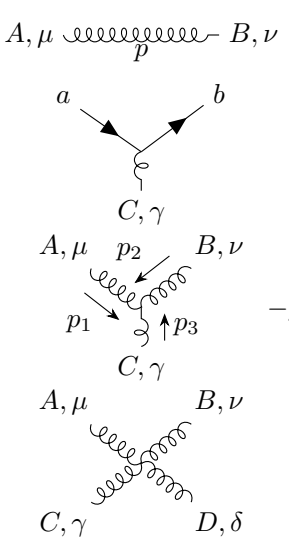
where

$$G^C_{\mu\nu} = \partial_\mu G^C_\nu - \partial_\nu G^C_\mu - g_s f^{ABC} G^B_\mu G^C_\nu. \tag{2.17}$$

The third term on the right-hand side of Eq.(2.17) indicates that QCD has triple and quartic gluon self-interactions.

However, unlike the Abelian QED case, the Lagrangian of Eq.(2.16) cannot give a perturbative theory without an additional gauge fixing term [39, p.8]. To define the propagator of the gluon field, a choice of gauge should be made:

$$\mathcal{L}_{\text{gauge-fixing}} = -\frac{1}{2\lambda}(\partial^\mu G^C_\mu)^2 \tag{2.18}$$



$$-\frac{i\delta_{AB}g_{\mu\nu}}{p^2+i\epsilon}$$

$$-ig_s(t^C)_{ab}\gamma^\gamma$$

$$-g_s f^{ABC}[(p_1-p_2)^\gamma g^{\mu\nu}+(p_2-p_3)^\mu g^{\nu\gamma}+(p_3-p_1)^\nu g^{\gamma\mu}]$$

$$-ig_s^2 f^{XAC}f^{XBD}[g^{\mu\nu}g^{\gamma\delta}-g^{\mu\delta}g^{\nu\gamma}]$$
$$-ig_s^2 f^{XAD}f^{XBC}[g^{\mu\nu}g^{\gamma\delta}-g^{\mu\gamma}g^{\nu\delta}]$$
$$-ig_s^2 f^{XAB}f^{XCD}[g^{\mu\gamma}g^{\nu\delta}-g^{\mu\delta}g^{\nu\gamma}]$$

Figure 2.3: QCD Feynman rules for gluons and quarks with $\lambda = 1$ for the covariant gauge fixing term.

which fixes the choice of covariant gauges with gauge parameter λ. Then to cancel unphysical degrees of freedom which would propagate in the covariant gauges, a ghost field term also needs to be added:

$$\mathcal{L}_{\text{ghost}} = \partial_\mu \eta^{A\dagger}(D^\mu_{AB}\eta^B) \tag{2.19}$$

where η is a ghost field obeying Fermi statistics.

In Figure 2.3, the QCD Feynman rules are shown specifically for the gluon propagator, and gluons and fermions interactions with fixing $\lambda = 1$ (Feynman gauge) [39, p.11].

2.1.3 Standard Model Lagrangian

The SM is described by a local SU(3)×SU(2)×U(1)$_\text{Y}$ gauge symmetry, where the weak and electromagnetic interactions are unified as *electroweak* interaction by a non-Abelian Yang-Mills SU(2)×U(1)$_\text{Y}$ group. The SU(2) group has gauge coupling g and gauge bosons W^α ($\alpha = 1, 2, 3$), and acts only on the flavour of the left-chiral fermions (refer to e.g. [37, p.40] for *chirality*). The U(1)$_\text{Y}$ group, of which the generator is given by the weak hypercharge (labeled Y), has gauge coupling g' and gauge boson B, and acts on both left- and right-chiral fermions but with different charges. After spontaneous symmetry breaking, the SU(2)×U(1)$_\text{Y}$ group becomes the U(1) group of QED (described in Section 2.1.1) where the photon is composed of a linear combination of the neutral W boson and B. The Z and $W^\pm$ bosons, which are composed of W^α, B after the breaking, are the vector bosons responsible for the weak interactions and have non-zero masses. The strong interaction is given in the SU(3) group as described in Section 2.1.2.

The SM Lagrangian density is given by the sum of a gauge, fermion, Higgs and Yukawa Lagrangian [37, p.257]:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_f + \mathcal{L}_\phi + \mathcal{L}_Y. \tag{2.20}$$

The gauge Lagrangian ($\mathcal{L}_{\text{gauge}}$) is defined by the field strength tensors of each group:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}G^C_{\mu\nu}G^{\mu\nu}_C - \frac{1}{4}W^\alpha_{\mu\nu}W^{\mu\nu}_\alpha - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{2.21}$$

where $G^C_{\mu\nu}$ is Eq.(2.17),

$$W^\alpha_{\mu\nu} = \partial_\mu W^\alpha_\nu - \partial_\nu W^\alpha_\mu - g\epsilon^{\alpha\beta\gamma}W^\beta_\mu W^\gamma_\nu, \tag{2.22}$$

and

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.23}$$

for SU(3), SU(2) and $U(1)_Y$, respectively. These include the gauge boson kinetic energy terms, and the three- and the four-point self-interactions for the SU(3) and SU(2) gauge bosons. The $U(1)_Y$ boson has no self-couplings.

The fermion terms of $\mathcal{L}_f$ involve 3 families of quarks and leptons, of which each consists of the left-chiral SU(2) doublets (up and down type quarks, or electron and neutrino type leptons) and the right-chiral singlets. Thus the SU(2) group is not invariant under the parity symmetry. The elements of each doublet transform into each other under the SU(2) symmetry. All the fermion fields except the right-handed neutrinos carry the weak hypercharge, $Y = Q - T_3$, where Q is the electric charge and T_3 is the third generator of the SU(2) group. Since the SU(2) and $U(1)_Y$ representations are chiral, no fermion mass terms are allowed*, and thus the fermion Lagrangian consists of gauge-covariant kinetic energy terms only.

The Higgs Lagrangian ($\mathcal{L}_\phi$) consists of a complex Higgs scalar field (ϕ) and the Higgs potential. The gauge covariant derivative allows three- and four-point interactions between the gauge and Higgs fields. The Higgs potential is described in fourth order of the Higgs fields. The vacuum expectation value of the Higgs field ($\langle 0|\phi|0\rangle$) is non-zero after spontaneous symmetry breaking, which then generates the SU(2) gauge boson masses. The potential gives the quartic self-coupling of the Higgs fields.

The Yukawa Lagrangian ($\mathcal{L}_Y$) represents the Yukawa couplings between the Higgs fields and the fermions. The Yukawa couplings are required to generate fermion masses by the spontaneous breaking of the chiral gauge symmetries, i.e., ultimately determine the fermion masses and mixing angles which are in turn free parameters in the SM. A Yukawa coupling vertex conserves the electric charge while chirality is flipped.

Explicit formulae for the fermion, Higgs and Yukawa Lagrangians can be found in e.g. [37, p.258-276]. The given formulae describe the unbroken phase of the SM and change after the spontaneous symmetry breaking. If a neutral component of the Higgs fields acquires a non-zero vacuum expectation value after the breaking, the $SU(2)\times U(1)_Y$ electroweak symmetry is broken to the U(1) symmetry of QED, generating masses of the bosons. W^α and B are combined to become the massive $W^\pm$ and Z boson and the massless photon (now denoted by γ), where Z and γ are composed with the weak angle defined to be

$$\sin\theta_W = \frac{g'}{g_Z}, \quad \cos\theta_W = \frac{g}{g_Z}, \quad \text{and} \quad g_Z = \sqrt{g'^2 + g^2}. \tag{2.24}$$

Then $e = g_W \sin\theta_W = g' \cos\theta_W$ where $g_W = g$. Including higher-order corrections, the boson masses are predicted to be $M_W \sim 80$ GeV and $M_Z \sim 91$ GeV. The weak scale, which is the vacuum expectation value, is determined to be ~ 246 GeV. The fermions have electric charges, $Q = Y + T_3$, after the breaking.

The SM interaction vertices between fermions and bosons are shown in Figure 2.4. The coupling constant (e, g_s, or $g_{W/Z}$) can also be expressed by a dimensionless constant, which is defined to be independent of the system of units used. For instance, the fine-structure constant [38, p.8]

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} \tag{2.25}$$

is introduced as the intrinsic strength for the electromagnetic interaction. The QCD interaction has intrinsically stronger strength with $\alpha_s \approx 1$. The intrinsic strength of the weak interaction is also greater than that of QED with $\alpha_{W/Z} \approx 1/30$, but due to the large mass of the associated bosons the effective strength is weaker.

*There is a possibility of Majorana mass terms for neutrinos.

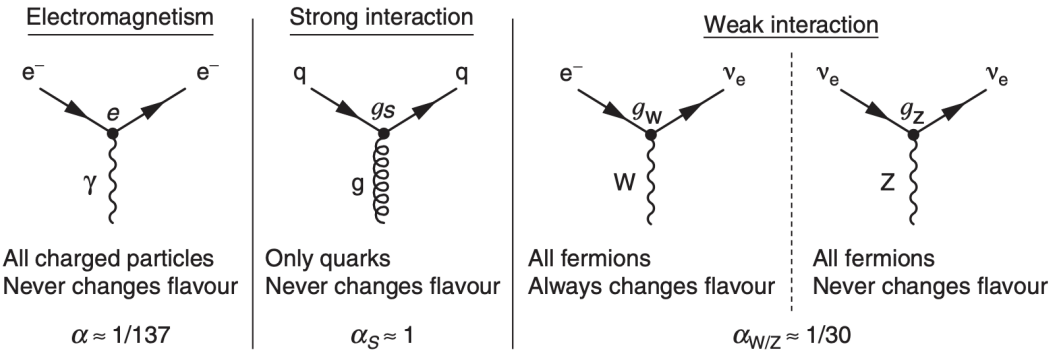
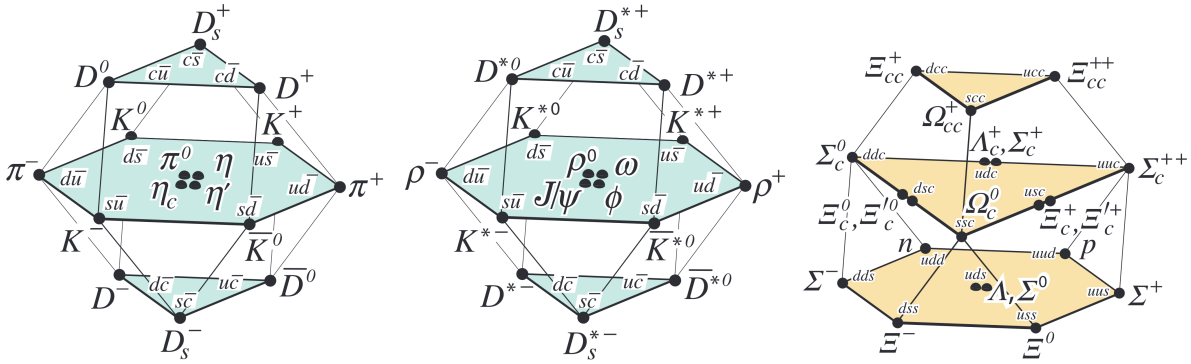


Figure 2.4: The SM vertices between fermions and bosons. Figure is taken from [38].



(a) Pseudoscalar mesons. (b) Vector mesons. (c) Baryons.

Figure 2.5: Hadrons made of u, d, s and c quarks, with figures taken from [40]. Particles denoted by D and with subsript c, including J/ψ of $c\bar{c}$, are called *charm hadrons*.

2.2 Heavy-quark fragmentation

Many experiments have attempted to detect free quarks, however, they never have been observed directly. To explain the non-observable free quarks, the hypothesis of *colour confinement* [38, p.248-253] is proposed, although there is no analytic proof of the concept yet. This hypothesis states that at low energies colour-charged particles cannot be isolated and are always confined to colour singlet states. The quarks carry colour charge, and their interactions with virtual gluons are attractive. These interactions squeeze the colour field between the quarks into a tube where the energy density between the quarks is constant at relatively large distance. This results in an energy proportional to the separation of the quarks. This means that it would require an infinite amount of energy to separate two quarks to infinity. Consequently, in detectable final states, the colour-charged particles are bound to each other in a *hadronic* state which is colourless and there is no colour field between hadrons. With the confinement hypothesis, gluons are also confined to colourless states and do not propagate over large distances unlike photons.

In other words, quarks are observed in experiments always as hadronic states. For instance, hadrons made of u, d, s and c quarks are shown in Figure 2.5 (further details of quark model in [40]). The hadrons containing at least one $c(\bar{c})$ in Figure 2.5 are called *charm hadrons* (further details of the charm hadrons in [41]). Beauty hadrons can be found in e.g. [42]. Top quarks decay before hadronization with short lifetimes.

The associated hadronization process can be described by *fragmentation functions* $D_i^h(x, \mu^2)$ $(i = q, \bar{q}, g)$ which represent a measure of the probability density that an outgoing parton produces a hadron h [42]. Here, x is the fraction of the parton's momentum transferred to the hadron and μ is factorization scale. The fragmentation functions satisfy the standard Altarelli-

Parisi equation (a.k.a, DGLAP evolution equation), of which initial values at a scale μ_0 of the order of the heavy-quark mass are perturbatively calculable [43]. However, the actual form of the fragmentation functions is non-perturbative at the pole mass scale and below. Inclusion of the non-perturbative effects is done by convoluting the perturbative result with a phenomenological non-perturbative form, so-called *non-perturbative fragmentation function* $D^{\mathrm{NP}}(x)$.

There are generally used D^{NP}s for both charm and beauty like e.g. Peterson [44] and Kartvelishvili [45] which have single non-perturbative parameters. Specifically, the Kartvelishvili function is defined as

$$D^{\mathrm{NP}}(x) = (\alpha_K + 1)(\alpha_K + 2)x^{\alpha_K}(1 - x), \tag{2.26}$$

where α_K is a non-perturbative parameter. The non-perturbative parameters need to be fitted together with some model of hard radiation, which can be either a shower Monte Carlo (MC), a leading-log or next-to-leading log calculation, or a fixed order calculation [42]. Depending on the convoluted model, the fitted parameters can differ. More dedicated parametrizations for charm are also available by so-called BCFY functions [4, 46] for each D^0, D^+ and D^{*+}, of which the underlying theory is introduced separately for pseudo-scalar and vector states (further details can be found in [46]). All these functions are described by a single parameter r. In this thesis, the BCFY functions were taken as the reference for kinematic distribution of charm hadrons (specifically for D^0 and D^{*+}). The non-perturbative parameter for these functions was extracted by fitting to measurements, of which examples will be discussed further in the next chapter.

The probability of a quark to fragment into a hadron state is determined by experiments, of which the results are given by so-called *fragmentation fractions*. Since the total charm-pair cross section in e^+e^- collisions ($\sigma_{e^+e^-\to c\bar{c}}$) can be precisely calculated in the SM [47] (refer also to the next section), the fragmentation fractions in e^+e^- collisions can be defined as

$$f(c \to H) = \frac{\sigma_{H_c}}{\sigma_{e^+e^-\to c\bar{c}}}, \tag{2.27}$$

where σ_{H_c} is the total charm-hadron cross section. In contrast, the kinematics of the final states cannot be clearly known with proton(s) in the initial states in *ep* and *pp* collisions. As a result, the production predictions cannot be calculated as precisely as in e^+e^- collisions, and the fragmentation fractions should be measured by considering all the final states. The fragmentation fractions are defined then as

$$f(c \to H) = \frac{\sigma_{H_c}}{\Sigma_{w.d.}\sigma_{H_c}}, \tag{2.28}$$

where the sum is given over all the known weakly-decaying ground states.

2.3 Heavy-quark cross section in electron-positron annihilation

By taking advantage of well known kinematics for the initial states in e^+e^- collisions, non-perturbative fragmentation functions and fractions have been mostly determined based on e^+e^- data. For instance, hadrons can be measured in variables of the scaled energy (x_E) or the scaled momentum (x_p) [48] defined to be

$$x_E = \frac{E}{E^{\mathrm{max}}}, \quad x_p = \frac{|\vec{p}|}{|\vec{p}^{\mathrm{max}}|} \tag{2.29}$$

where $E^{\mathrm{max}} = \sqrt{s}/2$ and $\vec{p}^{\mathrm{max}} = \sqrt{s/4 - m_H^2}$ (m_H is mass of hadron). This leads to measurements for the non-perturbative fragmentation functions (see Section 3.3). Furthermore, production cross sections in e^+e^- collisions can be predicted precisely compared to those in *ep* and *pp* collisions involving proton(s) as initial state. As a result, quantities like ratios of

the production cross sections between a lepton pair and hadrons are well known both experimentally and theoretically, which then can be used as reference for e.g. fragmentation fraction measurements. Since these measurements were taken as inputs for the extrapolation in Chapter 6, these quantities will be introduced further in this section.

Let me start with introducing predictions for a lepton-pair production represented by a QED annihilation process of $e^+e^- \to \mu^+\mu^-$. To compute cross sections, the Lorentz-invariant matrix element, $\mathcal{M}$, needs to be derived first. Using Feynman rules, $-i\mathcal{M}$ can be immediately derived by producting all elements present in Feynman diagram. For instance, $e^+e^- \to \mu^+\mu^-$ has a single Feynman diagram at the lowest order (i.e., Leading Order, LO), which is shown in Figure 2.6. The amplitude of Figure 2.6 can be written by following the QED Feynman rules shown in Figure 2.2:

$$i\mathcal{M} = \Big(u(p_1)(-i)e\gamma^\mu \bar{v}(p_2)\Big)\left(\frac{-ig_{\mu\nu}}{(p_1+p_2)^2}\right)\Big(\bar{u}(p_3)(-i)e\gamma^\nu v(p_4)\Big). \tag{2.30}$$

If the Mandelstam variables

$$\begin{aligned}(p_1+p_2)^2 &\equiv s,\\ (p_1-p_3)^2 &\equiv t \text{ and}\\ (p_1-p_4)^2 &\equiv u\end{aligned} \tag{2.31}$$

are introduced in the high energy limit ($\sqrt{s} \gg m_e, m_\mu$), the spin-averaged amplitude squared of Eq.(2.30) can be written as

$$\frac{1}{4}\sum |\mathcal{M}|^2 = 2e^4\left(\frac{t^2+u^2}{s^2}\right). \tag{2.32}$$

This results in the total cross section at the lowest order [38, p.137]:

$$\sigma = \frac{4\pi\alpha^2}{3s}. \tag{2.33}$$

In addition to the lowest-order diagram of Figure 2.6, there are an infinite number of higher-order diagrams resulting in the same final states. For example, one-loop corrections are shown in Figure 2.7. The matrix element is then given by the sum of all individual amplitude:

$$\mathcal{M} = \alpha M_{\mathrm{LO}} + \alpha^2 \sum_j M_{1,j} + \alpha^3 \sum_j M_{2,j} + \cdots, \tag{2.34}$$

where $M_{1,j}$ and $M_{2,j}$ are the amplitudes factored out of α with 4 and 6 interaction vertices, respectively. This results in the QED *perturbative* expansion in terms of α for the total amplitude ($|\mathcal{M}|^2$), where e.g. α^2, α^3 and α^4 terms refer to the LO, NLO and NNLO terms, respectively. Since α is less than 0.01, the perturbative theory results in the $\mathcal{O}(99\%)$ accuracy already at LO.

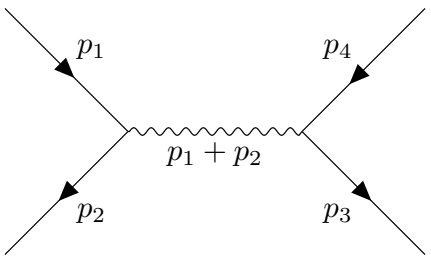


Figure 2.6: The lowest-order Feynman diagram for $e^+(p_2)e^-(p_1) \to \mu^+(p_4)\mu^-(p_3)$.

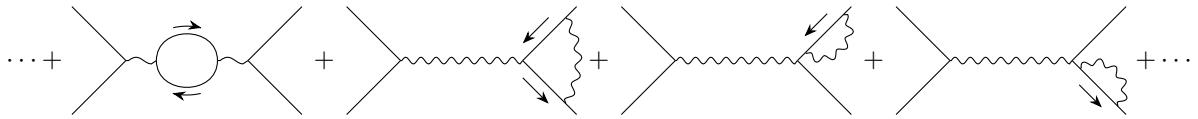


Figure 2.7: Examples of one-loop correction for $e^+e^- \to \mu^+\mu^-$.

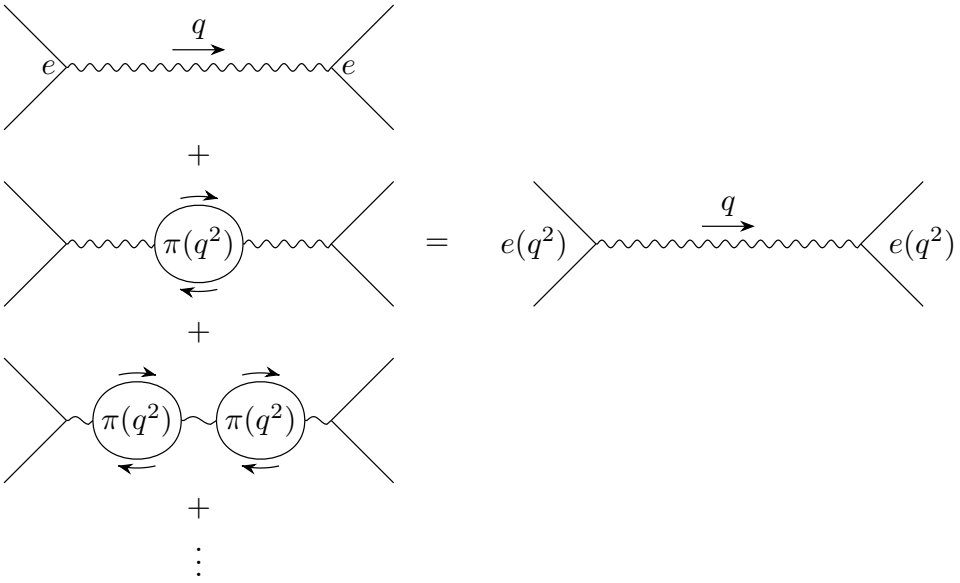


Figure 2.8: The photon self-energy terms absorbed into coupling constant.

In the case of QCD perturbation, $\alpha_s = g_s^2/4\pi$ requires much higher-order terms to achieve the same precision as QED predictions at LO.

However, the loop corrections on the photon propagator (the first term in Figure 2.7) result in virtual fermions and eventually give divergent results. The infinite series of these corrections are called the photon self-energy terms. The photon propagator with these self-energy terms is expressed by an effective formula [38, p.256]:

$$P \equiv \frac{e(q^2)^2}{q^2} = \frac{e^2}{q^2}\frac{1}{1 - e^2\Pi(q^2)} \tag{2.35}$$

where $\pi(q^2) = q^2\Pi^2$ is a correction factor of each loop. With the physical electron charge which is known at a reference scale μ, the infinities are renormalized away and the coupling strength is given by a running coupling constant

$$\alpha(q^2) = \frac{\alpha(\mu^2)}{1 - \alpha(\mu^2)\dfrac{1}{3\pi}\ln\left(\dfrac{q^2}{\mu^2}\right)}. \tag{2.36}$$

In this way, the infinities associated with the self-energy terms are absorbed into the definition of Eq.(2.36) as illustrated in Figure 2.8. With this running coupling, the perturbative theory now can be applied for the physical quantity.

In analogy with lepton-pair production, the LO amplitude of quark-pair production, $e^+e^- \to q\bar{q}$, can be calculated this time with different coupling constants. As a result, the total cross section at LO is [38, p.260]

$$\sigma(e^+e^- \to q\bar{q}) = 3 \times \frac{4\pi\alpha^2}{3s}Q_q^2 \tag{2.37}$$

where Q_q is the electric charge of quarks. Here the factor of 3 comes from the sum over the three possible colour combinations of the $q\bar{q}$ pair.

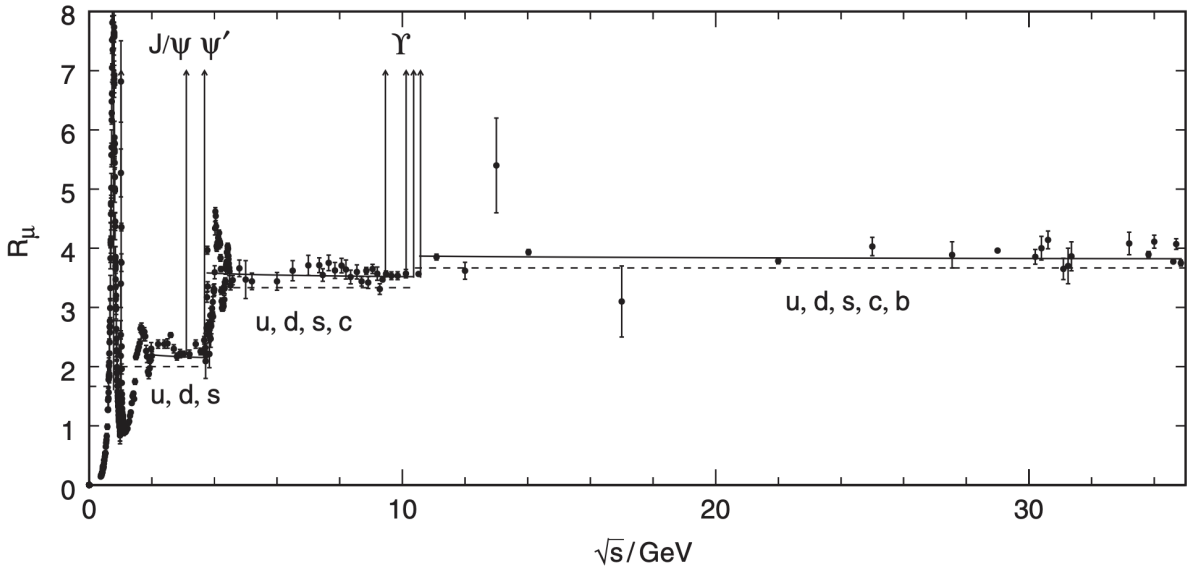


Figure 2.9: R_μ comparison with experimental measurements as a function of $\sqrt{s}$. Figure is taken from [38].

Then, the *inclusive* hadronic cross section, $\sigma(e^+e^- \to \text{hadrons})$, can be given by the sum of $\sigma(e^+e^- \to q\bar{q})$ (Eq.(2.37)) over the flavours [38, p.261]:

$$\sigma(e^+e^- \to \text{hadrons}) = \frac{4\pi\alpha^2}{3s} \times 3 \sum_{\text{flavours}} Q_q^2. \tag{2.38}$$

Here note that this is given at a center-of-mass energy $\sqrt{s} > 2m_q$. A more convenient formula can be given by taking a ratio relative to the lepton-pair cross section, e.g.:

$$R_\mu \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} = 3 \sum_{\text{flavours}} Q_q^2. \tag{2.39}$$

A comparison to experimental measurements as a function of $\sqrt{s}$ is shown in Figure 2.9. The dashed lines indicate the LO results by Eq.(2.39). For instance, above the thresholds of $c\bar{c}$ production (~ 3 GeV), Eq.(2.39) results in

$$R_\mu^c = 3 \times (\frac{1}{9} + \frac{4}{9} + \frac{1}{9} + \frac{4}{9}) = \frac{10}{3}, \tag{2.40}$$

where u, d, s and c are accounted for in the calculation.

However, as it is shown in Figure 2.9, the LO predictions (the dashed lines) disagree with the measurements at the level of approximately 10 %. To resolve this discrepancy, higher order corrections should be added. Examples of NLO corrections for this process are shown in Figure 2.10. Since the QED higher-order corrections are relatively small compared to the QCD corrections (α is much smaller than α_s as mentioned), the dominant correction comes from the QCD corrections. If the first-order QCD correction from $e^+e^- \to q\bar{q}g$ is added, the prediction is modified to [38, p.263]

$$R_\mu = 3 \times \left(1 + \frac{\alpha_s(Q^2)}{\pi}\right) \sum_{\text{flavours}} Q_q^2. \tag{2.41}$$

This then gives an excellent agreement with the measurements as shown in Figure 2.9 by the solid lines. Further higher order QCD corrections can be found in [49], although the effects are not significant. These well known quantities were used in [47] to verify the total charm cross

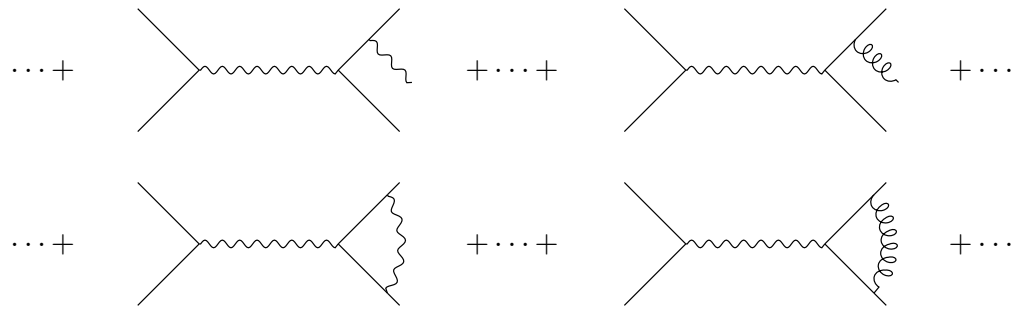


Figure 2.10: Examples of NLO corrections for the inclusive hadronic cross section $\sigma(e^+e^- \rightarrow$ hadrons).

section calculation which was taken as an input to measure fragmentation fractions in e^+e^- collisions according to Eq.(2.27). Here the fragmentation fraction measurements were used as inputs for the extrapolation in Chapter 6.

2.4 Heavy-quark cross section in hadron-hadron scattering

In this section, predictions for heavy-quark production in pp collisions will be discussed up to the highest orders known today. Predictions of the inclusive total cross section will be discussed up to NNLO in Section 2.4.1, which were taken for comparison to total charm cross section measurements and for constraining QCD parameters in Chapter 6. Differential cross section predictions up to NLO+NLL will be introduced in Section 2.4.2, which were taken for comparison to D^{*+} and D^0 measurements as a function of transverse momentum and rapidity in Chapter 5 and for the extrapolation of the measurements in Chapter 6.

In case of collisions involving (anti-)protons, the exact momentum and even the flavour of initial states are not clearly known due to colour confinement. Therefore, experimental inputs are required to account for the non-perturbative phenomena not only in final states, but also in initial states, unlike the previous e^+e^- case. PDFs, which were determined by fitting to experimental data, describe the probability density of finding a parton in the hadronic initial states.

In addition, the non-perturbative fragmentation functions and fractions are applied in the differential cross section predictions for the purpose of comparison to measurements. The differential cross section predictions were modified for the extrapolation to treat non-universal charm fragmentation, of which details will be given in Chapter 6.

2.4.1 Total cross section

A process of heavy-quark production in pp collisions can be given by

$$H_1(P_1) + H_2(P_2) \rightarrow Q(p_3) + \overline{Q}(p_4) + X \tag{2.42}$$

where X is an inclusive hadronic final state which also contains the beam remnants from the initial hadrons, H_1 and H_2. Here heavy quarks are distinguished from light quarks (u, d and s) and denoted by Q.

Based on the so-called QCD factorization, the perturbative and non-perturbative regime can be factorized at the so-called *factorization scale*, μ_f. This means that the cross section can be given by the convolution of the perturbative partonic cross section ($\hat{\sigma}_{ij}$) and the partially non-perturbative PDFs ($f_{i,j}$) for the initial partonic states i and j:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_f) f_j(x_2, \mu_f) \hat{\sigma}_{ij}(s, m, \mu_f, \alpha_s), \tag{2.43}$$

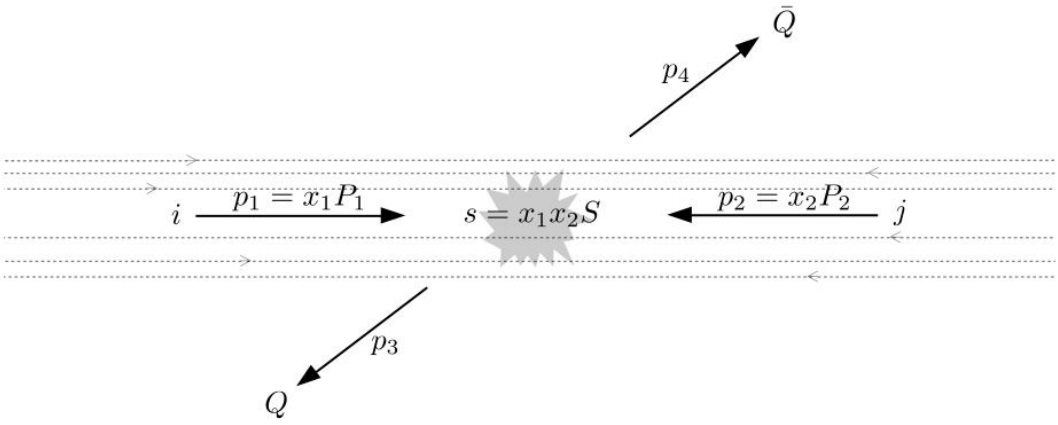


Figure 2.11: Diagram for heavy-quark production in hadron-hadron scattering. Two hadrons with four-momenta P_1 and P_2 collide at the hadronic center-of-mass energy $\sqrt{S}$. The hard scattering is induced by two initial partons i and j with four-momenta $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$, respectively. As a result one heavy quark Q and its anti-quark $\overline{Q}$ are produced with momenta p_3 and p_4, respectively. The dashed lines indicate partons which do not participate in the hard scattering.

where m indicates the heavy-quark mass. The initial partons with momenta, $x_1 P_1$ and $x_2 P_2$, participate in the hard scattering (refer to Figure 2.11) and contribute to the partonic cross section. With the initial partons being almost massless relative to the heavy final states, the partonic center-of-mass energy is $\sqrt{s} = \sqrt{x_1 x_2 S}$ where $\sqrt{S} = P_1 + P_2$ is the hadronic center-of-mass energy.

In the fixed order QCD perturbation theory, the partonic cross section $\hat{\sigma}$ in Eq.(2.43) can be written in terms of so-called scaling functions f_{ij} which are dimensionless [5]:

$$\hat{\sigma}_{ij}(s, m, \mu, \alpha_s) = \frac{\alpha_s^2}{m^2} f_{ij}(\rho, \mu) \tag{2.44}$$

where

$$f_{ij} = f_{ij}^{(0)} + 4\pi\alpha_s f_{ij}^{(10)} + (4\pi\alpha_s)^2 f_{ij}^{(20)} + \mathcal{O}(\alpha_s^3), \tag{2.45}$$

and $\rho = \dfrac{4m^2}{s}$. To eliminate ultraviolet divergences in the QCD perturbation theory, renormalization should be introduced, in analogy with the QED case in Section 2.3. As a result, a running coupling constant in QCD is introduced in leading order as

$$\alpha_s(q^2) = \frac{\alpha_s(\mu_r^2)}{1 + b_0 \alpha_s(\mu_r^2) \ln\left(\dfrac{q^2}{\mu_r^2}\right)}, \quad b_0 = \frac{33 - 2N_f}{12\pi} \tag{2.46}$$

where μ_r is *renormalization scale* (the derivation of this equation can be found in Appendix K).

For an arbitrary parameter μ, the cross section, which is the physical observable, should satisfy

$$\mu^2 \frac{d}{d\mu^2} \sigma = 0. \tag{2.47}$$

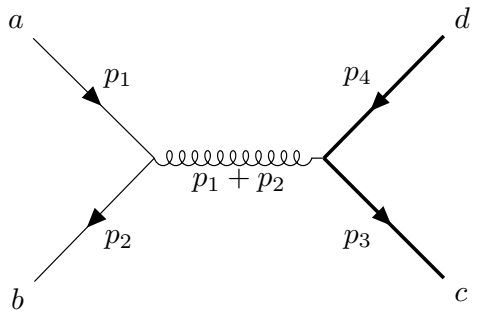


Figure 2.12: Feynman diagram for $q\bar{q} \to Q\overline{Q}$. Thin lines indicate light quarks and thick ones are heavy quarks. Indices, a, b, c and d denote colours.

A solution of Eq.(2.47) can be given with the two arbitrary parameters, μ_r and μ_f:

$$\begin{aligned}\alpha_s(\mu_f^2) = \alpha_s(\mu_r^2)\Bigg[1 - \alpha_s(\mu_r^2)\Bigg\{b_0 \ln\left(\frac{\mu_f^2}{\mu_r^2}\right)\Bigg\} \\ + \alpha_s^2(\mu_r^2)\Bigg\{b_0^2 \ln^2\left(\frac{\mu_f^2}{\mu_r^2}\right) - b_1 \ln\left(\frac{\mu_f^2}{\mu_r^2}\right)\Bigg\} \\ + \alpha_s^3(\mu_r^2)\Bigg\{3b_0 b_1 \ln^2\left(\frac{\mu_f^2}{\mu_r^2}\right)\Bigg\} + \mathcal{O}(\alpha_s^4)\Bigg]\end{aligned} \tag{2.48}$$

where $\mu_f \sim \mu_r$. The details to derive Eq.(2.48) can be found in Appendix K. The scaling functions are now expressed as [5]

$$\begin{aligned}f_{ij} = f_{ij}^{(0)} + 4\pi\alpha_s\Bigg\{f_{ij}^{(10)} + f_{ij}^{(11)} \ln\left(\frac{\mu_f^2}{m^2}\right) + 2\frac{b_0}{4\pi} f_{ij}^{(0)} \ln\left(\frac{\mu_r^2}{\mu_f^2}\right)\Bigg\} \\ + (4\pi\alpha_s)^2\Bigg\{f_{ij}^{(20)} + f_{ij}^{(21)} \ln\left(\frac{\mu_f^2}{m^2}\right) + f_{ij}^{(22)} \ln^2\left(\frac{\mu_f^2}{m^2}\right) \\ + \left(3\frac{b_0}{4\pi} f_{ij}^{(10)} + 2\frac{b_1}{(4\pi)^2} f_{ij}^{(0)}\right) \ln\left(\frac{\mu_r^2}{\mu_f^2}\right) + 3\left(\frac{b_0}{4\pi}\right)^2 f_{ij}^{(0)} \ln^2\left(\frac{\mu_r^2}{\mu_f^2}\right) \\ + 3\frac{b_0}{4\pi} f_{ij}^{(11)} \ln\left(\frac{\mu_r^2}{\mu_f^2}\right) \ln\left(\frac{\mu_f^2}{m^2}\right)\Bigg\} + \mathcal{O}(\alpha_s^3)\end{aligned} \tag{2.49}$$

with Eq.(2.48).

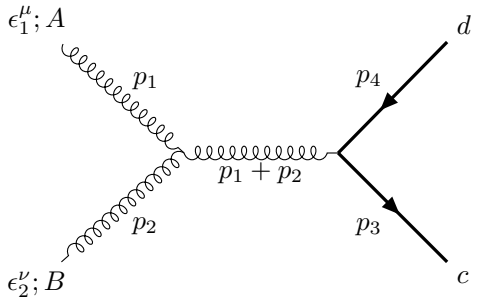
The partonic cross section at LO includes two subprocesses: quark-antiquark annihilation and gluon-gluon fusion.

First, the quark-antiquark annihilation includes one Feynman diagram which is shown in Figure 2.12. Following the Feynman rules shown in Figure 2.3, the matrix element of Figure 2.12 is given by

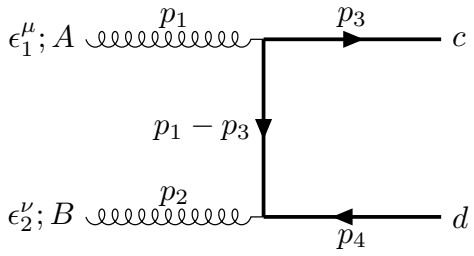
$$i\mathcal{M} = \Big(u(p_1)(-i)g_s t_{ba}^A \gamma^\mu \bar{v}(p_2)\Big)\left(\frac{-i}{(p_1+p_2)^2}\right)\Big(\bar{u}(p_3)(-i)g_s t_{cd}^A \gamma_\mu v(p_4)\Big), \tag{2.50}$$

where $t_{ij}^A \equiv c_i^\dagger t^A c_j$ $(i,j = 1, \cdots, 3)$ with $c_{i,j}$ are colour vectors for quarks. It was shown in [50] that this results in the initial quark spin- and colour-averaged amplitude-squared

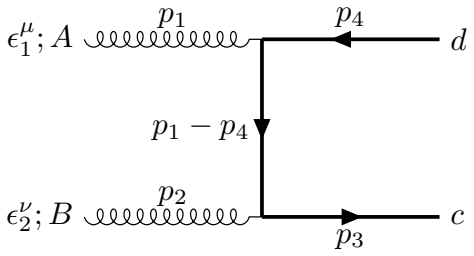
$$\frac{1}{4}\sum |\mathcal{M}|^2 = \frac{4(4\pi\alpha_s)^2}{s^2}(t_1^2 + u_1^2 + 2m^2 s) \tag{2.51}$$



(a) s-channel



(b) t-channel



(c) u-channel

Figure 2.13: Feynman diagrams for the gluon-gluon fusion. Thick lines indicate the heavy quarks. Indices c and d are for the singlet colours and A and B are for the octet colours. ϵ indicates the gluon polarization where $\epsilon_i \equiv \epsilon(p_i),\ i = 1, 2$.

with kinematic variables defined to be

$$\begin{aligned}(p_1 + p_2)^2 &\equiv s,\\ (p_1 - p_3)^2 - m^2 &\equiv t_1 \text{ and}\\ (p_1 - p_4)^2 - m^2 &\equiv u_1.\end{aligned} \tag{2.52}$$

Eventually, $\hat{\sigma}_{q\bar{q}}$ at LO can be expressed as [5]

$$\hat{\sigma}_{q\bar{q}} = \alpha_s^2 \frac{4\pi^3}{27} \frac{1}{s} \beta(3 - \beta^2) \tag{2.53}$$

where $\beta = \sqrt{1 - \rho}$.

In case of the gluon-gluon fusion process, three Feynman diagrams which are shown in Figure 2.13 need to be considered. The amplitude for these diagrams is

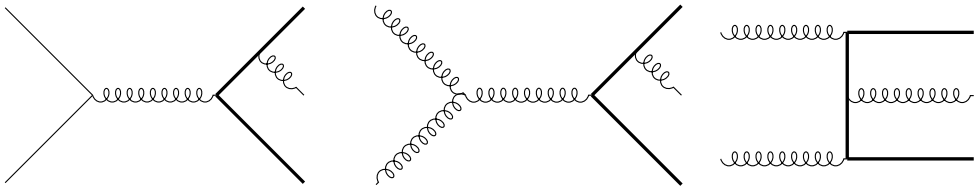
$$\begin{aligned}i\mathcal{M}_s = &- g_s f^{ABC}[g^{\mu\nu}(p_1 - p_2)^\rho + g^{\nu\rho}(2p_2 + p_1)^\mu - g^{\rho\mu}(2p_1 + p_2)^\nu]\epsilon_1^\mu \epsilon_2^\nu\\ &\times \left(\frac{-i}{(p_1 + p_2)^2}\right) \bar{u}(p_3)(-i)g_s t_{cd}^C \gamma^\rho v(p_4)\end{aligned} \tag{2.54}$$

for the s-channel,

$$i\mathcal{M}_t = \bar{u}(p_3)(-i)g_s t_{ce}^A \gamma^\mu \epsilon_1^\mu \left[\frac{i(\not{p}_1 - \not{p}_3 + m)}{(p_1 - p_3)^2 - m^2}\right](-i)g_s t_{ed}^B \gamma^\nu \epsilon_2^\nu v(p_4) \tag{2.55}$$

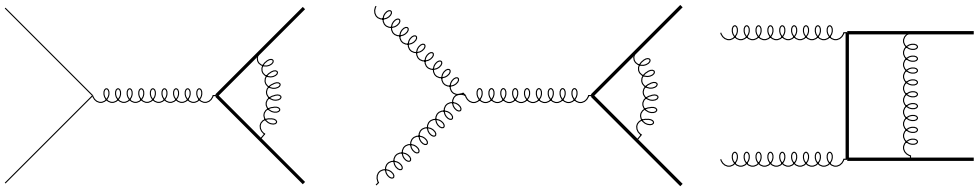
for the t-channel and

$$i\mathcal{M}_u = \bar{u}(p_3)(-i)g_s t_{ce}^B \gamma^\nu \epsilon_2^\nu \left[\frac{i(\not{p}_1 - \not{p}_4 + m)}{(p_1 - p_4)^2 - m^2}\right](-i)g_s t_{ed}^A \gamma^\mu \epsilon_1^\mu v(p_4) \tag{2.56}$$



(a) $q\bar{q}$-annihilation. (b) gg-fusion.

Figure 2.14: Examples of real emissions.



(a) $q\bar{q}$-annihilation. (b) gg-fusion.

Figure 2.15: Examples of virtual corrections.

for the u-channel, where ϵ is the gluon polarization. These result in six amplitude-squared terms for all the possible combination of the s, t and u channels, which in turn can give e.g. a differential cross section [50]

$$s^2 \frac{d^2\hat{\sigma}_{gg}}{dt_1 du_1} = \pi\alpha_s^2 \delta(s + t_1 + u_1) \times \left[\frac{-3}{8}\left(\frac{t_1^2 + u_1^2}{s^2} + \frac{4m^2}{s}\right) + \frac{1}{6}\left(\frac{u_1}{t_1} + \frac{t_1}{u_1} + \frac{9m^4 + 4m^2 s}{t_1 u_1}\right) - \frac{2}{3}\frac{m^4 s^2}{t_1^2 u_1^2}\right]. \quad (2.57)$$

$\hat{\sigma}_{gg}$ at LO can be expressed as [5]

$$\hat{\sigma}_{gg} = \frac{\pi^3}{48}\frac{1}{s}\left\{(33 - 18\beta^2 + \beta^4)\ln\left(\frac{1+\beta}{1-\beta}\right) - 59\beta + 31\beta^3\right\}. \quad (2.58)$$

Higher order predictions beyond LO include gluon real emissions and virtual corrections. Examples for NLO corrections with real emissions and virtual corrections are shown in Figure 2.14 and 2.15, respectively. Furthermore, quark-gluon scattering process starts to contribute beyond LO, of which examples are shown in Figure 2.16. To complete the NLO prediction, at least total 37 master integrals are required [51], which is a challenging computation. The NLO prediction is completely known at NLO to date for the inclusive total cross section, of which analytic results can be found in [51].

In case of gluon real emission, if a gluon is emitted with low energy (*soft*) or in the same direction as the original quark (*collinear*), large logarithmic contributions appear. The logarithms of the form $\alpha_s^n \ln^m \beta$ ($m \leq 2n$ and $\beta \to 0$ at threshold limit) [52] can be resummed

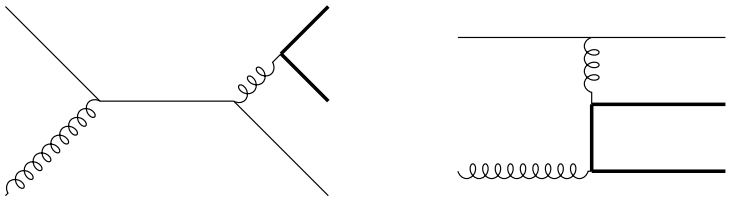


Figure 2.16: Examples of qg-scattering at $\mathcal{O}(\alpha_s^3)$.

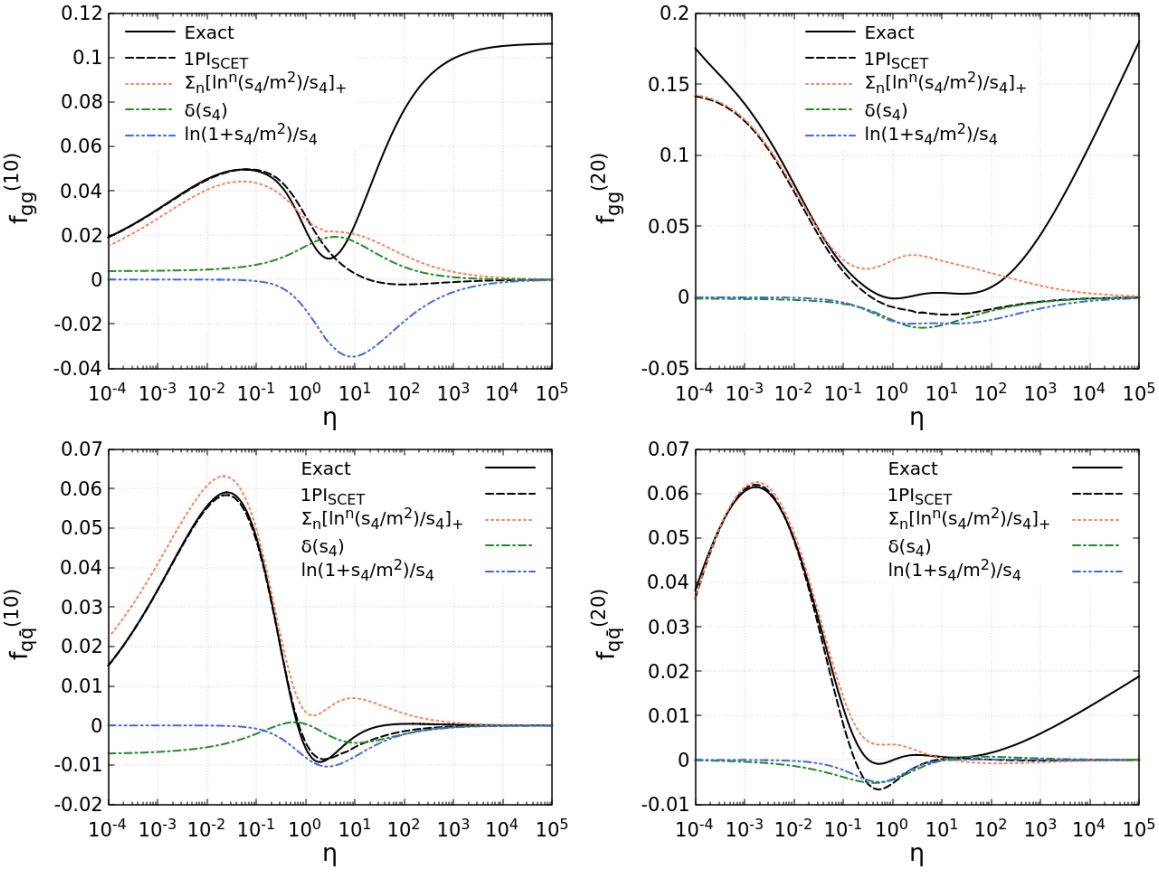


Figure 2.17: Comparison of approximate (labeled as "$1PI_{SCET}$") and complete (labeled as "Exact") NNLO for gg-fusion and $q\bar{q}$ annihilation. The complete NNLO includes both the threshold and high energy enhancement, while the approximate NNLO contains threshold corrections only. $\eta = 1/\rho - 1$. Figures are from [50].

using the soft and collinear factorization [53,54]. This resummation can improve the agreement with respect to the experimental result where low energy contributions dominate. However, for high energy experiments, e.g., at the LHC, large logarithmic contributions appear apart from threshold-enhanced logarithms. In the high energy regime, partons scatter at a partonic center-of-mass energy which is much larger than the mass scale ($s \gg m$) and $\rho \to 0$. In this limit, resummation of the high-energy logarithms can be introduced in the framework of PDFs unintegrated in the transverse momentum k_T [55] and the concept of k_T-factorization [55–60]. This resummation accounts for the high energy contribution. These logarithmic corrections can be used e.g. to approximate NNLO.

Today, the prediction of the total cross section is completely known up to NNLO. In other words, all coefficient functions up to NNLO in Eq.(2.49) are known; $f_{ij}^{(10)}$ and $f_{ij}^{(11)}$ in [61], $f_{q\bar{q}}^{(20)}$ in [62, 63], $f_{gg}^{(20)}$ in [5, 64], $f_{qg}^{(20)}$ in [65], and $f_{ij}^{(21)}$ and $f_{ij}^{(22)}$ in [66]. In Figure 2.17, this complete NNLO (the solid black curve) is compared to approximate NNLO (the dashed black curve labeled as "$1PI_{SCET}$") which includes the threshold corrections only, to show the effects of threshold or high energy enhancement. The approximate NNLO is derived by integrating differential cross section provided in [52], of which detailed calculations were shown in [50]. Figure 2.17 shows that the two predictions agree well at the low energy scale (i.e., η is small where $\eta = 1/\rho - 1$). At the high energy scale (η is large), the approximate NNLO has zero contribution by definition, while the complete NNLO contains the significant high energy corrections. The effect of NNLO correction at high energy is more significant in charm production compared to

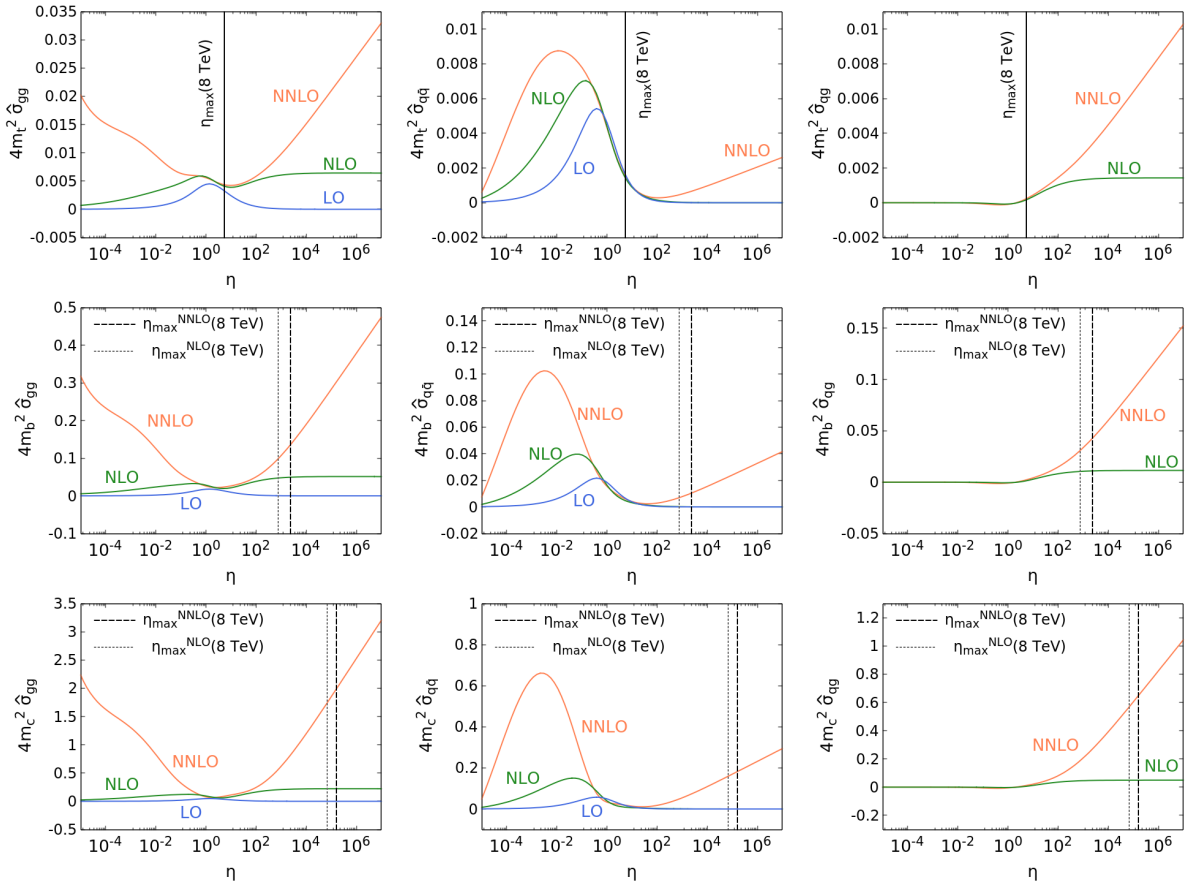


Figure 2.18: Partonic total cross sections of gg, $q\bar{q}$ and qg processes at LO, NLO and NNLO with top (top), beauty (middle) and charm (bottom) pole masses. $\eta = 1/\rho - 1$, and $\eta_{\max}$ refers to a reference scale indicating 95% of total cross section. Figures are from [50].

beauty and top ones. In Figure 2.18, the cross sections of gg, $q\bar{q}$ and qg processes are shown at LO, NLO and NNLO with different masses of charm, beauty and top. Referring to $\eta_{\max}$, which indicates 95% of the total cross section as a reference scale, contributions of the high energy corrections especially at NNLO are significant in charm production. To date, the predictions are known up to NNLO also for the differential cross section in case of top [1] and beauty [2] production, while those for charm production are still missing. It is shown that predictions of e.g. total beauty cross section derived from the theory given in [2] agree well with the complete NNLO predictions (refer to [2]). At the end of this thesis, the total charm cross section measured from LHC experiments will be compared with the complete NNLO prediction.

2.4.2 Differential cross section

With the non-perturbative fragmentation function and fraction, the differential cross section is given for the heavy-quark hadron (H_Q) production as

$$d\sigma_{H_Q} = f_{H_Q} \cdot (d\sigma_Q \otimes D^{\mathrm{NP}}_{Q\to H_Q}), \quad d\sigma_Q = f_i f_j \otimes d\hat{\sigma}_{ij}. \tag{2.59}$$

where f_{H_Q} is the fragmentation fraction, $d\sigma_Q$ is the differential cross section of heavy-quark production and $D^{\mathrm{NP}}_{Q\to H_Q}$ is the non-perturbative fragmentation function which is factorized out from $d\sigma_Q$, again using the QCD factorization theorem.

To date, the highest available order for differential cross section predictions as a function of transverse momentum (p_T) or (pseudo-)rapidity ($(\eta)y$) for charm is NLO+NLL which is also

known as FONLL [4, 43]. In this thesis, the prediction of the differential cross section is thus provided by FONLL for comparisons to and extrapolation of measurements in Chapter 5 and 6. FONLL is based on resummation of the logarithms of the form $\alpha_s^2(\alpha_s \log p_T/m)^k$ (Leading Log, LL) and $\alpha_s^3(\alpha_s \log p_T/m)^k$ (Next-to-Leading Log, NLL) with μ_r and μ_f defined of the order of p_T [43]. This resummation is based on the evolution of perturbative fragmentation functions via the standard Altarelli-Parisi equation (refer also to Section 2.2). [43] says:

> The perturbative fragmentation functions, evolved up to any scale $\mu \sim p_T$ via the Altarelli-Parisi equations, can be used to evaluate heavy-quark cross sections in the large-transverse-momentum region by convoluting them with short-distance cross sections for massless partons, subtracted in the $\overline{\text{MS}}$ scheme. The heavy quark is also treated as a massless active flavour, and therefore also appears in the parton distribution functions of the colliding hadrons and in the evolution of the strong coupling constant.

In this thesis, the theoretical prediction of charm hadron (D^{*+} or D^0) production was calculated with $m_c = 1.5$ GeV and $\mu_r = \mu_f = \mu_0$ where $\mu_0 = \sqrt{m_c^2 + p_T^2}$. PDF sets were given by CTEQ6.6 [67] as a proxy of PROSA_VFNS; a modern PDF set PROSA [68] was fitted including low-x charm measurements without assuming universality of charm fragmentation but uncertainty sets are provided only for a Fixed-Flavour-Number Scheme (FFNS), while FONLL requires a Variable-Flavour-Number Scheme (VFNS). The BCFY function [46] was taken as the non-perturbative fragmentation function.

Then, the theoretical uncertainty was determined by

$$\Delta_\pm = \sqrt{\Delta_{\pm,\text{scales}}^2 + \Delta_{\pm,\text{mass}}^2 + \Delta_{\pm,\text{PDF}}^2}, \tag{2.60}$$

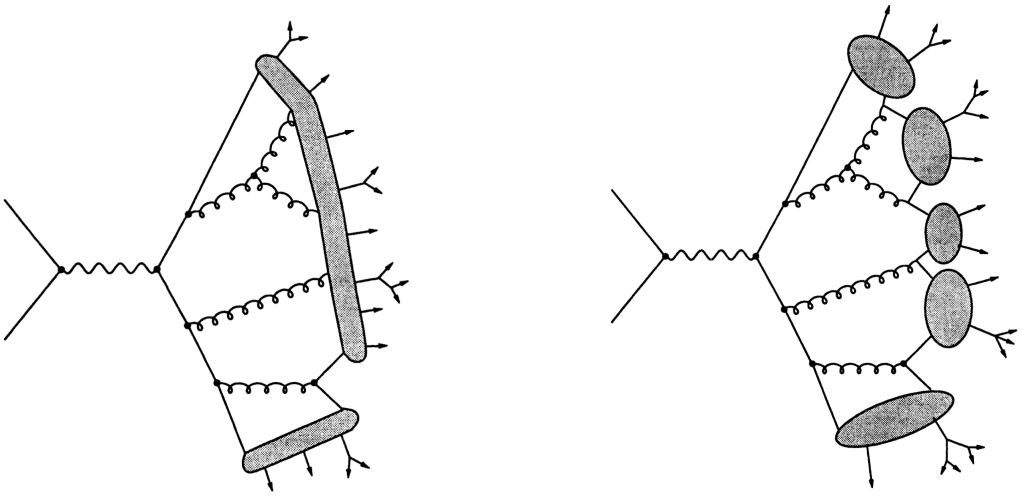
where $\Delta_{\pm,\text{scales}}$, $\Delta_{\pm,\text{mass}}$ and $\Delta_{\pm,\text{PDF}}$ are scale, mass and PDF uncertainty, respectively. Up to the highest order known today, the theoretical uncertainty is dominated by the scale uncertainty. The scale uncertainty was calculated by varying each scale by a factor 2 up and down and taking the envelope. In other words, cross sections were calculated for seven scale sets within a constraint $0.5 \leq \mu_f/\mu_r \leq 2$, which are

$$\begin{aligned}(\mu_f, \mu_r) \in \{&(\mu_0, \mu_0), \\ &(2\mu_0, 2\mu_0), (2\mu_0, \mu_0), (\mu_0, 2\mu_0), \\ &(0.5\mu_0, 0.5\mu_0), (0.5\mu_0, \mu_0), (\mu_0, 0.5\mu_0)\},\end{aligned} \tag{2.61}$$

and the scale uncertainty was determined by the maximum and minimum deviations. The mass uncertainty was determined with $m_c = 1.3$ and 1.7 GeV by taking the envelope again. The CTEQ6.6 PDF uncertainty is defined as asymmetric error:

$$\begin{aligned}\Delta\mathcal{O}_+ &= \sqrt{\sum_{k=1}^{n/2} \Big\{\max\big[\mathcal{O}_{2k-1} - \mathcal{O}_0, \mathcal{O}_{2k} - \mathcal{O}_0, 0\big]\Big\}^2} \\ \Delta\mathcal{O}_- &= \sqrt{\sum_{k=1}^{n/2} \Big\{\max\big[\mathcal{O}_0 - \mathcal{O}_{2k-1}, \mathcal{O}_0 - \mathcal{O}_{2k}, 0\big]\Big\}^2},\end{aligned} \tag{2.62}$$

with 44 error sets in total.



(a) String model. (b) Cluster model.

Figure 2.19: Parton shower with hadronization model, e.g., for $e^+e^- \to$ hadrons, with figures taken from [39].

2.5 Parton shower and hadronization model

In QCD, final state quarks generate an indefinite number of gluon emissions losing part of their energy until they eventually become hadrons. An approximate perturbative treatment of this parton cascade, *parton shower* [39, p.157-158], is introduced with resummation of soft and collinear gluon emissions. As mentioned, there are contributions when a soft gluon is emitted or when a gluon splits into two almost collinear partons. The leading contributions of these soft and collinear emissions and the corresponding virtual corrections can be identified and summed to all orders. This improves the convergence of the perturbative series. The soft and collinear enhancements are associated with parton branchings on outgoing lines of Feynman diagrams, which lead to parton showers implemented in QCD Monte Carlo (MC) program. QCD dynamics at a scale larger than Λ, which is typically order of 1 GeV as the perturbative regime, is described by the parton shower in MC. Then the dynamics at lower scale than Λ is treated by a non-perturbative model of hadronization process.

For the hadronization process, there are typically two models interfaced to MC event generators: string [69,70] and cluster [71] model. In the string model [39, p.187-188], for example the produced quark-pair in e^+e^- annihilation loses energy to the colour field and the string breaks up with forming separate colourless states as similar to the independent fragmentation of the simple quark-antiquark system. During the evolution of the parton shower, strings are connected between the various endpoints, and the gluon which at the end of the shower determines the jet final states. Each string segment then breaks up into hadrons as shown in Figure 2.19a. On the other hand, in the cluster model [39, p.188-190], the gluons split non-perturbatively into quark pairs to form colourless clusters after the perturbative jet development. Each cluster then decays into hadrons accounting for the multiplicities of the various kinds of hadrons observed in e^+e^- final states (see Figure 2.19b). Among the general-purpose MC generators, PYTHIA [30], HERWIG [72] and SHERPA [73], PYTHIA is based on the string model and the other two are based on the cluster model.

3 Experimental Particle Physics and Heavy-Quark Production

Particle physics experiments are designed to detect and identify particles produced in high-energy collisions [38]. Only a few of the particles which can be produced are stable: electron, proton, photon and, effectively undetectable, the neutrino. The particles with lifetimes long enough to propagate over several metres when produced in high-energy collisions can be directly detected, which include muon, neutron, charged pion, and charged kaon. In contrast, short-lived particles which decay before they travel a significant distance from the point of production, are detected only via their decay products.

3.1 Particle colliders

High energy experiments can be achieved by colliding two beams in accelerators where the particles collide in the center-of-mass frame. With the colliding beams, the laboratory frame is almost coincident with the center-of-mass frame, thus the center-of-mass energy can coincide with the sum of the energies of the two colliding beams. The center-of-mass energy determines the types of particles to be studied and discovered. The resulting event rates are determined by the instantaneous luminosity $\mathcal{L}$ of a machine. The number of interactions is then

$$N = \sigma \int \mathcal{L}(t)dt \tag{3.1}$$

where σ is the cross section for a given process and $\mathcal{L}$ is integrated over the livetime of the operation of the machine. The instantaneous luminosity can be calculated [38, p.27] by

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

where n_1 and n_2 are the numbers of particles in the colliding bunches of frequency f, and σ_x and σ_y are the root-mean-square horizontal and vertical beam sizes. This relation is assuming that the beams have a Gaussian profile and collide head-on. In practice, an accurate a priori calculation of $\mathcal{L}$ is not possible since the transverse profiles of the beams are not known precisely. Thus a cross section measurement is given with a reference process of σ_{ref} and N_{ref}:

$$\sigma = \sigma_{\text{ref}}\frac{N}{N_{\text{ref}}}. \tag{3.3}$$

To achieve high energies, only charged stable particles are used such that colliding beam machines are restricted to accelerating eletron, proton and their antiparticles. In this thesis, the high energy measurements are discussed specifically with experiments from the e^+e^- collider LEP [74, 75], the e^+e^- colliders (also known as B-factories) CESR [76], DORIS [77], SLAC [78] and KEKB [79], the ep collider HERA [80], and the pp collider LHC [81].

All of these colliders are based on circular accelerators such as synchrotrons except the linear accelerator at SLAC (B-factory). Depending on the experiments, the details of the accelerator complex are different, although a typical synchrotron layout is given in the following.

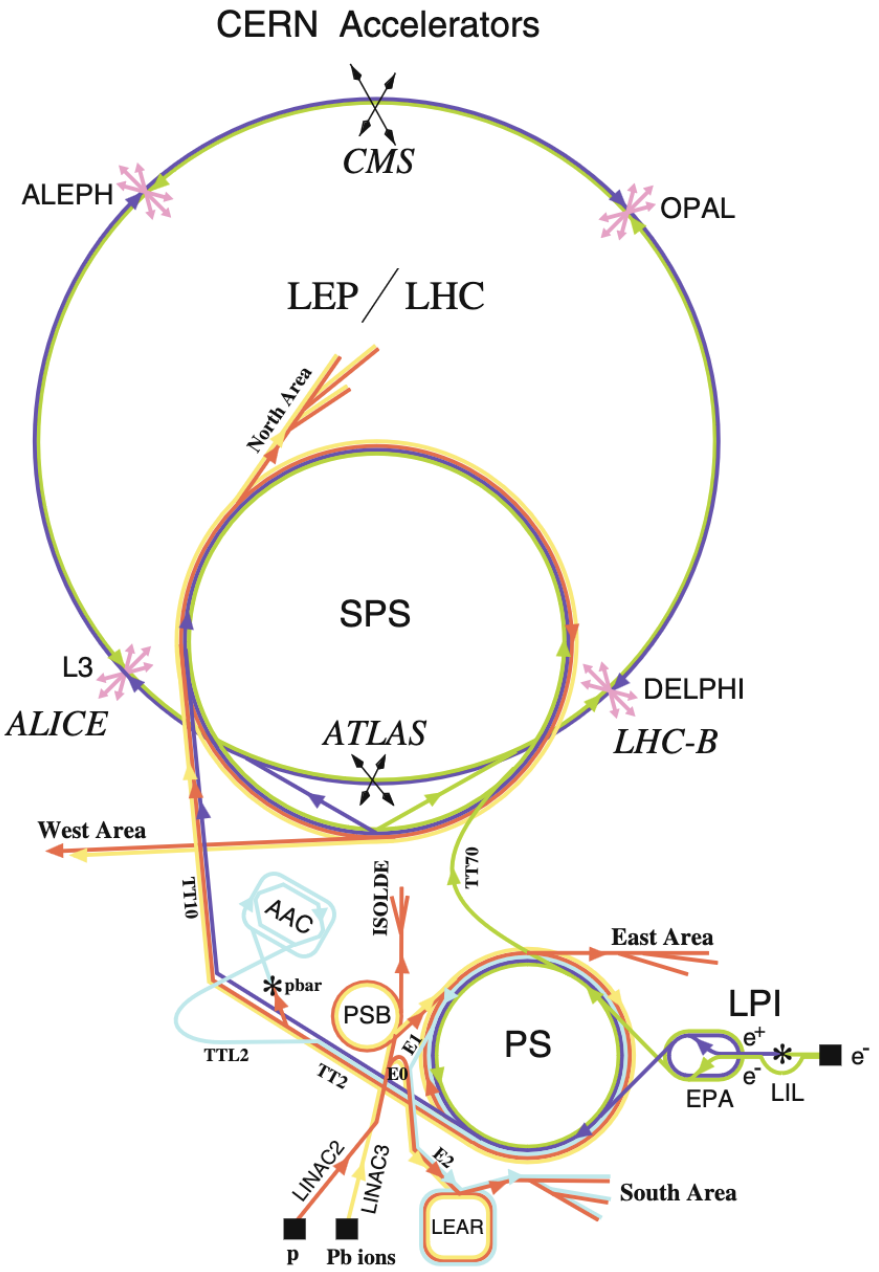
In synchrotrons, particles are preaccelerated with an electrostatic and a linear accelerator. These preaccelerated particles are then injected into a ring, which is surrounded by a set of (superconducting) magnets to deflect the particles. The injected particles are further accelerated through Radio-Frequency (RF) electric fields in resonant cavities between the magnets. This confines the bunches of particles to stay on the same orbit of the ring with gradually increasing magnetic field.

Specifically, the CERN accelerator complex (Figure 3.1) was constructed for both the LEP and the LHC colliders. First it was operated for the LEP collider until 2000 [82, p.53]. In the first phase, a linear accelerator (LIL) and the Electron and Positron Accumulater (EPA) were built to accelerate electrons and positrons up to 600 MeV. In the second phase, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) were upgraded to accelerate the particles up to 3.5 GeV and 20 GeV, respectively, by making electrons and positrons to move in opposite directions. Then in the third phase, the large ring of 27 km was put into operation. At LEP, four bunches of electrons and positrons were accelerated and circulated in the collider for a few hours. In the ring, the bunches were kept vertically separated in one set of four regions and brought to collide in the other set of four regions. To obtain higher possible luminosity, the beams were focused to reduce the transverse dimensions before collisions. Then, the ALEPH, DELPHI, L3 and OPAL detectors were installed around each of the four interaction points. The LEP magnets were made of iron and concrete of low cost, and very a low magnetic field of $0.2 - 0.4$ Tesla was used to reduce the energy loss through bremsstrahlung (see also later). In the first phase, normal RFs were used, which were later replaced by superconducting RF to reach energies up to $130 - 209$ GeV. Eventually, the maximum beam energy was 104.6 GeV with an integrated luminosity of 0.275 fb^{-1} [42, p.544].

In 2000, LEP was dismantled in order to use the tunnel for the LHC collider. The LHC uses a total of $\sim 10,000$ superconducting dipole magnets. Due to the acceleration of protons, the synchrotron radiation is weaker compared to the LEP collider using electron beams. A linear accelerator, Linac4, generates negative hydrogen ions. Then the Proton Synchrotron Booster (PSB) strips out electrons from the ions to leave one proton only. The protons are then accelerated and injected into the PS and the SPS to increase the energies further. Protons are accelerated from 450 GeV to 6.8 TeV while the magnetic fields are increased from 0.54 to 7.7 Tesla. The LHC Run 1 started in 2009 and ended in 2012, where pp collisions were produced at $\sqrt{s} = 0.9$, 7 and 8 TeV with an integrated luminosity of ~ 0.3 nb^{-1}, 6.1 and 23.3 fb^{-1} [42, p.546], respectively. After the first long shutdown, the LHC Run 2 followed in 2015-2018 at $\sqrt{s} = 13$ TeV with an integrated luminosity of 160 fb^{-1} [42, p.546]. In the current Run 3, which started in 2021 (with commisioning runs), the center-of-mass energy was successfully increased up to 13.6 TeV, while some commisioning runs were taken at $\sqrt{s} = 0.9$ TeV [83]. Further upgrades are scheduled for the High-Luminosity LHC (HL-LHC) to deliver an integrated luminosity up to 3000 fb^{-1} at $\sqrt{s} = 14$ TeV. The large detectors, ALICE, ATLAS, CMS and LHCb for the LHC are placed in the four interaction points and used for various experimental studies.

3.2 Particle detection

Particles lose energy when passing through matter in various processes. For instance, all charged relativistic particles passing through a medium lose energy by ionisation of the atoms. The particles collide with the electrons of the atoms inelastically, and the energy loss per path length unit ($\frac{dE}{dx}$) is given by the Bethe-Bloch formula [42]. In Figure 3.2, the mean rates of the energy loss ($\langle -\frac{dE}{dx} \rangle$) are shown as a function of particle momentum through liquid hydrogen, gaseous helium, carbon, aluminum, iron, tin, and lead. In the case of the muon, the ionisation is the dominant energy-loss process except at very high energies (e.g., $\gtrsim 100$ GeV) and thus the muons travel a significant distance even in dense materials. This means that the muons are highly penetrating to pass through the entire detector and leave a trail of ionisation. In contrast,



LEP: Large Electron Positron collider
SPS: Super Proton Synchrotron
AAC: Antiproton Accumulator Complex
ISOLDE: Isotope Separator OnLine DEvice
PSB: Proton Synchrotron Booster
PS: Proton Synchrotron
LPI: Lep Pre-Injector
EPA: Electron Positron Accumulator
LIL: Lep Injector Linac
LINAC: LINear ACcelerator
LEAR: Low Energy Antiproton Ring

Figure 3.1: Schematic layout of the CERN accelerators. The figure is from [75].

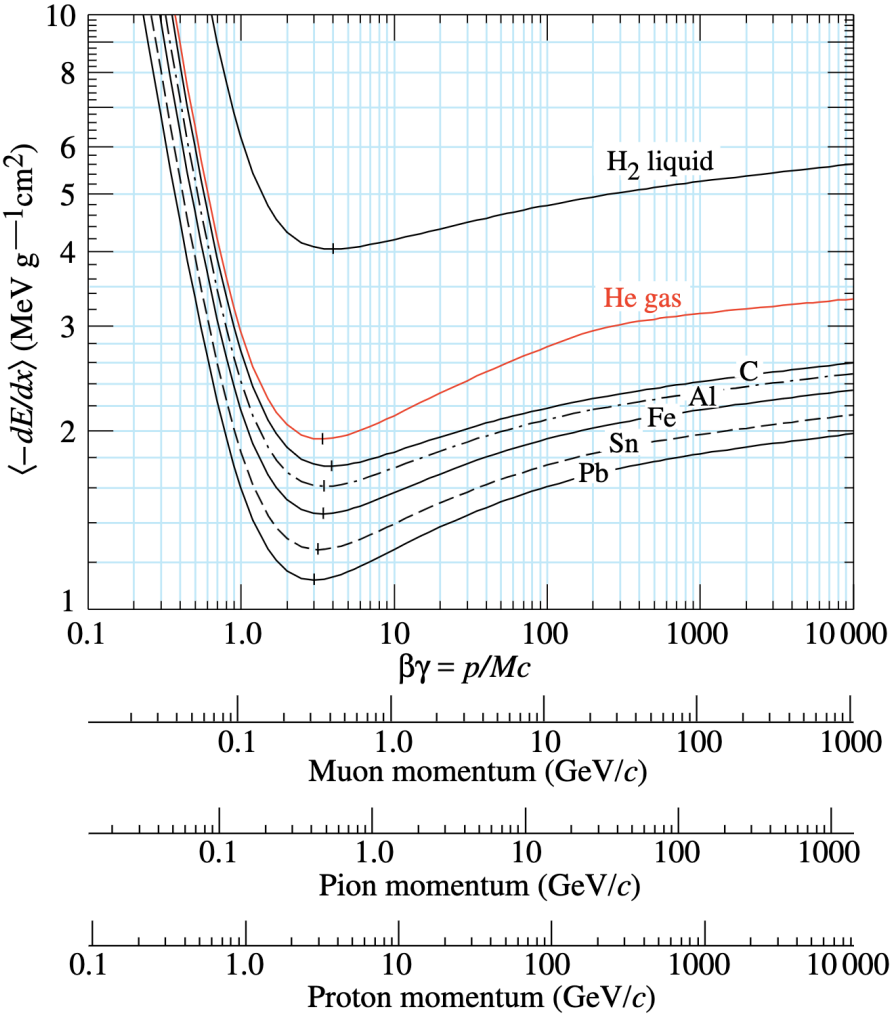


Figure 3.2: Mean energy loss rate in liquid hydrogen, gaseous helium, carbon, aluminum, iron, tin, and lead. Figure is taken from [42].

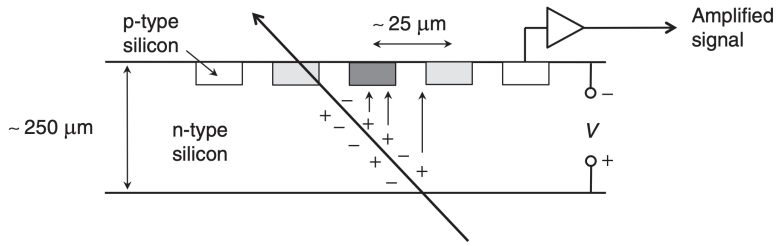


Figure 3.3: Example of a silicon tracking sensor. Electron-hole pairs, which are created by a traversing particle through the silicon, drift by a potential V. Triggered electronic signals are amplified and transferred to be analyzed. Figure is taken from [38].

electrons lose energy significantly also via the bremsstrahlung process. Charged particles radiate a photon in the electrostatic field of a nucleus above a critical energy, and the energy loss rate is inversely proportional to the particle mass squared (m^2) (see e.g. [38, p.18] and [82, p.20]). Thus the energy loss rate of an electron by bremsstrahlung increases by $(m_e/m_\mu)^2$ relative to that of a muon. Detailed discussions of the energy loss for high-energy experiments can be found further in e.g. [42, p.549-564].

Since the charged particle leaves a trail of ionised atoms and liberated electrons through the medium, the trajectory of a charged particle can be reconstructed by detecting this ionisation. Various detectors are optimized depending on the type of radiation in a given energetic range. There are some parameters to characterize detector operations; typically detector efficiency (ϵ), dead time, and time/spatial/energy resolution [82, p.31]. Specifically, the detector efficiency is the probability that a detector records a radiation, which is given by

$$\epsilon = \frac{N_{\text{rec}}}{N} \tag{3.4}$$

where N_{rec} is the recorded particles and N is the number of particles which traverse the detector. This is usually determined by MC simulation studies. Among various detectors introduced in high-energy experiments, semiconductor tracking detectors will be discussed in this chapter and a few more in the next chapter. More details also can be found in e.g. [42, p.565] and [82, p.30-42].

The semiconductor detectors provide outstanding performances in terms of all three of time, spatial and energy resolution also in combination. The ATLAS and CMS experiments at the LHC, for example, use the semiconductor technology implemented in silicon pixels and strips to track and reconstruct charged particles. When a charged particle traverses a doped silicon wafer, electron-hole pairs are created by the ionisation [38, p.15-16] as shown in Figure 3.3. The holes drift in the direction of the electric field produced by a potential (V) applied across the silicon, and are collected by p-n junctions. Then the information is transformed into electric signals to be analyzed. The trajectory of the charged particle track can be reconstructed by constructing tracking detectors with several cylindrical surfaces of silicon wafers as shown in the left panel of Figure 3.4. If a large solenoid surrounds the tracking detectors to produce a uniform magnetic field with the magnetic flux density B, the trajectory of the particle forms a helix with a radius R and an angle λ (Figure 3.4) by the Lorentz force:

$$p \cos\lambda \propto \mathrm{B}R \tag{3.5}$$

where p is the particle momentum. Therefere, if the parameters, R and λ, are determined by measurements in the tracking detectors, p can be reconstructed. Details of the semiconductor technologies implemented in the CMS detector will be introduced in the next chapter.

As mentioned, a high-energy charged particle radiates photons by the bremsstrahlung process and can produce electron-positron pairs. This process continues to produce a cascade

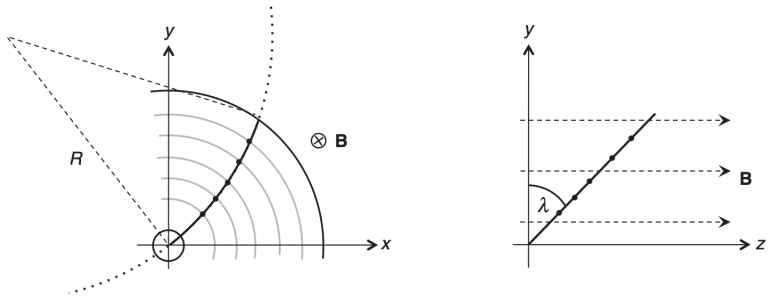


Figure 3.4: Track reconstruction of a charged particle in a silicon detector. The curvature parameters, R and λ, which are measured by the detectors, determine the particle momentum. Figures are taken from [38].

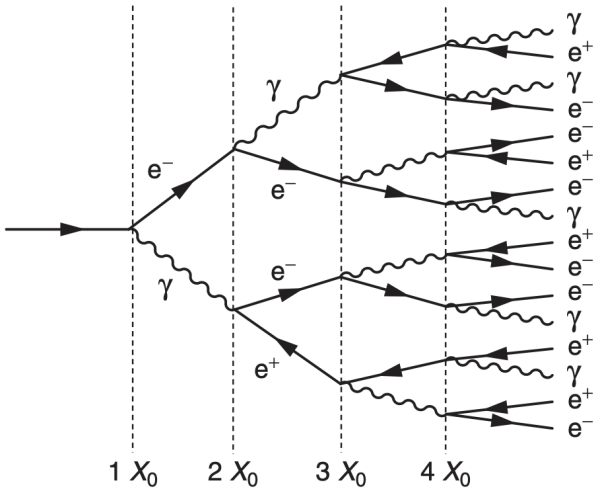


Figure 3.5: Electromagnetic shower induced by the bremsstrahlung process. Figure is taken from [38].

of photons, electrons and positrons (i.e., electromagnetic shower). Similarly, a primary high-energy photon also produces a shower as shown in Figure 3.5. The energies of the electromagnetic shower and the hadronic shower are measured in electromagnetic calorimeter and hadron calorimeter, respectively. With the relatively large distance of nuclear interactions, the hadron calorimeter occupies a large volume compared to the electromagnetic one.

3.3 Heavy-quark production measurements

Since quarks are observed always as hadron states in experiments due to colour confinement, non-perturbative hadronization/fragmentation cannot be predicted by theory and thus should be determined by measurements. Furthermore, recently the non-universality of charm (and beauty) fragmentation was reported from the LHC experimenets. Therefore, providing measurements of heavy-quark production is essential not only for tests of QCD but also to understand the fragmentation mechanism and to provide proper fragmentation inputs. Direct measurements of the kinematic spectrum of fragmentation (the fragmentation function) can be possible in e^+e^- collisions. Meanwhile, *ep* or *pp* data provided as a function of transverse momentum or (pseudo-)rapidity can be an indirect observation of the fragmentation effects. The average effects of fragmentation have been shown by measuring the fragmentation fractions from all three collisions. This section is devoted to introducing such measurements, of which most will be used as inputs especially for the studies given in Chapter 6.

As mentioned, the initial state kinematics is well known in e^+e^- collisions compared to other collisions involving protons. Therefore, heavy quark production can be measured as a function of the scaled energy or momentum defined in Eq.(2.29), which then gives measurements of the fragmentation function. For example, D^0, D^+ and D^{*+} measurements as a function of x_p from BELLE [48] ($\sqrt{s} \sim M(\Upsilon)$) are shown in Figure 3.6a, 3.6b and 3.6c, respectively. D^{*+} measurements as a function of x_E from ALEPH [84] ($\sqrt{s} \sim M(Z)$) are also shown in Figure 3.6d. As shown in these figures, almost the full phase space is covered by the given variables. This gives direct measurements of the charm fragmentation spectra. Specifically, the common parameter, r, of the three non-perturbative fragmentation functions, D^0, D^+ and D^{*+} BCFY, was extracted by fitting the BELLE data (refer to [46]). The results are shown in Figure 3.6a, 3.6b and 3.6c by solid lines. A fit result was also provided with the ALEPH data in Figure 3.6d. The BCFY functions describe the e^+e^- data well including the CLEO data [85], of which the fit results also can be found in [46]. However, the fitted parameters are quite different between the $\sqrt{s} \sim M(\Upsilon)$ and the $\sqrt{s} \sim M(Z)$ results. The FONLL prediction for the charm production in pp collisions is given with the fitted parameter to the ALEPH data.

Charm and beauty production have been measured also as a function of p_T and/or (η)y of hadrons. In particular, charm production has been measured from the LHC experiments at various center-of-mass energies. For instance, the overall phase space coverage provided with D meson measurements is shown in Figure 3.7. All these measurements show no significant differences with theory* predictions. It was shown that specifically, the D^0, D^+ and D^{*+} measurements are consistent with the upper edge of the FONLL theoretical uncertainty band [23].

The HERA experiments have also provided charm measurements as a function of p_T and η, and examples from H1 are shown for D^0, D^+, D_s^+ and D^{*+} [86] in Figure 3.8. Here, each measurement is provided with respect to the corresponding fragmentation fractions. This figure shows that there is no significant difference between the D mesons for the kinematic spectrum. Meanwhile, ALICE directly measured the ratios between D meson spectra as a function of p_T at $\sqrt{s} = 5$, 7 and 13 TeV [7], which are shown in Figure 3.9. All these measurements show that the charm fragmentation to D mesons is independent of the kinematics.

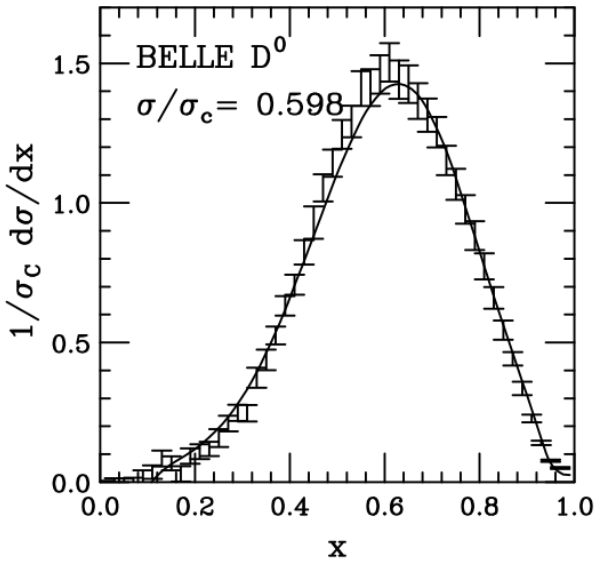
Contrary to the meson-to-meson ratios, recent measurements showed that baryon-to-meson ratios are strongly dependent on p_T in pp collisions [10, 15, 22]. Figure 3.10 shows the Λ_c^+/D^0 measurements as a function of p_T at $\sqrt{s} = 5$, 7 and 13 TeV from ALICE [15] and at $\sqrt{s} = 5$ TeV from CMS [22]. The $\Xi_c^{0,+}/D^0$ measurements at $\sqrt{s} = 5$, 7 and 13 TeV from ALICE [15,87] are shown in Figure 3.11. The p_T dependence looks consistent between Λ_c^+/D^0 and $\Xi_c^{0,+}/D^0$, which shows indirectly that there is no significant dependence on p_T in the baryon-to-baryon ratios.

ALICE provided also measurements of ratios between $\sqrt{s} = 13$ TeV and $\sqrt{s} = 5$ TeV as a function of p_T (Figure 3.12). It shows a clear p_T dependence of the ratios, but all the hadron states show a similar spectrum.

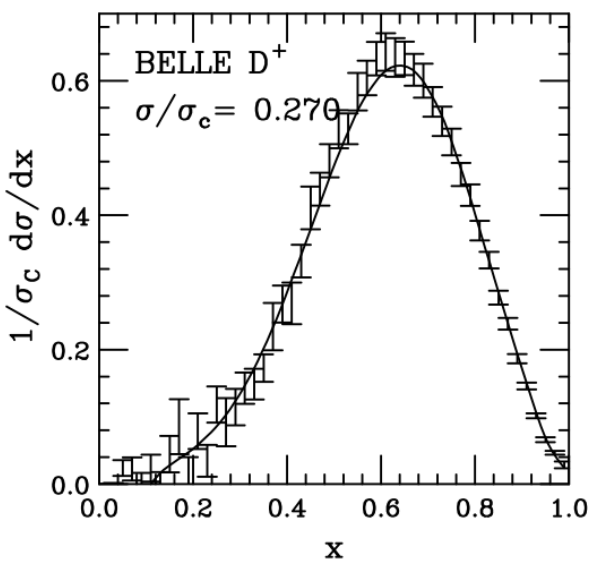
The B-factory experiments provided measurements of $\sigma(e^+e^- \to H_c)$ or $\sigma(e^+e^- \to H_c) \times BR(H_c \to \text{daughters})$ where BR is the branching ratio [47]. Following Eq.(2.27), these measurements were used to extract the fragmentation fractions with the total charm cross section prediction which is verified with experimentally and theoretically well known R_μ shown in Section 2.3. In the case of LEP, the fragmentation fractions were extracted by measuring the products of the partial decay width (Γ) of the Z into $c\bar{c}$, the fragmentation fraction and the branching ratio [47]: $R_c \times f(c \to H_c) \times BR(H_c \to \text{daughters})$ where

$$R_c = \frac{\Gamma(Z \to c\bar{c})}{\Gamma(Z \to \text{hadrons})}. \tag{3.6}$$

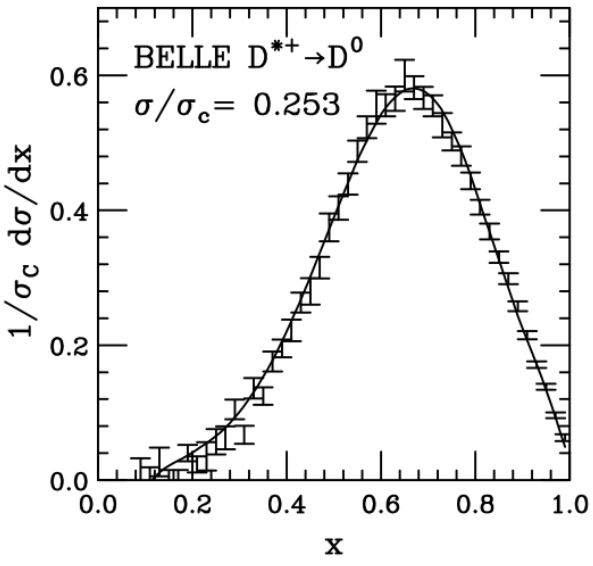
*The theory here is based on the universality assumption, which is not expected to agree with pp data. Further discussions come later.



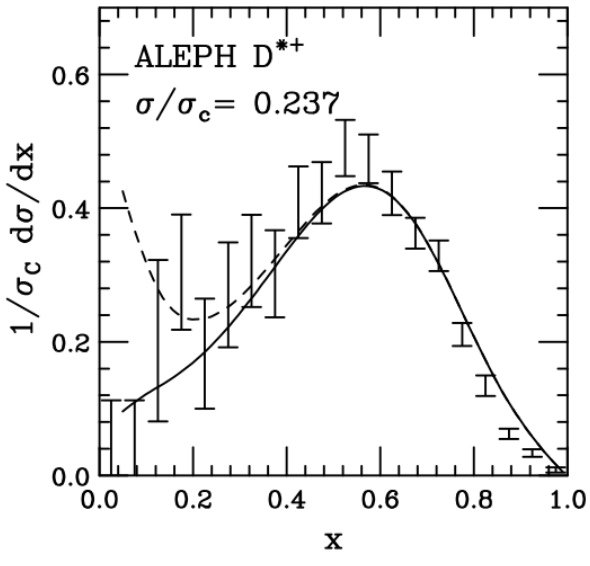
(a) D^0 measurements as a function of x_p (x).



(b) D^+ measurements as a function of x_p (x).



(c) D^{*+} measurements as a function of x_p (x).



(d) D^{*+} measurements as a function of x_E (x).

Figure 3.6: D meson measurements as a function of x and the best fit results with the BCFY functions. Figures are taken from [46].

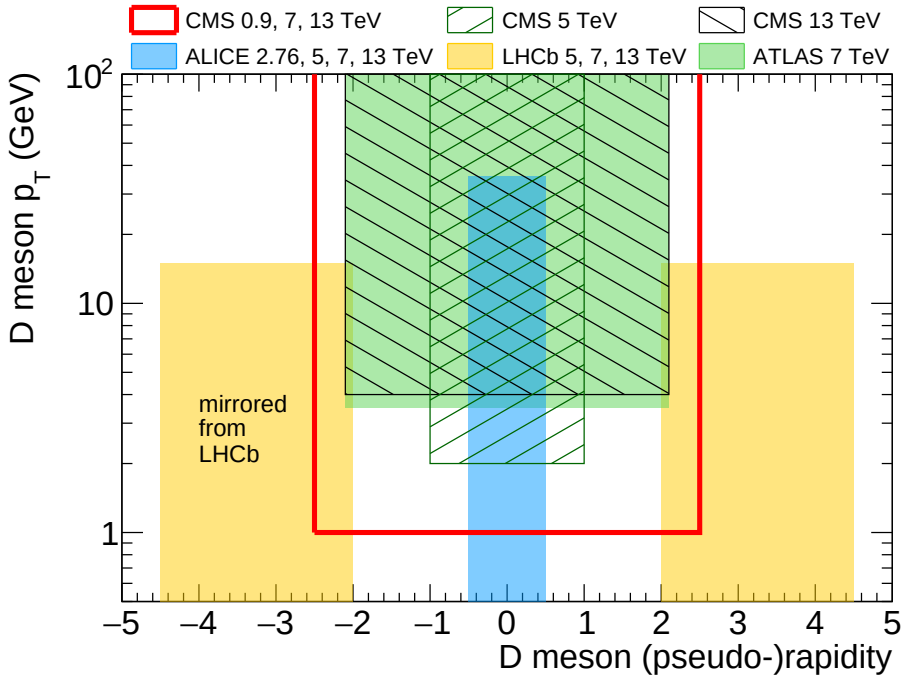


Figure 3.7: Phase space for D meson measurements in the LHC experiments. The box highlighted by red lines indicates the phase space covered for D^{*+} (and D^0) measurements at $\sqrt{s} = 7$, 0.9 and 13 TeV from CMS, which are not published yet and will be discussed in detail in Chapter 5. The other boxes are from ALICE [6–15], ATLAS [19], CMS [20–23] and LHCb [16–18].

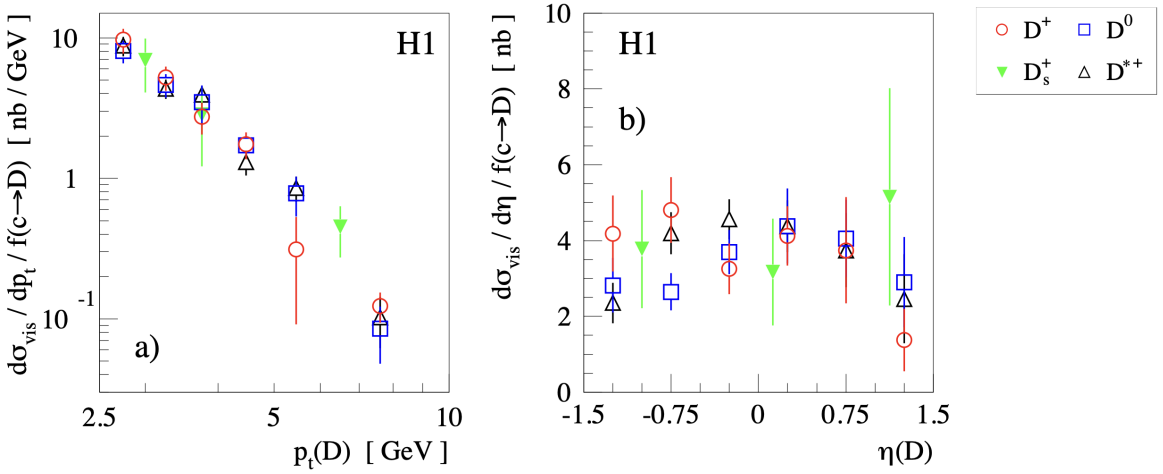


Figure 3.8: D^0, D^+, D_s^+ and D^{*+} cross section measurements from H1, with figures taken from [86]. Each measurement was divided by the respective fragmentation fraction ($f(c \to D)$).

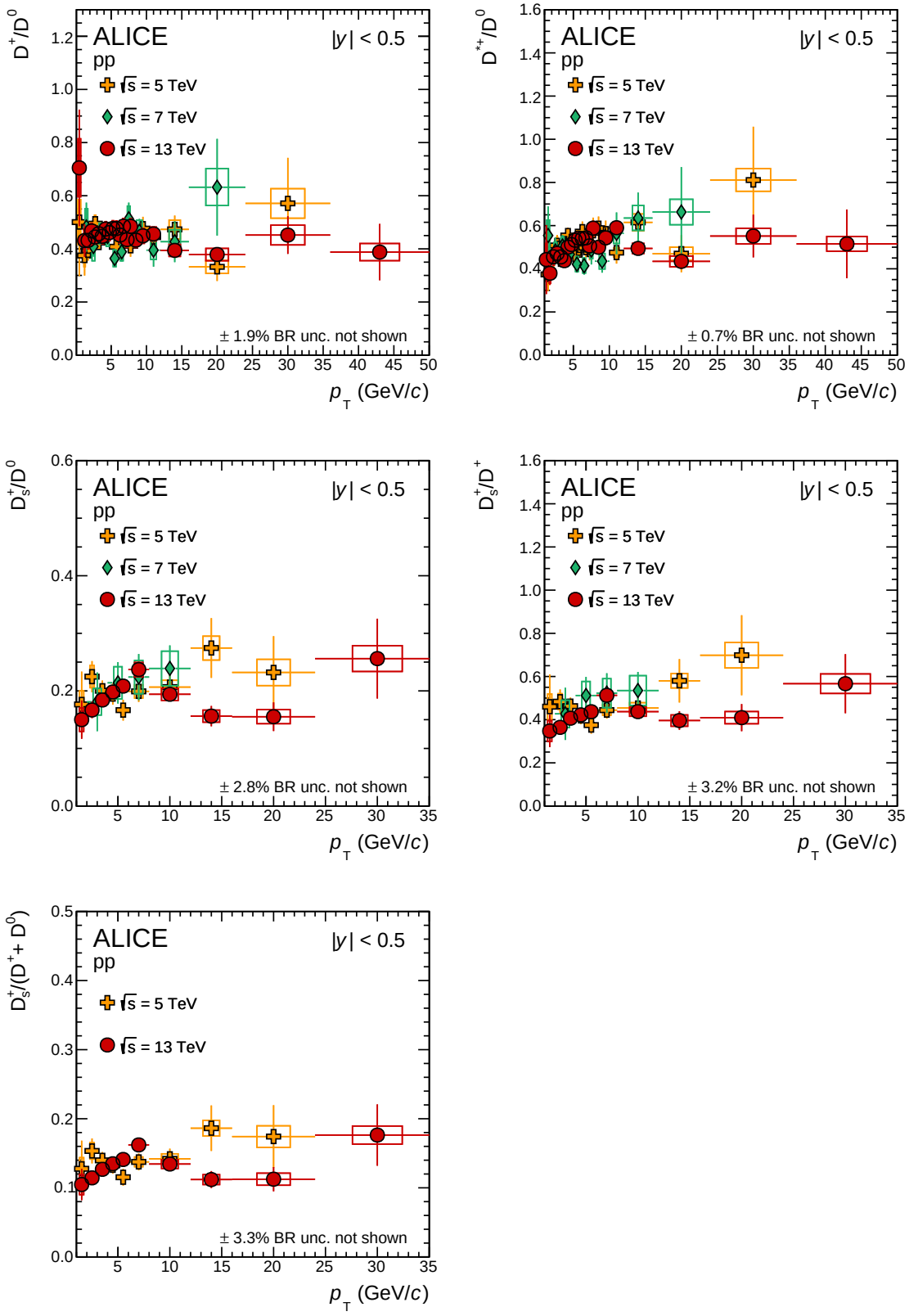


Figure 3.9: D^+/D^0, D^{*+}/D^0, D_s^+/D^0 and D_s^+/D^+ measurements from ALICE experiment. Figures are taken from [15] (the figures only for $\sqrt{s} = 5$ and 7 TeV can be found in [7]).

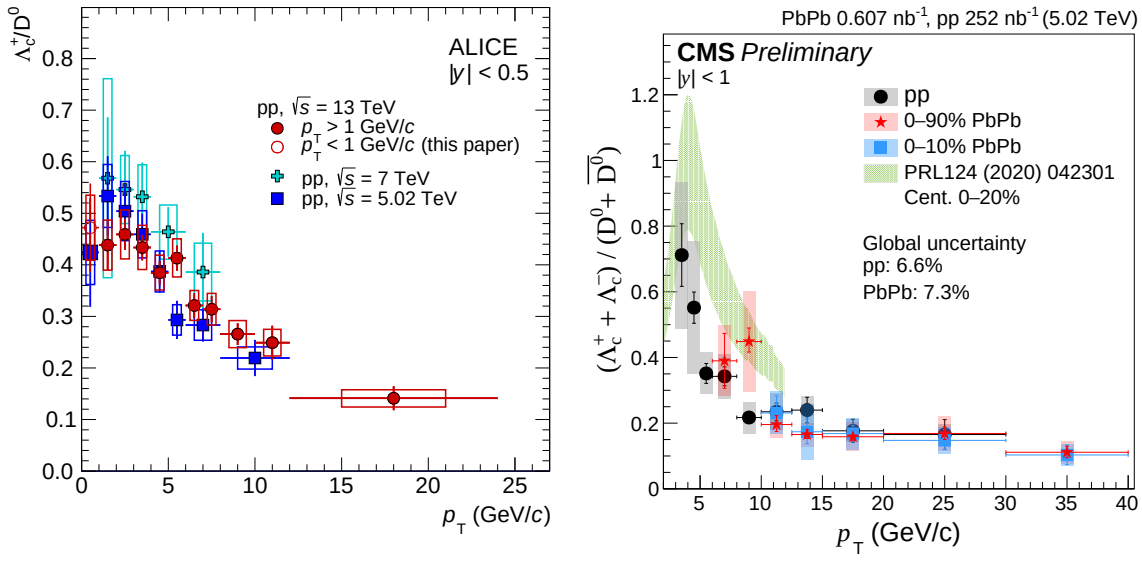


Figure 3.10: Λ_c^+/D^0 measurements as a function of p_T from ALICE (left) and CMS (right), with figures taken from [15] and [22], respectively.

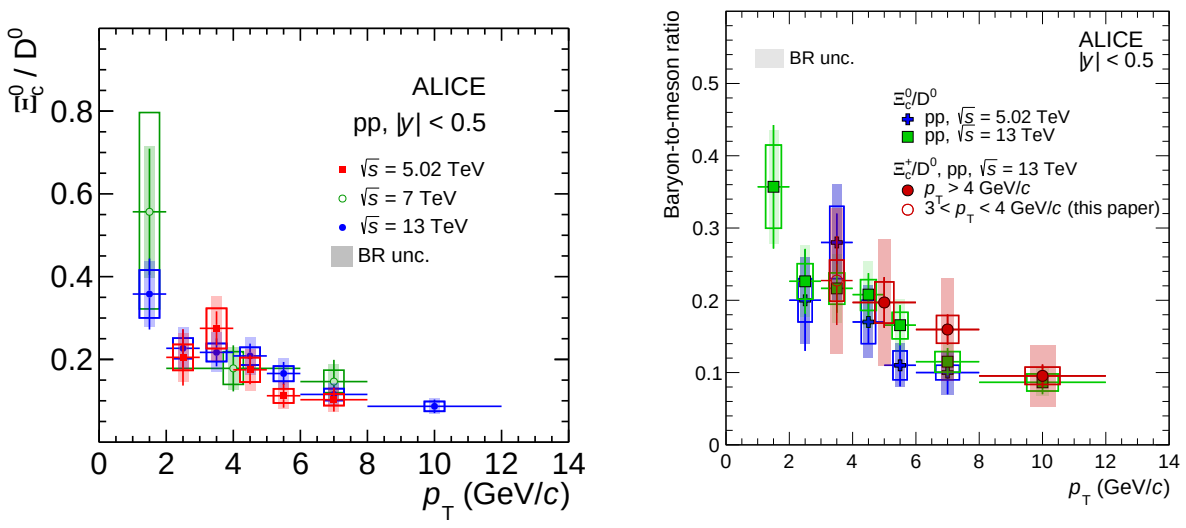


Figure 3.11: $\Xi_c^{0,+}/D^0$ measurements as a function of p_T from ALICE. Figures are taken from [15, 87].

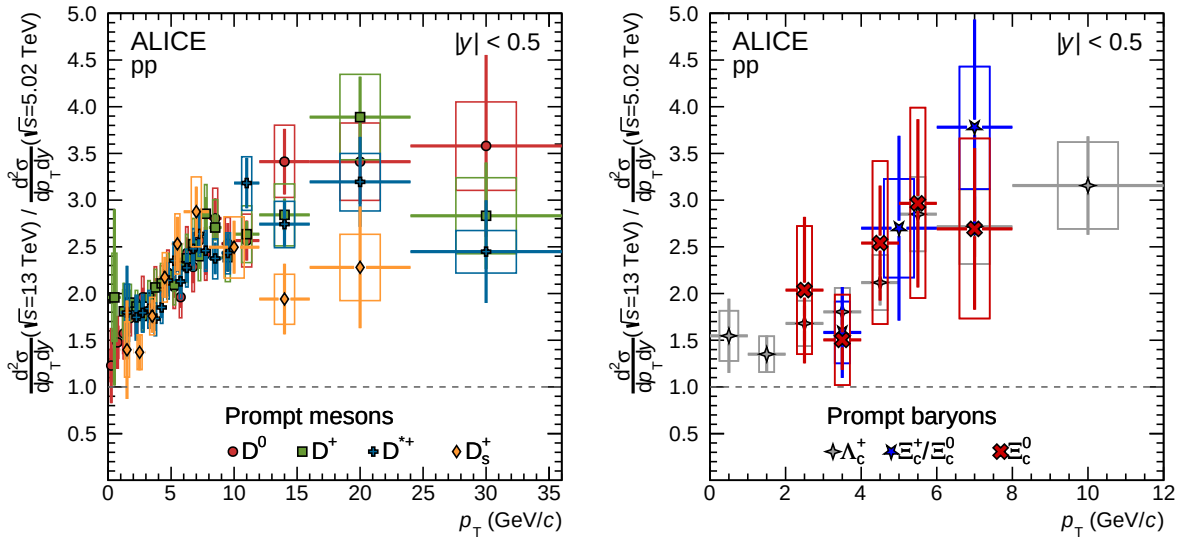


Figure 3.12: Production ratios between $\sqrt{s} = 5$ TeV and $\sqrt{s} = 13$ TeV, which are measured from ALICE. Figures are taken from [15].

R_c also is well known both experimentally and theoretically.

In ep and pp collisions, the fragmentation fractions have been measured following Eq.(2.28). In the case of ep collisions, however, the other ground states except D^0, D^+, D_s^+ and Λ_c^+ have not been measured yet, and the fragmentation fraction measurements have been provided with some assumptions for the non-measured states. For example, an assumption provided in [47] is based on $f(s \to H)$ measurements ($H = \Lambda^0, \Xi^-$ and Ω^-) in e^+e^- collisions, which gives

$$\sigma(\Xi_c^0) + \sigma(\Xi_c^+) + \sigma(\Omega_c^0) \sim 0.14\sigma(\Lambda_c^+). \tag{3.7}$$

The charm fragmentation fractions have been measured mostly from e^+e^-/ep collisions. No significant discrepancy has been reported between the measurements in e^+e^- and ep collisions (see e.g. [88]). Thus it has been assumed that the fragmentation is independent of the collision system including pp collisions. Recent reports from LHC experiments, specifically from ALICE [15, 25], however, show large differences on the fragmentation fractions between e^+e^-/ep and pp collisions (Figure 3.13). Especially the Λ_c^+ fragmentation fraction shows a big discrepancy of $\sim 5\sigma$ between pp and e^+e^-/ep collisions, while the overall meson fractions in pp collisions are smaller compared to the other collisions.

Similar phenomena were observed also from beauty hadron production. The production fractions have been measured for beauty, e.g. from the LEP, B-factories, and LHC experiments [89]. Especially, it is shown that the Λ_b^0/B^0 measurements from the LHC, which show a clear p_T dependence, are asymptotically consistent with the LEP measurement at high p_T (Figure 3.14). In this figure a fit is provided for the LHCb data using an exponential function, and the fit results agree well with the LEP value positioned at an approximate p_T in Z decays. Furthermore, the ratio of the strange meson to other mesons has been measured in beauty production with better precision compared to charm production. Specifically, the B_s^0/B^0 ratios have been measured by LHCb [90] and CMS [91], which show a clear p_T dependence (Figure 3.15). In contrast, Figure 3.16 shows that the B^0/B^+ ratio is p_T independent. Also, no significant (pseudo-)rapidity dependence has been reported for any ratios of the production fractions (see the right panels in Figure 3.15 and in Figure 3.16, and also Figure 3.17).

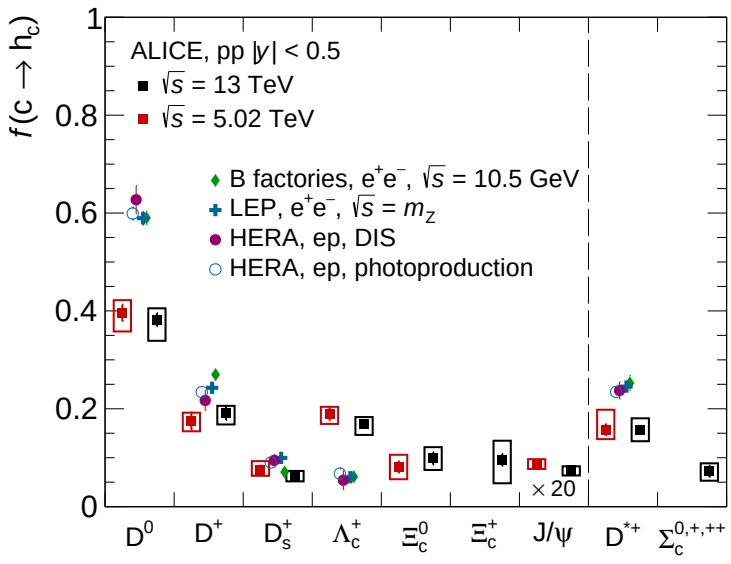


Figure 3.13: Charm fragmentation fractions measured from pp collisions at $\sqrt{s} = 5$ and 13 TeV (the figure from [15]). 5-6σ differences have been observed especially for D^0 and Λ_c^+.

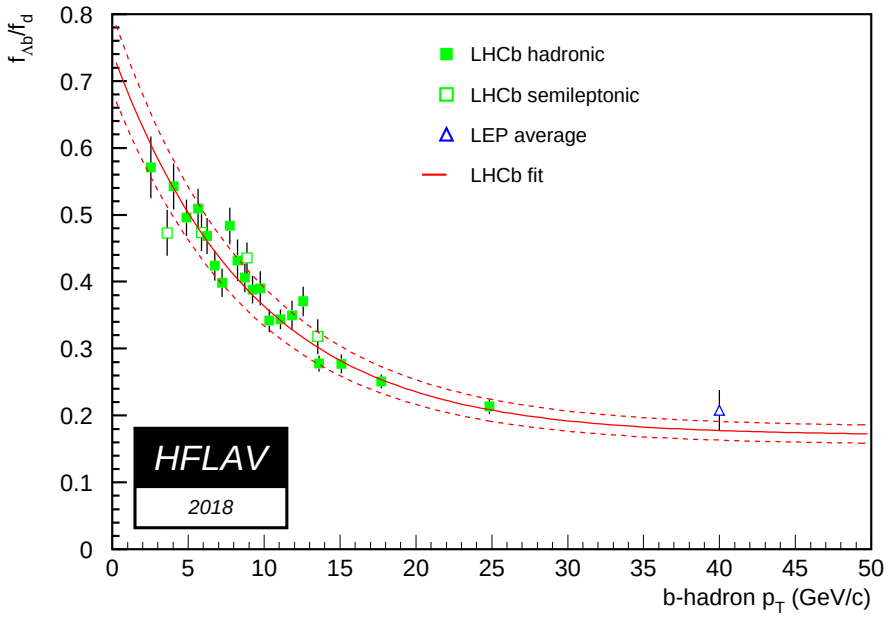


Figure 3.14: Measurements of the ratio of Λ_b^0 (f_{Λ_b}) to B^0 (f_d) with figure taken from [89]. A fit to the LHCb data is given by an exponential function and the LEP data is placed at an approximate p_T in Z decays.

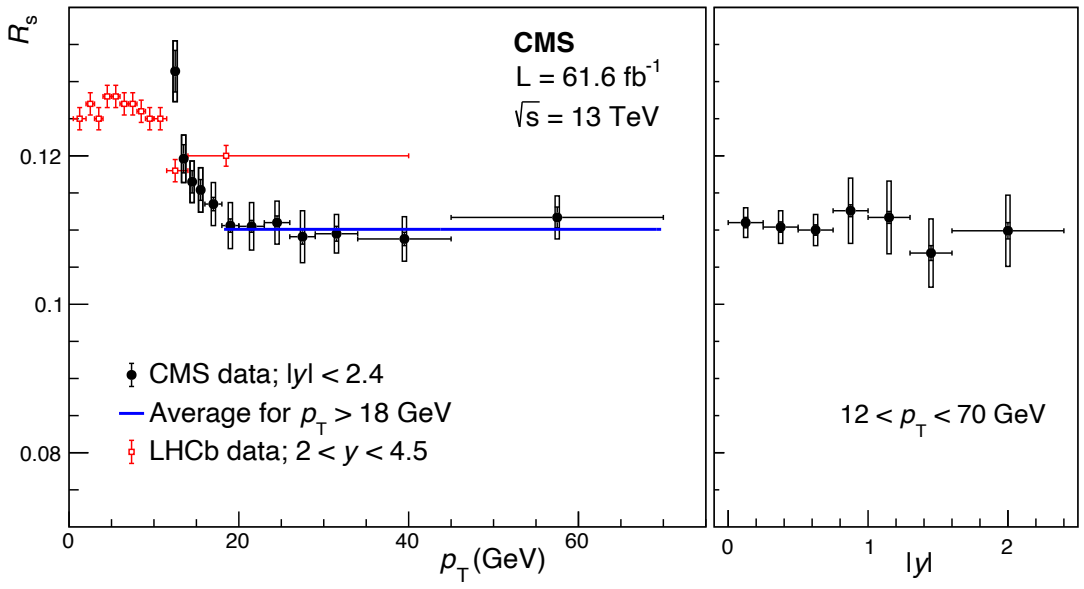


Figure 3.15: B_s^0/B^+ measurements as a function of p_T (left) and $|y|$ (right) from CMS [91]. The LHCb measurements [90] are also plotted in the left figure. Figure is taken from [91].

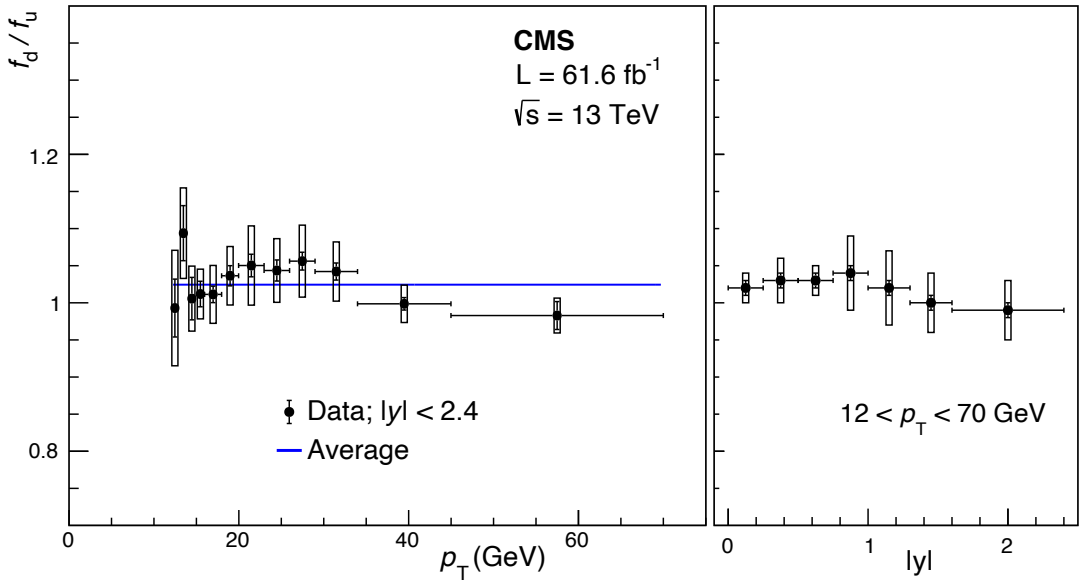


Figure 3.16: B^0/B^+ measurements as a function of p_T (left) and $|y|$ (right) from CMS. Figure is taken from [91].

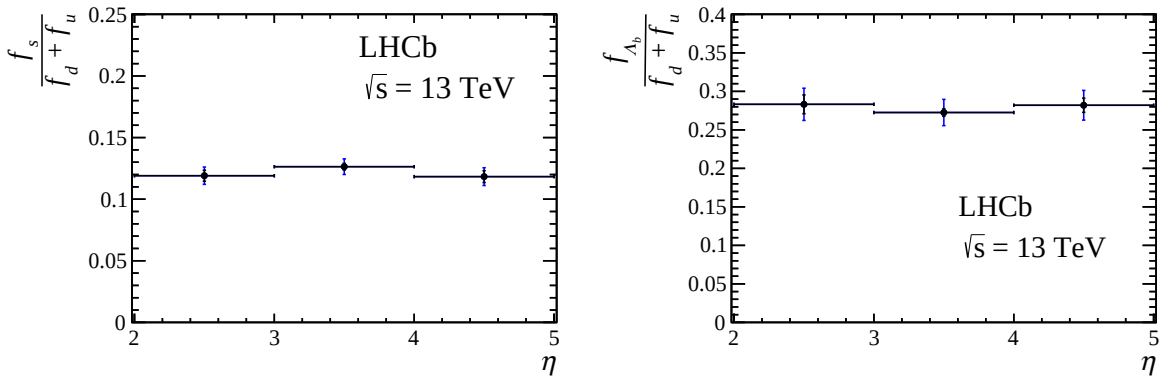


Figure 3.17: $B_s/(B^0 + B^+)$ (left) and $\Lambda_b/(B^0 + B^+)$ (right) measurements at $\sqrt{s} = 13$ TeV from LHCb. Figure is taken from [92].

4 | CMS Detector in LHC

4.1 Introduction

The CMS (Compact Muon Solenoid) detector [93–97], of which a schematic view is shown in Figure 4.1, has an overall length of 22 m, a diameter of 15 m, and weighs 14 000 tons [97]. The detector is designed to trigger on and identify electrons, muons, photons, and (charged and neutral) hadrons. It has a superconducting solenoid of 6 m internal diameter and 12.5 m length that provides a magnetic field of 3.8 T with a stored energy of 2.2 GJ [97]. Within the magnetic volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, of which each is composed of a barrel and two endcap sections. Outside the solenoid, muons are measured in gas-ionization detectors which are embedded in the steel flux-return yoke. The details of the CMS layers are introduced in Section 4.2.

Events, including simulated events, are selected and reconstructed with a data acquisition and a two-tier trigger system, using a dedicated CMS software workflow. Data from triggers and subdetectors are processed through the data acquisition system, and the trigger system is introduced to reduce the recorded event rate down to a few kHz using the first- and second-level trigger, called L1 trigger and HLT, respectively. Specifically, during Run 1, 100 kHz of L1 accepted events was reduced to around 400 Hz by the HLT before data storage [95]. This has been changed such that during Run 3, the L1 trigger and HLT operate at typical output rates of 110 kHz and 5 kHz, respectively [97]. Further details of the data collection and its analysis will be discussed in Section 4.3.

4.2 CMS layers

4.2.1 CMS coordinate

The coordinate system adopted by CMS is introduced as described in [93]:

> The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC. Thus, the z-axis points along the beam direction toward the Jura mountains from LHC Point 5. The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The polar angle θ is measured from the z-axis. Pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. Thus, the momentum and energy measured transverse to the beam direction, denoted by p_T and E_T, respectively, are computed from the x and y components. The imbalance of energy measured in the transverse plane is denoted by E_T^{miss}.

The coordinate system is illustrated in Figure 4.2.

4.2.2 Superconducting magnet

CMS chose a large superconducting solenoid for the purpose of having a strong magnet. The CMS solenoid is made of a high-purity aluminium-stabilised conductor [93], and designed to

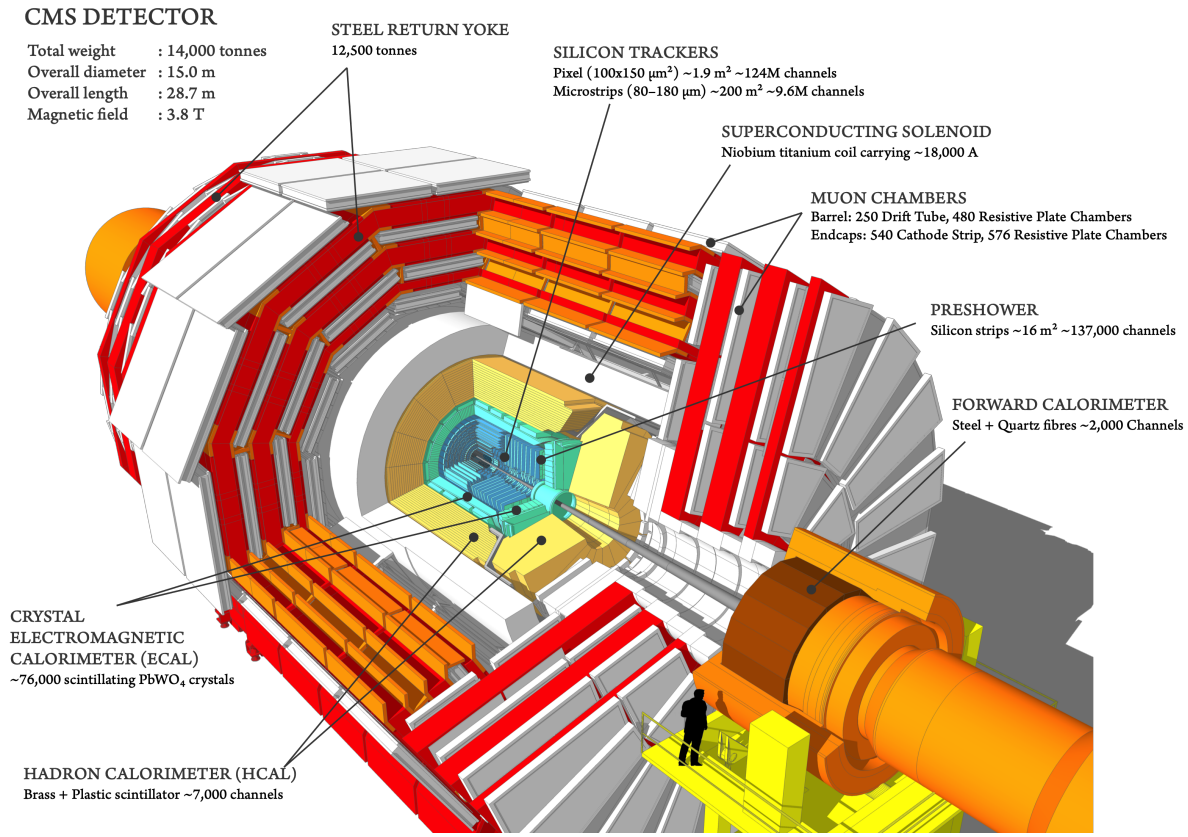


Figure 4.1: Schematic drawing of the CMS detector. Figure is taken from [98].

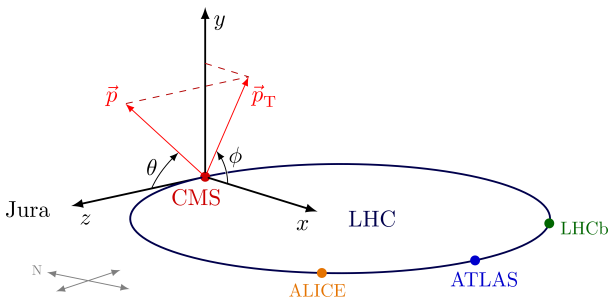


Figure 4.2: The CMS coordinate system. Figure is taken from [99].

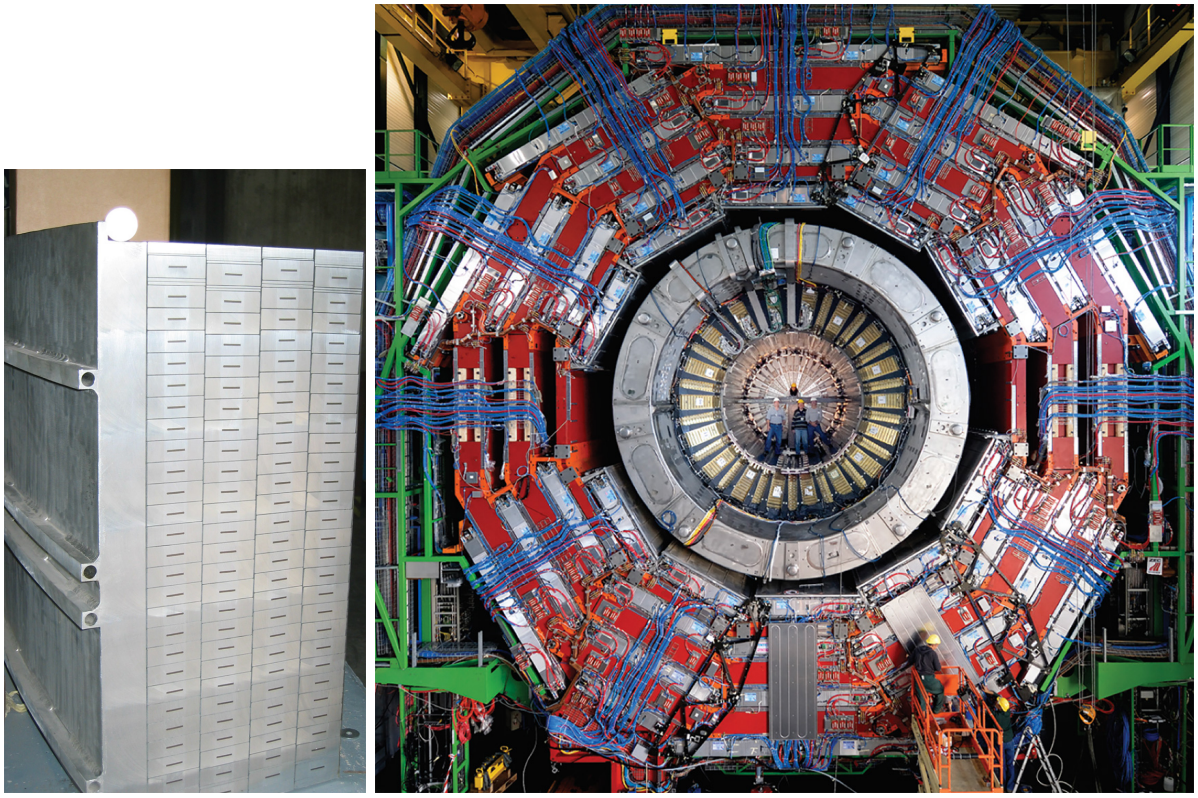


Figure 4.3: The left figure shows a 4-layer coil-winding structure of the CMS solenoid coil. The right figure shows a transverse view of the CMS solenoid. The solenoid (silver-coloured ring) is surrounded by a solid steel yoke (red) comprising three concentric layers interspersed with muon chambers (silver). Figures are taken from [100].

reach a 4 T field. The solenoid coil is composed of a 4-layer winding to account for ampere-turns required for generating the 4 T field [94] as shown in the left panel of Figure 4.3. Eventually the solenoid was assembled in a cylinder shape with 6 m diameter, 12.5 m length, and 220 tons weight [100]. To make it superconducting, this is surrounded by the cryostat that keeps it at a temperature of 4 K by the vacuum system providing the thermal insulation inside the cryostat (no longer superconducting above 10 K) [101]. Next, a solid steel yoke (red in the right panel of Figure 4.3) surrounds the solenoid (silver-coloured ring), comprising three concentric layers interspersed with muon chambers (silver). The yoke is composed of 6 endcap disks and 5 barrel wheels [94]. The magnetic flux return through the yoke is shown in the right side of Figure 4.4.

4.2.3 Tracker

The CMS tracking detectors are composed of pixels or strips in different regions depending on particle flux: pixel detectors are placed in the region closest to the interaction vertex where the particle flux is the highest ($r \lesssim 10$ cm), the remaining regions are covered by strip detectors, where the particle flux is low enough to use silicon microstrip detectors [93] ($r \gtrsim 10$ cm). The region of the strip detectors is again devided into two regions at $r \sim 20$ cm, where the region of $r \gtrsim 20$ cm is covered by larger-pitch silicon microstrips and called *outer tracker*.

Pixel detector

The extremely high particle fluxes require the innermost tracking layers to be composed of pixel devices. [103] says:

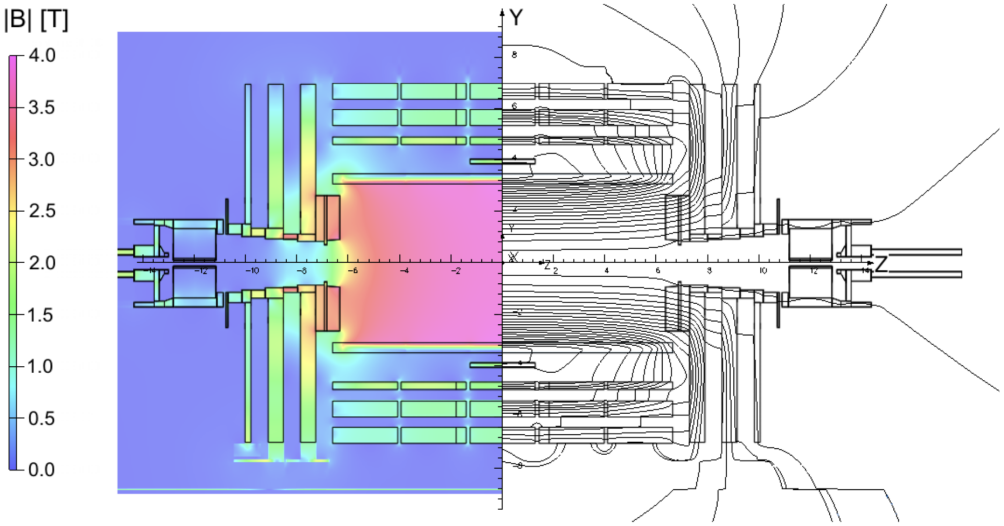


Figure 4.4: Prediction of magnetic flux density (B) and field lines on a longitudinal section of the CMS detector. Figure is taken from [102].

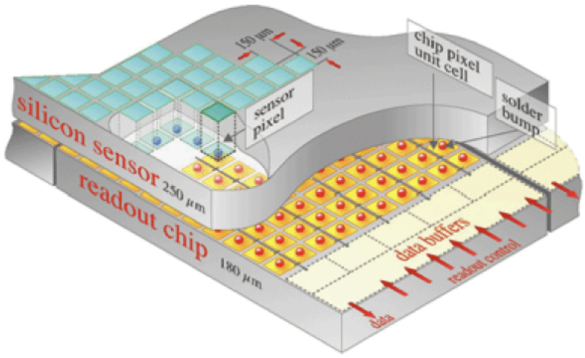


Figure 4.5: Schematic view of a pixel detector element. Each sensor pixel is connected via a solder bump to a pixel unit cell on the readout chip, where the signal is amplified, and the hit data are stored on the edge of the chip where they wait for trigger confirmation [103]. Figure is taken from [104]. This figure was introduced in 1998, and the size of the pixel to date is given by 100×150 μm^2. The thickness of the layers differs depending on position, and can be found in [105].

> The layers are composed of modular detector units. These modules consist of a thin, segmented sensor plate with highly integrated readout chips connected to them using the bump bonding technique (see Figure 4.5). The chips are connected through bond wires to hybrid circuits, which distribute the readout control and clock signals and where the data signals are collected. Kapton cables connected to the hybrids transmit the signals to and from a periphery situated at the outer region of the pixel system frame where detector control chips and electro-optical converters for optical signal transmission are located. The modules are attached to cooling frames, with the cooling tubes being an integral part of the mechanical structure.

The sensors for the CMS-pixel detector adopt the so-called n-on-n concept, which consist of high dose n-implants introduced into a high resistance n-substrate [94]. The size of the pixel is given by 100×150 μm^2 in r-ϕ and z, respectively, which is driven by the desired impact parameter resolution, the occupancy is of the order 10^{-4} per pixel and LHC bunch crossing [94].

The first CMS pixel detector was installed in 2008 and used in Run 1, as described in [97]:

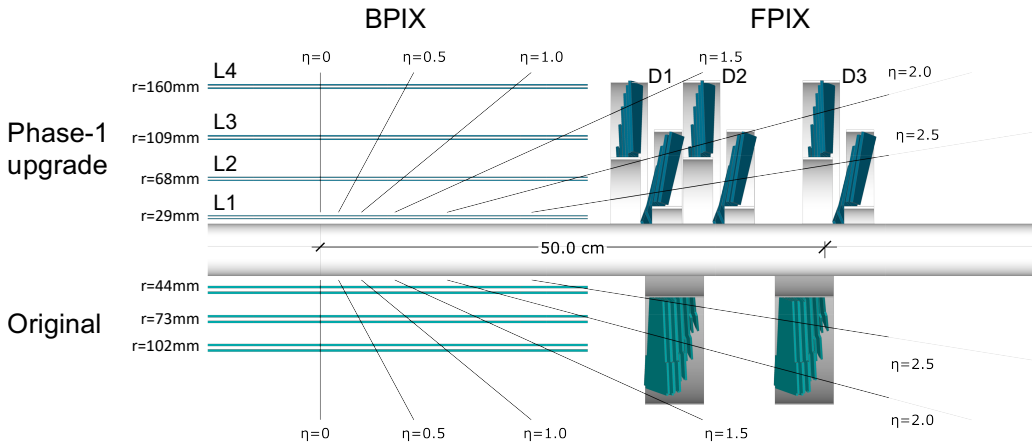


Figure 4.6: Longitudinal view of the Phase 1 pixel detector compared to the previous detector layout, with figure taken from [97].

> The first CMS pixel detector [94], installed in 2008, consisted of three barrel layers at radii of 44, 73, and 102 mm and two endcap disks on each end at distances of 345 and 465 mm from the detector center. It provided three-point tracking for charged particles and performed very well during Run 1. However, already in Run 1 the instantaneous luminosity delivered by the LHC exceeded the design value of 1×10^{34} cm^{-2} s^{-1}, which resulted in a pixel detector readout inefficiency. In order to maintain good tracking performance, this pixel detector was replaced with a more efficient and robust four-point tracking system. In addition, the radius of the beam pipe was reduced in 2014 from 30 to 23 mm, which allowed the innermost pixel layer to be placed closer to the interaction point. The improved pixel detector was installed at the beginning of 2017.

The pixel detector was upgraded after Run 1, and the so-called *Phase 1 pixel detector* (details can be found in [105, 106]) is described by [97] as:

> The new detector, referred to as the Phase 1 pixel detector [105], consists of four barrel layers (L1–L4) at radii of 29, 68, 109, and 160 mm, and three disks (D1–D3) on each end at distances of 291, 396, and 516 mm from the center of the detector. The layouts of the two detectors, the original and the upgraded one, are compared in Figure 4.6. The new layout provides four-hit coverage, instead of three, for tracks up to an absolute pseudorapidity of 3.0.

Strip detector

The Silicon Strip Tracker (SST), together with the pixel detector, provides measurements of charged particle trajectories up to a pseudorapidity of $|\eta| < 2.5$ [97], which is shown in Figure 4.7. The SST is composed of single-sided p-on-n micro-strip sensors, and has [97]

> ten layers in the barrel region with four layers in the Tracker Inner Barrel (TIB) and six layers in the Tracker Outer Barrel (TOB). The TIB is supplemented with three Tracker Inner Disks (TID) at each end. In the forward regions, the detector consists of Tracker EndCaps (TEC). Each TID is composed of three rings of modules and each TEC is composed of up to seven rings. In TIB, TID, and in rings 1-4 of the TECs, sensors with a thickness of 320 μm are used, while in TOB and in rings 5–7 of the TECs, 500 μm thick sensors are used. The modules in the barrel layers measure r and ϕ coordinates, while the modules in the TECs and TIDs are oriented to measure the coordinates in ϕ and z.

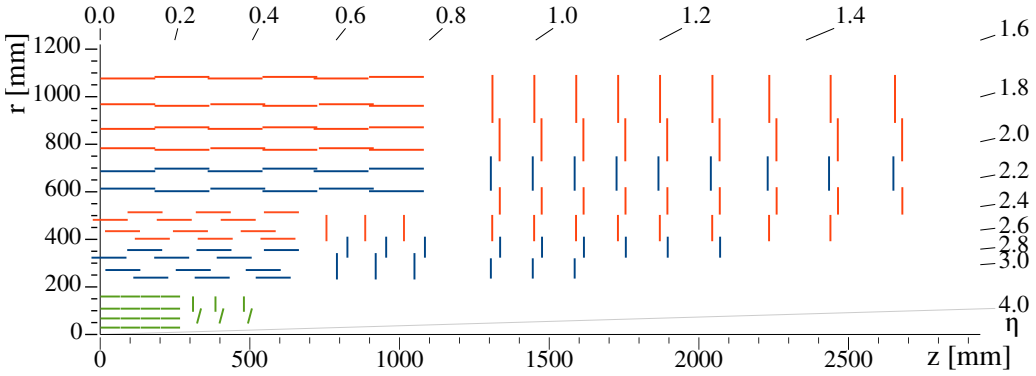


Figure 4.7: Schematic view of one quadrant in the r-z view of the CMS tracker: single-sided and double-sided strip modules are depicted as the red and blue segments, respectively. The pixel detector is shown in green. Figure is taken from [97].

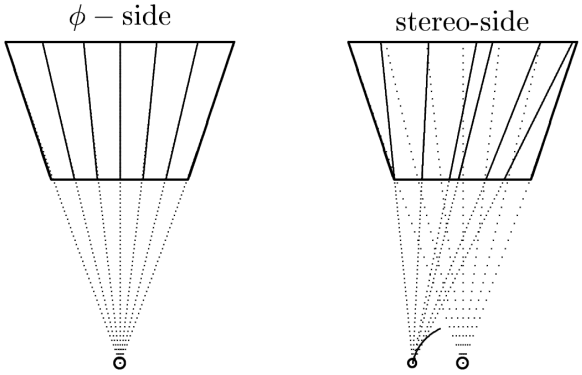


Figure 4.8: The layout of the strips of stereo module, which figure taken from [103].

In the four layers in the barrel and three rings in the endcaps, stereo modules are used [97]. The stereo modules are composed of a pair of wedge-shaped detectors, coupled back-to-back [103] (see Figure 4.8). In the front device, radial strips point to the beam line (ϕ-view), and in the device on the back, the strips are tilted by 100 mrad (stereo-view), thus pointing to a point on a concentric ring around the beam line [103]. The stereo modules provide coarse measurements of an additional coordinate (z in the barrel and r in the endcaps) [97].

4.2.4 Electromagnetic calorimeter

The Electromagnetic CALorimeter (ECAL), is designed to identify electrons and photons, and measure their positions and energies [97]. Electrons or photons are reconstructed from energy deposits using algorithms that constrain the clusters to the size and shape expected. The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons is generally better in the barrel region than in the endcaps, and also depends on the bremsstrahlung energy emitted by the electron as it traverses the material in front of the ECAL. The ECAL also provides information on the arrival time of the electrons and photons which can be used in physics analyses, such as searches for long-lived particles, and contributes to the reconstruction of jets and missing transverse momentum (p_T^{miss}).

The ECAL, of which a layout is shown in Figure 4.9, is made of lead tungstate ($PbWO_4$) scintillating crystals. [93] says about the crystals:

> These crystals have short radiation (X_0 = 0.89 cm) and Moliere (2.2 cm) lengths, are fast (80% of the light is emitted within 25 ns) and radiation hard (up to 10

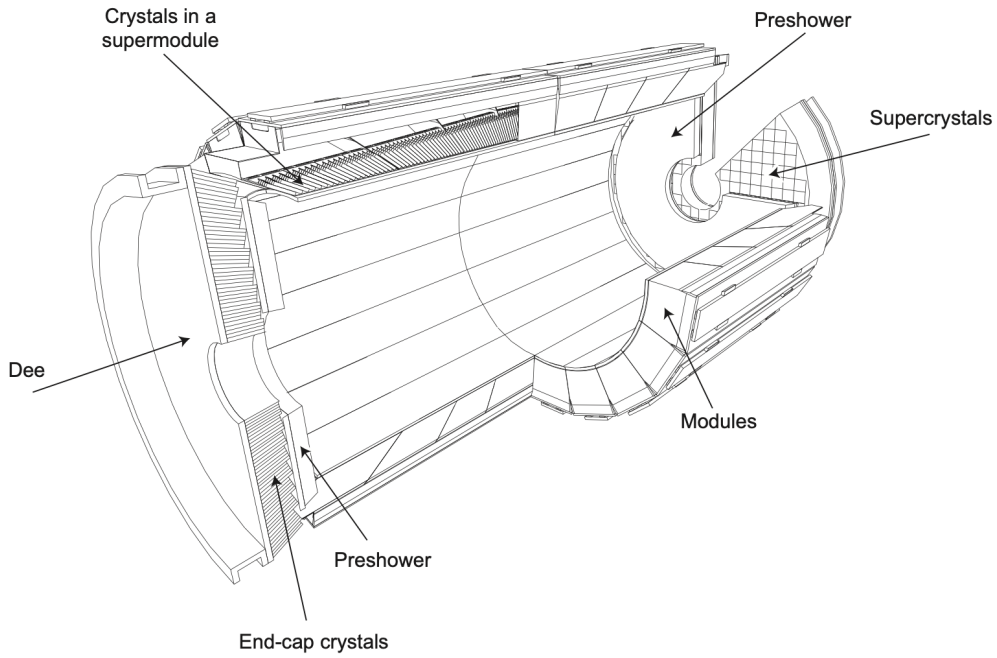


Figure 4.9: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. Taken from [94].

> Mrad). However, the relatively low light yield (30γ/MeV) requires use of photodetectors with intrinsic gain that can operate in a magnetic field. Silicon Avalanche PhotoDiodes (APDs) are used as photodetectors in the barrel and Vacuum PhotoTriodes (VPTs) in the endcaps. In addition, the sensitivity of both the crystals and the APD response to temperature changes requires a temperature stability (the goal is 0.1°C). The use of $PbWO_4$ crystals has thus allowed the design of a compact calorimeter inside the solenoid that is fast, has fine granularity, and is radiation resistant.

The ECAL has the barrel and endcap section. First, the barrel section (EB), which has an inner radius of 129 cm, is described in [93] as:

> It is structured as 36 identical "supermodules," each covering half the barrel length and corresponding to a pseudorapidity interval of $0 < |\eta| < 1.479$. The crystals are quasi-projective (the axes are tilted at 3° with respect to the line from the nominal vertex position) and cover 0.0174 (i.e. 1°) in $\Delta\phi$ and $\Delta\eta$. The crystals have a front face cross-section of $\sim 22 \times 22$ mm^2 and a length of 230 mm, corresponding to 25.8 X_0.

On the other hand, the endcap (EE) section is placed at a distance of 314 cm from the vertex, covering a pseudorapidity range of $1.479 < |\eta| < 3.0$ [93]: The EEs

> are each structured as 2 "Dees", consisting of semi-circular aluminium plates from which are cantilevered structural units of 5×5 crystals, known as "supercrystals." In the ECAL TDR [107] the basic mechanical unit was envisaged to hold 6×6 crystals. The change was accommodated by a corresponding increase in the lateral size of the crystals. The endcap crystals, like the barrel crystals, off-point from the nominal vertex position, but are arranged in an x-y grid (i.e. not an $\eta - \phi$ grid). They are all identical and have a front face cross section of 28.6×28.6 mm^2 and a length of 220 mm (24.7 X_0). A preshower device is placed in front of the crystal

> calorimeter over much of the endcap pseudorapidity range. The active elements of this device are 2 planes of silicon strip detectors, with a pitch of 1.9 mm, which lie behind disks of lead absorber at depths of 2 X_0 and 3 X_0.

4.2.5 Hadronic calorimeter

The Hadron CALorimeter (HCAL) contributes to the identification of hadrons and the measurement of their properties, and also aids in the reconstruction of jets and p_T^{miss}, and the identification of electrons and photons. Referring to [93], the design of the HCAL is influenced by the choice of magnet parameters since most of the CMS calorimetry is located inside the magnet coil and surrounds the ECAL system. [93] says:

> An important requirement of HCAL is to minimize the non-Gaussian tails in the energy resolution and to provide good containment and hermeticity for the E_T^{miss} measurement. Hence, the HCAL design maximizes material inside the magnet coil in terms of interaction lengths. This is complemented by an additional layer of scintillators, referred to as the Hadron Outer (HO) detector, lining the outside of the coil. Brass has been chosen as absorber material as it has a reasonably short interaction length, is easy to machine and is non-magnetic. Maximizing the amount of absorber before the magnet requires keeping to a minimum the amount of space devoted to the active medium. The tile/fibre technology makes for an ideal choice. It consists of plastic scintillator tiles read out with embedded WaveLength-Shifting (WLS) fibres. The WLS fibres are spliced to high-attenuation-length clear fibres outside the scintillator that carry the light to the readout system.

The HCAL comprises 4 subdetectors: the Hadron Barrel (HB), Hadron Endcap (HE), hadron outer, and Hadron Forward (HF) calorimeters (refer to Figure 4.10). The HB and HE cover the pseudorapidity regions $|\eta| < 1.392$ and $1.305 < |\eta| < 3.0$ [97], respectively. [97] says:

> The HB and HE primarily use brass as the absorber, except for the inner and outer layers of HB, which are constructed from steel. The HB absorber is shown in Figure 4.11 (left). The signals are produced in plastic scintillating tiles (Figure 4.11, right), and the resulting blue light is shifted to green via embedded wavelength-shifting fibers. The towers in HB(HE) have up to 17(18) scintillator layers, as shown in Figure 4.10. Sequential layers are grouped into "depth" segments: the light from the layers in a given depth segment is optically summed and read out by a single photodetector. Clear plastic fibers send the signal to the Hybrid PhotoDetectors (HPDs) in the original design or Silicon PhotoMultipliers (SiPMs) after the upgrades. The segmentation is a tower structure in $\eta - \phi$ space. The η segmentation is indicated by the black solid lines in Figure 4.10. The towers are referenced using integer indices ieta and iphi, where the ieta assignments are given in the figure and iphi runs from 0 to 71, corresponding to the 72 divisions in ϕ. Physically, the scintillators are arranged in "megatiles", which are trays that support an array of scintillator tiles, along with the fibers that route the light to the photodetectors.

The HO and HF were built to complement the HB and HE. [94] says:

> The hadron calorimeter barrel is radially restricted between the outer extent of the electromagnetic calorimeter (R = 1.77 m) and the inner extent of the magnet coil (R = 2.95 m). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an outer hadron calorimeter or tail catcher is placed outside the solenoid complementing the barrel calorimeter. Beyond $|\eta| = 3$, the forward hadron calorimeters placed at 11.2 m from the interaction point extend

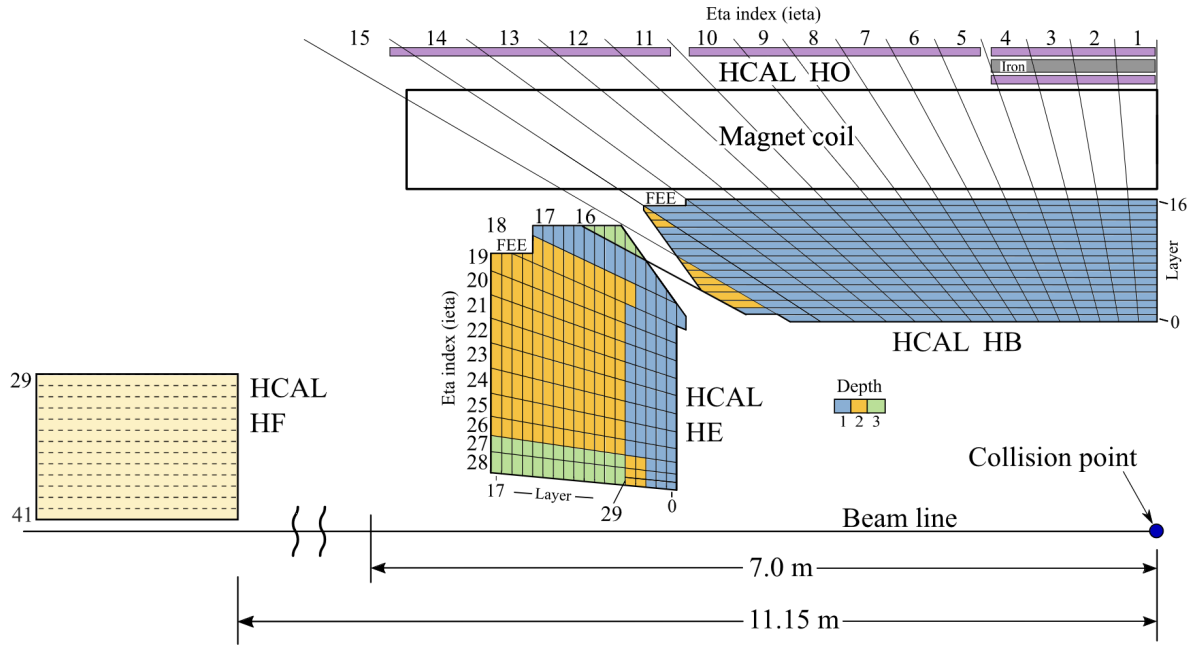


Figure 4.10: Schematic view of the HCAL as of 2016, showing the positions of its four major components: HB, HE, HO, and HF. The layers marked in blue are grouped together as "depth 1," i.e., the signals from these layers of a given tower are optically summed and read out by a single photodetector. Similarly, the layers shown in yellow and green are combined as depths 2 and 3, respectively, and the layers shown in purple are combined for HO. The notation "FEE" denotes the locations of the HB and HE frontend electronics readout boxes. The solid black lines, roughly projective with the interaction point, denote the η divisions in the tower $\eta - \phi$ segmentation, and the numbers at the edge of the tower denote the ieta index. Taken from [97].

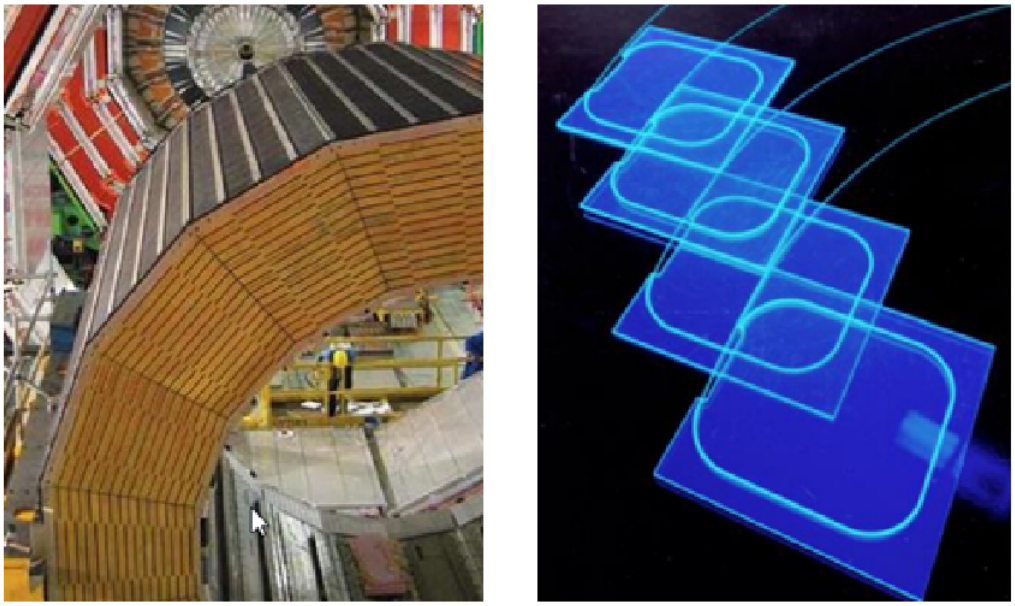


Figure 4.11: Brass absorber for the hadron barrel calorimeter HB (left) and scintillating tiles with wavelength shifting fibers used as the active media in the barrel, endcap, and outer hadron (right). Figures are taken from [97].

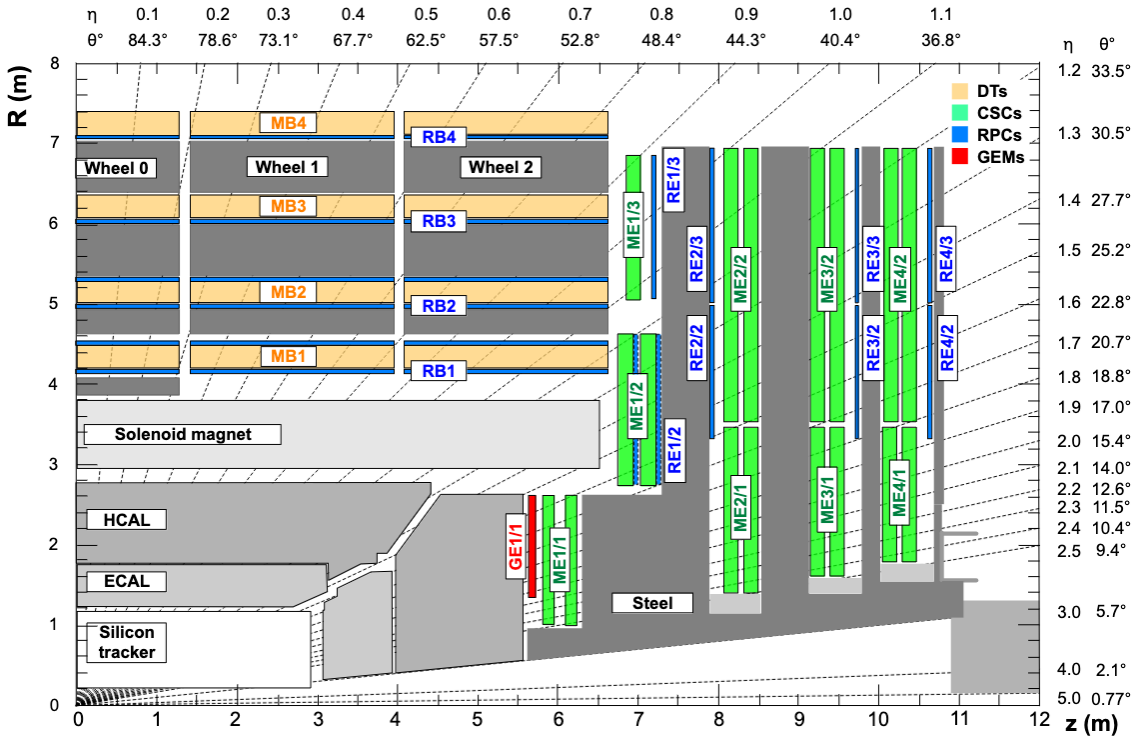


Figure 4.12: Longitudinal view of the CMS Muon system. The interaction point is in the lower left corner. The locations of the various muon stations are shown in color: drift tubes (DTs) with labels MB, cathode strip chambers (CSCs) with labels ME, resistive plate chambers (RPCs) with labels RB and RE, and gas electron multipliers (GEMs) with labels GE. The M denotes muon, B stands for barrel, and E for endcap. The magnet yoke is represented by the dark gray areas. Taken from [97].

> the pseudorapidity coverage down to $|\eta| = 5.2$ using a Cherenkov-based, radiation-hard technology.

4.2.6 Muon detector

Muons produced by beam collisions are measured in three different regions: in the inner tracker, after the coil, and in the return flux yoke, as shown in Figure 4.12. Depending on the radiation environments, different types of gaseous detectors are used to identify and measure muons.

First, Drift Tube (DT) chambers are used in the barrel region ($|\eta| < 1.2$) where the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low [93]. [97] describes the DT detector as:

> The DTs consist of chambers formed by multiple layers of long rectangular tubes that are filled with an Ar and CO_2 gas mixture. An anode wire is located at the center of each tube, whereas cathode and field-shaping strips are positioned on its borders. They create an electric field that induces an almost uniform drift of ionization electrons produced by charged particles traversing the gas. The charged-particle trajectory is determined from the arrival time of the currents generated on the anode wires of the readout.

Second, Cathode Strip Chambers (CSC) are deployed in the 2 endcaps, covering the region up to $|\eta| < 2.4$, where the muon rate as well as the neutron induced background rate is high, and the magnetic field is also high [93]. [97] also says:

> The CSCs are made of layers of proportional wire chambers with orthogonal cathode strips and are operated with a gas mixture of Ar, CO_2, and CF_4. Signals are generated on both anode wires and cathode strips. The finely segmented cathode strips and fast readout electronics provide good timing and spatial resolution to trigger on and identify muons.

Third, Resistive Plate Chambers (RPC) are used in both the barrel and the endcap regions. They are described in [97] as

> The RPCs comprise two detecting layers of high-pressure laminate plates that are separated by a thin gap filled with a gas mixture of $C_2H_2F_4$, i-C_4H_{10}, and SF_6. The electronic readout strips are located between the two layers, and the high voltage is applied to high-conductivity electrodes coated on each plate. The detectors are operated in avalanche mode to cope with the high background rates. Due to their excellent time resolution, they ensure a precise bunch-crossing assignment for muons at the trigger level.

In addition to these three systems, a Gas Electron Multiplier (GEM) detector has been recently added since Run 3 [97]:

> The key feature of the GEM is a foil consisting of a perforated insulating polymer surrounded on the top and bottom by conductors. A voltage difference is applied on the foils producing a strong electric field in the holes. The GEM is operated with a gas mixture of Ar and CO_2. When the gas volume is ionized, electrons are accelerated through the holes and read out on thinly separated strips. This structure allows for high amplification factors with modest voltages that provide good timing and spatial resolution, and can be operated at high rate.

4.3 Data analysis

4.3.1 Data acquisition and trigger system

The Data AcQuisition (DAQ) and trigger systems are responsible for online event selection. They consist of the detector electronics, the Level-1 (L1) trigger processors (calorimeter, muon, and global), the readout network, and an online event filter system (processor farm) that executes the software for the High-Level Triggers (HLTs) [93].

More specifically, the DAQ is responsible for the readout of all detector data for events accepted by the L1 trigger; the building of complete events from subdetector event fragments; the operation of the filter farm cluster running the HLT; and the transport of event data selected by it to the permanent storage in the Tier 0 computing center [97], which is the primary center located at CERN, supplemented by worldwide Tier 1 and Tier 2 centers. The DAQ consists of custom-built electronics reading out event fragments; a data-concentrator network transporting the fragments to the surface; a cluster of readout and event-building servers inter-connected via the event-building network; the filter-farm cluster of multicore servers connected by the data network running the HLT software; a distributed storage system where event data selected by HLT filtering are buffered; and a transfer system connected to the Tier 0 center via a high-speed network including the Trigger Control and Distribution System (TCDS), which distributes timing to the trigger and subdetector electronics, and implements trigger control logic as well as the Trigger Throttling System (TTS) [97].

Level-1 Trigger

The L1 trigger selects approximately 100 kHz of events based on coarse information from the calorimeters and the muon detectors. It consists of electronics responsible for making a fast

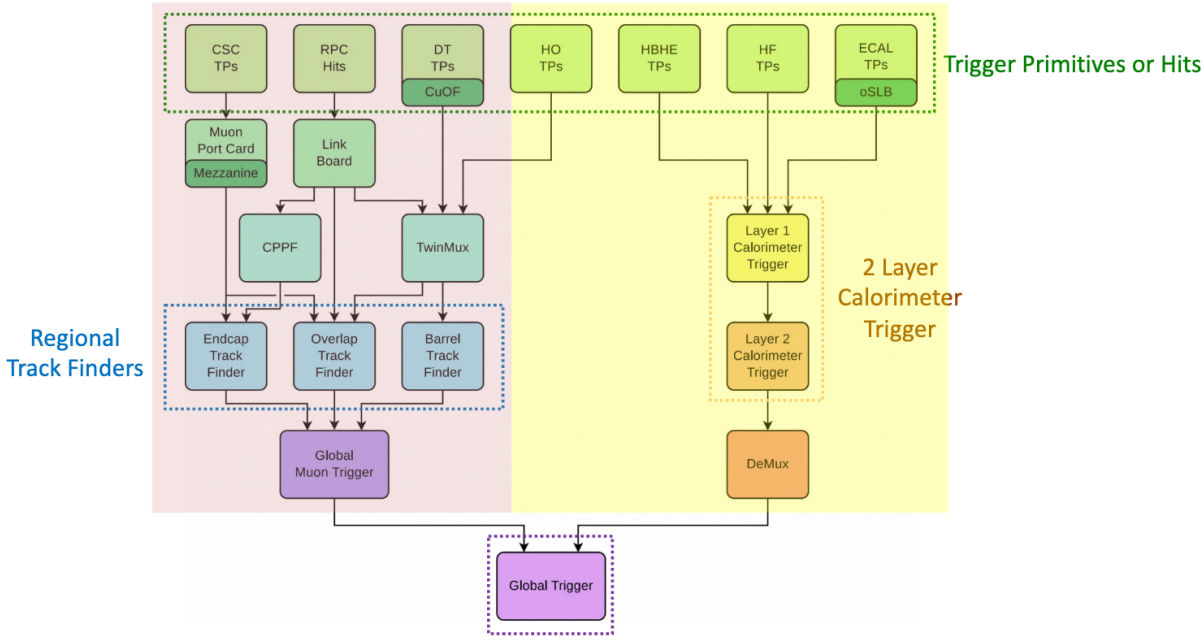


Figure 4.13: Level-1 trigger system, with figure taken from [108].

selection of events based on the presence of high-energy particles in the detector. [97] says:

> The L1 trigger receives energy and position information, so-called Trigger Primitives (TPs), from the calorimeters and the muon detectors. The TPs are evaluated by a trigger processor, which is composed of custom-built electronics and Field Programmable Gate Array (FPGA) devices that perform the trigger decision based on a set of predefined trigger algorithms. The L1 trigger operates at trigger rates of about 110 kHz. During LS2, the L1 trigger was upgraded to also process TPs that are designed to select long-lived particles.

Overall the L1 trigger system is shown in Figure 4.13. The L1 trigger [97]

> comprises calorimeter and muon trigger systems that provide jets, e/γ, hadronic τ, and muon candidates, along with calculations of energy sums, to the Global Trigger (GT). At the GT, the trigger decision is generated, based on the multiplicity and kinematic information of the various candidate trigger objects. The trigger configuration is implemented in a trigger "menu" comprised of several hundred "seed" algorithms. Upon a positive GT decision, the full detector data are read out for further filtering in the HLT.

The calorimeter layer 1 trigger [97]

> receives TPs from ECAL, HCAL, and HF, calibrates them, combines the ECAL and HCAL TPs into single Trigger Towers (TTs), and transmits the TTs to layer 2 for further processing. Calorimeter TPs for triggered events are readout to DAQ, and used in the Data Quality Monitoring (DQM) system, where they provide the input to the software emulator, such that online and emulated data can be compared in real time for monitoring purposes.

The calorimeter layer 2 [97]

> receives calibrated TTs from layer 1, reconstructs jet, e/γ, and τ candidates, and computes energy sums. The energies of jet, e/γ, and τ candidates are calibrated

> as a function of p_T and η, and isolation and ID criteria are applied to e/γ and τ candidates. Pileup mitigation is applied to all objects to reduce the rates while maintaining high efficiencies.

The L1 muon trigger [97]

> receives TPs from four partially overlapping muon subdetectors: DT, CSC, RPC, and GEM. The L1 muon trigger system reconstructs muon tracks and provides measurements of muon track parameters using TPs which provide position, timing, and quality information from detector hits.

In the Barrel Muon Track Finder (BMTF) [97],

> the kBMTF algorithm reconstructs muons in the barrel region using TPs received from the DTs and RPCs via the TwinMux concentrator cards, and has been used online since 2018. It is the successor of the original BMTF algorithm, used between 2016 and 2018, and based on an approximate Kalman filter algorithm in which muon tracks are reconstructed from detector hits starting from the outermost muon station and propagating inwards while updating the track parameters.

The Overlap Muon Track Finder (OMTF) builds muon tracks using the TPs from the DTs and RPCs in the barrel and CSCs and RPCs in the endcap using "a Bayes classifier algorithm" [97]. The Endcap Muon Track Finder (EMTF) builds muon tracks using TPs from CSCs, RPCs, and GEMs in both endcaps [97].

High-Level triggers

The HLT, which runs on a farm of commercial computer nodes integrated with the DAQ data flow, processes fully assembled events, applying algorithms similar to those used in offline reconstruction, and selects a few kHz of events for storage on disk [97]. The HLT data processing uses the concept of *paths* to structure its workflow [97]:

> These paths are sequences of algorithmic steps designed to reconstruct physics objects and make selections based on specific physics requirements. Steps within a path are typically organized in ascending order of complexity, reconstruction refinement, and physics sophistication. For example, the resource-intensive track reconstruction process is usually carried out after completing a series of initial reconstruction and selection steps involving the data from the calorimeters and muon detectors.

The HLT selects data for storage through the application of a trigger *menu*, where the collection of individual HLT paths is configured [97]:

> The trigger path definitions, physics object thresholds, and rate allocations are set to meet the physics objectives of the experiment. In 2022, the HLT menus for pp data taking typically contained around 600 paths. This includes the primary HLT paths for analysis as well as paths for calibration and efficiency measurements that are typically looser than the primary paths. These latter HLT paths are often *prescaled*, i.e., only a fraction of the events that pass the requirements are actually accepted, to limit processing time and storage rate. Different trigger menus are used for the recording of heavy-ion collision data.

In this thesis, the so-called *B parking* data sample is used to measure charm production at $\sqrt{s} = 13$ TeV (see the next chapter). Data parking is introduced since Run 1, which is defined as [97]

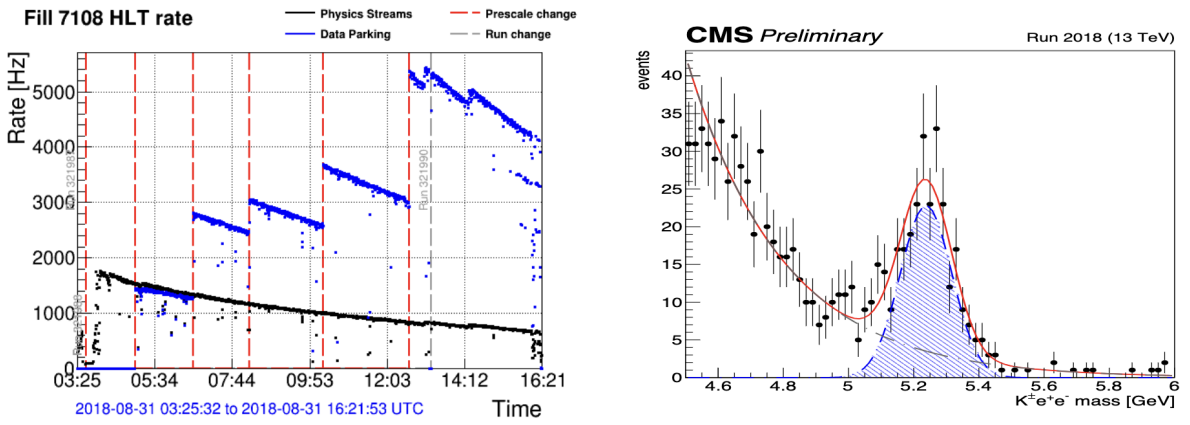


Figure 4.14: The left figure shows the trigger rates of the CMS HLT system as a function of time during an LHC fill in 2018, for the physics (the black data points) and B parking (the blue data points) streams. The right figure is an example of reconstruction on B parking data; $B^+ \to K^+(J/\psi \to)e^+e^-$ reconstruction on a small fraction of the B parking data sample. Figures are taken from [109].

> One limiting factor for the HLT output rate is the bandwidth of the prompt event reconstruction at Tier 0. An alternative approach to increase the amount of data available for physics analysis is to increase the storage rate on disk, while delaying the reconstruction of the data until a later time, when the necessary computing resources are available. The reconstruction can be scheduled during a year-end technical stop or a long shutdown, for instance. This concept, known as *data parking*, was already implemented in Run 1 and Run 2 to record additional data for B physics and other studies. During an LHC fill, as the luminosity decreases, the bandwidth to trigger additional events increases, and the trigger thresholds for data-parking events are gradually relaxed to record parking data.

Specifically in 2018 (the 13 TeV data which will be used in the next chapter) [97],

> the collection of $b\bar{b}$ events was enhanced by tagging and storing events containing at least one displaced muon, e.g., from a semileptonic B decay. The p_T threshold at the HLT for a single isolated muon was 24 GeV for the standard physics menu. For the parked B data the p_T threshold was as low as 7 GeV and the HLT output rate reached 5 kHz at the end of fills, enabling CMS to accumulate about 10^{10} b hadrons [109].

In Figure 4.14, the trigger rates for the B parking stream in 2018 are shown.

4.3.2 Event reconstruction

The reconstruction process in CMS can be divided into 3 steps, corresponding to local reconstruction within an individual detector module, global reconstruction within a whole detector, and combination of these reconstructed objects to produce higher-level objects [93]:

> The reconstruction units providing local reconstruction in a detector module use as input real data from the DAQ system or simulated data representing the real data. These data in either case are called "digis". The output from the reconstruction units are "RecHits," reconstructed hits which are typically position measurements (from times or clusters of strips or pixels) in tracking-type detectors (Muon and Tracker

systems) and calorimetric clusters in Calorimeter systems. The RecHits are added to the event as EDProducts, and used as the input to the global reconstruction.

In the global reconstruction step information from the different modules of a subdetector are combined, although information from different subdetectors is not. For example, Tracker RecHits are used to produce reconstructed charged particle tracks and Muon RecHits are used to produce candidate muon tracks. Once again, the objects produced are added to the event as EDProducts.

The final reconstruction step combines reconstructed objects from individual subdetectors to produce higher-level reconstructed objects suitable for high-level triggering or for physics analysis. For example, tracks in the Tracker system and tracks in the Muon system are combined to provide final muon candidates, and electron candidates from the Calorimeter system are matched to tracks in the Tracker system.

Track and vertex reconstruction

In particular, track reconstruction can be described in 3 major steps: seed generation, pattern recognition, and track fitting and smoothing. The seed generation provides initial trajectory candidates for the full track reconstruction [93]:

A seed must define initial trajectory parameters and errors. They can be obtained externally to the Tracker, using inputs from other detectors, but the precision of initial trajectory parameters obtained in such a way is, in general, poor. Another way is to construct seeds internally. In this case each seed is composed from the set of reconstructed hits that are supposed to come from 1 charged particle track. Since 5 parameters are needed to start trajectory building, at least 3 hits, or 2 hits and a beam constraint, are necessary. If the beam constraint is used it is removed during the final fit. Hits that are seed constituents are provided by the dedicated reconstruction.

The pattern recognition is based on a combinatorial Kalman filter* method [93]:

First, a dedicated navigation component determines which layers are compatible with the initial seed trajectory. The trajectory is then extrapolated to these layers according to the equations of motion of a charged particle in a magnetic field, accounting for multiple scattering and energy loss in the traversed material.

Since several hits on the new layer may be compatible with the predicted trajectory, several new trajectory candidates are created, 1 per hit. In addition, 1 additional trajectory candidate is created, in which no measured hit is used, to account for the possibility that the track did not leave any hit on that particular layer. This *fake* hit is called an "invalid hit".

Each trajectory is then updated with the corresponding hit according to the Kalman filter formalism. This update can be seen as a combination of the predicted trajectory state and the hit in a weighted mean, as the weights attributed to the measurement and to the predicted trajectory depend on their respective uncertainties.

* [110] says: A Kalman filter is mathematically equivalent to a global least-squares minimization, which is the optimal estimator when the model is linear, all random noise is Gaussian and there are no outlying measurements. In that case, the estimators are unbiased and have minimum variance, residuals and pulls of estimated quantities have Gaussian distributions and the value of the objective function (the function which is minimized) at the minimum obeys a χ^2 distribution. For non-linear models or non-Gaussian noise, it is still the optimal linear estimator. It is nevertheless very sensitive to outliers, and the bias in the fitted parameters is proportional to the bias in the outlying measurements

> All resulting trajectory candidates are then grown in turn to the next compatible layer(s), and the procedure is repeated until either the outermost layer of the tracker is reached or a "stopping condition" is satisfied. In order not to bias the result, all trajectory candidates are grown in parallel. To avoid an exponential increase of the number of trajectory candidates, the total number of candidates is truncated at each layer. To limit the number of combinations, and hence to avoid an exponential increase thereof, only a limited number of these are retained at each step, based on their normalized χ^2 and number of valid and invalid hits.

The track fitting and smoothing steps are required as [93]:

> For each trajectory, the building stage results in a collection of hits and in an estimate of the track parameters for each trajectory. However, the full information is only available at the last hit of the trajectory and the estimate can be biased by constraints applied during the seeding stage. Therefore the trajectory is refitted using a least-squares approach, implemented as a combination of a standard Kalman filter and smoother.
>
> The Kalman filter is initialized at the location of the innermost hit with an estimate obtained during seeding. The corresponding covariance matrix is scaled by a large factor in order to avoid any bias. The fit then proceeds in an iterative way through the list of hits. For each valid hit the position estimate is re-evaluated using the current values of the track parameters: information about the angle of incidence increases the precision of the measurement especially in the pixel modules. The track parameters and their covariance matrix are updated with the measurement and the trajectory is propagated to the surface associated with the next hit. The track parameters and their covariance matrix are modified according to the estimates for energy loss and multiple scattering at the target surface and the sequence is repeated until the last hit is included.
>
> This first filter is complemented with a smoothing stage: a second filter is initialized with the result of the first one (except for the covariance matrix, which is scaled with a large factor) and run backward toward the beam line. At each hit the "updated" parameters of this second filter, which contain all information from the outermost hit up to and including the current hit, are combined with the "predicted" parameters of the first filter, i.e., the information from the innermost hit outward, but excluding the current hit.
>
> This filtering and smoothing procedure yields optimal estimates of the parameters at the surface associated with each hit and, specifically, at the first and the last hit of the trajectory. Estimates on other surfaces, e.g., at the impact point, are then derived by extrapolation from the closest hit.

On the other hand, vertex reconstruction usually involves 2 steps, vertex finding and vertex fitting [93]:

> Vertex finding involves grouping tracks into vertex candidates. The vertex-finding algorithms can be very different depending on the physics case (primary or secondary vertex finding, reconstruction of exclusive decays, etc.). Vertex fitting involves determining the best estimate of the vertex parameters (position, covariance matrix, and track parameters constrained by the vertex position and their covariances) for a given set of tracks, as well as indicators of the fit quality (total χ^2, number of degrees of freedom, or track weights).

The most often used algorithm for vertex fitting is the Kalman filter.

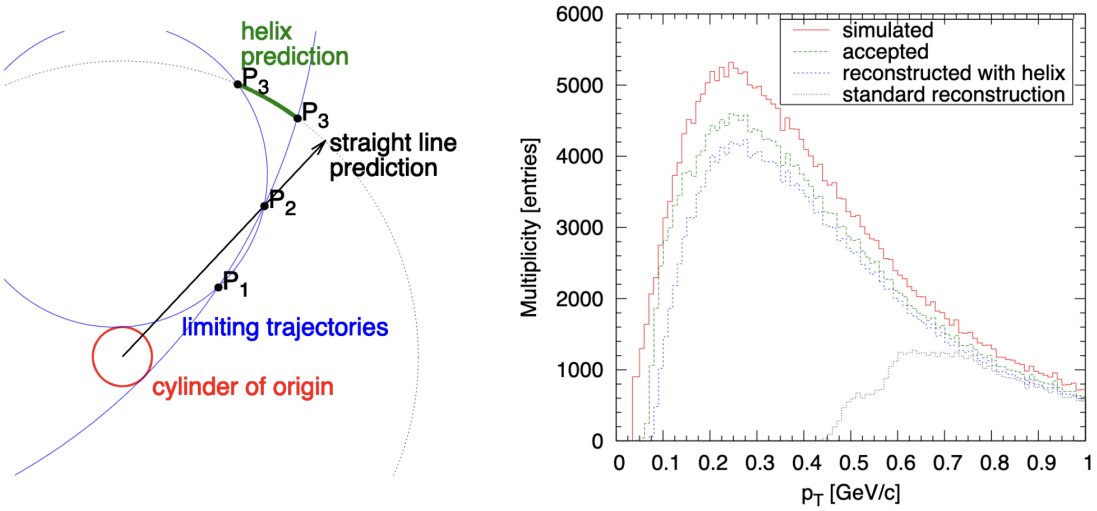


Figure 4.15: Left: Schematic comparison of the standard straight line prediction and the new helix prediction for finding the third hit. Right: Transverse momentum distributions of the charged particles: simulated (solid red), accepted (green dashed) and reconstructed, with the standard method (dotted black) or with the new helix method (dotted blue). Taken from [111].

For the measurement of charm production in this thesis, the lowest possible momentum is of importance to be reconstructed. In case of 2010 data, which are used for charm production measurements at $\sqrt{s} = 7$ TeV (see the next chapter), low p_T tracks were reconstructed using a special algorithm. This is an improved tracking algorithm which reconstructs tracks down to 0.1 GeV/c, using just the three pixel layers, with the modified hit triplet finding and cleaning procedures [111]. The track finding procedure [111]

> starts by pairing two hits from different layers (see Figure 4.15). During the search for the third hit, the following requirements must be fulfilled: the track must come from the cylinder of origin (given by its radius, half-length and position along the beam-line); the p_T of the track must be above the minimal value $p_{T,\mathrm{min}}$; and the track must be able to reach the layer where the third hit may be located. In the small volume of the pixel detector the magnetic field is practically constant and the charged particles propagate on helices. The projection of a helix or a cylinder onto the transverse plane is a circle. Each requirement defines a region of allowed track trajectories. They are enclosed by a pair of limiting circles which can be constructed using simple geometrical transformations. A third hit candidate is accepted if its position is within a region which takes into account the expected multiple scattering.

Then, the triplet cleaning is required as [111]:

> While high p_T tracks are relatively clean, uncorrelated hit clusters can often be combined to form fake low p_T tracks. However, a cluster contains more information than its position. The geometrical shape of the hit cluster depends on the angle of incidence of the particle: bigger angles will result in longer clusters. We can, thus, check whether the measured shape of the cluster is compatible with the predicted angle of incidence of the track; if any of the hits in the triplet is not compatible, the triplet is removed from the list of track candidates.

4.3.3 Detector simulation

Primary physics processes are generated in CMS by programs such as PYTHIA [30, 31] and MadGraph [112], and the output particle information is converted into the standard HepMC

format [113]. The detector simulation employs models of various types of particle interactions with materials, called "physics lists", to propagate particles through the detector [114]. This simulation is based on the GEANT4 software [115, 116]. It is augmented with computationally efficient techniques, such as shower libraries for the forward calorimeters, and specific identification criteria per particle type and detector region for neutrons, which guarantee high fidelity of the simulation [97]. Subsequently, the response from detector electronics is simulated in a dedicated digitization step to produce output signals, and then various reconstruction algorithms are applied to those signals [114]:

> The detector simulation step is the most expensive in terms of CPU usage, consuming 40% of the total computing budget of CMS. The other steps in producing Monte Carlo samples–event generation, digitization, reconstruction, and analysis–together consume 45%, with reconstruction as the largest contributor. The remainder of the CPU time is used to process observed data. Within the detector simulation step, evaluating the geometry and magnetic field propagation uses 60% of the time; electromagnetic physics models use 15%; hadronic physics use 10%; and other components, including CMS-specific operations, use the remaining 15%.

4.3.4 Software and data tiers

The HLT, reconstruction, simulation and analysis are introduced in the framework of the CMS software CMSSW. The framework [117]

> is centered around the concept of an Event. A data processing job is composed of a series of algorithms (e.g., a track finder or track fitter) that run in a particular order. The algorithms only communicate via data stored in the Event. To facilitate testing, all data items placed in the Event are storable to ROOT/IO using POOL. This allows one to run a partial job (e.g., just track finding) and check the results without having to go through any further processing steps. In addition, the POOL/ROOT files generated by the framework are directly browseable in ROOT. This allows one to accomplish simple data analyses without any additional tools. More complex studies can be supported in ROOT just by loading the appropriate shared libraries that contain the dictionaries for the stored objects.

Data and simulation processing workflows are broken into several steps, each defined by the output data structures per event it produces, referred to as a *data tier*: a single executable process may produce multiple outputs corresponding to different data tiers, which reduces I/O operations when the necessary data structures are already in process memory [97]. The data tiers in use for centrally produced simulation and reconstruction workflows (refer to Figure 4.16) are defined by [97] as the followings:

- GEN; Intermediate and outgoing stable ($c\tau \gtrsim 1$ cm) particles from the collision simulation. May include Les Houches accord event (LHE) data from the matrix-element generator, if applicable.

- SIM; Detailed description of energy deposits left by stable outgoing particles in the detector material. Two options are available: a highly-accurate GEANT4-based application (Full MC); and a parametric fast simulation application (Fast MC), which trades accuracy for a 100-fold decrease in detector simulation time or 10-fold decrease in total CPU time per simulated event. The level of inaccuracy introduced by Fast MC is typically a difference of less than 10% in final analysis observables

- DIGI; Digitized detector readout or simulation thereof. In simulation, the effect of additional collision events (pileup) is folded into the event description in this step. In Run

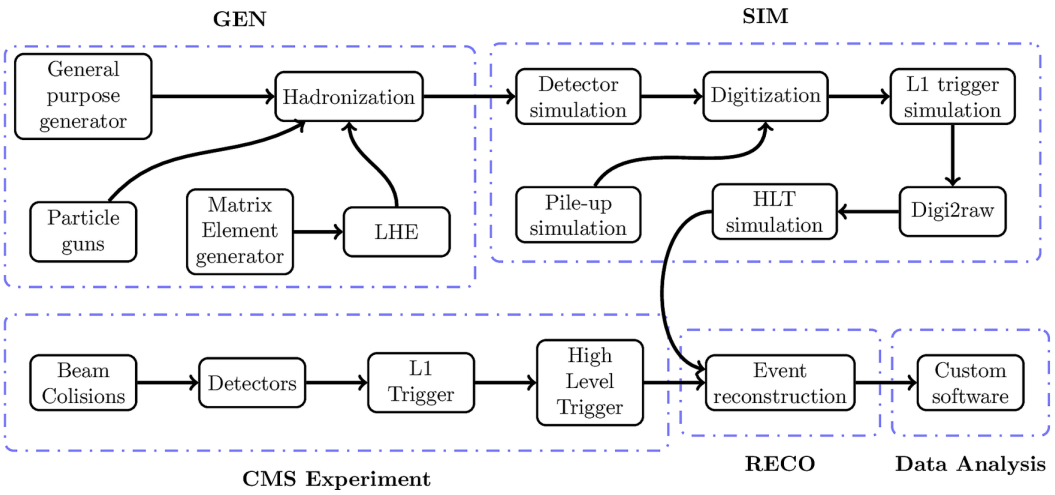


Figure 4.16: Overview of the main steps in the production chain of simulated data. Figure is taken from [118].

2, a 'premixing' technique was introduced, where the additional events are summed in a separate processing step and then applied to the simulated primary event

- RAW; Packed detector readout data
- RECO; Detailed description of calibrated detector hits and low-level physics objects
- AOD; Reduced description of calibrated detector hits and low-level physics objects, uncalibrated high-level physics objects
- MiniAOD; Reduced low-level physics objects and calibrated high-level physics objects. A truncated floating-point representation is used for most object attributes. Introduced for Run 2 to reduce the number of analyses requiring AOD inputs
- NanoAOD; Compact data format containing only high-level physics object attributes stored as (arrays of) primitive data types. Introduced during Run 2 to reduce the number of analyses requiring MiniAOD inputs

An overview of the main steps in the production chain of the simulated data is illustrated in Figure 4.16. The smaller-size analysis formats are key to reducing both the overall amount of data stored by CMS and analysis processing time per event; in Run 2, approximate event sizes in each format are 400 kB for AOD (Analysis Object Data), 40 kB for MiniAOD, and 1–2 kB for NanoAOD [97].

5 D^{*+} and D^0 Production Measurements in CMS

5.1 Introduction

In this chapter, physics analyses in CMS specifically for charm are introduced by measuring D^{*+} and D^0. The excited-state charmed meson, D^{*+}, has the best significance of signal to background, of which the most probable decay chains are given by $D^{*+} \to D^0\pi_s^+$ and $D^0 \to K^-\pi^+$ with branching fractions, 0.677 ± 0.005 and 0.03947 ± 0.00030 [42], respectively. Here π_s^+ specifies a π^+ which has the special kinematics constrained by the difference between $M_{D^{*+}} = 2010.26 \pm 0.05$ MeV and $M_{D^0} = 1864.84 \pm 0.05$ MeV [42]. Due to the small phase space window with $M_{\pi^+} = 139.57039 \pm 0.00018$ MeV [42], the momentum of π_s^+ is lower compared to D^{*+} and D^0. Thus π_s^+ has a larger curvature compared to the others as illustrated in the left panel of Figure 5.1. This π_s^+ distinguishes K^- and π^+ for the signal even without particle identification. Furthermore, since D^0 and π_s^+ are produced almost at rest in the D^{*+} rest frame, $p_T^{D^0}$ and $p_T^{\pi_s^+}$ scale with their relative masses (see the right panel of Figure 5.1), and the direction of D^0 and π_s^+ essentially coincide. $M_{D^{*+}} - M_{D^0}$ has a good resolution with strongly suppressed combinatorial background. D^0 has the largest fragmentation fraction among the weakly-decaying ground states.

Taking the advantage of the D^{*+} decay introduced above, D^{*+} was considered to be measured first, which resulted in more detailed discussions being provided for this thesis compared to the D^0 measurement. In CMS, especially for the purpose of measuring the total charm cross section, the largest phase space possible was taken to measure D^{*+} at different center-of-mass energies, $\sqrt{s} = 7$ [32, 119], 5 [120], 0.9 [121] and 13 TeV. In particular, the 7, 0.9 and 13 TeV analysis will be discussed in this chapter, which I actually worked for. In the 7 TeV analysis, D^{*+} has been measured on 2010 pp data, which was started by my previous colleague, Nur Zulaiha Binti Jomhari in [119] where preliminary prompt D^{*+} cross sections and the total charm cross section were presented. Further studies to complete systematic uncertainties have been provided after the previous study and will be discussed in this chapter more in detail: the determination of the non-prompt contamination, the signal extraction systematics introducing a fit method, and pileup/era dependence. Similarly D^{*+} has been measured also at $\sqrt{s} = 0.9$ and 13 TeV. An illustration of the tracker and reconstruction was performed with the D^{*+} reconstruction at $\sqrt{s} = 0.9$ TeV for the early Run 3 [83], which was the very first charm production publicly shown at $\sqrt{s} = 0.9$ TeV ever in LHC. Afterwards, the data were kindly taken over by Luca Della Penna to provide a detailed analysis for D^{*+} cross section measurement [121], of which results were used as input to derive the total charm cross section at $\sqrt{s} = 0.9$ TeV in the next chapter and thus will also be briefly introduced in this chapter. In the meantime, while writing this thesis, the Run 3 samples have been combined with the 2010 0.9 TeV samples, which increases the statistics slightly. The 13 TeV analysis was performed on 2018 pp data starting with collecting almost 200 billion pileup collisions. Separately from D^{*+}, measurements of the D^0 cross section were started with these data. Preliminary prompt D^{*+} cross sections at $\sqrt{s} = 13$ TeV were obtained and will be shown in this chapter. A very first attempt to measure prompt

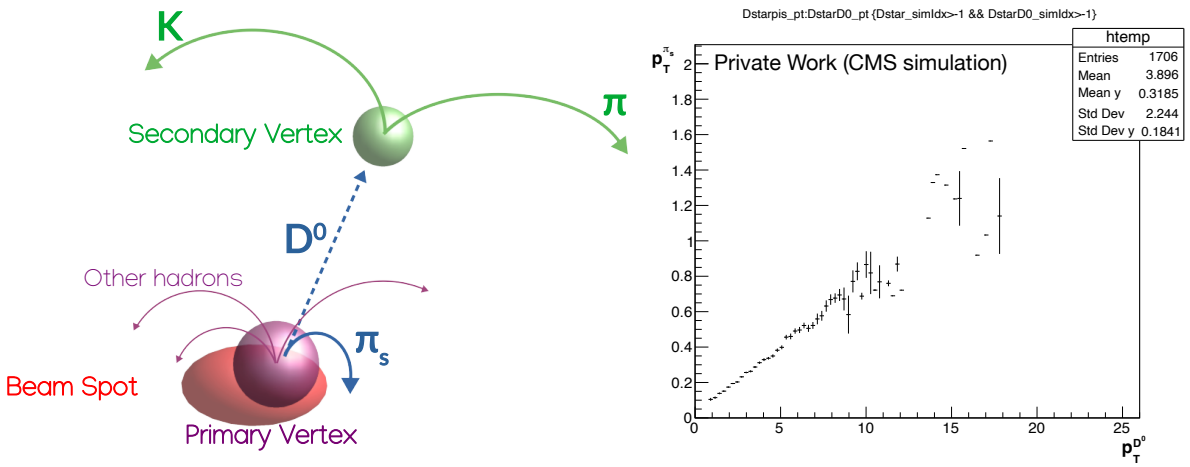


Figure 5.1: An illustration of decay chains, $D^{*\pm} \to D^0\pi_s^\pm \to K^\mp\pi^\pm\pi_s^\pm$ (left), and a demonstration of strongly correlated D^0 transverse momentum ($p_T^{D^0}$) and $\pi_s^\pm$ transverse momentum ($p_T^{\pi_s}$) (right). D^0 and $\pi_s^\pm$ in the right figure are decay products of $D^{*\pm}$ which was reconstructed with matching generated $D^{*\pm}$ and D^0 on a MC sample.

D^0 cross sections will also be discussed. Apart from the cross section measurements, a special study has been also introduced at $\sqrt{s} = 13$ TeV to measure the π_s^+ tracking efficiency by taking the ratio between the separately reconstructed D^{*+} and D^0, which is discussed in Appendix C.

Eventually, the D^{*+} cross sections at $\sqrt{s} = 7$ and 0.9 TeV were taken as inputs for total charm cross section measurements in the next chapter, which were derived by introducing the non-universal charm fragmentation.

5.2 Samples and event selection

Starting from the 7 TeV analysis, D^{*+} has been measured on 2010 data with the effective luminosity of 3.0 nb^{-1} [119, 122]. For the π_s^+ reconstruction which requires the lowest possible p_T, the 2010 samples of the special low p_T tracking [111] were taken to perform the analysis as mentioned in the previous chapter, which are listed in Table 5.1. These datasets were recorded during the 2010 commissioning and Run A-B data taking periods, of which all are available as Open Data [123]. In Table 5.1, the second column shows the number of events survived with a so-called JSON selection certified by CMS. All events except in ZeroBias and MinimumBias samples were provided with the JSON selection. The numbers are slightly updated compared to the ones in [119], although the rounded effective luminosity ($\mathcal{L}$ in Table 5.1) of the total remains the same as before. The effective luminosities were calculated by taking the number of collisions (denoted by (N)MB in Table 5.1) instead of events, as done in [119]:

$$\mathcal{L} = \frac{\text{(N)MB}}{\sigma_{\text{eff}}}, \tag{5.1}$$

where $\sigma_{\text{eff}} = 52.1 \pm 2.1$ mb. Here σ_{eff} is the effective cross section determined in a data driven way using unprescaled triggers (further details can be found in [124]). (N)MB was introduced to maximize statistics. With the limited statistics of these samples, all the pileup collisions were considered for D^{*+} events. In case of the ZeroBias and MinimumBias samples, all good Primary Vertex (PV) collisions were collected, which are called MB collisions. A good PV is defined with the following criteria [119, 124]:

- vertex is from primary vertex collection (with beamspot constraint)

Table 5.1: List of 2010 7 TeV data and MC samples. Data samples have the updated numbers compared to the previous study [119]. The effective luminosity of the total remains the same as before.

Data sample	#Events (w. JSON)	(N)MB	$\mathcal{L}$ [nb^{-1}]
/ZeroBias/Commissioning10-May19ReReco-v1/RECO	129,186,198	646,507	0.012
/ZeroBias/Run2010A-Apr21ReReco-v1/AOD	34,923,622	9,884,270	0.190
/MinimumBias/Commissioning10-May19ReReco-v1/RECO	46,553,963	32,316,473	0.620
/MinimumBias/Run2010A-Apr21ReReco-v1/AOD	103,848,957	25,951,045	0.498
/MinimumBias/Run2010B-Apr21ReReco-v1/AOD	40,785,403	16,092,376	0.309
/MuOnia/Run2010A-Apr21ReReco-v1/AOD	29,583,697	4,259,684	0.082
/MuOnia/Run2010B-Apr21ReReco-v1/AOD	22,677,534	20,388,793	0.391
/Mu/Run2010A-Apr21ReReco-v1/AOD	20,850,125	6,052,663	0.116
/Mu/Run2010B-Apr21ReReco-v1/AOD	26,718,043	15,131,362	0.290
/MuMonitor/Run2010A-Apr21ReReco-v1/AOD	11,367,866	740,827	0.014
/MuMonitor/Run2010B-Apr21ReReco-v1/AOD	10,085,889	1,836,674	0.035
/EG/Run2010A-Apr21ReReco-v1/AOD	47,187,984	5,034,671	0.097
/Electron/Run2010B-Apr21ReReco-v1/AOD	27,856,626	16,598,770	0.319
/EGMonitor/Run2010A-Apr21ReReco-v1/AOD	25,665,131	3,324	0.000
/EGMonitor/Run2010B-Apr21ReReco-v1/AOD	8,961,541	1,457,837	0.028
Total	586,252,579	156,395,276	3.001
MC sample	#Events		$\mathcal{L}$ [nb^{-1}]
/D0Kpi_pT0toInf_TuneZ2star_7TeV-pythia6-evtgen /LowPU2010DR42-NoPU2010_DR42_START42_V17B-v2/AODSIM	5,801,549		20.37

- vertex is valid (fit converged reasonably)
- vertex is not fake (not an empty vertex)
- number of degrees of freedom, ndof > 4
- $|\text{vertex} - \text{beamspot}| < 15$ cm in z-direction
- $|\text{vertex} - \text{beamspot}| < 2$ cm in xy-plane.

Alternatively, so-called Next-to-Minimum-Bias (NMB) collisions, which are again pileup collisions with good PV, were collected from the lepton trigger samples. The PV associated to triggering leptons can be uniquely identified and NMB collisions are defined to be separated from trigger vertex by at least 1 or 3 cm depending on the vertex-association. All tracks and vertices related to triggering collisons (denoted by TVx collisions in Figure 5.2) and events in which the trigger vertex could not be uniquely defined were removed. Typical pileup in these data is ~ 2. The MC main sample to evaluate the reconstruction efficiency is given by Minimum-Bias events enriched by the decay $D^0 \to K^{\mp}\pi^{\pm}$, which is shown also in Table 5.1.

The 13 TeV analysis was performed on the 2018 B parking data samples [109] as mentioned in the previous chapter, which were collected with the integrated luminosity ~ 40 fb^{-1} [125]. The samples are listed in Table 5.2. These samples were provided with a set of single-muon triggers, and thus contain both TVx and NMB collisions. However, the beauty and charm contributions differ much such that TVx and NMB collisions are dominated by beauty and charm production, respectively. In this thesis, for the purpose of measuring the charm cross section, only the NMB collisions were considered on these datasets. The HLTs used to select collisions are shown in Table 5.3. For instance, the p_T distributions of D^0 candidates are compared with and without separation of the TVx and NMB collisions in Figure 5.3. It is clearly shown that the NMB collisions are free from the muon trigger bias. The data typically have a pileup of ~ 20 and thus the NMB collisions are ~ 200 billion with the raw event number

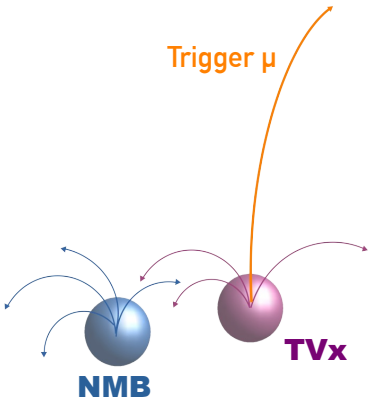
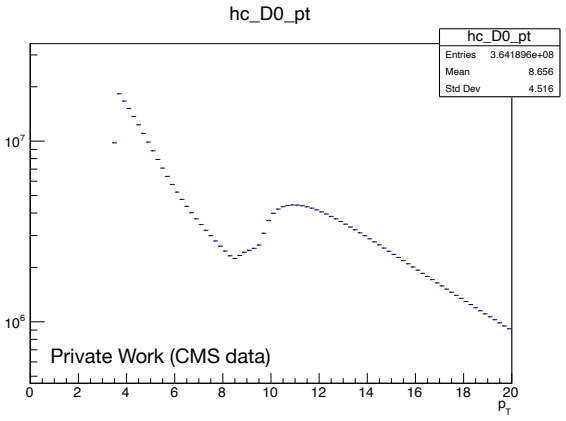
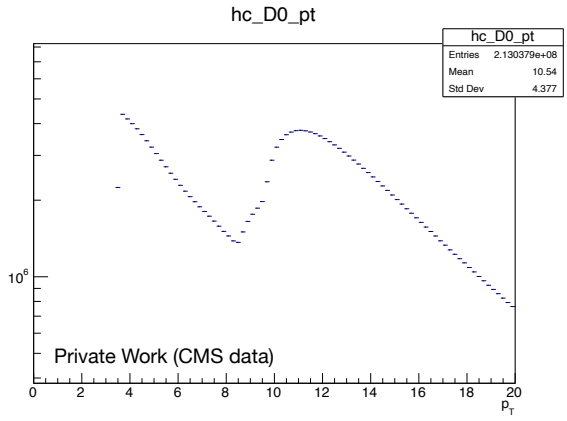


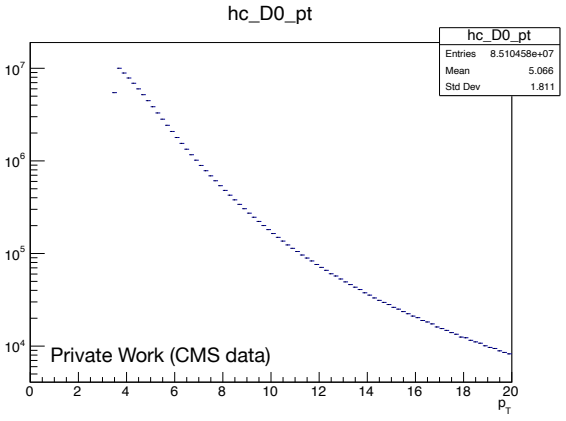
Figure 5.2: NMB and TVx collision definition in a muon trigger sample.



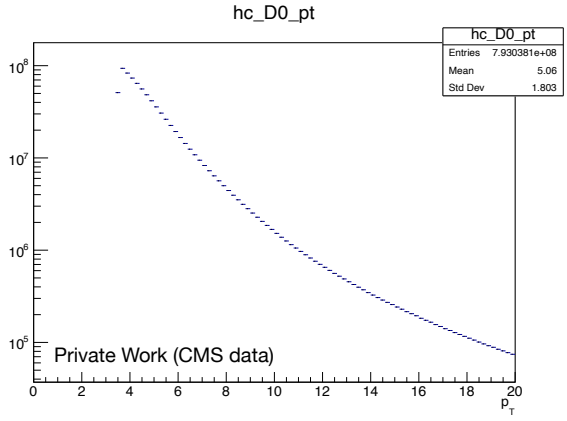
(a) Without the TVx and NMB separation.



(b) TVx collisions only.



(c) NMB collisions only.



(d) NMB collisions only.

Figure 5.3: p_T distributions of D^0 candidates on the 2018 B parking data. (a), (b) and (c) show Run A statistics only as an illustration. (d) shows NMB collisions with full statistics of the parking data, which were used eventually for the 13 TeV results.

Table 5.2: List of 2018 13 TeV ParkingBPH UL data samples. "#Events" refers to the raw number of events, while "w. JSON" column shows the number of events with applying a JSON selection provided by Table 5.4. Due to a technical issue on the storage, a few % of the originally collected events were lost, and "w. JSON survived" shows the numbers which survived after all.

Sample	#Events	w. JSON	w. JSON survived	NMB
/ParkingBPH1/Run2018A -20Jun2021_UL2018-v1/AOD	205,962,879	192,447,960	187,442,212	3,214,374,252
/ParkingBPH2/Run2018A -20Jun2021_UL2018-v1/AOD	205,968,583	192,467,153	167,638,775	2,892,117,178
/ParkingBPH3/Run2018A -20Jun2021_UL2018-v1/AOD	205,928,713	192,418,621	184,952,613	3,173,432,860
/ParkingBPH4/Run2018A -20Jun2021_UL2018-v1/AOD	205,977,925	192,466,177	178,702,567	3,066,599,547
/ParkingBPH5/Run2018A -20Jun2021_UL2018-v1/AOD	205,932,887	192,422,866	184,001,070	3,156,536,075
/ParkingBPH6/Run2018A -20Jun2021_UL2018-v1/AOD	205,981,846	192,471,696	168,380,141	2,880,631,407
/ParkingBPH1/Run2018B -20Jun2021_UL2018-v1/AOD	252,798,836	239,039,492	234,197,576	4,007,474,752
/ParkingBPH2/Run2018B -20Jun2021_UL2018-v1/AOD	252,730,676	238,496,243	234,770,668	4,016,906,231
/ParkingBPH3/Run2018B -20Jun2021_UL2018-v1/AOD	252,699,513	238,881,017	234,205,865	4,006,834,540
/ParkingBPH4/Run2018B -20Jun2021_UL2018-v1/AOD	252,618,514	238,838,252	234,676,702	4,016,197,772
/ParkingBPH5/Run2018B -20Jun2021_UL2018-v1/AOD	252,697,594	238,929,690	235,074,299	4,020,463,684
/ParkingBPH6/Run2018B -20Jun2021_UL2018-v1/AOD	98,671,902	97,734,889	96,646,365	1,643,967,729
/ParkingBPH1/Run2018C -20Jun2021_UL2018-v1/AOD	226,781,174	219,959,688	215,908,882	4,117,558,658
/ParkingBPH2/Run2018C -20Jun2021_UL2018-v1/AOD	226,763,111	220,027,080	216,552,961	4,130,731,639
/ParkingBPH3/Run2018C -20Jun2021_UL2018-v1/AOD	226,766,133	216,518,932	212,591,487	4,048,952,891
/ParkingBPH4/Run2018C -20Jun2021_UL2018-v1/AOD	226,740,205	220,298,126	216,163,822	4,125,400,837
/ParkingBPH5/Run2018C -20Jun2021_UL2018-v1/AOD	226,759,786	216,034,030	212,011,280	4,050,282,710
/ParkingBPH1/Run2018D -20Jun2021_UL2018-v1/AOD	1,632,671,054	1,632,225,048	1,597,950,002	26,200,725,308
/ParkingBPH2/Run2018D -20Jun2021_UL2018-v1/AOD	1,629,032,833	-	1,618,869,051	26,606,672,011
/ParkingBPH3/Run2018D -20Jun2021_UL2018-v1/AOD	1,629,408,802	1,628,529,909	1,596,921,608	26,240,491,899
/ParkingBPH4/Run2018D -20Jun2021_UL2018-v1/AOD	1,629,471,771	-	1,601,986,567	26,322,333,442
/ParkingBPH5/Run2018D -20Jun2021_UL2018-v1/AOD	1,624,678,097	-	1,596,307,746	26,226,480,125
Total	11,877,042,834	-	11,425,952,259	192,165,165,547
$\mathcal{L}$ [pb^{-1}]				3

Table 5.3: List of HLTs used to select events on the 13 TeV B parking samples.

Trigger path
HLT_Mu12_IP6
HLT_Mu10p5_IP3p5
HLT_Mu9_IP6
HLT_Mu9_IP5
HLT_Mu8p5_IP3p5
HLT_Mu8_IP6
HLT_Mu8_IP5
HLT_Mu8_IP3
HLT_Mu7_IP4

of ~ 10 billion (explicit numbers can be found in Table 5.2). The effective luminosity for these samples was calculated temporarily with a MC MinimumBias cross section $\sigma = 78.42$ mb [126] with an arbitrarily assigned scale factor*, 0.8, which results in 3 pb^{-1}. This luminosity will have to be reevaluated with more dedicated studies in the future. A basic quality selection was also applied for these samples with a JSON file in Table 5.4.

Table 5.4: JSON applied for 13TeV ParkingBPH UL samples.

JSON file
`Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt`

The reconstruction efficiencies in the 13 TeV analysis were evaluated with a $D^0 \to K^{\mp}\pi^{\pm}$ filtered MC sample, which was produced with a PileUp (PU) profile flat from 0 to 75. For further studies of e.g. pileup systematics, an inclusive MinimumBias MC sample was also considered in this thesis. Especially this sample was used to provide templates for the D^0 background fit (see later). The effective luminosity of these samples was calculated with cross section $\sigma = 0.544$ mb [127] and $\sigma = 78.42$ mb [126] for the D^0 and MinimumBias MC sample, respectively. The MC samples are listed in Table 5.5.

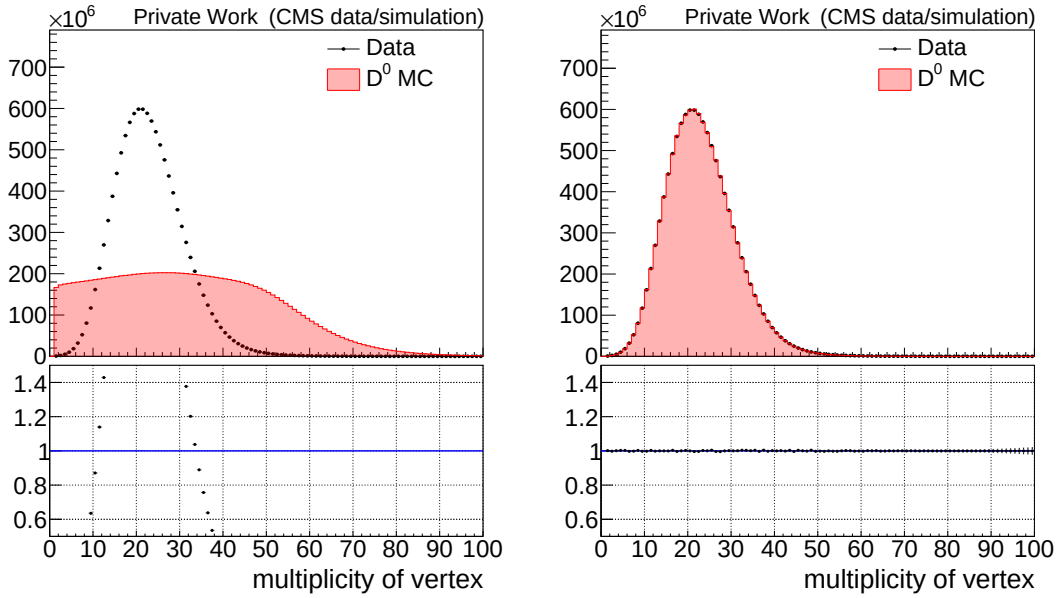
Table 5.5: List of 13 TeV UL MC samples used in this thesis.

Sample	#Events	$\mathcal{L}$ [nb^{-1}]
`/MinBias_TuneCP5_13TeV-pythia8` `/RunIISummer20UL18RECO-NoPU_106X_upgrade2018_realistic_v11_L1v1-v2` `/AODSIM`	499,793,400	6.37
`/D0ToKPi_SoftQCDinelastic_TuneCP5_13TeV_pythia8-evtgen` `/RunIISummer20UL18RECO-FlatPU0to75_106X_upgrade2018_realistic_v11_L1v1-v4` `/AODSIM`	99,180,526	182.32

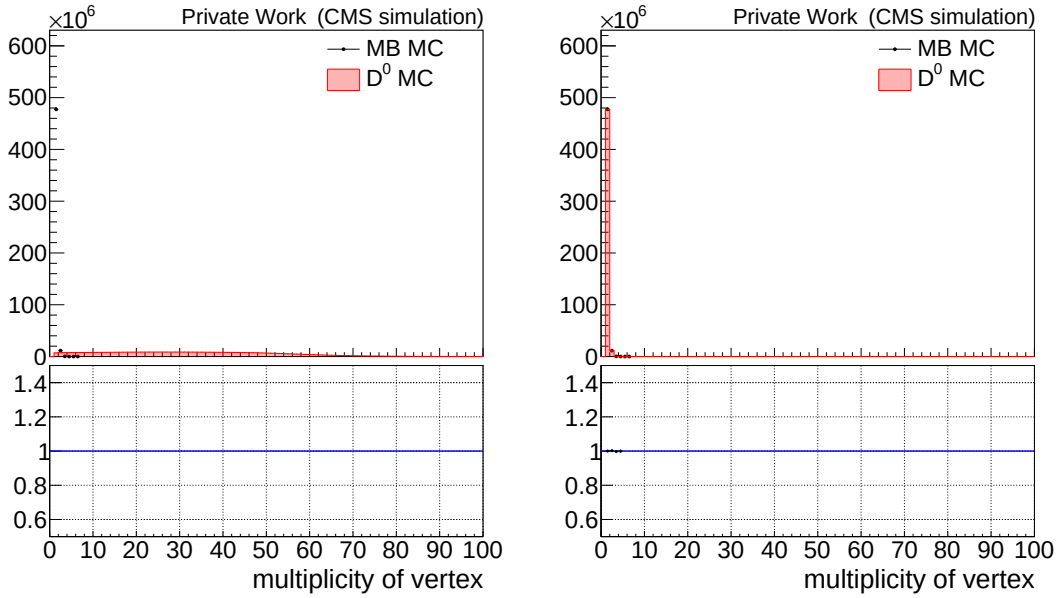
Pileup distributions on the 13 TeV samples are shown in Figure 5.4. For the evaluation of the reconstruction efficiency, all the events in the D^0 MC sample were reweighted by a comparison of the pileup distribution between the data and MC sample (see Figure 5.4a). As a reference, the D^0 MC events were also reweighted for the MinimumBias MC pileup distribution as shown in Figure 5.4b, and used as a cross check for pileup effects (see later).

The 0.9 TeV analysis was performed on the early Run 3 data in 2021 and 2022 before moving to $\sqrt{s} = 13.6$ TeV and the early Run 1 data in 2010, of which samples are listed in Table 5.6. The D^{*+} cross section measurements introduced in this chapter are based only on the Run 3 samples, which were taken as inputs for the total charm cross section measurements at $\sqrt{s} = 0.9$ TeV in the next chapter. The updated results including the 2010 data can be found in Appendix

*This scale factor was temporarily chosen such that the D^{*+} cross sections match well with the ones of the previously published LHC measurements in the end.



(a) Comparison between data and D^0 MC.



(b) Comparison between MinimumBias MC and D^0 MC.

Figure 5.4: Pileup distributions on the 13 TeV samples before (left) and after (right) reweighting.

Table 5.6: List of 900 GeV data and MC samples. $\mathcal{N}$ in data samples refers to a series of datasets from 0 to 9 and the event numbers correspond to the total of each series.

Data sample	#Events	$\mathcal{L}$ [nb^{-1}]
/MinimumBias$\mathcal{N}$/Commissioning2021-900GeVmkFit_FEVT-v1/RECO	41,230,571	1.1
/MinimumBias$\mathcal{N}$/Run2022A-PromptReco-v1/AOD	274,720,502	2.1
/MinimumBias/Commissioning10-07JunReReco_900GeV/AOD /MinimumBias/Run2010A-Apr21ReReco-v1/AOD /ZeroBias/Run2010A-Apr21ReReco-v1/AOD	26,433,766	0.22
MC sample	#Events	$\mathcal{L}$ [nb^{-1}]
/MinBias_D0Filter_TuneCP5_900GeV-pythia8-evtgen /Commissioning900DR-122X_mcRun3_2021_realistic_forpp900GeV_v2-v1 /AODSIM	1,893,632	48.29
/D0Kpi_pT0toInf_TuneZ2star_900GeV-pythia6-evtgen /LowPU2010DR42-NoPU2010_DR42_900GeV_START42_V17B-v2/AODSIM /AODSIM	4,857,763	68.21

A. The evaluation of the reconstruction efficiency was provided with an inclusive MinimumBias MC sample enriched by the $D^0 \to K^{\mp}\pi^{\pm}$ filter again (refer to Table 5.6).

5.3 Event reconstruction

To reconstruct $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ (including the charge conjugate), three tracks should be selected for the final states. The selection was provided in two steps; (i) candidates were preselected without kinematic dependence and (ii) then tighter cuts were applied to the preselected candidates in two different categories divided at $p_T = 3.5$ GeV. In the second step, the *lower* p_T region ($p_T^{D^{*+}} < 3.5$ GeV) has larger combinatorial background compared to the *higher* p_T region ($p_T^{D^{*+}} > 3.5$ GeV), thus tighter cuts were applied in the lower p_T region. Similarly $D^0 \to K^-\pi^+$ (including the charge conjugate) was reconstructed by selecting two tracks. Except for a few cuts, K^- and π^+ candidates share the same cuts for D^{*+} and D^0 reconstruction in the preselection step. The D^0 analysis cuts were not fully optimized yet, and almost the same cuts were applied in two p_T categories of the second step. The preselection cuts are summarized in Table 5.7 for both D^{*+} and D^0. The final selection(the second step) cuts are summarized for D^{*+} and D^0 in Table 5.8 and 5.9, respectively.

Basically cuts on p_T and the impact parameters with respect to the PV were applied for each track, while no (pseudo-)rapidity cuts were applied for the full fiducial coverage. In case of the 7 and 0.9 TeV analysis, dE/dx cuts were added for K candidates with $p^K < 1.5$ GeV:

$$\frac{0.6}{|\vec{p}^K|^2} + 2 < \frac{dE}{dx} < \frac{1.0}{|\vec{p}^K|^2} + 3.5. \tag{5.2}$$

Additionally, further cuts were applied to select D^0 candidates (independent of whether D^0 is from D^{*+} or not). The D^0 has a non-negligible decay length, and the decay length was calculated by

$$\mathrm{dl} = \frac{\vec{p}^{D^0} \cdot \vec{d}^{\mathrm{PV,SV}}}{|\vec{p}^{D^0}|} \tag{5.3}$$

where $d^{\mathrm{PV,SV}}$ is the distance between the primary vertex and the Secondary Vertex (SV) which is in this case the D^0 vertex. The angle between $\vec{p}^{D^0}$ and $\vec{d}^{\mathrm{PV,SV}}$ is defined as ϕ such that

$$\cos\phi = \frac{\mathrm{dl}}{|\vec{d}^{\mathrm{PV,SV}}|}. \tag{5.4}$$

Table 5.7: Preselection cuts for D^{*+} and D^0 in the 7, 0.9 and 13 TeV analysis. The distance between A and B is denoted by $d_{xy}^{A,B}$ and $d_z^{A,B}$ in the xy-plane and z-direction, respectively. The last 6 rows were added at preselection level due to a storage issue with the large statistics of the 13 TeV data, which were introduced to reduce the file size down to $\sim 20\%$. However, all these cuts were loosely assigned to have barely any effect on the final selection.

	D^{*+}			D^0
	7 TeV	0.9 TeV	13 TeV	13 TeV
p_T^{π}, p_T^K	> 0.3 GeV	> 0.5 GeV	> 0.3 GeV	
$d_{xy}^{K,\pi}$, $d_z^{K,\pi}$	< 0.5 cm			
$d_z^{D^0,\mathrm{PV}}$	< 2 cm			
$p_T^{D^0}$	> 0.9 GeV	> 1.4 GeV		
$D^0 p_T^{frac}$	> 0.15			
	$p_T^{D^0} < 0.9$ GeV	$p_T^{D^0} < 1.5$ GeV		
M_{D^0}	$\in [1.5, 2.3]$ GeV			$\in [1.68, 2.05]$ GeV
$\lvert \mathrm{dl}_{xy} \rvert$				> 0.02 cm
$d_{xy}^{\pi_s,D^0}$, $d_z^{\pi_s,D^0}$	< 2 cm			
$p_T^{\pi_s}/p_T^{D^0}$	$\in [0.03, 0.20]$			
$M_{D^{*+}} - M_{D^0}$	($\in [0.143426, 0.147426]$ GeV or $M_{D^0} \in [1.82484, 1.90484]$ GeV) and < 0.17 GeV			
$d_{xy}^{K,\mathrm{PV}}$, $d_{xy}^{\pi,\mathrm{PV}}$			< 0.2 cm	
$d_{xy}^{\pi_s,D^0}$			< 0.4 cm	
$d_z^{K,\mathrm{PV}}$, $d_z^{\pi,\mathrm{PV}}$			$< 0.15/\sin\theta$ cm	
$d_z^{\pi_s,D^0}$			$< 0.3/\sin\theta$ cm	
$\cos\phi$			> 0.7 or $\cos\phi_{xy} > 0.7$	
dl^{sig}			> -1 or $\mathrm{dl}_{xy}^{sig} > -1$	> 2.5 or $\mathrm{dl}_{xy}^{sig} > 2.5$

Table 5.8: Final selection for D^{*+}. The distance between A and B is denoted by $d_{xy}^{A,B}$ and $d_z^{A,B}$ in the xy-plane and z-direction, respectively. In addition to the cuts listed in this table, dE/dx cuts also were applied into the 7 and 0.9 TeV analysis (see text).

	$p_T^{D^{*+}} < 3.5$ GeV			$p_T^{D^{*+}} > 3.5$ GeV		
	7 TeV	13 TeV	0.9 TeV	7 TeV	13 TeV	0.9 TeV
p_T^π	> 0.5 GeV					
p_T^K		> 0.5 GeV			> 0.75 GeV	
$d_{xy}^{K,\mathrm{PV}}$, $d_{xy}^{\pi,\mathrm{PV}}$	< 0.15 cm					
$d_{xy}^{\pi_s,\mathrm{PV}}$	< 0.3 cm					
$d_z^{K,\mathrm{PV}}$, $d_z^{\pi,\mathrm{PV}}$	$< 0.1/\sin\theta$ cm					
$d_z^{\pi_s,\mathrm{PV}}$	$< 0.2/\sin\theta$ cm					
M^{D^0}	$\in [1.84, 1.89]$ GeV			$\in [1.85, 1.88]$ GeV		
$\cos\phi$	> 0.8					
dl^{sig}	(> 1.5 and $D^{*+}p_T^{frac} > 0.15$) or > 3 or (> 2 and $\cos\phi > 0.995$)			(> -1 and $D^{*+}p_T^{frac} > 0.15$) or > 2		
$D^0 p_T^{frac}$	> 0.1					
$D^{*+}p_T^{frac}$			> 0.1		> 0.03	
$M_{D^{*+}} - M_{D^0}$	$\in [0.14440, 0.14664]$ GeV					

Table 5.9: Final selection for D^0 in the 13 TeV analysis. The distance between A and B is denoted by $d_{xy}^{A,B}$ and $d_z^{A,B}$ in the xy-plane and z-direction, respectively.

	$p_T^{D^0} < 3.5$ GeV	$p_T^{D^0} > 3.5$ GeV
p_T^π, p_T^K	> 0.5 GeV	
$d_{xy}^{K,\mathrm{PV}}$, $d_{xy}^{\pi,\mathrm{PV}}$	> 0.15 cm	
$d_z^{K,\mathrm{PV}}$, $d_z^{\pi,\mathrm{PV}}$	$> 0.1/\sin\theta$ cm	
dl^{sig}	> 3.5	> 4
$\cos\phi$	> 0.99	

A significance of the decay length is also calculated by considering the uncertainty of dl ($\mathrm{dl}^{\mathrm{err}}$):

$$\mathrm{dl}^{\mathrm{sig}} = \frac{\mathrm{dl}}{\mathrm{dl}^{\mathrm{err}}}. \tag{5.5}$$

In the hypothesis of originating from a common vertex, the vertex is refitted with rejecting low $p_T^{D^0}$ and p_T^{frac}. p_T^{frac} is defined to be the fraction of $p_T^{D^0}(p_T^{D^{*+}})$ to the (scalar) sum of p_T of all tracks within 5 mm in $\mathrm{d_{xy}}$ and $\mathrm{d_z}$ from the average position of K and π.

The last rows in Table 5.7 were added due to a storage issue with the large statistics of the 13 TeV data, which were introduced to reduce the file size down to $\sim 20\%$. Overall it was determined to be loose enough for the final cuts to come. The additional cuts on the variable dl_{xy}^{sig} and $\cos\phi_{xy}$ were checked that in the end it reduces 1 out of 5209961 (example D^{*+} candidates by the ParkingBPH6 Run B sample), but does not change at all the D^{*+} signal (180986 for the example statistics).

In the case of the 7 TeV analysis, this reconstruction and selection resulted in overall very good agreements between data and MC distributions within the given statistics. For example, p_T^{frac} and PV track multiplicity distributions are shown in Figure 5.5. Further distributions can be found in [122]. In contrast, notable discrepancies are observed in the p_T^{frac} distributions on the 0.9 TeV samples [121]. To resolve the discrepancies, the variables on the MC were rescaled compared to those on the data as shown in Figure 5.6. For the others good agreements were observed, and further comparisons can be found in [121].

Discrepancies on the p_T^{frac} and track multiplicity distributions were observed also on the 13 TeV samples but not as significant as those on the 0.9 TeV samples. It turns out that the p_T^{frac} distributions are strongly correlated with the pileup distributions. The p_T^{frac} and track multiplicity distributions are shown before and after applying reweights on the D^0 MC pileup distribution (refer to Figure 5.4) compared to the data and the MinimumBias MC sample in Figure 5.7 and 5.8, respectively. In both cases, the pileup reweighting resolves the discrepancies for the p_T^{frac} distributions, and it is more dramatic in the comparison between the MinimumBias and D^0 MC sample with larger pileup difference. On the other hand, the discrepancies on track multiplicity between the data and the MC samples were not resolved with the pileup reweighting (it appears that the multiplicity distribution is almost uncorrelated with the pileup one), and could not be cured by rescaling either. No further investigation is provided for this discrepancy in this thesis. The other distributions are reasonably matched (see Appendix E).

5.4 Signal extraction

5.4.1 Background subtraction and fit for D^{*+}

As mentioned, K^- and π^+ can be distinguished by the existence of the π_s^+ in the D^{*+} signal reconstruction. As a result, the charges of K and π can be uniquely assigned relative to that of π_s. The signal including combinatorial background is reconstructed by $D^{*\pm} \to D^0\pi_s^\pm \to K^\mp\pi^\pm\pi_s^\pm$, which is called *right charge* combination. $D^{*\pm} \to K^\pm\pi^\pm\pi_s^\mp$, which is called *wrong charge* combination, gives the combinatorial background only and is signal- and reflection-free with minimal charge correlation bias. For instance, $M_{D^{*+}} - M_{D^0}$ distributions are shown in Figure 5.9, where the blue(red) histograms refer to the right(wrong) charge combination. The D^{*+} signal can be extracted from the mass distributions using the *background subtraction method* which can be applied in a robust way compared to a fit method. In the subtraction method, the number of the wrong charge events is subtracted from that of the right charge, which leaves the number of signal events only. Here before the subtraction, a scale factor is applied to the wrong charge, which is determined by the ratio between the wrong and right charge events counted within the *side bands* (which are defined by the gray areas in Figure 5.9). Then the signal number is counted only within the *signal bands* which are defined with pink

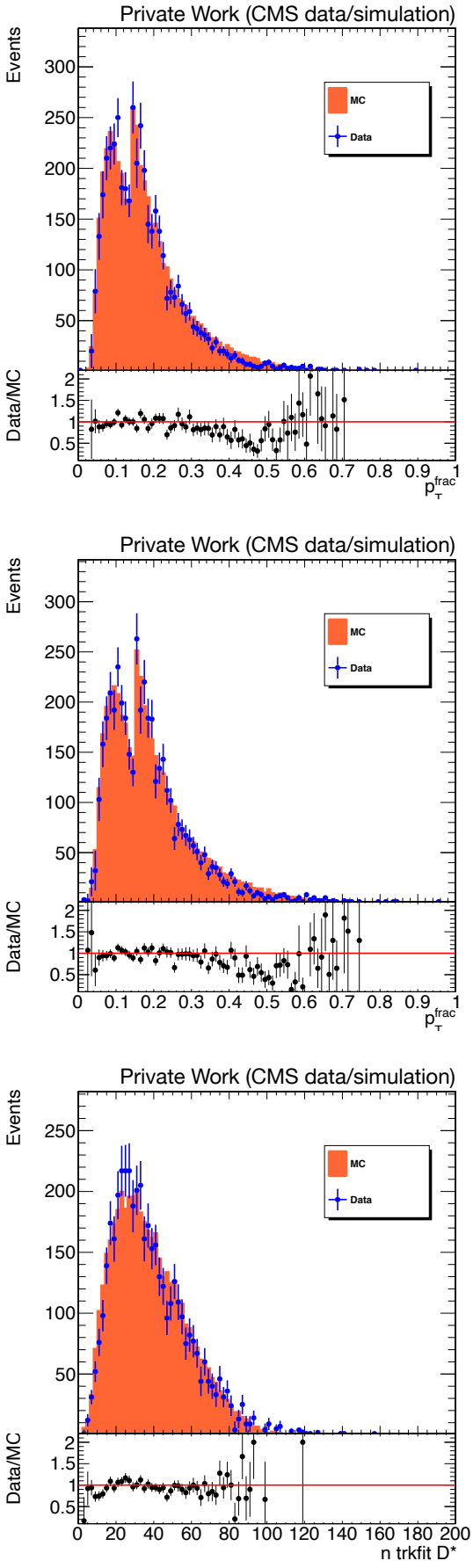


Figure 5.5: p_T^{frac} of D^0 (top) and D^{*+} (middle), and the number of tracks fitted to the D^{*+} PV (bottom) in the 7 TeV analysis. Figures are taken from [122].

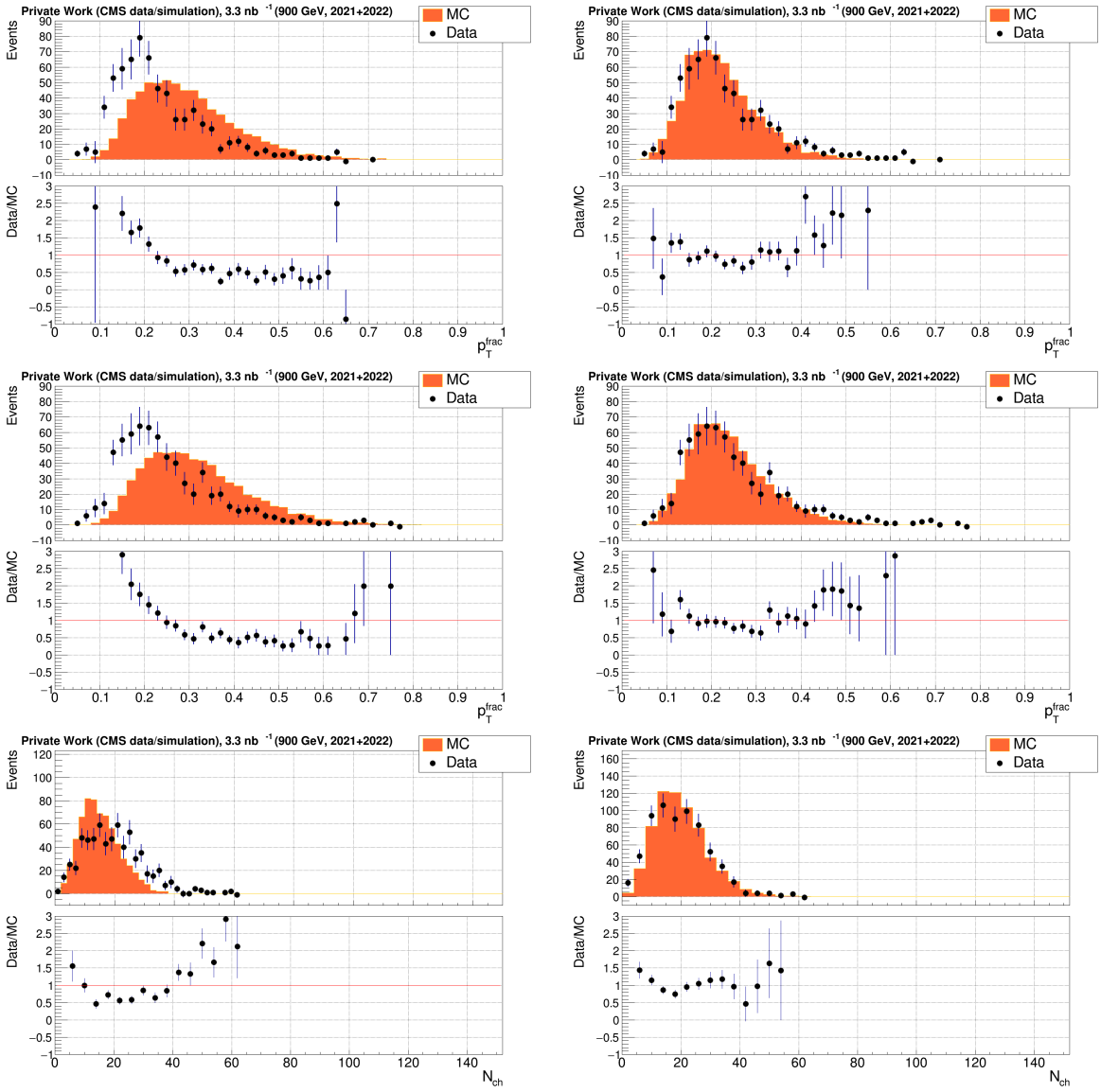
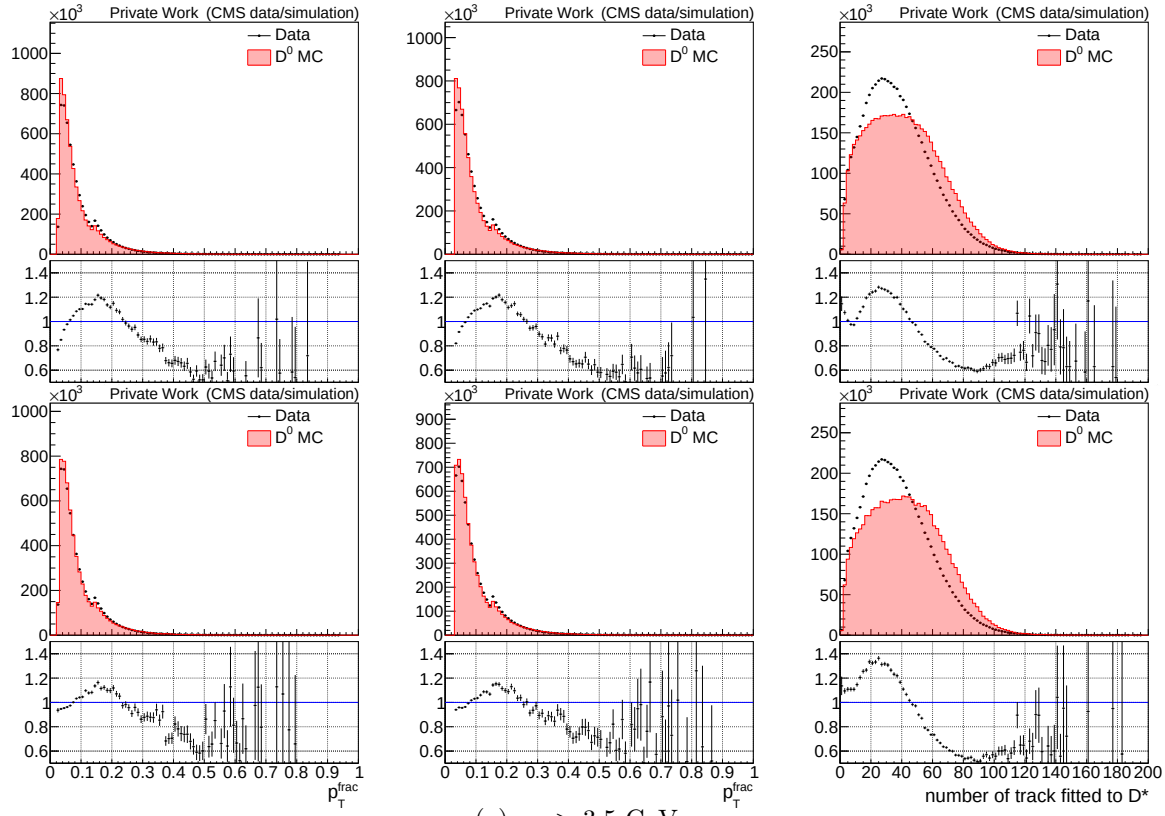
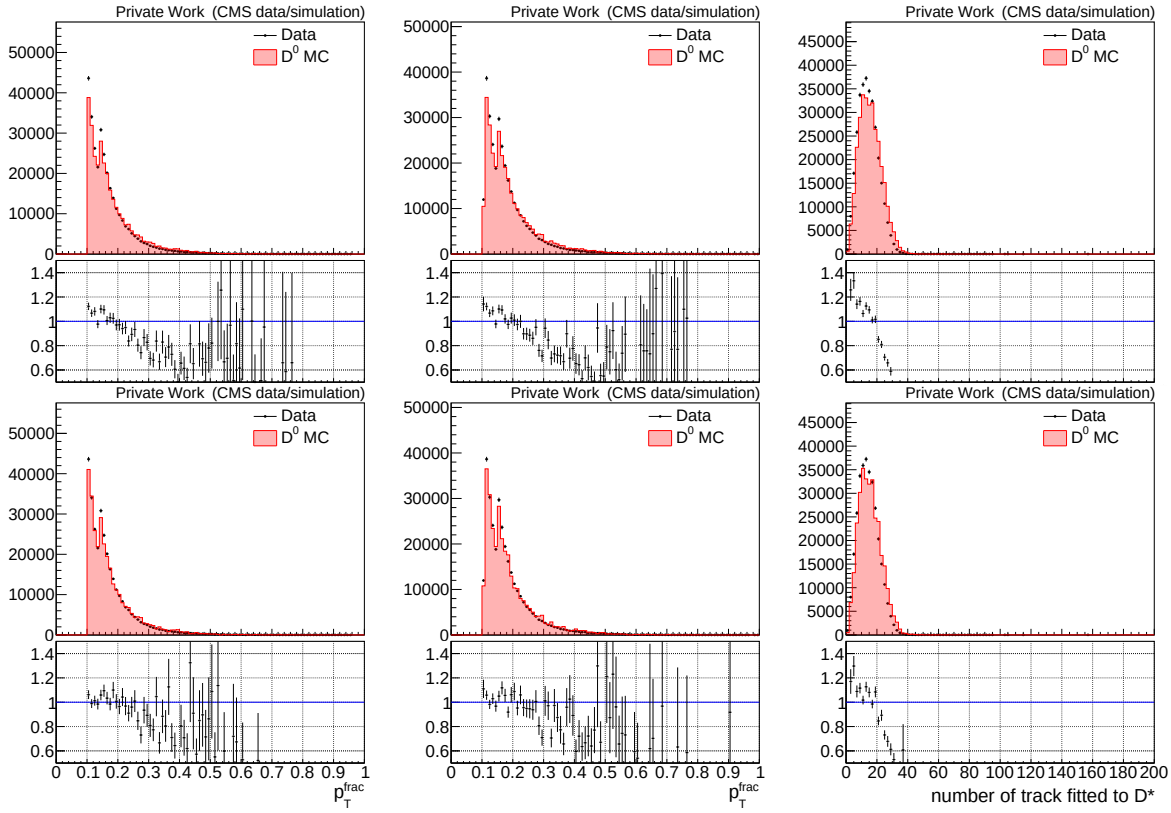


Figure 5.6: p_T^{frac} of D^0 (top) and D^{*+} (middle), and the number of tracks fitted to the D^{*+} PV (bottom) in the 0.9 TeV analysis. The left(right) figures are before(after) rescaling the variables. Figures are from [121].

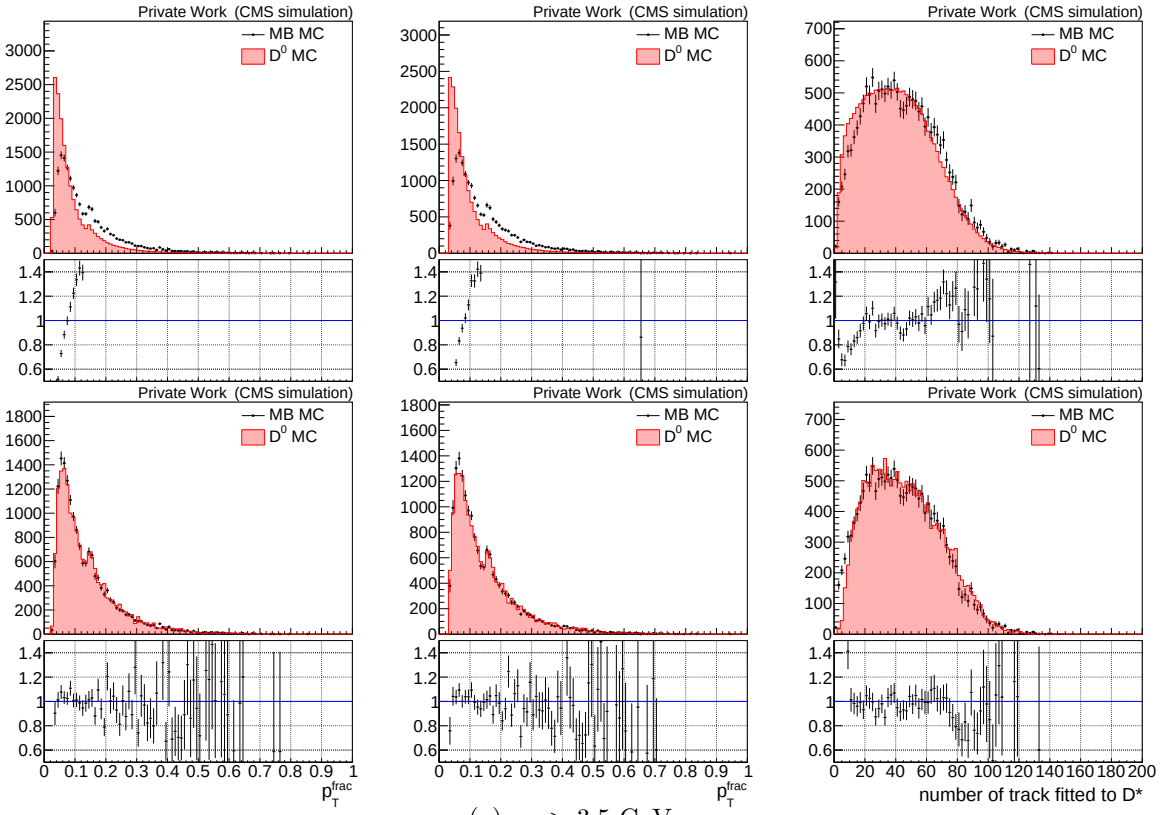


(a) $p_T > 3.5$ GeV.

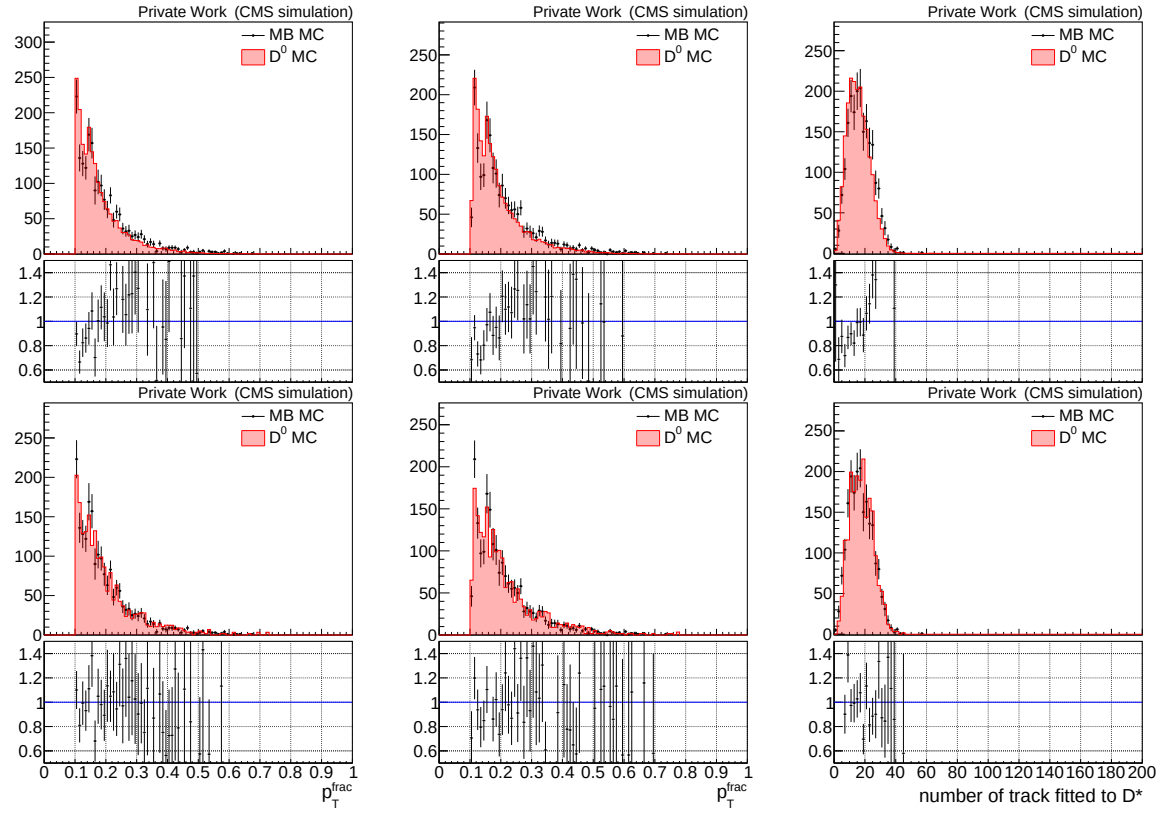


(b) $p_T < 3.5$ GeV.

Figure 5.7: p_T^{frac} of D^0 (left) and D^{*+} (middle), and the number of tracks fitted to the D^{*+} PV (right) before (upper) and after (lower) reweighting the D^0 MC pileup distribution to the data one.



(a) $p_T > 3.5$ GeV.



(b) $p_T < 3.5$ GeV.

Figure 5.8: p_T^{frac} of D^0 (left) and D^{*+} (middle), and the number of tracks fitted to the D^{*+} PV (right) before (upper) and after (lower) reweighting the D^0 MC pileup distribution to the MinimumBias MC one.

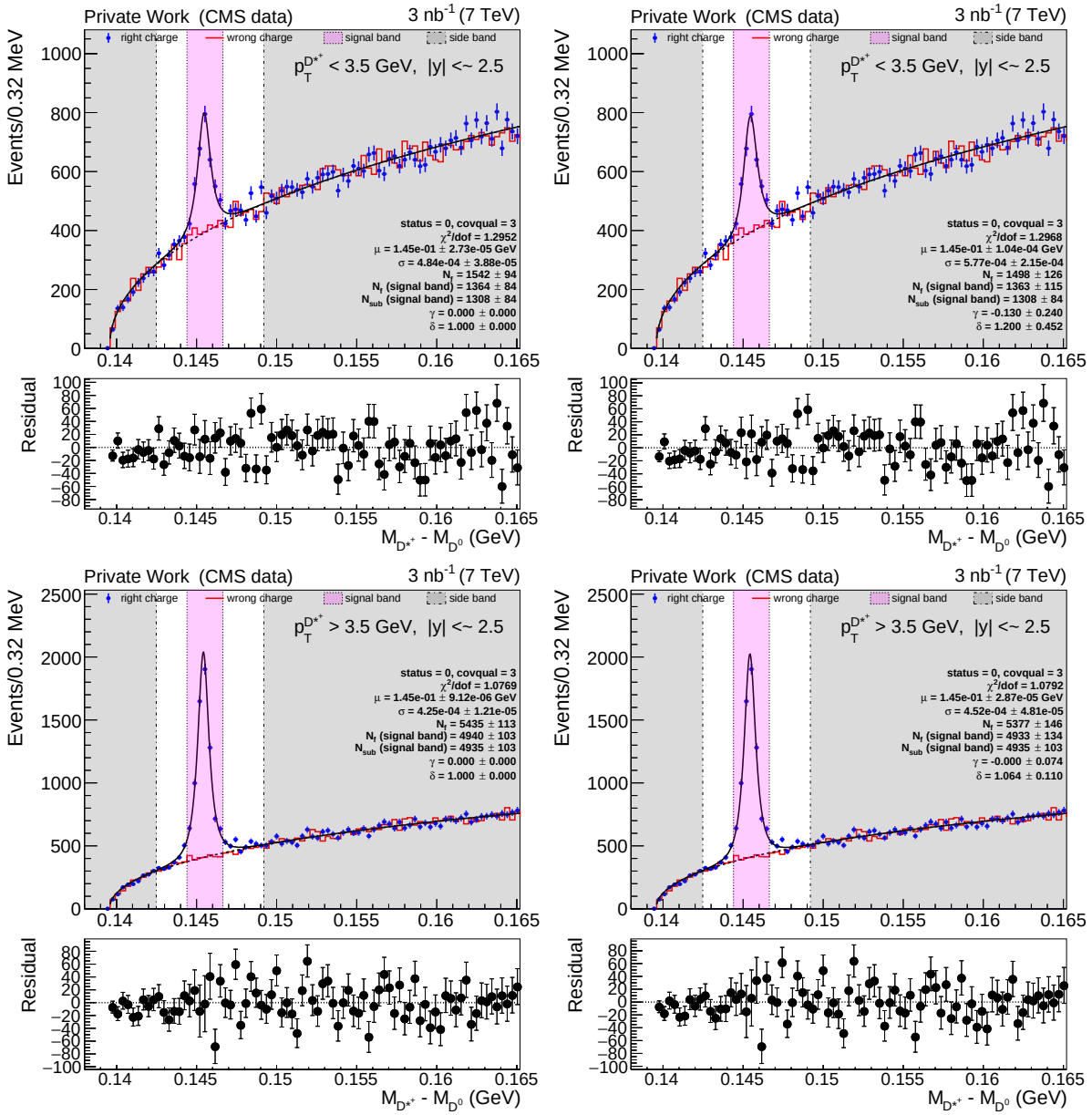


Figure 5.9: $M_{D^{*+}} - M_{D^0}$ distributions at $\sqrt{s} = 7$ TeV in $p_T^{D^{*+}} < 3.5$ GeV (upper) and $p_T^{D^{*+}} > 3.5$ GeV (lower). The black curves are fits given to the right charge using a Johnson and a threshold function for signal and background, respectively, and the red-dashed curves show the background function only. The signal function is given with fixed (left) and free (right) γ and δ parameter of the Johnson function.

colour.

The signal can be extracted also using a fit method. In this thesis, a maximum likelihood fit is introduced with a negative log likelihood

$$-\ln\big[\mathrm{Pois}(N_{S+B}, N_O) \cdot \prod_i \mathcal{P}(x_i|\mu, \sigma, N_S, a_1, a_2, N_B)\big], \tag{5.6}$$

where Pois refers to a Poisson distribution with the expected (N_{S+B}) and the observed (N_O) number of events for the right charge. $\mathcal{P}$ represents a probability density function for the right charge, which is defined by

$$\mathcal{P}(x|\mu, \sigma, N_S, a_1, a_2, N_B) = \frac{N_S}{N_{S+B}}\mathcal{J}(x|\mu, \sigma) + \frac{N_B}{N_{S+B}}\mathcal{T}(x|a_1, a_2). \tag{5.7}$$

The first(second) term in Eq.(5.7) represents the probability density function for the signal(background), which is normalized by the fraction of the expected number of signal(background) events, $N_S(N_B)$. The signal is described by a *Johnson* function

$$\mathcal{J} = \frac{1}{\sigma\sqrt{2\pi}} \frac{\delta}{\sqrt{1 + (\frac{x-\mu}{\sigma})^2}} \exp\left[-\frac{1}{2}\left(\gamma + \delta \sinh^{-1}\left(\frac{x-\mu}{\sigma}\right)\right)^2\right] \tag{5.8}$$

where μ and σ represent the center and width of the signal, respectively. Here, γ set to be non-zero results in an asymmetric distribution and δ determines the width of the signal in addition to σ. In other words, $\gamma = 0$ and $\delta = 1$ give a symmetric distribution of which the width is determined only by σ. For example, the right charge in the left panel of Figure 5.9 is fitted with fixed $\gamma = 0$ and $\delta = 1$, while in the right panel the fit is given with letting the parameters to be free. As it can be seen from Figure 5.9, no significant improvement was observed when using free γ and δ and therefore for a better convergence, γ and δ were fixed by default at 0 and 1, respectively. The background is described by a so-called *threshold* function

$$\mathcal{T} \propto (x - m_\pi)^{a_1} \cdot \exp\left[a_2(x - m_\pi)\right] \tag{5.9}$$

where $m_\pi = 0.13957$ GeV. As a result, $\mathcal{P}$ has a total of 6 free parameters, $\mu, \sigma, N_S, a_1, a_2$ and N_B.

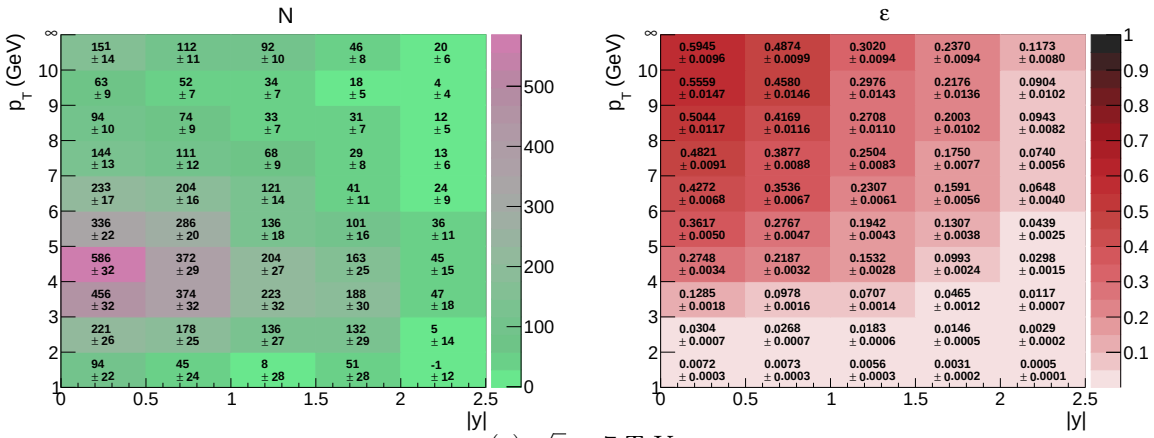
The signal yields of D^{*+} on data were determined using the background subtraction method by default, with a bin width 1 GeV in p_T and 0.5 in $|y|$ including overflow bins of $p_T > 10$ GeV. The reconstruction efficiency was calculated by

$$\frac{N^{\text{reco\&true}}}{N^{\text{true}}} \tag{5.10}$$

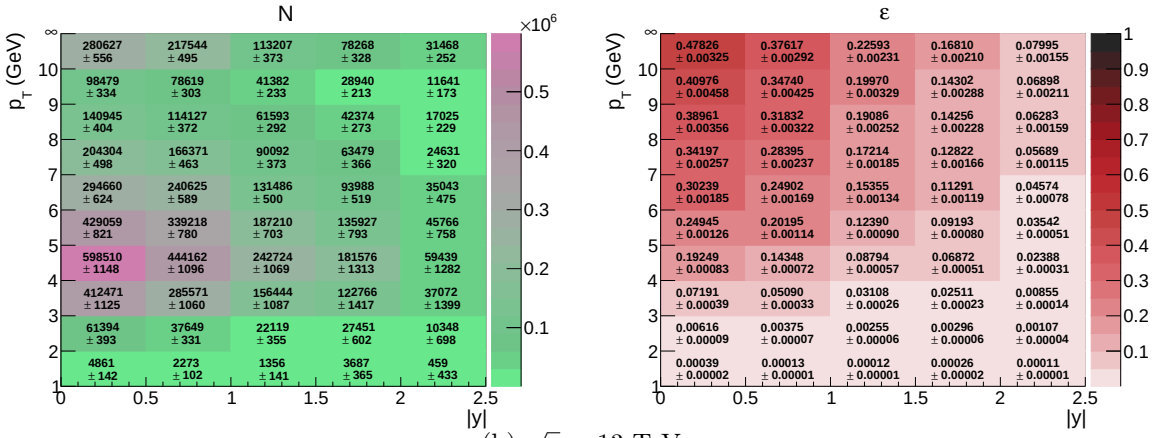
where $N^{\text{reco\&true}}$ is the number of reconstructed D^{*+} in the bin with matching generated D^{*+} without bin requirement and N^{true} is the number of generated D^{*+} in the bin. $N^{\text{reco\&true}}$ is determined also using the background subtraction method. The signal yields and the efficiencies are summarized in Figure 5.10a, 5.10b and 5.10c for the 7, 13 and 0.9 TeV analysis, respectively.

5.4.2 Fit on D^0 mass distribution

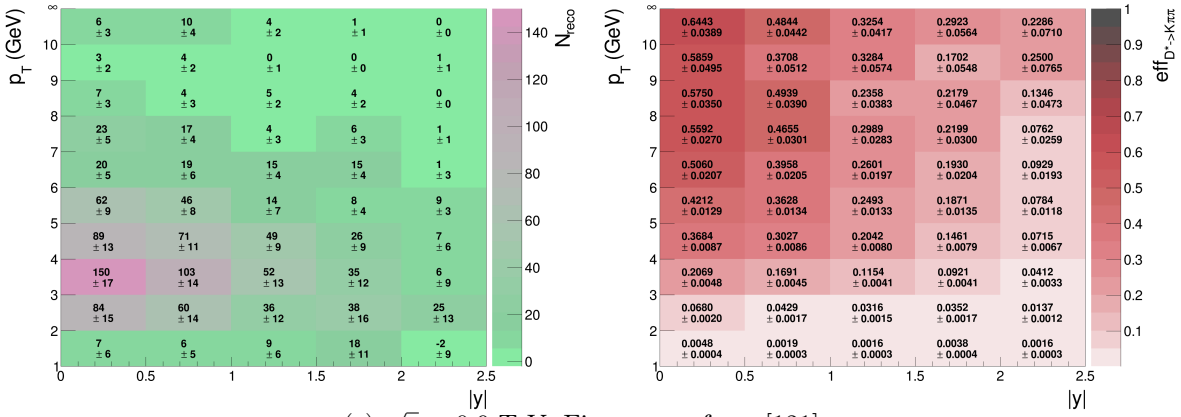
The signal of the $D^0 \to K^\mp\pi^\pm$ reconstruction on the 13 TeV samples was extracted by fitting the M_{D^0} distribution. In the CMS detector, K and π candidates are essentially not distinguished. Thus the reconstruction has a so-called K-π *swapped signal* as background in addition to combinatorial background. For instance, the black and blue histograms in Figure 5.11a were reconstructed by $D^0 \to K^\mp\pi^\pm$ (narrow signal) and $D^0 \to \pi^\mp K^\pm$ (swapped signal) with matching the generated K and π on the MC samples. Compared to the narrow signal, the swapped



(a) $\sqrt{s} = 7$ TeV.

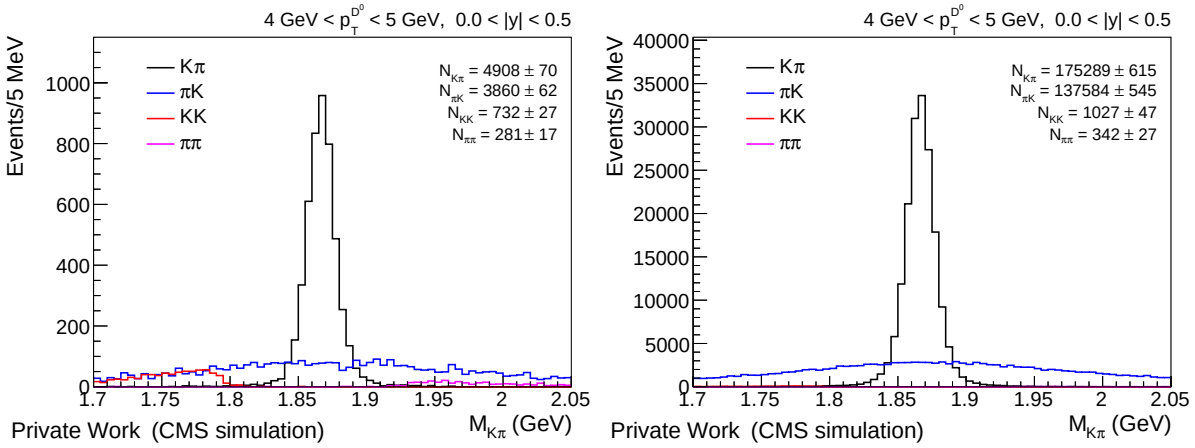


(b) $\sqrt{s} = 13$ TeV.

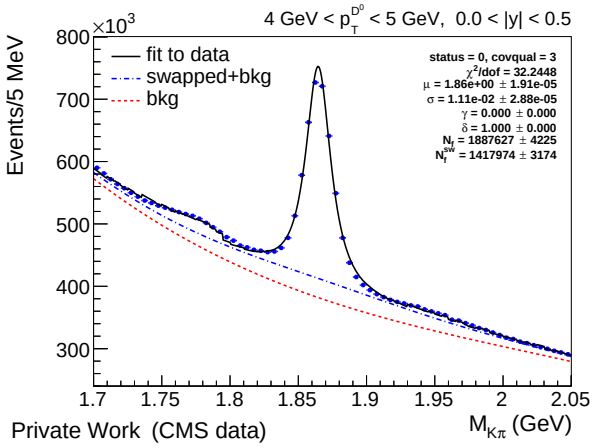


(c) $\sqrt{s} = 0.9$ TeV. Figures are from [121].

Figure 5.10: D^{*+} signal yields on data (left) and the reconstruction efficiency determined with MC (right). Essentially nothing is expected to be measured in $1 < p_T < 2$ GeV and $2 < |y| < 2.5$, so that these bins were excluded for the cross section measurements in the end.



(a) D^0 signal ($K\pi$), the K-π swapped signal (πK), the KK background (KK) and the $\pi\pi$ background ($\pi\pi$) reconstructed by matching the generated K and π on the MinimumBias MC sample (left) and the D^0 MC sample (right).



(b) $D^0 \to K^{\mp}\pi^{\pm}$ reconstruction on the data.

Figure 5.11: Example of D^0 mass distribution on the 13 TeV samples. The mass distribution in (b) is fitted using a Johnson function for the narrow and swapped signal and a Chebyshev polynomial function for combinatorial background (bkg). The width of the swapped signal is fixed from that of the swapped signal on the MinimumBias MC. The KK and $\pi\pi$ backgrounds were fitted using histogram templates from the MinimumBias MC. Templates for partially reconstructed final states are still missing.

signal has a much wider width, while the two should have the same number of events in the full mass range. Furthermore, non-negligible KK and $\pi\pi$ contributions were also observed on the data (refer to Figure 5.11b) and the MinimumBias MC sample (refer to the left panel of Figure 5.11a), which should also be fitted properly as background. The backgrounds, meanwhile, are very small on the D^0 MC since this sample is produced with a $D^0 \to K^{\mp}\pi^{\pm}$ filter (refer to the right panel of Figure 5.11a).

Similary to the D^{*+} fit, a maximum likelihood fit is introduced also for the D^0 mass fit with a negative log likelihood

$$-\ln\big[\mathrm{Pois}(N_{S+B}, N_O) \cdot \prod_i \mathcal{P}(x_i|\mu, \sigma, N_S, a_1, a_2, a_3, N_B)\big], \tag{5.11}$$

where N_{S+B} and N_O are the expected and observed number of events, respectively, for the D^0 mass distribution. $\mathcal{P}$ is composed of the Johnson function for the signals, a Chebyshev polynomial of the first kind for the combinatorial background and histogram templates from MC for the KK and $\pi\pi$ background:

$$\begin{aligned}\mathcal{P}(x|\mu, \sigma, N_S, a_1, a_2, a_3, N_B) =& \frac{N_S}{N_{S+B}}\mathcal{J}(x|\mu, \sigma) + \frac{N_S}{N_{S+B}}\mathcal{J}_{sw}(x|\mu) + \\ & \frac{N_B}{N_{S+B}}\mathcal{C}(x|a_1, a_2, a_3) + \frac{N_{kB}}{N_{S+B}}\mathcal{T}_k(x) + \frac{N_{pB}}{N_{S+B}}\mathcal{T}_p(x).\end{aligned} \tag{5.12}$$

Here $\mathcal{J}$ and $\mathcal{J}_{sw}$ are given by Eq.(5.8) and N_S is the number of the signal events which is the same in the narrow and swapped signal. Furthermore, the swapped signal function ($\mathcal{J}_{sw}$) is defined to have the same center (μ) as the one of the narrow signal. The width of the swapped signal is fixed by that on the MC sample. $\mathcal{C}$ is the third-order Chebyshev polynomial composed of

$$T_0(x) = 1, \quad , T_1(x) = x, \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{5.13}$$

with coefficients applied as

$$a_i \cdot T_i(x), \quad i = 1, \cdots, 3, \tag{5.14}$$

and N_B is the number of combinatorial background events. $\mathcal{T}_k$ and $\mathcal{T}_p$ are functions defined by template histograms extracted directly from MC for the KK and $\pi\pi$ backgrounds, respectively. The KK and $\pi\pi$ event numbers (N_{kB} and N_{pB}, respectively) are fixed relative to N_S based on the MC ratio, i.e.,

$$N_{kB} = N_S \times \frac{N_{kB}^{mc}}{N_S^{mc}}, \quad N_{pB} = N_S \times \frac{N_{pB}^{mc}}{N_S^{mc}}. \tag{5.15}$$

As a result, $\mathcal{P}$ has a total of 7 free parameters, $\mu, \sigma, N_S, a_1, a_2, a_3$ and N_B.

The fit results on the data for all bins can be found in Appendix D.2. The reconstruction efficiencies of D^0 were calculated also based on Eq.(5.10) with the D^0 MC. $N^{\mathrm{reco\&true}}$ is determined by fitting the narrow signal using the first term only in Eq.(5.12), of which results can be found in Appendix D.2. The signal yields and the efficiencies are summarized in Figure 5.12.

5.5 Cross section

The $D^{*+}(D^0)$ cross section is determined by the following formula

$$\sigma = \frac{N}{\mathcal{L} \cdot \varepsilon \cdot BR} \tag{5.16}$$

where N and ε are the signal yields on data and the reconstruction efficiencies, respectively, which can be found in Figure 5.10 and 5.12. $\mathcal{L}$ is the effective luminosity of the data sample, which can be found in Table 5.1, 5.2 and 5.6. BR is the branching ratio of the corresponding decay chain(s). In the case of D^{*+}, BR was given by 0.0267 ± 0.0003. BR was given by 0.0395 ± 0.0003 for D^0. The cross sections are summarized in Figure 5.13 and 5.14 for D^{*+} and D^0, respectively.

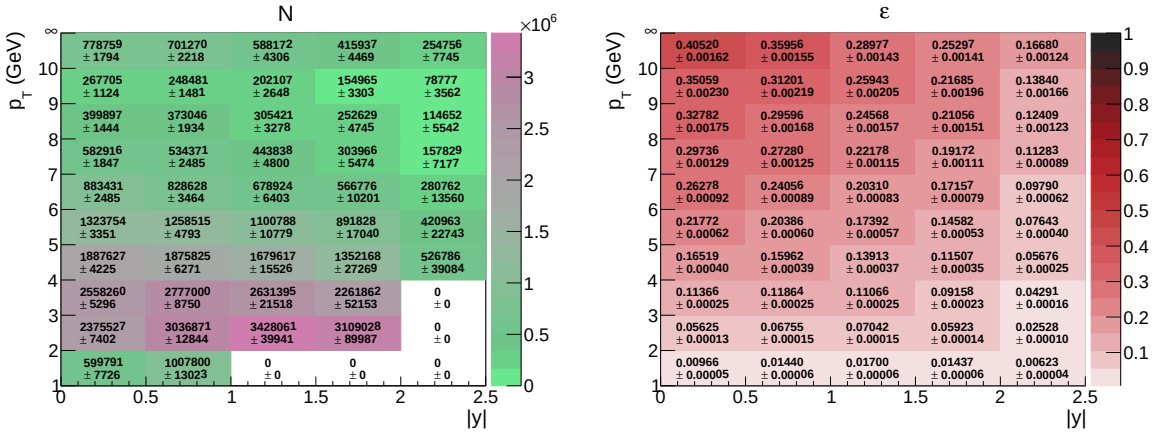


Figure 5.12: D^0 signal yields on the data (left) and the reconstruction efficiency on the D^0 MC (right) in the 13 TeV analysis. In the left figure, the bins where the fit does not give a convergent result are quoted by 0.

5.6 Non-prompt contamination

The $D^{*+}(D^0)$ reconstruction includes both prompt and non-prompt $D^{*+}(D^0)$. The prompt $D^{*+}(D^0)$ is produced directly from a charm quark, while the non-prompt one is a decay product of a b hadron. For the purpose of measuring charm production, the non-prompt D^{*+} needs to be separated from the prompt one. Using the so-called D^0 *DCA* [128] distribution, the non-prompt contribution can be separated based on data, as the prompt cross sections were measured at $\sqrt{s} = 5$ TeV in [120, 128]. However, this data-based approach cannot give reasonable results for e.g. the 7 TeV data, of which statistics is not enough. In the previous 7 TeV study [119], the preliminary prompt cross sections were given by a charm fraction[†], 0.90 ± 0.05, which was determined with an assumption based on the 5 TeV data-driven results, and applied without kinematic dependence. After the previous study, I developed a theory-inspired approach, which was also used as an evaluation of systematics on the non-prompt contamination in [23]. The charm fractions in the 7 and 13 TeV analysis were determined based on this approach, although the 13 TeV sample has enough statistics[‡].

In the theory-inspired approach, charm fractions were determined based on Pythia bin-by-bin. First, beauty fractions were extracted by Pythia and then normalized to data as shown in Figure 5.15. The data points were derived by taking a ratio of the non-prompt measurement to the mixed $D^{*+}(D^0)$ measurement. The $D^{*+}(D^0)$ measurement was taken from the previous sections, while the non-prompt measurement was derived by translating the non-prompt D^0 measurement at $\sqrt{s} = 5$ TeV [128] using the FONLL ratio:

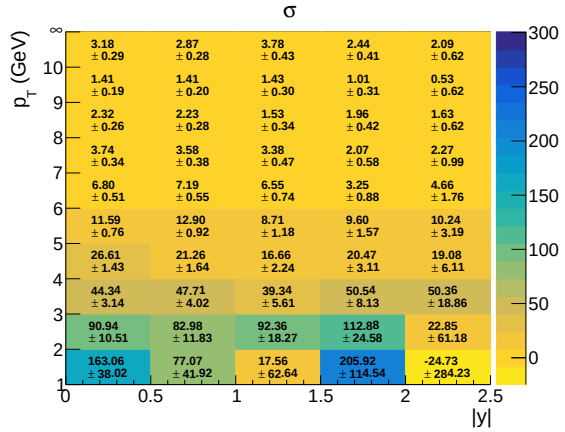
$$\text{CMS b} = \text{CMS b(5 TeV } D^0) \times \frac{\text{FONLL b}}{\text{FONLL b(5 TeV } D^0)} \tag{5.17}$$

where b stands for the non-prompt production cross section. The FONLL b was derived by a convolution of the b hadron FONLL prediction with a kinematic Pythia decay table shown in Figure 5.16 which shows a b hadron decaying into $D^{*+}(D^0)$ including migrations in both p_T and $|y|$.

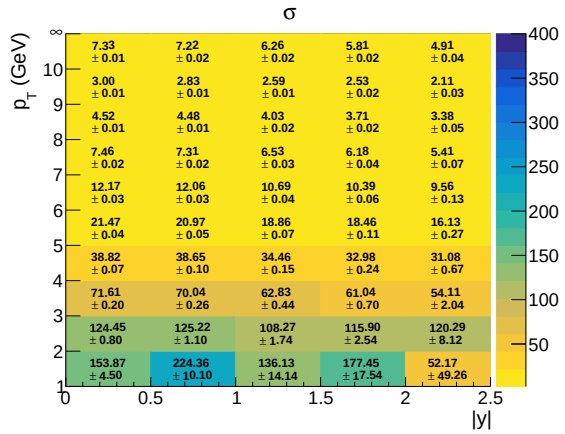
After the normalization, the Pythia ratio of non-prompt to prompt describes the data well within the assigned uncertainty of 30% for D^{*+} (see Figure 5.17a and 5.17c) and 60% for D^0 (see Figure 5.17d). Here, a much larger uncertainty was assigned for D^0 as a conservative

[†]In this thesis, the charm(beauty) fraction refers to a ratio of prompt(non-prompt) to prompt plus non-prompt production.

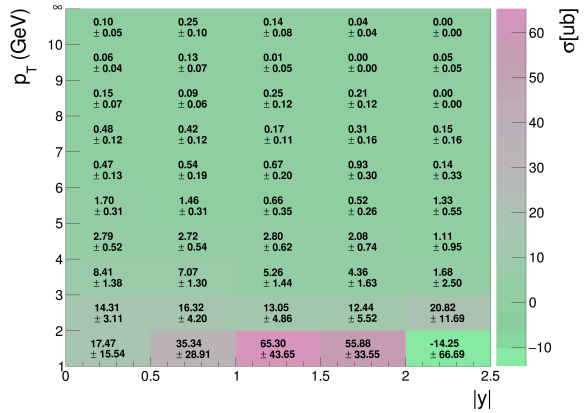
[‡]The 13 TeV non-prompt contamination determined in this thesis could be a reference to the one which would be determined with the data-based approach in the future.



(a) The 7 TeV results with statistical uncertainties only.



(b) The 13 TeV results with statistical uncertainties only.



(c) The 0.9 TeV results (figure taken from [121]) with part of systematic uncertainties (see later) in addition to statistical uncertainties.

Figure 5.13: D^{*+} cross sections. Unit is [μb]. The bins of $1 < p_T < 2$ GeV and $2 < |y| < 2.5$ show results which are not real measurements.

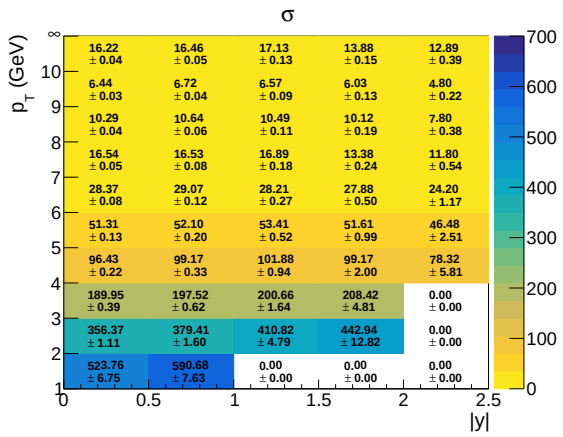


Figure 5.14: D^0 cross sections at $\sqrt{s} = 13$ TeV with statistical uncertainties only. Unit is [μb]. The bins where the fit does not give a convergent result are quoted by 0.

uncertainty by covering the FONLL prediction also well. Meanwhile ALICE has both non-prompt and prompt D^0 measurements at $\sqrt{s} = 5$ TeV [9], and the ratio was thus cross-checked against the ALICE measurements. The 7 TeV Pythia describes also the ALICE measurements well within the 30% assigned uncertainty (see Figure 5.17b). The assigned uncertainty on the ratio was then propagated into a systematic uncertainty of the charm cross section, resulting in roughly 1-3 % for D^{*+} and 3-6 % for D^0 depending on the phase space. The explicit numbers can be found in Figure 5.18 bin-by-bin.

The determined charm and beauty fractions are summarized in Figure 5.19. By normalizing the Pythia beauty fraction, the reconstruction efficiency needs to be corrected accordingly. The corrected efficiencies are currently provided for the D^{*+} results only, which are summarized in Figure 5.20. Comparing the corrected efficiency (Figure 5.20) with the previous one shown in Figure 5.10 and 5.12, it turns out that the correction is almost negligible ($\mathcal{O}(1\%)$).

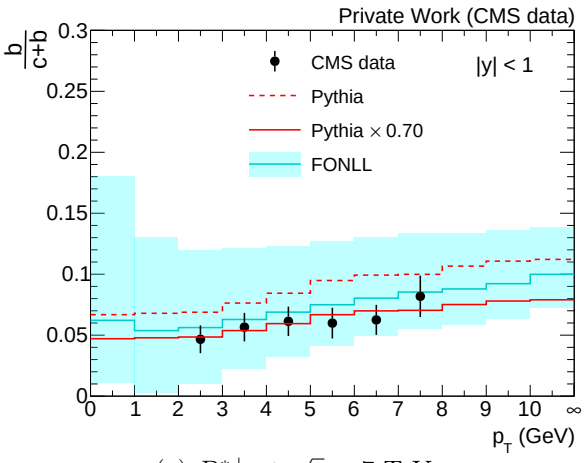
Since it is not clear whether the Pythia decay tables (produced at $\sqrt{s} = 5$ TeV) can be extrapolated also to $\sqrt{s} = 0.9$ TeV, this approach was not introduced into the 0.9 TeV analysis. Charm fractions in the 0.9 TeV analysis were thus determined by taking the Pythia fractions without normalization, assigning very conservative uncertainties directly by the beauty fractions. The preliminary fractions are shown in Figure 5.21.

5.7 Systematic studies

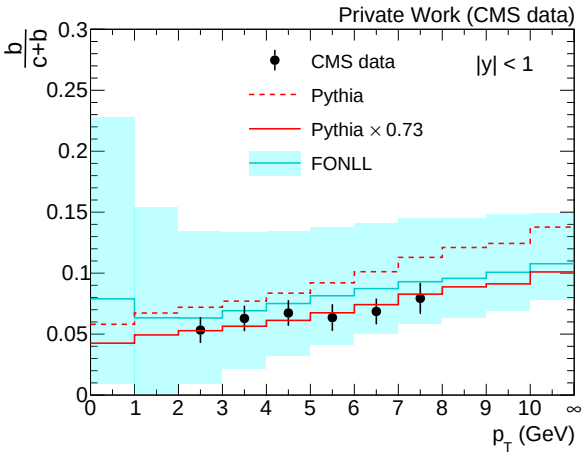
This section presents mainly the studies done to complete the systematics first in the 7 TeV analysis. The systematic studies of the 13 and 0.9 TeV analysis are not complete yet, although, the partial status of it will be briefly summarized at the end of this section.

5.7.1 Signal extraction systematics at 7 TeV

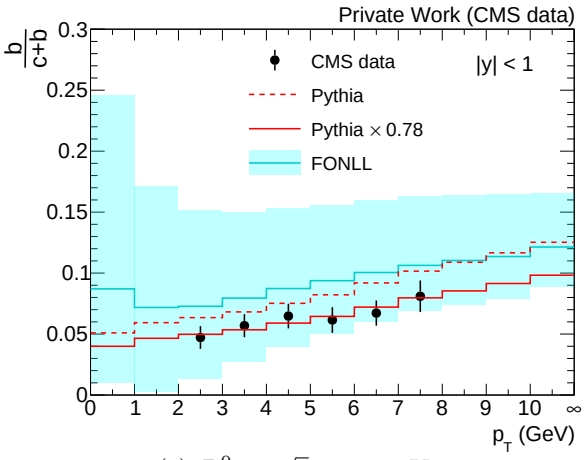
Many further systematic studies on signal extraction were provided with the 7 TeV samples. As a validation of using the background subtraction method and to determine systematics of the signal band definition, the D^{*+} signal was extracted bin-by-bin also using the fit method. However, as it can be seen from Figure 5.22, a few bins (highlighted by the red box) have very poor statistics such that the so-called *unconstrained* fit, which is defined by Eq.(5.6), can not give a reasonable result. To derive a more reasonable fit result, a so-called *constrained* fit was introduced using the D^0 MC sample, which has about 7 times larger statistics compared to the data (refer to Table 5.1). The constrained fit is given by a negative log likelihood, which is



(a) D^{*+} at $\sqrt{s} = 7$ TeV.



(b) D^{*+} at $\sqrt{s} = 13$ TeV.



(c) D^0 at $\sqrt{s} = 13$ TeV.

Figure 5.15: Beauty fraction comparison between data, Pythia and the FONLL prediction. The Pythia shape (the red dashed line) is normalized to the CMS data (the black points), which results in the red bold line.

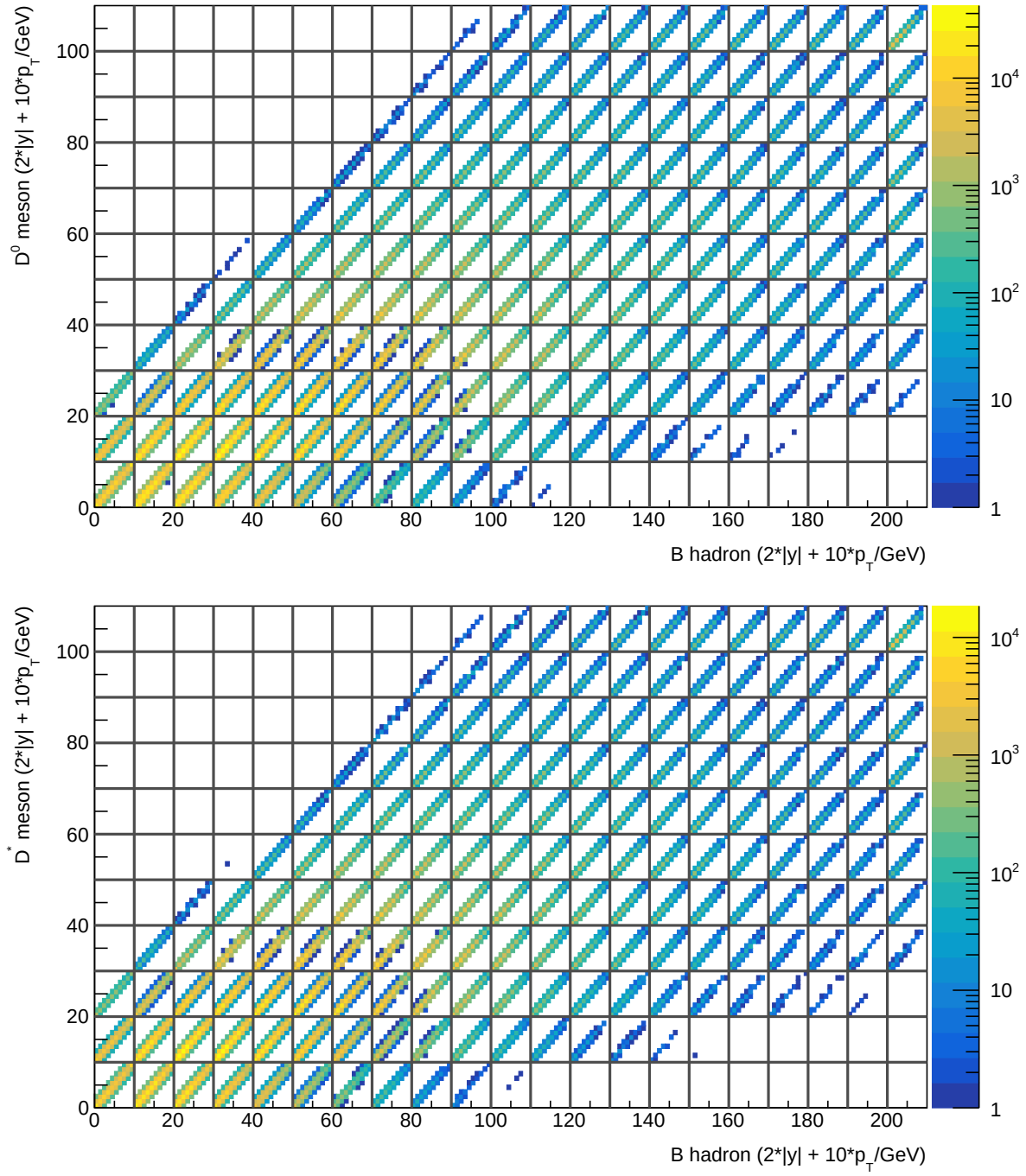
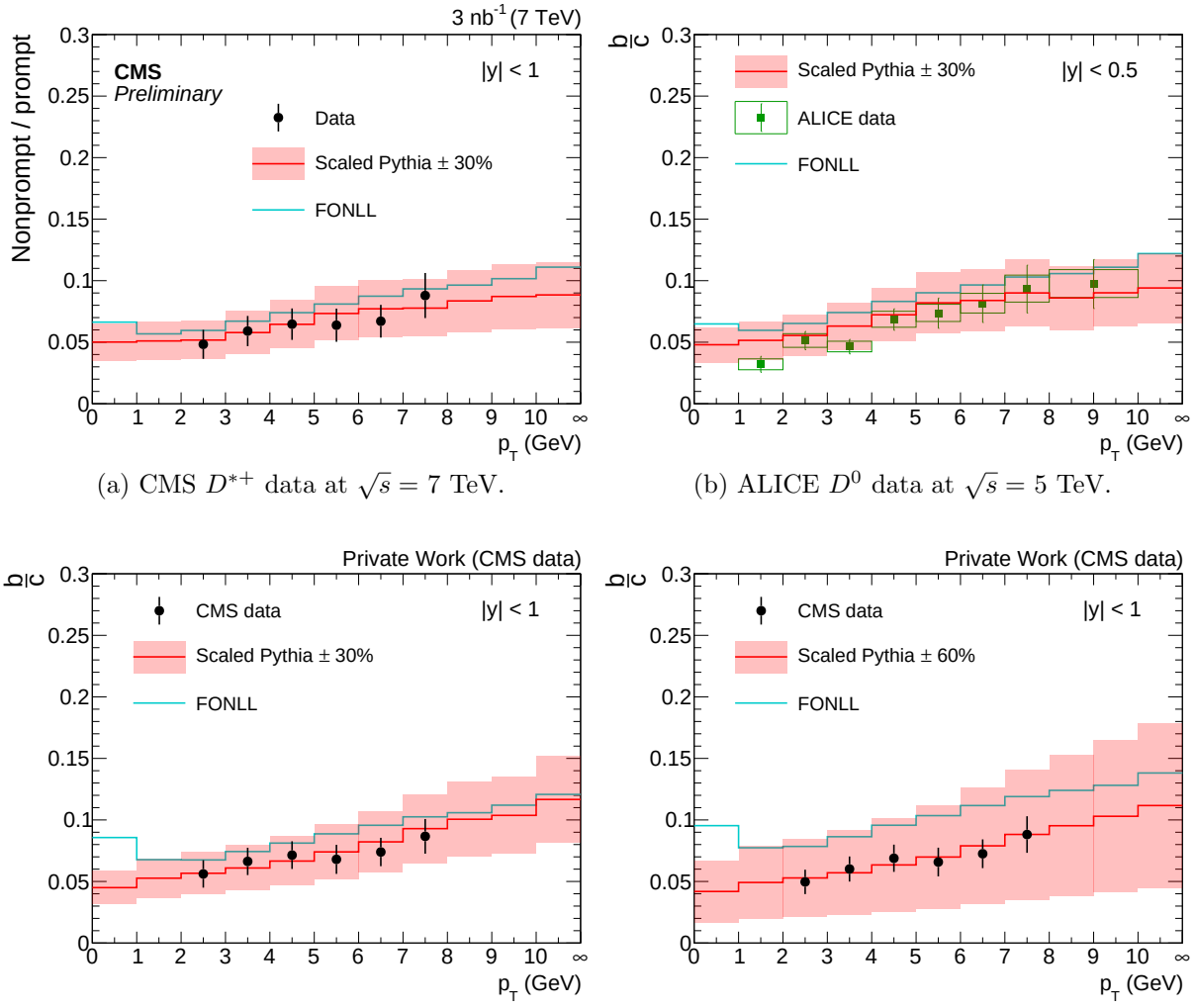


Figure 5.16: p_T and $|y|$ bin association tables of b hadrons decaying into D^0 (upper) and D^{*+} (lower) from Pythia.



(a) CMS D^{*+} data at $\sqrt{s} = 7$ TeV.

(b) ALICE D^0 data at $\sqrt{s} = 5$ TeV.

(c) CMS D^{*+} data at $\sqrt{s} = 13$ TeV.

(d) CMS D^0 data at $\sqrt{s} = 13$ TeV.

Figure 5.17: Ratio of non-prompt to prompt D meson.

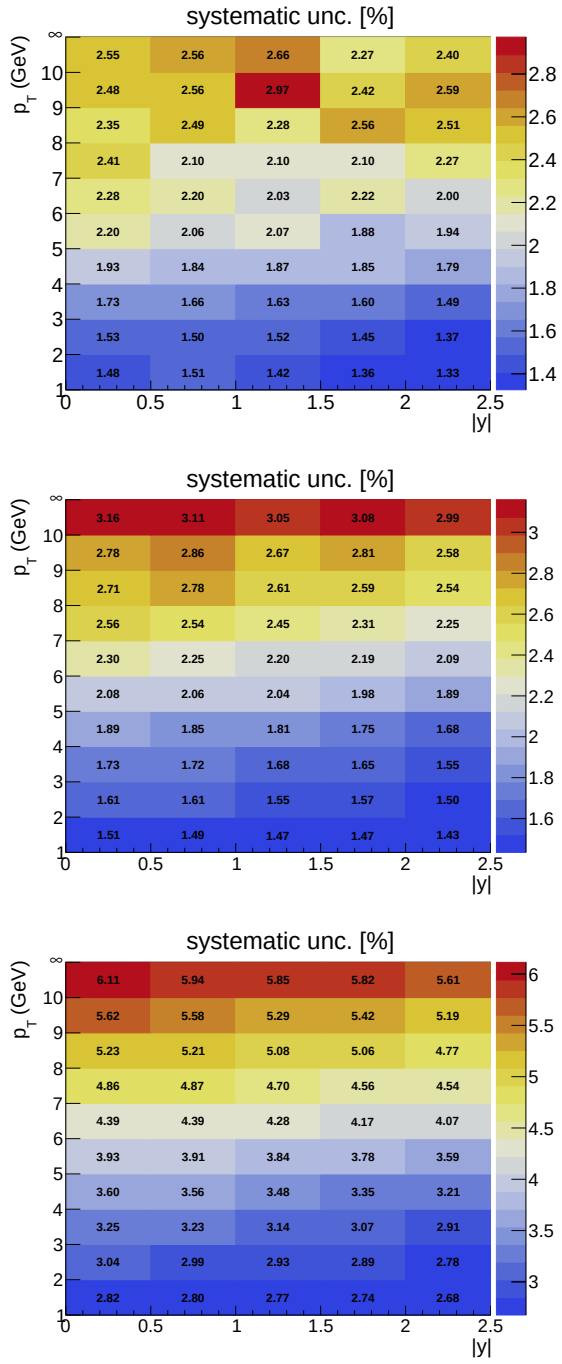


Figure 5.18: Systematic uncertainty of beauty fraction subtraction for D^{*+} at $\sqrt{s} = 7$ TeV (top) and 13 TeV (middle), and D^0 at $\sqrt{s} = 13$ TeV (bottom). The 7 TeV results are final, while the 13 TeV ones need to be reevaluated.

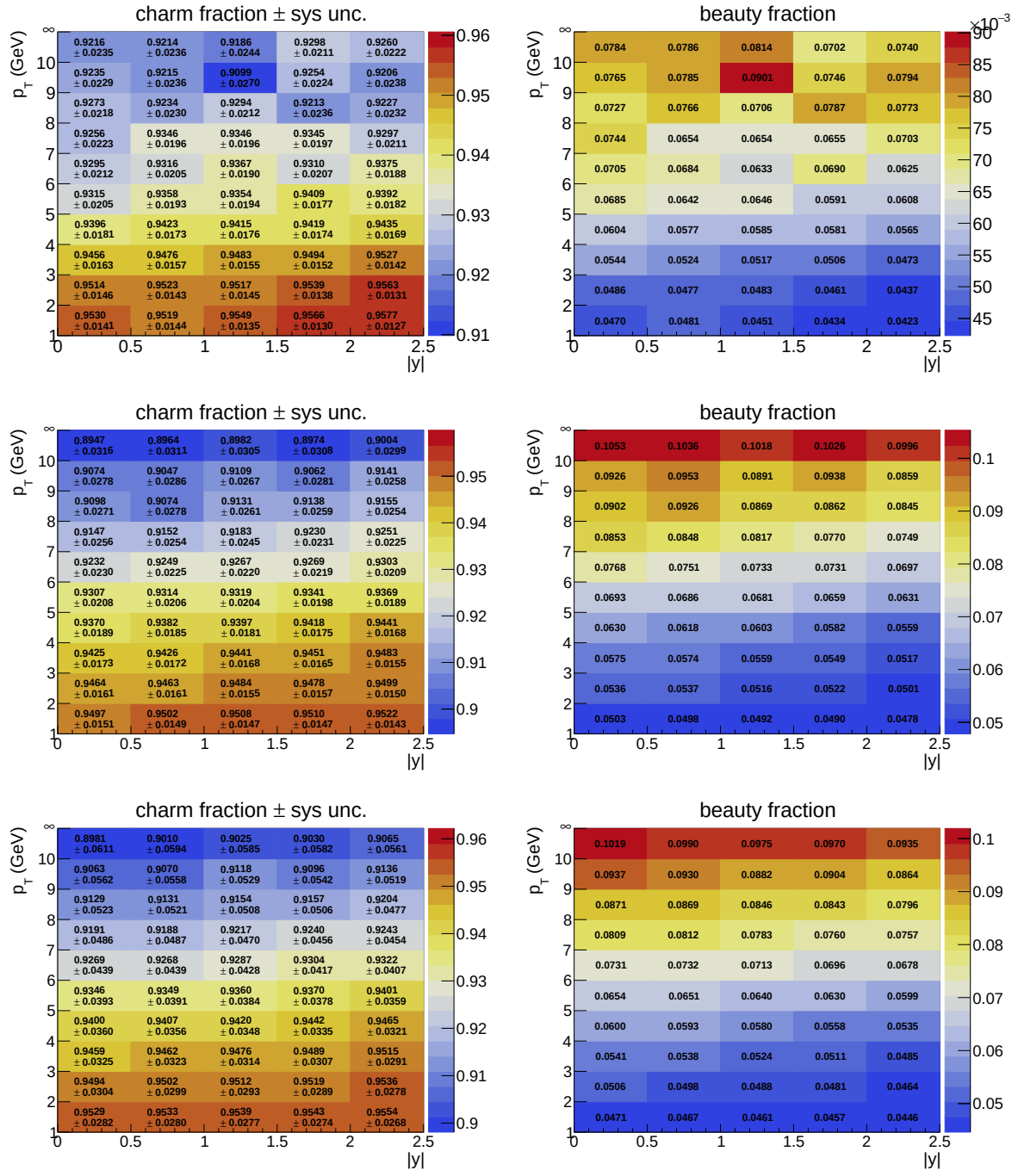


Figure 5.19: Charm (left) and beauty (right) fractions after normalizing Pythia to the 7 TeV D^{*+} data (top), 13 TeV D^{*+} data (middle) and 13 TeV D^0 data (bottom). The charm fractions are given with systematic uncertainties determined by Figure 5.18.

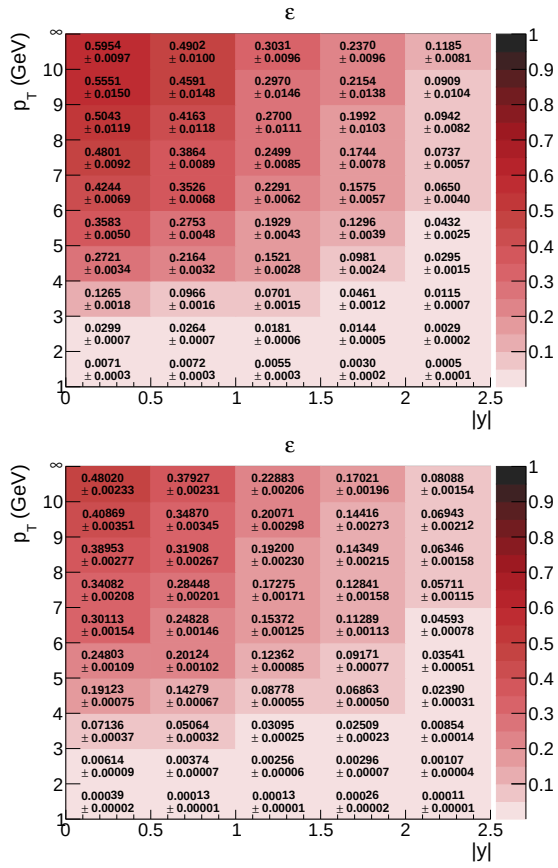


Figure 5.20: The corrected reconstruction efficiencies for D^{*+} at $\sqrt{s} = 7$ TeV (upper) and 13 TeV (lower).

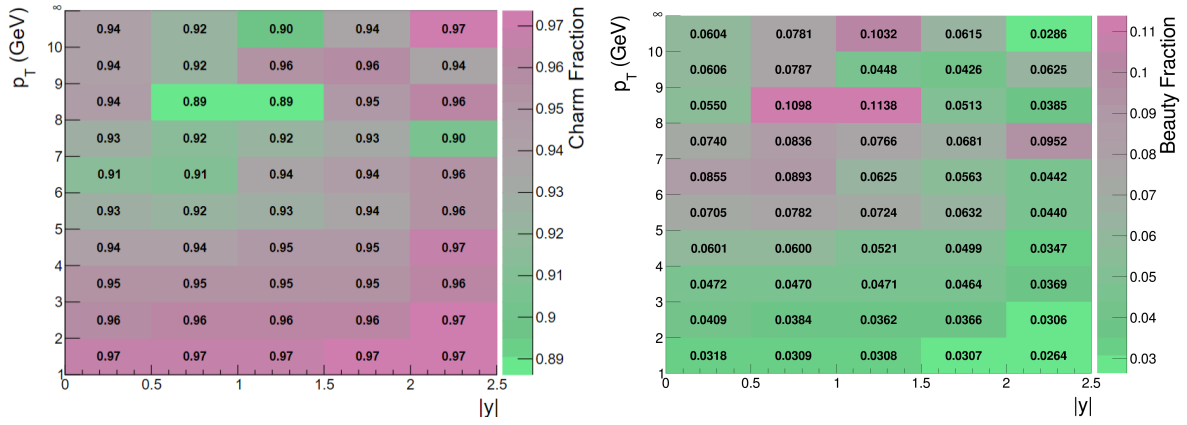


Figure 5.21: Charm (left) and beauty (right) fractions for D^{*+} at $\sqrt{s} = 0.9$ TeV. The beauty fractions were propagated as systematic uncertainties on the prompt cross sections. Figures are taken from [121].

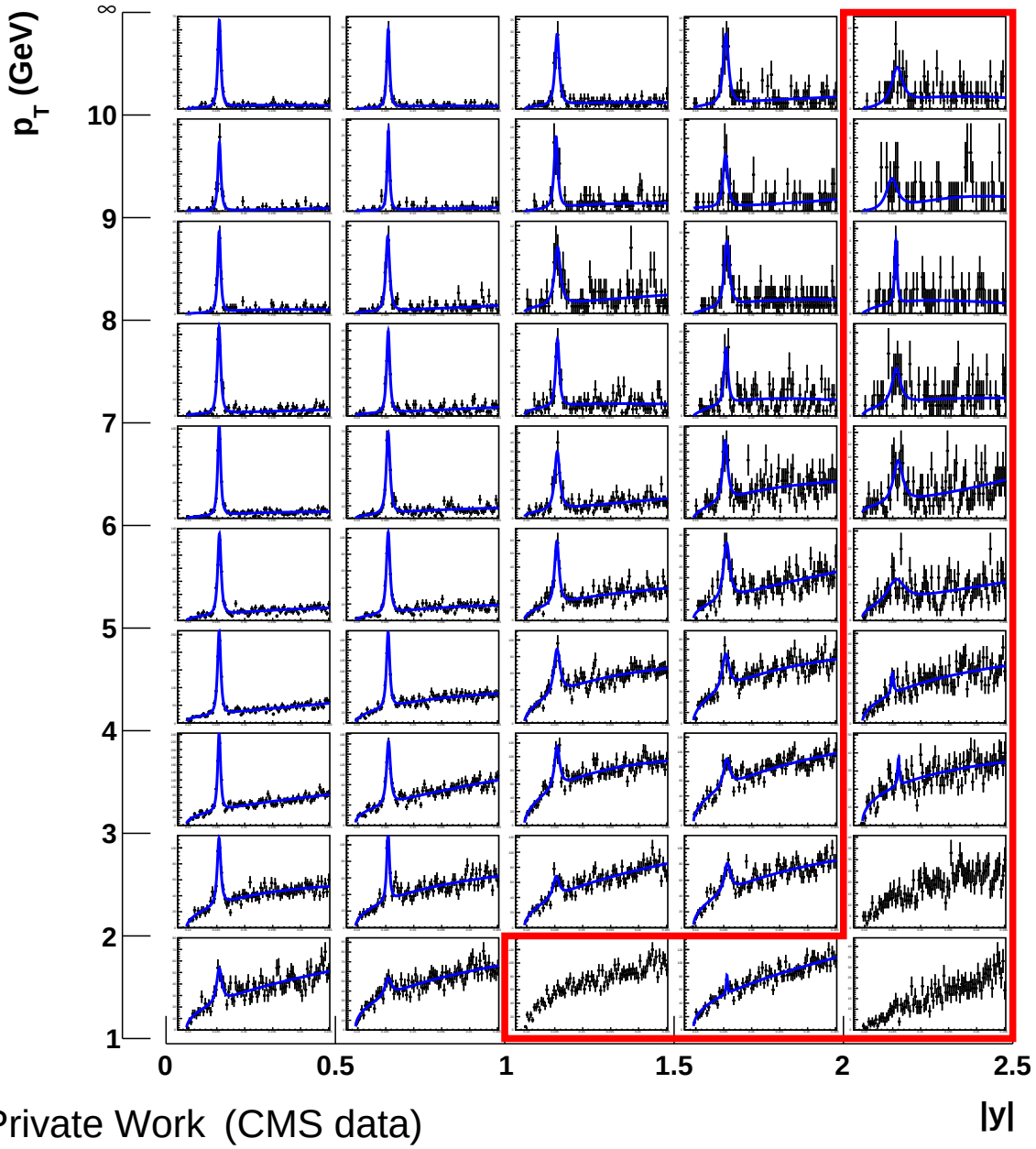


Figure 5.22: D^{*+} signal on the 7 TeV data. The blue curves show fits of the right charge histograms, which were provided by Eq.(5.6). The red boundary indicates the region where the unconstrained fit method does not work well.

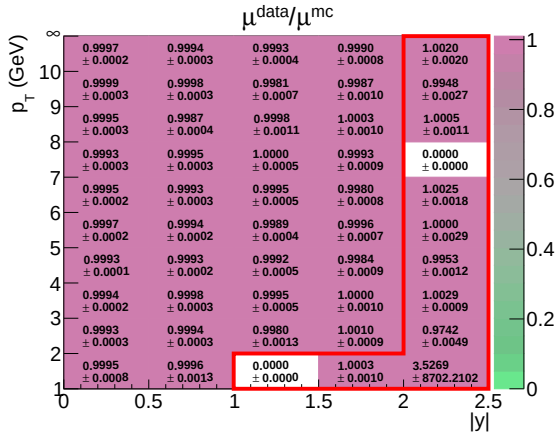


Figure 5.23: Ratio of the center of the signal between the data and MC sample. These were derived by fits with Eq.(5.6). The red box indicates the region where the fit might fail.

defined as a simultaneous fit of the three distributions, the right and wrong charge on the data and the right charge on the MC:

$$
\begin{aligned}
&- \ln(\mathrm{Pois}(N^r_{S+B}, N^r_O) \cdot \mathcal{G}_\mu(\mu|\mu^{mc}, \sigma_\mu) \cdot \mathcal{G}_\sigma(\sigma|\sigma^{mc}, \sigma_\sigma) \cdot \prod_i \mathcal{P}_r(x_i|\mu, \sigma, N^r_S, a_1, a_2, N^r_B)) \\
&- \ln(\mathrm{Pois}(N^w_B, N^w_O) \cdot \prod_j \mathcal{P}_w(x_j|a_1, a_2, N^w_B)) \\
&- \ln(\mathrm{Pois}(N^{mc}_{S+B}, N^{mc}_O) \cdot \prod_k \mathcal{P}_{mc}(x_k|\mu^{mc}, \sigma^{mc}, N^{mc}_S, a^{mc}_1, a^{mc}_2, N^{mc}_B)).
\end{aligned} \tag{5.18}
$$

The first and second line of Eq.(5.18) are the likelihoods for the right and wrong charge, respectively, on the data ($r(w)$ stands for the right(wrong) charge), while the third line is given for the MC fit. The wrong charge likelihood doubles the statistics and provides a constraint for the background function not biased by the signal. $\mathcal{P}_r$ and $\mathcal{P}_{mc}$ are described by Eq.(5.7), while $\mathcal{P}_w$ is given by the second term on the right-hand side of Eq.(5.7). The right charge likelihood contains two normal Gaussians ($\mathcal{G}_\mu$ and $\mathcal{G}_\sigma$), which constrain the center and width of the signal on data relative to those on MC. Here the normal distribution is expressed by

$$
\mathcal{G}(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{5.19}
$$

The parameters σ_μ and σ_σ in the constraining terms ($\mathcal{G}_\mu$ and $\mathcal{G}_\sigma$) were determined based on results of independent fits on the data and MC, which were extracted using Eq.(5.6). The individual results were taken to derive ratios of the center and the width of the signal between the data and the MC; μ/μ^{mc} and σ/σ^{mc} which are summarized in Figure 5.23 and 5.24, respectively.

As shown in Figure 5.23, the center of the signal shows consistent results in the entire phase space, excluding the poor statistics bins again highlighted by the red box. Therefore, an average over the full phase space (except the red-box phase space) was taken to determine μ^{mc} and σ_μ, which turned out to be 0.9994 ± 0.0001. The central value of this number was rounded and taken to correct the center of $\mathcal{G}_\mu$ as

$$
\mu^{mc} \to 0.999\mu^{mc} \tag{5.20}
$$

with $\sigma_\mu = 0.001\mu^{mc}$.

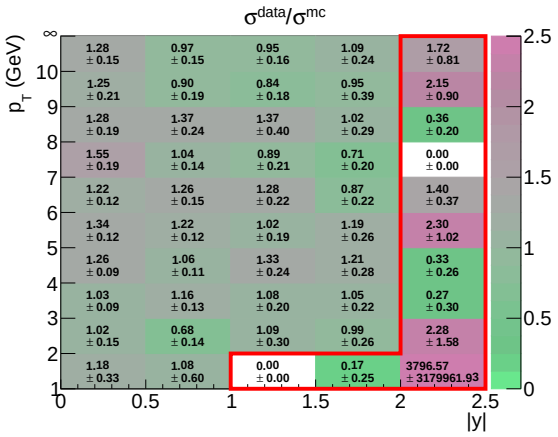


Figure 5.24: Ratio of the width of the signal between the data and MC sample. These were derived by fits with Eq.(5.6). The red box indicates the region where the fit might fail.

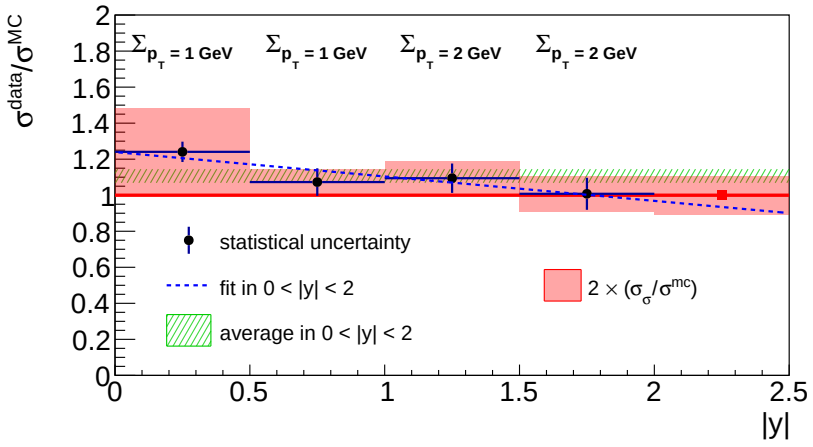


Figure 5.25: Estimation of a value for the bin $2 < |y| < 2.5$. The blue dashed line is a linear fit and the green band is the average of the black circle points in the range $0 < |y| < 2$. The ratio for the bin $2 < |y| < 2.5$ was determined to be 1, which was allowed in the constrained fit to vary within the red band at 68 % confidence level.

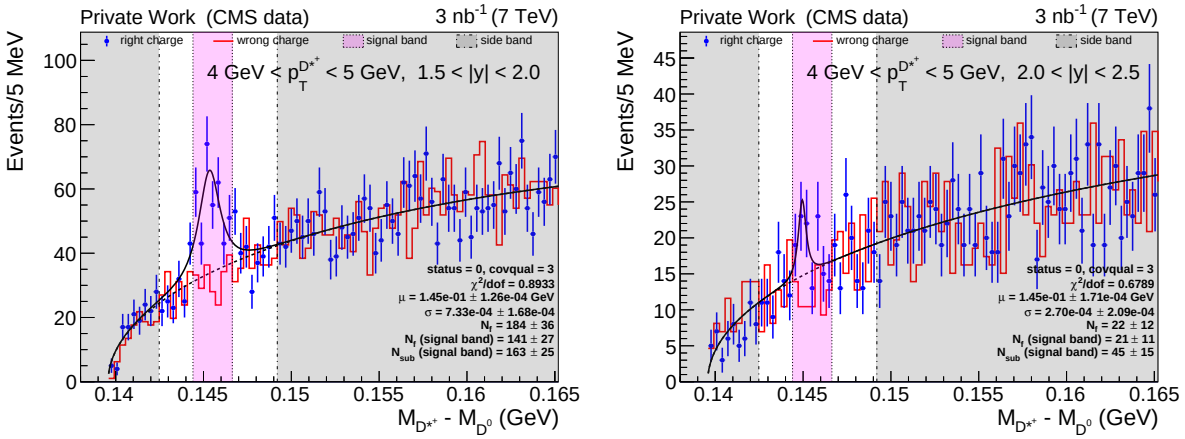
In the case of the width, the ratios are $|y|$ dependent, thus the values were averaged for each $|y|$ bin, while again the poor statistics bins were excluded from consideration. The results are

$$\begin{aligned} 0.0 < |y| < 0.5 &: 1.24 \pm 0.06 \\ 0.5 < |y| < 1.0 &: 1.07 \pm 0.08 \\ 1.0 < |y| < 1.5 &: 1.09 \pm 0.08 \\ 1.5 < |y| < 2.0 &: 1.01 \pm 0.09 \end{aligned} \tag{5.21}$$

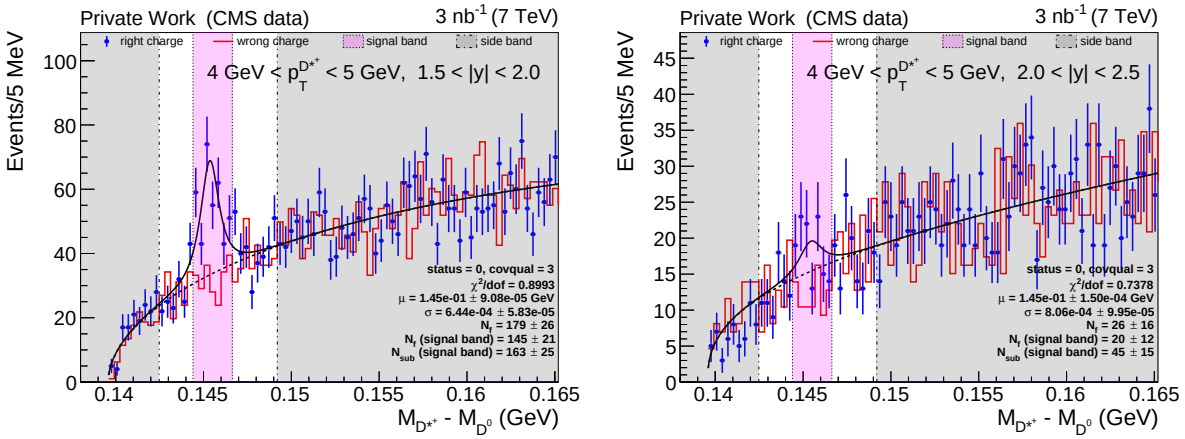
where the uncertainties are statistical uncertainties. Instead, for the bin $2 < |y| < 2.5$, an extrapolation was introduced as shown in Figure 5.25. A linear fit and an average are introduced to the four black-circle points (the four numbers of Eq.(5.21)), which are shown by the blue dashed line and the green band, respectively. A reference point indicating $\sigma = \sigma^{mc}$ is placed between the blue line and the green band in the bin $2 < |y| < 2.5$. Therefore, 1.00 is taken to be the ratio for the bin $2 < |y| < 2.5$. With these numbers, σ^{mc} and σ_σ were determined to be

	$0 < \|y\| < 0.5$	$0.5 < \|y\| < 1$	$1 < \|y\| < 1.5$	$1.5 < \|y\| < 2$	$2 < \|y\| < 2.5$
σ^{mc}	1.24 σ^{mc}	1.07 σ^{mc}	1.09 σ^{mc}	1.01 σ^{mc}	1.00σ^{mc}
σ_σ	0.24 σ^{mc}	0.07 σ^{mc}	0.09 σ^{mc}	0.10 σ^{mc}	0.11 σ^{mc}

.



(a) Unconstrained fit results. The black curves are fits to the right charge, which are provided by Eq.(5.6). The fit result of the right figure is unreasonable compared to the left (e.g., the width in the right figure is much more narrow by a factor ~ 3 compared to the one in the left figure).



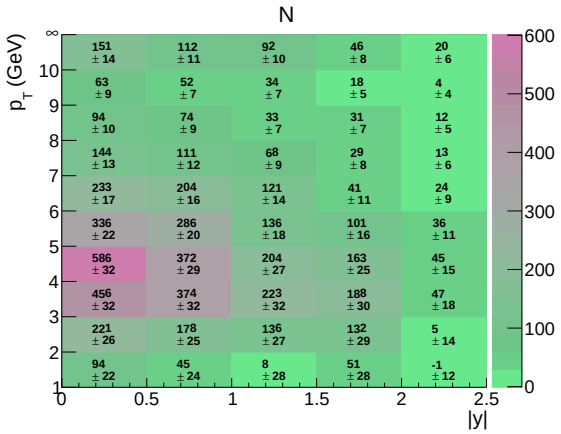
(b) Constrained fit results. The black curves are fits to the right charge, which are provided by Eq.(5.18). The fit result of the right figure appears to be within a reasonable range of what is expected.

Figure 5.26: $M_{D^{*+}} - M_{D^0}$ distributions and fit results for the bin $1.5 < |y| < 2$ (left) and $2 < |y| < 2.5$ (right) in the range $4 < p_T < 5$ GeV.

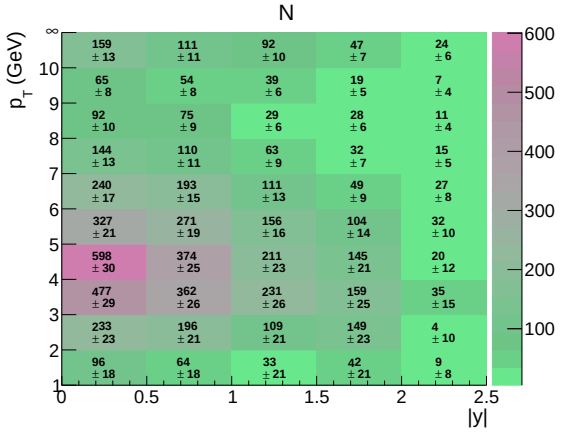
By default, the difference between the data and MC width is used to determine σ_σ, which in turn results in the red bands for the three points in the range $0 < |y| < 1.5$ in Figure 5.25. The red bands cover well both the fitted and the averaged values. The other two points in the range $1.5 < |y| < 2.5$ are already very close to $\sigma = \sigma^{mc}$ and thus the uncertainties were determined by the maximum distance to the fitted and the averaged value, which are shown again by the corresponding red bands.

With these determined parameters, the constrained fit of Eq.(5.18) was applied to the samples. An example of the fit results on one of the poor statistics bins is shown in Figure 5.26b. All the individual results bin-by-bin can be found in Appendix D.1.

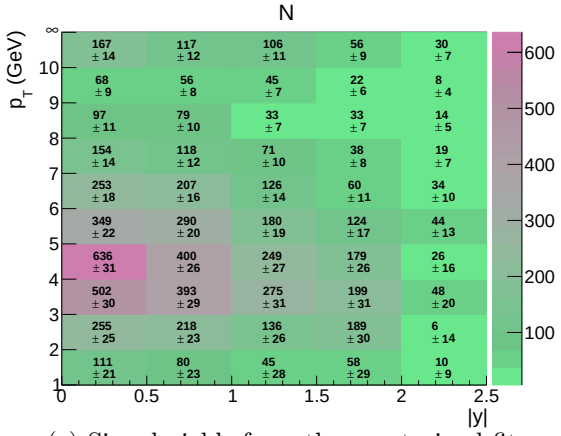
The signal yields were extracted from the constrained fit results and compared to the ones obtained from the background subtraction method, which are shown in Figure 5.27. To compare with the subtraction results (Figure 5.27a), the fit results were provided for the numbers of the signal truncated to the signal band (Figure 5.27b) in addition to the full signal results (Figure 5.27c). Figure 5.27a and 5.27b show consistent results, and it was decided that the background subtraction was used by default and the systematics with the signal band was further studied based on the fit results.



(a) Signal yields from the background subtraction method.



(b) Signal yields from the constrained fit. The signal region were truncated to the signal band of the background subtraction.



(c) Signal yields from the constrained fit.

Figure 5.27: The number of D^{*+} signal events on the 7 TeV data.

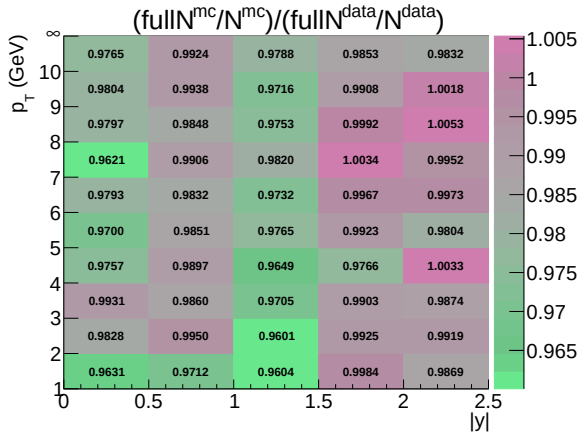


Figure 5.28: Double-ratios between the 7 TeV data and MC sample for the number of the signal events with and without the truncation of the signal region.

To determine the systematics of the tail outside the signal band, double-ratios of the results with and without the signal truncation between the data and the MC were provided, which are shown in Figure 5.28. Based on what is shown in Figure 5.28, a 3 % uncertainty was assigned to account for the signal tail systematics.

5.7.2 Pileup and era dependence at 7 TeV

To determine pileup effects on the 7 TeV cross sections, the reconstruction efficiency was compared between pileup and no-pileup MC samples. For this comparison, the inclusive MinimumBias MC samples, which are listed in Table 5.10, were taken for the study. Events on

Table 5.10: 2010 7 TeV MinimumBias MC samples used to determine pileup systematics. ⟨PU⟩ is the average of the number of good primary vertices per D^{*+} candidate.

MC Sample	#Events	⟨PU⟩
/MinBias_TuneZ1_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	9,994,400	1.010
/MinBias_TuneZ2star_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	19,866,400	1.018
Total	29,860,800	1.015
/MinBias_TuneZ1_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_PU_S0_START42_V17B-v1/AODSIM	9,994,400	2.875
/MinBias_TuneZ2star_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_PU_S0_START42_V17B-v2/AODSIM	19,868,000	2.909
Total	29,862,400	2.899

each pileup and no-pileup sample in the list were added up with a weight 1. Still the statistics of these combined samples is much lower than the data, thus the reconstruction efficiency was compared in the two categories, $p_T < 3.5$ GeV and $p_T > 3.5$ GeV, which is shown in Figure 5.29. With pileup, the reconstruction efficiency is lower by $5.2 \pm 6.7\%$ in the lower p_T region and larger by $3.6 \pm 2.9\%$ in the higher p_T region.

The pileup effect was also checked in the reconstructed kinematic distributions of inclusive MinimumBias tracks. In Figure 5.30, p_T and η distributions are compared between the pileup and no-pileup MC samples. Part of the differences could potentially originate from the MC tune for the pileup simulations which might be different from the main simulated MinimumBias events. The fluctuations shown in Figure 5.30 are typically within ~ 5 % in p_T and much less than those in η. The regions where the deviations are large have almost no statistics and lie outside the kinematic region used in this analysis. Thus it is assumed that there is no significant

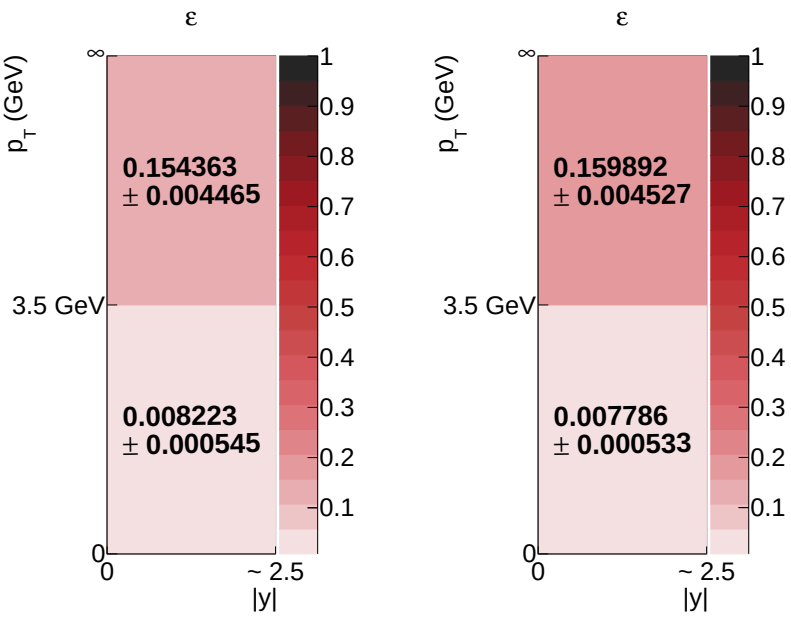


Figure 5.29: Reconstruction efficiencies calculated from no-pileup (left) and pileup (right) MC samples in Table 5.10.

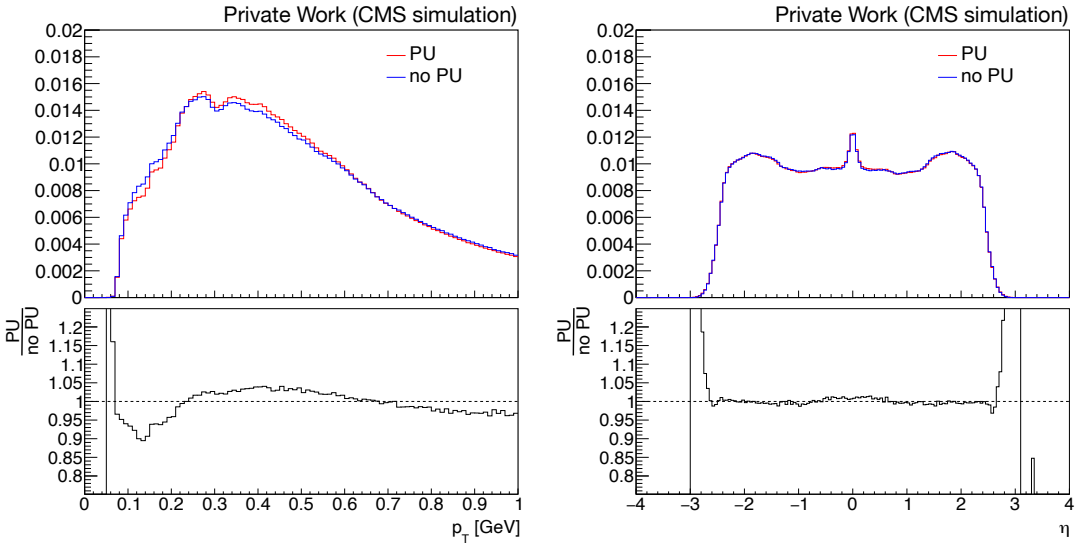


Figure 5.30: Comparison of track p_T (left) and η (right) distributions between the pileup and the no-pileup MC sample. The statistics are a few % of the TuneZ2star sample.

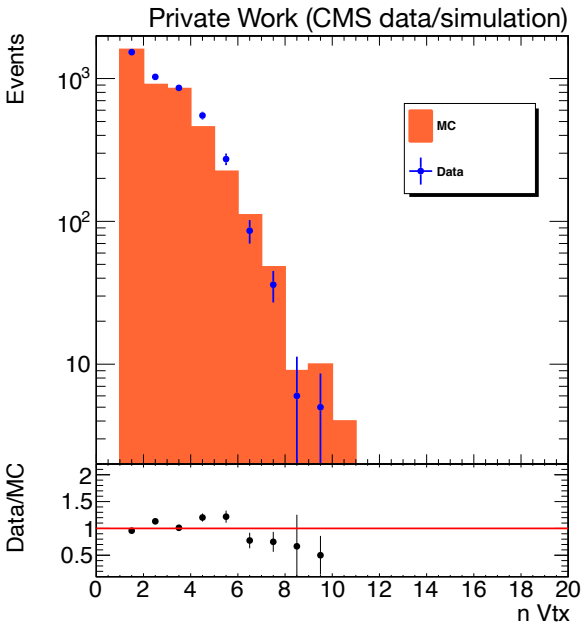


Figure 5.31: Pileup distributions of D^{*+} signal events. Figure is taken from [122].

pileup dependence for the shape in η, while effects of $\mathcal{O}(5\%)$ may be expected in p_T. This is also qualitatively consistent with what is observed in the reconstruction efficiency in Figure 5.29. Therefore, a systematic uncertainty of $\pm 5\%$ was added globally accounting for pileup effects without correction on the central value of the cross sections.

It turns out that the data distributions are well described by a mixture of the pileup and no-pileup MC samples with a relative weight to be 2.6:1, which is consistent with the fraction of data from low pileup MinimumBias/ZeroBias runs and NMB events from standard trigger runs. For example, the pileup distributions of D^{*+} signal events show good agreement between the data and MC in Figure 5.31.

The data samples listed in Table 5.1 have different $\langle \mathrm{PU} \rangle$ from the early runs of 2010 to Run B. In an additional Table 5.11, the number of good primary vertices per D^{*+} candidate in (N)MB events, which refers to $\langle \mathrm{PU} \rangle$, are shown. To determine the era dependence of the cross sections, the datasets were separated into 7 subsamples with similar effective NMB luminosity. Then ratios of the number of D^{*+} signal events to the effective luminosity ($N/\mathcal{L}$) were determined for each subsample first without pileup correction. Calculating a χ^2/ndof for the individual results to be compatible with their average(total) gives 9.9/6 in the lower p_T region and 10.8/6 in the higher p_T region. This changes into 8.9/6 in the lower p_T region and 13.8/6 in the higher p_T region with a pileup correction applied. The pileup correction was assumed to scale according to a formula [122]:

$$c = 1 + a \langle \mathrm{PU} \rangle (\langle \mathrm{PU} \rangle - 1) \tag{5.22}$$

With $\langle 2.9 \rangle$ of the pileup MC samples in Table 5.10, a is determined to be $+0.0094$ with $c = +1.052$ in the lower p_T region (5.2% decreased efficiency), and -0.0065 with $c = 0.964$ in the higher p_T region (3.6% increased efficiency). Empirically assigning an additional uncorrelated systematic uncertainty of 7% for each of the 7 subsamples, the χ^2/ndof becomes 8.6/6(7.6/6) in the lower p_T region and 4.3/6(5.8/6) in the higher p_T region without(with) the pileup correction. The sum of χ^2/ndof turns out to be 13/12, which is consistent with 1/dof independent of whether the pileup correction is applied or not. Detailed calculations given to derive these χ^2/ndof results can be found in Table 5.11.

For the 7 independent samples, this gives an uncertainty of $\sqrt{7}\%$. With slightly varying

Table 5.11: Determination of the pileup effects on the era dependence of the D^{*+} cross sections at $\sqrt{s} = 7$ TeV. Data samples are listed in the same order as in Table 5.1. $\langle$PU$\rangle$ is the average of the number of good primary vertices per D^{*+} candidate, and N is the number of D^{*+} signal events. σ refers to the deviation from the total, which is divided by the corresponding data uncertainty, and $\sigma(+7\%)$ shows the same but with an empirically assigned 7% uncertainty. χ^2/ndof at the bottom of this table was derived by each σ value.

Sample	$\langle$PU$\rangle$	$\mathcal{L}$ [nb^{-1}]	before PU correction						after PU correction					
			lower p_T			higher p_T			lower p_T			higher p_T		
			N/$\mathcal{L}$ [nb]	σ	$\sigma(+7\%)$	N/$\mathcal{L}$ [nb]	σ	$\sigma(+7\%)$	N/$\mathcal{L}$ [nb]	σ	$\sigma(+7\%)$	N/$\mathcal{L}$ [nb]	σ	$\sigma(+7\%)$
ZB C10	0.846	0.202	391 ± 109	-0.41	-0.40	1708 ± 134	+0.48	+0.36	395	-0.50	-0.48	1690	+0.59	+0.44
ZB A	1.741													
MB C10	1.041	0.620	503 ± 61	+1.10	+0.95	1689 ± 76	+0.59	+0.32	503	+0.89	+0.77	1689	+1.03	+0.55
MB A	1.773	0.498	512 ± 66	+1.15	+1.01	1639 ± 84	-0.06	-0.04	517	+1.03	+0.90	1623	+0.14	+0.08
MB B	2.494	0.309	505 ± 87	+0.79	+0.73	1757 ± 107	+1.06	+0.69	520	+0.82	+0.75	1722	+1.04	+0.69
MO A	2.991	0.473	478 ± 70	+0.60	+0.54	1457 ± 85	-2.20	-1.41	516	+0.96	+0.85	1384	-2.67	-1.76
MO B	3.598													
Mu A	2.927	0.455	299 ± 73	-1.88	-1.80	1534 ± 86	-1.28	-0.80	323	-1.73	-1.65	1457	-1.79	-1.15
Mu B	3.663													
MM A	2.500													
MM B	3.630													
EG A	3.079	0.444	318 ± 72	-1.64	-1.57	1793 ± 92	+1.62	+0.96	343	-1.47	-1.40	1703	+1.00	+0.61
El B	3.449													
EGM A	0.000													
EGM B	3.614													
Total	2.422	3.001	436 ± 28			1644 ± 34			449			1611		
χ^2/ndof				9.9/6	8.6/6		10.8/6	4.3/6		8.9/6	7.6/6		13.8/6	5.8/6

luminosity, eventually an additional systematic uncertainty of 3% was assigned to account for the era and NMB dependence. This uncertainty was applied as correlated between all bins, since all measurement bins are from the same average sample.

5.7.3 Systematic uncertainties

The following summarizes the complete(partial) set of systematic uncertainties in the 7 TeV(13 TeV and 0.9 TeV) analysis:

MC statistical uncertainties were propagated as systematic uncertainties for the evaluation of the reconstruction efficiencies, which were shown in Figure 5.10 and Figure 5.12 for each analysis.

Luminosity uncertainty assigned for the 7 TeV analysis is **4%** by accounting for the calculation of the effective luminosity of the NMB samples and the luminosity measurement for 2010 7 TeV data [129]. Official uncertainties of the Run 3 luminosity at $\sqrt{s} = 0.9$ TeV are not available at the time of writing this thesis and a conservative uncertainty of **10%** is currently used in the 0.9 TeV analysis [121], which will be reevaluated.

Branching fraction uncertainties with 0.0267 ± 0.0003 for D^{*+} and 0.0395 ± 0.0003 for D^0, were propagated as systematic uncertainty of **1%** in the 7 and 13 TeV (for both D^{*+} and D^0 cross section) analysis and **1.5**% in the 0.9 TeV analysis.

Tracking efficiency uncertainties were assigned in the 7 TeV analysis to be 2.3% for K and π tracks based on measurements in [130], while the π_s uncertainty was estimated to be 5% to account for the reconstruction efficiency dropping down for $p_T \lesssim 1$ GeV to about a half of its maximum. When assuming the tracks are maximally correlated, a total of **8%** is determined for the 7 TeV D^{*+} reconstruction.

Non-prompt contamination systematics in taking Pythia to determine beauty fractions was studied compared to measurements as described in Section 5.6 in the 7 and 13 TeV analysis. Accordingly the systematic uncertainties of the non-prompt contamination were propagated to the prompt cross section, 1-3% for D^{*+} and 3-6% for D^0 depending on phase space as shown in Figure 5.18. In case of the 0.9 TeV analysis, the beauty fractions determined by Pythia without normalization were propagated as the systematic uncertainties, 3-11% depending on phase space as shown in Figure 5.21. The uncertainties in the 13 and 0.9 TeV analysis are temporarily assigned and should be reevaluated. The 7 TeV uncertainty is the final one.

Pileup systematics in the 7 TeV analysis was determined in Section 5.7.2, and as a result, **5%** and **3%** uncertainties were assigned to account for pileup effects on the reconstruction efficiencies and era dependence, respectively. In the case of 0.9 TeV, the MC sample used for the reconstruction efficiency evaluation was produced based on the 2021 data and thus describes well its pileup distribution. A discrepancy between the 2021 MC sample and the 2022 data was studied and systematic uncertainties turned out to be **less than 1%** except for $p_T > 7$ GeV bins where the efficiencies were corrected by 0.001. In the case of 13 TeV, the pileup distribution of the D^0 MC sample was reweighted to that of the data as shown in Figure 5.4, for which the distributions of the number of primary vertices per event were used to account for both the D^{*+} and D^0 reconstruction at once. However, it was observed that the average pileup for D^{*+} events (the number of primary vertices per D^{*+} candidate) was shifted a bit to each other between the data and MC (see Appendix E), i.e., the average pileup is ~ 22 on the MC and ~ 25 on the data. Further studies to determine the shifted pileup effects are required.

Signal tail systematics was studied further in detail using the contrained fit method in the 7 TeV analysis as described in Section 5.7.1. Accordingly a **3%** uncertainty was assigned. For the 13 and 0.9 TeV analysis, the systematic uncertainties are not yet evaluated.

Multiplicity rescaling systematics in the 0.9 TeV analysis, which was induced by introduction of a rescaling factor to resolve discrepancies of p_T^{frac} variables between the MC and the data (refer to Figure 5.6), was studied. As a result, uncertainties of **1-3%** depending on phase space were determined.

Meanwhile, discrepancies observed between the data and MC sample in the 13 TeV analysis, and systematic effects of those on cross section measurement, were not fully understood yet, which should be properly investigated; the shifted pileup distribution, discrepancies in the number of tracks fitted to D^{*+} candidates in the higher p_T region, $\sim 10\%$ discrepancies observed with e.g. p_T^{frac} variables, and discrepancies observed on π_s^+ tracking efficiency measurements (see Appendix C).

The D^0 signal extraction is totally based on the fit method, and thus further detailed systematic studies should be provided for e.g. the definition of the signal function and systematics of the background template definition.

5.8 Prompt D^{*+} cross sections

Finally, prompt D^{*+} cross sections were determined at $\sqrt{s} = 7$ TeV (with the complete systematic uncertainty), 13 TeV (with incomplete systematic uncertainty) and 0.9 TeV (with incomplete systematic uncertainty), which are shown in Figure 5.32 - 5.34. The 7 and 13 TeV cross sections are compared with ALICE measurements from [13, 15] in $|y| < 0.5$ and LHCb measurements from [17, 18] in $2 < |y| < 2.5$. Good agreement was observed between the D^{*+} measurements (apart from the fact that the effective luminosity in the 13 TeV analysis is not determined yet). The 0.9 TeV cross sections shown in Figure 5.34 are the first cross section measurements ever introduced for charm production at this center-of-mass energy, and thus there is no measurement available to be compared with.

Each measurement is also compared to Pythia (the dashed lines) and FONLL predictions (the blue bands). Note that the FONLL prediction is provided based on the charm fragmentation universality, while it has been reported recently that the charm fragmentation is non-universal. Therefore it is not expected to be consistent between the measurements and the FONLL prediction in the figures, but the effects of the non-universality are not significant in D meson production compared to the given theory uncertainties. Further detailed discussions about the non-universality will be presented in the next chapter, and total charm cross sections will be derived by taking these measurements shown in this section and applying the non-universal charm fragmentation.

The very preliminary results of measuring prompt D^0 cross sections at $\sqrt{s} = 13$ TeV are shown in Appendix B.

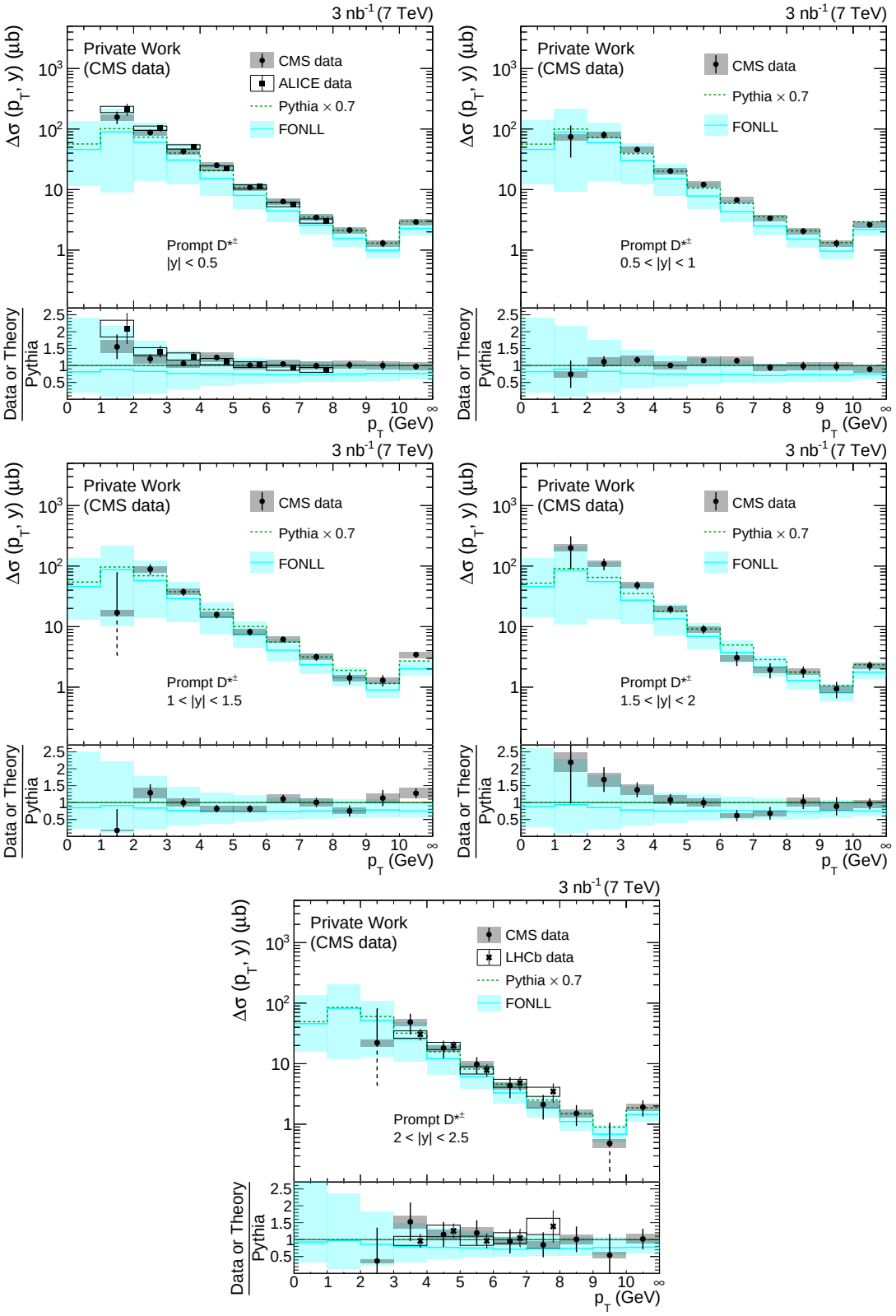


Figure 5.32: Prompt $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The vertical error bars indicate the statistical uncertainties, and the error boxes show the (complete) systematic uncertainties.

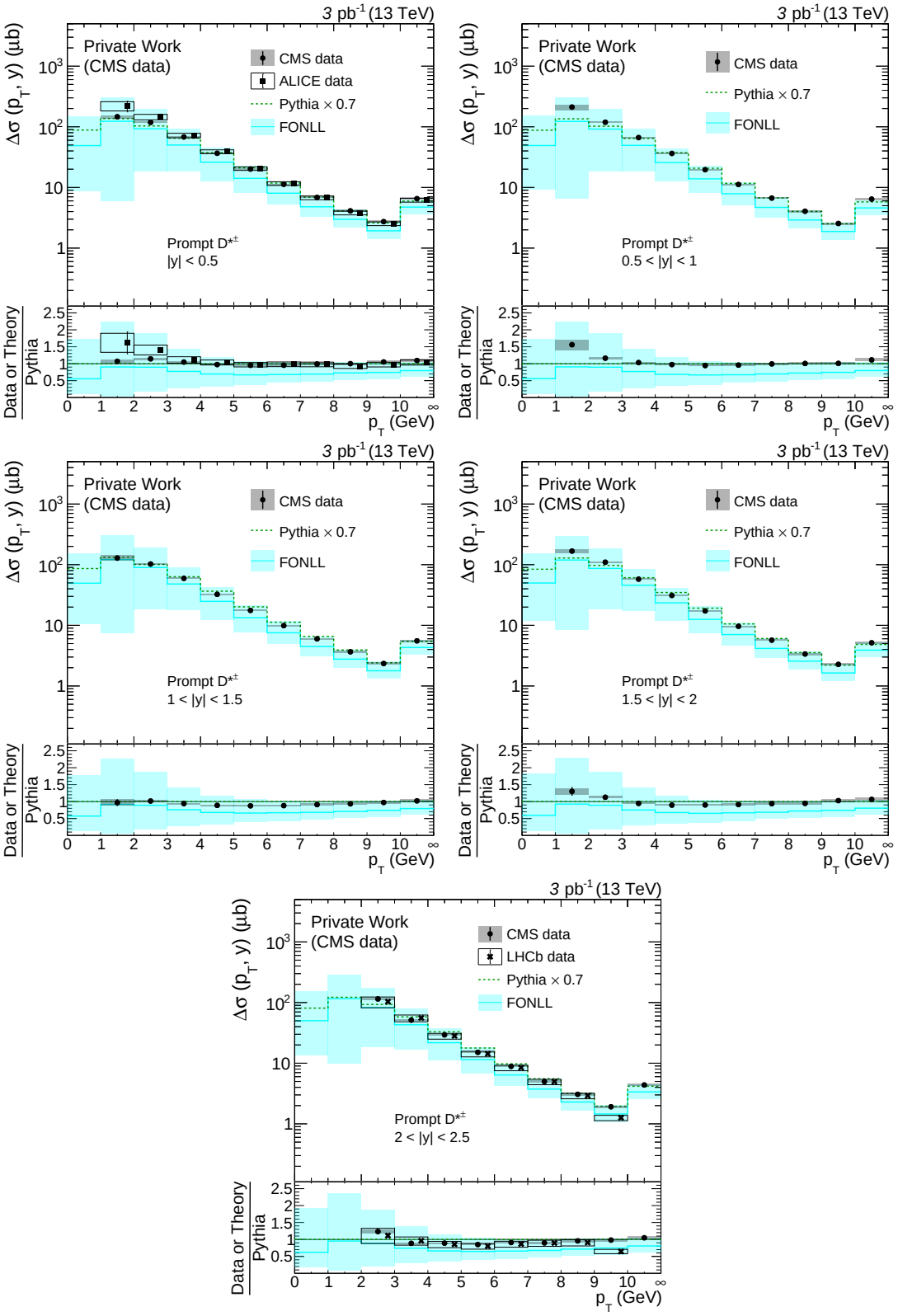


Figure 5.33: Prompt $D^{*\pm}$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The vertical error bars indicate the statistical uncertainties, and the error boxes show the systematic uncertainties. The systematic uncertainties of CMS are incomplete.

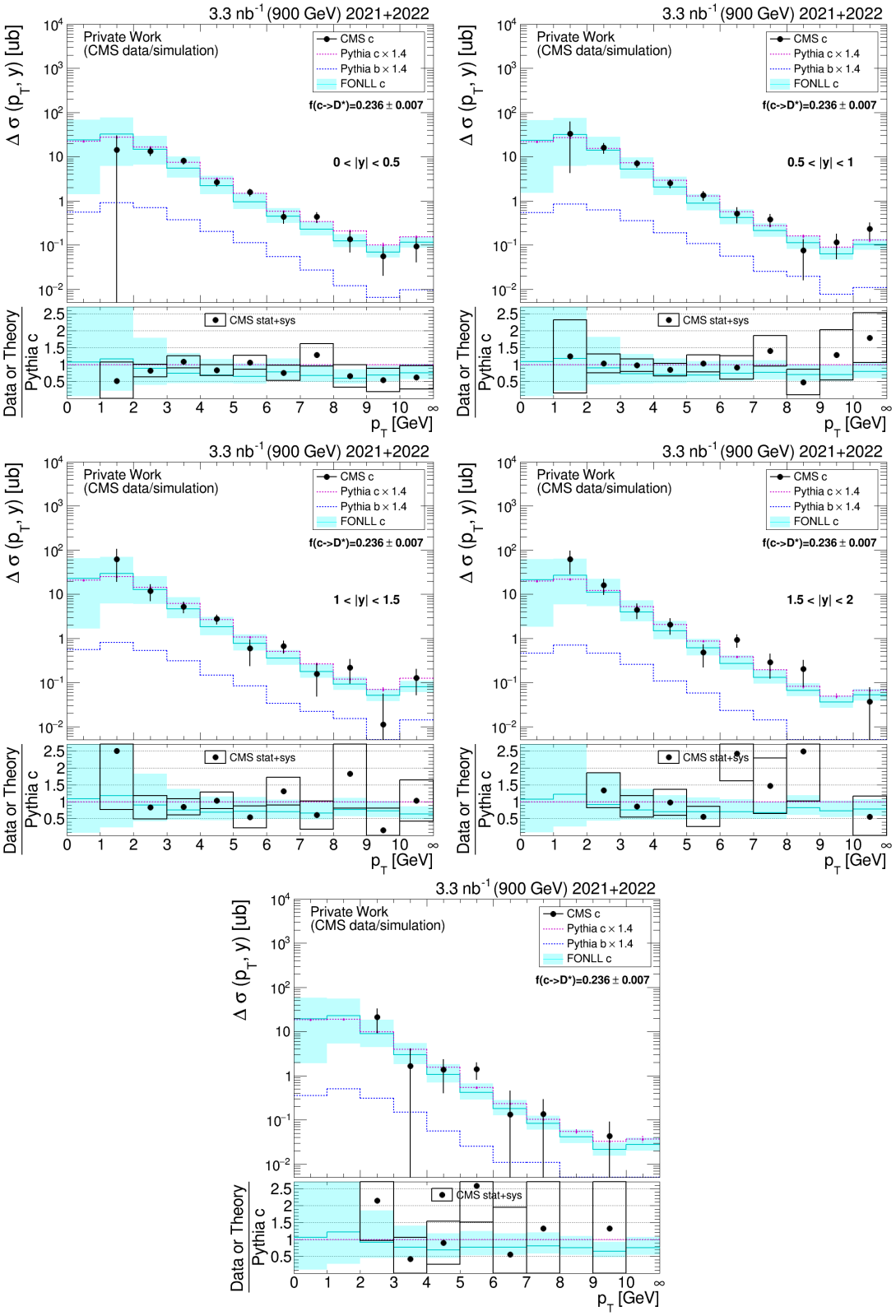


Figure 5.34: Prompt $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The systematic uncertainties are incomplete. Figures are taken from [121].

6 Charm Fragmentation Study and Total Charm Cross Section

6.1 Introduction

Recent LHC measurements reported that the charm fragmentation fractions disagree between e^+e^-/ep and pp collisions [15, 25] with up to $\sim 5\sigma$ discrepancies and the baryon-to-meson ratio in charm production has a clear p_T dependence at low p_T [10, 15, 21] (refer to also Section 3.3). A significant p_T dependence was observed also in beauty production as shown in the left panel of Figure 6.1 where the fit to the LHCb measurements gives a result consistent with the LEP value at high p_T. Similarly, the charm baryon-to-meson ratios at the LHC are asymptotically close to e.g. the LEP data at high p_T, as shown in the right panel of Figure 6.1. Note that the LEP point is added at the approximate p_T in Z decays as done in the left figure. To show the asymptotic agreement with the LEP data at high p_T, a fit (the red curve) is performed to the ALICE data at $\sqrt{s} = 13$ TeV including the LEP point, using an exponential function as the one used in the left figure and the resulting parametrization is $0.083 + \exp\left[-0.748 + (-0.095) \times p_T\right]$. This shows also good agreement with the ALICE and CMS data at $\sqrt{s} = 5$ TeV. This fit is provided only for a demonstration purpose and thus the fit results will not be discussed further or used anywhere else in this thesis. Although, it is clear that the baryon-to-meson production ratios in pp collisions agree well with the asymptotic p_T dependence and with the LEP value at high p_T, both in the charm and beauty production.

The study given in this chapter starts from these observations. First of all, phenomenological results from e^+e^-/ep data cannot be directly applied to pp collisions in an universal way as done so far. With the universality assumption, the charm fragmentation is independent of either kinematics or the collision system. The charm fragmentation fraction has been measured precisely from e^+e^-/ep collisions and indeed no significant discrepancies have been reported between the two. Thus fragmentation universality has been assumed so far and applied also to pp collisions. However, since it has been shown above that the charm fragmentation is clearly dependent on the collision system (Section 3.3) with a clear kinematic dependence in pp collisions, this *non-universality* should be treated properly with pp data. This is particularly important for the derivation of the total charm cross section by extrapolating measured fiducial cross sections of hadrons.

To treat the non-universal charm fragmentation, in this thesis, a p_T dependent correction is introduced to derive so-called p_T *dependent production fractions* instead of using the fragmentation fractions (defined to be independent of kinematics). Since the LHC measurements show consistency with LEP data at high p_T, the LEP data (more generally, e^+e^- data, see later) is taken to be a value at high p_T with a p_T dependent correction introduced for the lower p_T region. As a starting point, the study is performed with the most simple assumption, that the meson-to-meson and baryon-to-baryon ratios are still universal, i.e., independent of either the collision system or kinematics (refer to Section 3.3 and see more later). Under this assumption, the p_T-dependence of the baryon-to-meson ratio can be defined by a general form. It is also assumed that the charm fragmentation is independent of (pseudo-)rapidity (refer to Figure 3.17). With these assumptions, the p_T dependent correction is determined by one of the well measured

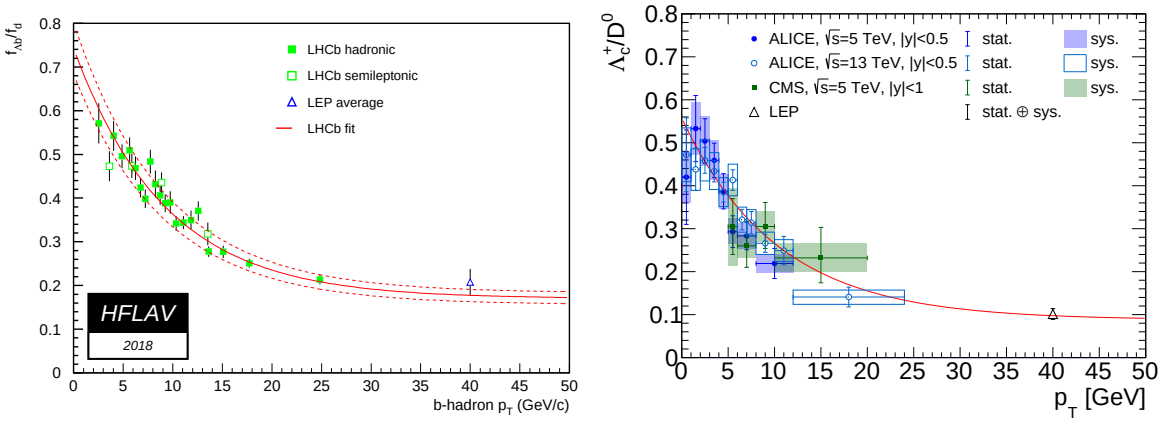


Figure 6.1: Measurements of the ratio of Λ_b^0 to B^0 (left figure) and the ratio of Λ_c^+ to D^0 (right figure). The left figure is from [89], where the exponential fit to LHCb data gives consistent results to LEP at high p_T. In the right figure, the ALICE and CMS points are collected from [10, 15] and [21], respectively, and LEP points are derived based on the numbers given in [47]. Here the fit in the right figure is given only for a demonstration purpose to show consistency with what is observed from beauty production using the same exponential function.

baryon-to-meson ratios as a function of p_T. Using data directly, the study is introduced without the need to assume any particular non-universality fragmentation model. The derivation of the p_T dependent production fractions by applying the p_T dependent correction will be discussed further in Section 6.2.1.

This p_T dependent production fraction is then applied to derive total charm cross section measurements at the LHC. To extrapolate the fiducial hadronic cross sections to the full phase space, FONLL was taken but modified by applying the p_T dependent production fractions. This is called *ddFONLL* in this thesis. More detailed discussions about this phenomenological approach for the extrapolation are given in Section 6.2.2.

Eventually, total charm-pair cross section measurements are presented at various center-of-mass energies ($\sqrt{s}$) at the LHC in Section 6.3.2. Providing these measurements allows for a comparison to the NNLO theory (refer to Section 2.4). Furthermore, the theory is totally free from fragmentation inputs at the total cross section level. Thus, the measurements provided as a function of $\sqrt{s}$ can be used to constrain QCD parameters like the charm mass and the low-x part of PDFs (some PDF sets show still very large uncertainties at the low x, i.e., the large $\sqrt{s}$). The first example of applying the results to constrain the QCD parameters will be also shown in Section 6.3.2.

6.2 Non-universal charm fragmentation and extrapolation

As introduced above, the study starts with the simplest assumptions made to apply the non-universal charm fragmentation to the extrapolation:

Assumption 1 *meson-to-meson and baryon-to-baryon ratios are universal, i.e., independent of kinematics and the collision system,*

and

Assumption 2 *baryon-to-meson ratios are dependent on transverse momentum, while independent of (pseudo-)rapidity.*

Besides what have been shown already above, additional cross checks on these assumptions will be provided in the following sections.

6.2.1 p_T dependent charm hadron production fractions

In this thesis, p_T dependent production fractions ($\tilde{f}(p_T)$) are defined for pp collisions as cross-section fractions of each hadron state relative to the sum of all the weakly-decaying ground states ($w.d.$) as a function of p_T:

$$\tilde{f}_{H_c}(p_T) = \frac{d\sigma_{H_c}}{\Sigma_{w.d.} d\sigma_{H_c}}, \tag{6.1}$$

where $d\sigma$ is the p_T-differential cross-section and p_T is the transverse momentum of each hadron. The weakly-decaying ground states* are taken to be D^0, D^+, D_s^+, Λ_c^+, Ξ_c^0, Ξ_c^+ and Ω_c^0.

To derive $\tilde{f}(p_T)$, so-called p_T dependent factors ($F(p_T)$) are applied to the fragmentation fractions of e^+e^-/ep collisions (f^{uni}):

$$\tilde{f}_{D^0}(p_T) = f^{uni}_{D^0} F_{MS}(p_T), \tag{6.2}$$

$$\tilde{f}_{D^+}(p_T) = f^{uni}_{D^+} F_{MS}(p_T), \tag{6.3}$$

$$\tilde{f}_{D_s^+}(p_T) = f^{uni}_{D_s^+} F_{MS}(p_T), \tag{6.4}$$

$$\tilde{f}_{\Lambda_c^+}(p_T) = f^{uni}_{\Lambda_c^+} F_{BY}(p_T), \tag{6.5}$$

$$\tilde{f}_{\Xi_c^0}(p_T) = f^{uni}_{\Xi_c^0} F_{BY}(p_T), \tag{6.6}$$

$$\tilde{f}_{\Xi_c^+}(p_T) = f^{uni}_{\Xi_c^+} F_{BY}(p_T), \tag{6.7}$$

and

$$\tilde{f}_{\Omega_c^0}(p_T) = f^{uni}_{\Omega_c^0} F_{BY}(p_T), \tag{6.8}$$

where the same factors $F_{MS}(p_T)$ and $F_{BY}(p_T)$ are applied to each meson and baryon state, respectively, making use of the simplest assumption mentioned above. By definition, the sum of the production fractions for all the weakly-decaying ground states is unity:

$$f^{uni}_{MS} F_{MS}(p_T) + f^{uni}_{BY} F_{BY}(p_T) = 1 \tag{6.9}$$

where f^{uni}_{MS} is the sum of all meson fractions:

$$f^{uni}_{D^0} + f^{uni}_{D^+} + f^{uni}_{D_s^+}, \tag{6.10}$$

and f^{uni}_{BY} is the sum of all baryon fractions:

$$f^{uni}_{\Lambda_c^+} + f^{uni}_{\Xi_c^0} + f^{uni}_{\Xi_c^+} + f^{uni}_{\Omega_c^0}. \tag{6.11}$$

Then the relation between meson and baryon modifiers can be given by

$$F_{BY}(p_T) = \frac{1 - f^{uni}_{MS} F_{MS}(p_T)}{f^{uni}_{BY}}. \tag{6.12}$$

To determine $F_{MS}(p_T)$ and $F_{BY}(p_T)$, the most precise measurements of the ratio of baryon to meson are to be taken and those are the measurements of the ratio of Λ_c^+ to D^0 for the time being. Denoting the p_T-dependent cross-section ratio of Λ_c^+ to D^0 by $R(p_T)$, a relation between $R(p_T)$ and $\tilde{f}(p_T)$ can be given as

$$R(p_T) \equiv \frac{\tilde{f}_{\Lambda_c^+}(p_T)}{\tilde{f}_{D^0}(p_T)} = \frac{f^{uni}_{\Lambda_c^+} F_{BY}(p_T)}{f^{uni}_{D^0} F_{MS}(p_T)} = C\left(\frac{1}{f^{uni}_{MS} F_{MS}(p_T)} - 1\right), \tag{6.13}$$

*Here multi-charm states are neglected.

where a constant term C is defined by:

$$C \equiv \frac{f^{uni}_{\Lambda_c}}{f^{uni}_{D^0}} \frac{f^{uni}_{MS}}{f^{uni}_{BY}}. \tag{6.14}$$

As a result, $F_{MS}(p_T)$ and $F_{BY}(p_T)$ can be determined by the fragmentation fractions measured in e^+e^-/ep collisions and $R(p_T)$:

$$F_{MS}(p_T) = \frac{1}{f^{uni}_{MS}} \frac{C}{R(p_T) + C} \tag{6.15}$$

and

$$F_{BY}(p_T) = \frac{1}{f^{uni}_{BY}} \left(1 - \frac{C}{R(p_T) + C}\right). \tag{6.16}$$

Therefore, the p_T dependent production fractions can be described, e.g., for D^0 and Λ_c^+ as

$$\tilde{f}_{D^0}(p_T) = \frac{f^{uni}_{D^0}}{f^{uni}_{MS}} \times \frac{C}{R(p_T) + C} \tag{6.17}$$

and

$$\tilde{f}_{\Lambda_c^+}(p_T) = \frac{f^{uni}_{\Lambda_c^+}}{f^{uni}_{BY}} \times \left(1 - \frac{C}{R(p_T) + C}\right) \tag{6.18}$$

by inserting Eq.(6.15) into Eq.(6.2) and Eq.(6.16) into Eq.(6.5), respectively.

The f^{uni}s are to be extracted from either e^+e^- or ep data. However, as described in Section 2.2, the definition of fragmentation fraction is different in e^+e^- and ep collisions. In this thesis, avoiding additional arguments about the assumption Eq.(3.7) applied for charm fragmentation fraction measurements, the e^+e^- values only were used.

Then in this thesis, the fragmentation fraction sum of all the other states not yet measured (Ξ_c^0, Ξ_c^+ and Ω_c^0) in e^+e^- collisions is assumed to be

$$\begin{aligned} f(c \to \Xi_c^0) + f(c \to \Xi_c^+) + f(c \to \Omega_c^0) \\ \equiv 1 - f(c \to D^0) + f(c \to D^+) + f(c \to D_s^+) + f(c \to \Lambda_c^+) \end{aligned} \tag{6.19}$$

so that the sum of all the known weakly-decaying ground states is unity. The fragmentation fractions measured in e^+e^- collisions were taken from [47]. The collected e^+e^- fragmentation fractions for D^0, D^+, D_s^+ and Λ_c^+ (including D^{*+}) can be found in Table 6.1. The fractions of the rest (Ξ_c^0, Ξ_c^+ and Ω_c^0) were derived by Eq.(6.19) where the uncertainty was calculated under the assumption that all the measured fraction uncertainties are fully uncorrelated (the detailed calculation of the uncertainty propagation is explained in Appendix I).

In Figure 6.2, the fragmentation fractions of e^+e^- data are compared to the ALICE measurements at $\sqrt{s} = 5$ TeV [25] and 13 TeV [15]. The $f(c \to \Xi_c + \Omega_c)$ (here $\Xi_c + \Omega_c$ indicates $\Xi_c^0 + \Xi_c^+ + \Omega_c^0$) of the 5 TeV ALICE data is the double of the measured $f(c \to \Xi_c^0)$ to account for the additional Ξ_c^+ contribution where the Ω_c^0 contribution is already accounted for in the uncertainties as described in [25]. Recently the 5 TeV ALICE results were updated in [15], but the differences are not significant in the present context. Thus, here the numbers are still based on [25] (i.e., the numbers are the same as the ones used for the results presented in [28]). The $f(c \to \Xi_c + \Omega_c)$ of the 13 TeV ALICE data[†] is given by the sum of the measured $f(c \to \Xi_c^0)$ and $f(c \to \Xi_c^+)$, where the Ω_c^0 contribution is accounted for in the uncertainties as described in [15]. The $f(c \to \Xi_c + \Omega_c)$ uncertainties of the 5 and 13 TeV ALICE data were calculated again by assuming that the measurement uncertainties are fully uncorrelated.

Based on the fragmentation fractions shown in Figure 6.2, the ratios of meson to D^0 and baryon to Λ_c^+ are derived as shown in Figure 6.3. The uncertainties were calculated again

[†]The fragmentation fractions of the 13 TeV ALICE data [15] were measured by counting the J/ψ contribution also. However, the contribution is less than 1%, which is neglected in this thesis.

$f(c \to H_c)$	LEP	B-factory	e^+e^- averaged
D^0	0.547 ± 0.022	0.577 ± 0.024	0.562 ± 0.016
D^+	0.227 ± 0.010	0.264 ± 0.014	0.245 ± 0.009
D_s^+	0.093 ± 0.008	0.069 ± 0.005	0.081 ± 0.005
D^{*+}	0.237 ± 0.006	0.247 ± 0.014	0.242 ± 0.008
Λ_c^+	0.056 ± 0.007	0.053 ± 0.003	0.054 ± 0.004
$\Xi_c^0 + \Xi_c^+ + \Omega_c^0$	0.078 ± 0.026	0.037 ± 0.028	0.058 ± 0.019

Table 6.1: Fragmentation fractions from e^+e^- data extracted from [47]. The fractions of $\Xi_c^0 + \Xi_c^+ + \Omega_c^0$ were derived so that the sum of fragmentation fractions of all the weakly-decaying ground states is unity by assuming that measurement uncertainties are fully uncorrelated.

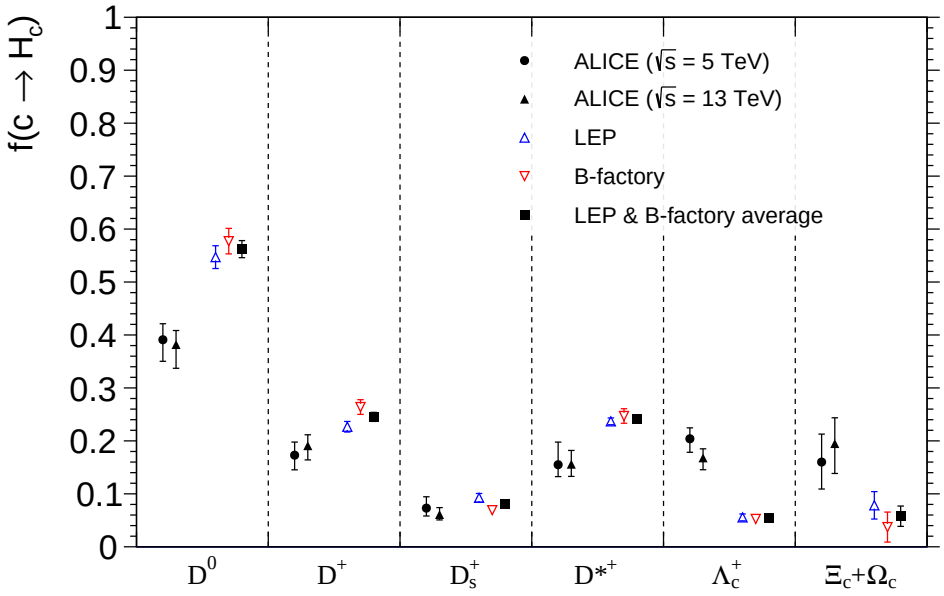


Figure 6.2: Fragmentation fractions measured from the ALICE [15, 25] and e^+e^- experiments [47]. The label $\Xi_c + \Omega_c$ indicates $\Xi_c^0 + \Xi_c^+ + \Omega_c^0$.

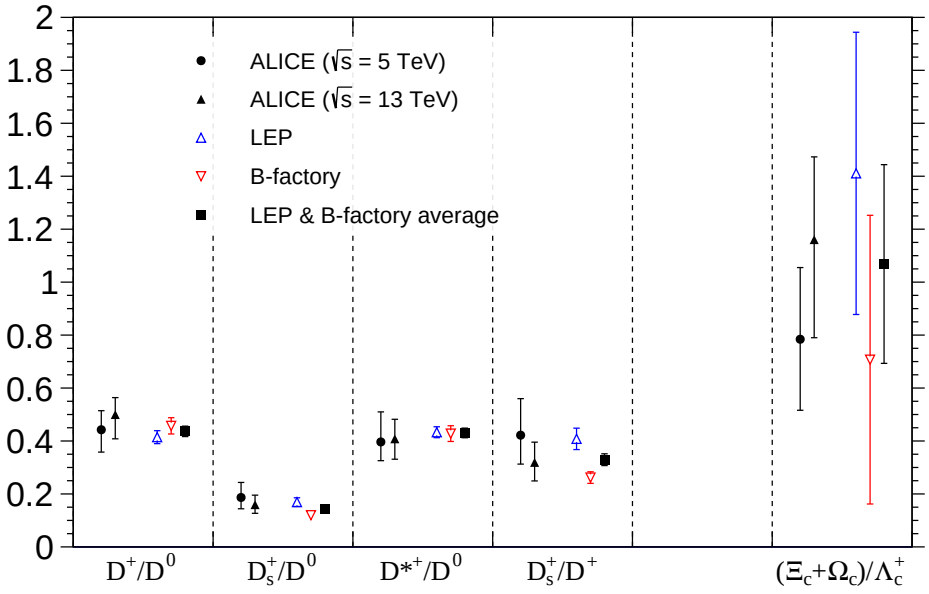


Figure 6.3: Comparison of meson-to-D^0, D_s^+-to-D^+ and baryon-to-Λ_c^+ ratios of fragmentation fractions between e^+e^- and pp collisions. The label $\Xi_c + \Omega_c$ indicates $\Xi_c^0 + \Xi_c^+ + \Omega_c^0$. The uncertainties were derived under the assumption that the measurement uncertainties are fully uncorrelated.

assuming that all the measurement uncertainties are fully uncorrelated. In particular, the propagation of the measurement uncertainties for $\frac{f_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{f_{\Lambda_c^+}}$ of e^+e^- data can be found in Appendix I. These ratios show consistency with the assumption that the meson-to-meson and baryon-to-baryon ratios are universal.

The ratios of Λ_c^+ to D^0 to be used as $R(p_T)$ were collected from the measurements as a function of p_T at $\sqrt{s} = 5$ TeV from ALICE [10] and CMS [21], and at $\sqrt{s} = 13$ TeV from ALICE [15]. As an asymptotic value at high p_T, the e^+e^- averaged numbers shown in Table 6.1 were used. In the case of 5 TeV, since the ALICE measurements are more precise in the lower p_T region, by default the ALICE points were used if applicable, otherwise the CMS points were taken instead. The $R(p_T)$ values used for this thesis are explicitly written in Table 6.2. The values in the range $0 < p_T < 8$ GeV were taken from the ALICE measurements as they are and the value of $8 < p_T < 10$ GeV was taken from the ALICE measurement of $8 < p_T < 12$ GeV. For $p_T > 10$, the CMS measurement in the range $10 < p_T < 20$ GeV was taken for $10 < p_T < 20$ GeV and the averaged e^+e^- value was used as the $p_T > 20$ GeV point. For the extrapolation which will be introduced in the next section, the two values were summed to give an overflow bin $p_T > 10$ GeV by applying weights determined based on the FONLL predictions. Tentative uncertainties for these weights are negligible compared to the measurement uncertainties in the end. The statistical and systematic uncertainties of the ALICE and CMS measurements were summed in quadrature. Similarly, $R(p_T)$ at $\sqrt{s} = 13$ TeV is shown in Table 6.3. All the other points except the overflow bin were collected directly from the ALICE measurements. The overflow bin is given by a combined point of the ALICE measurements in the range $10 < p_T < 24$ and the e^+e^- point defined for $p_T > 24$, applying again weights determined based on the FONLL prediction.

Using the numbers in Table 6.2 or Table 6.3 as $R(p_T)$ and the averaged numbers in Table 6.1 as f^{uni}s, $F_{MS}(p_T)$ and $F_{BY}(p_T)$ were derived with Eq.(6.15) and Eq.(6.16) and are shown in Figure 6.4. Since the ratios of Λ_c^+ to D^0 are asymptotically identical to the ratio in e^+e^- collisions at high p_T by construction, $F_{MS}(p_T)$ and $F_{BY}(p_T)$ are asymptotically unity at high

[GeV]	$R(p_T)$	
$0 < p_T < 1$	$0.420 + 0.125 - 0.125$	
$1 < p_T < 2$	$0.533 + 0.098 - 0.098$	
$2 < p_T < 3$	$0.504 + 0.078 - 0.077$	
$3 < p_T < 4$	$0.459 + 0.061 - 0.061$	
$4 < p_T < 5$	$0.387 + 0.057 - 0.057$	
$5 < p_T < 6$	$0.293 + 0.048 - 0.047$	
$6 < p_T < 8$	$0.283 + 0.044 - 0.043$	
$8 < p_T < 10$	$0.219 + 0.041 - 0.041$	
$10 < p_T < 20$	$0.232 + 0.078 - 0.067$	$0.223 + 0.074 - 0.063$
$p_T > 20$	$0.096 + 0.007 - 0.007$	

Table 6.2: $R(p_T)$ at $\sqrt{s} = 5$ TeV. The third column of $p_T > 10$ GeV was derived by the sum of the two values in the second column applying weights determined based on the FONLL predictions.

[GeV]	$R(p_T)$	
$0 < p_T < 1$	$0.472 + 0.029 - 0.106$	
$1 < p_T < 2$	$0.438 + 0.068 - 0.069$	
$2 < p_T < 3$	$0.459 + 0.056 - 0.056$	
$3 < p_T < 4$	$0.434 + 0.050 - 0.050$	
$4 < p_T < 5$	$0.385 + 0.041 - 0.041$	
$5 < p_T < 6$	$0.413 + 0.044 - 0.044$	
$6 < p_T < 7$	$0.321 + 0.038 - 0.038$	
$7 < p_T < 8$	$0.314 + 0.040 - 0.040$	
$8 < p_T < 10$	$0.266 + 0.033 - 0.033$	
$10 < p_T < 12$	$0.249 + 0.042 - 0.042$	$0.189 + 0.033 - 0.034$
$12 < p_T < 24$	$0.141 + 0.028 - 0.029$	
$p_T > 24$	$0.096 + 0.007 - 0.007$	

Table 6.3: $R(p_T)$ at $\sqrt{s} = 13$ TeV. The third column of $p_T > 10$ GeV was derived by the sum of the three values in the second column applying weights determined based on the FONLL predictions.

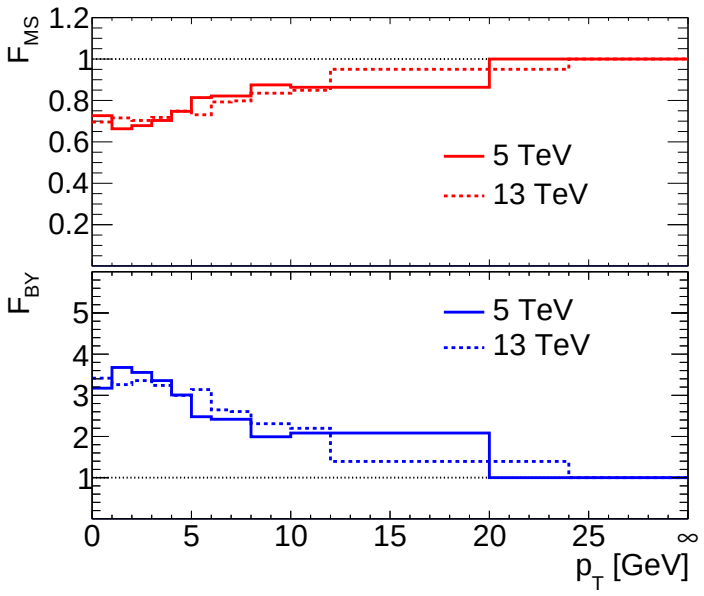


Figure 6.4: $F_{MS}(p_T)$ (red histogram) and $F_{BY}(p_T)$ (blue histogram) at $\sqrt{s} = 5$ (solid line) and 13 (dashed line) TeV. These are asymptotically close to 1 at high p_T by definition, Eq.(6.9).

p_T by definition, Eq.(6.9). Lastly, $\tilde{f}_{D^0}(p_T)$ and $\tilde{f}_{\Lambda_c^+}(p_T)$ were derived by Eq.(6.2) and (6.5), of which results are shown in Figure 6.5. The detailed calculation of the $\tilde{f}$ uncertainties can be found in Appendix I.

For the $\tilde{f}$ uncertainties, an additional systematic uncertainty was assigned to account for the assumption that meson-to-meson and baryon-to-baryon ratios are consistent between e^+e^- and pp collisions and independent of kinematics. Precise measurements in beauty production from LHC experiments, showed that B_s^0/B^+ has a clear p_T dependence at low p_T, which is asymptotically flat (refer to Figure 3.15) at high p_T. No precise measurement is available to show such a clear p_T dependence yet for D mesons (refer to Figure 3.9). Therefore, an additional uncertainty was assigned to account for possible p_T dependence of the D_s^+/D^0 and D_s^+/D^+ ratios, by covering the ALICE uncertainties of $0.14 < D_s^+/D^0 < 0.24$ and $0.33 < D_s^+/D^+ < 0.56$, as shown in Figure 6.6a. This uncertainty covers well also the ratios measured as a function of p_T (see Figure 6.6b).

In addition to $\tilde{f}$s for the ground states, $\tilde{f}$ for D^{*+} was also derived in order to extrapolate the CMS D^{*+} measurements introduced in the previous chapter. To define a proper $\tilde{f}$ for D^{*+}, D^{*+}/D^0 measurements from ALICE at $\sqrt{s} = 5$ and 7 TeV [7] were compared with a prediction extracted based on e^+e^- data, as shown in Figure 6.7. The dashed curve is the prediction which was derived by FONLL using the BCFY functions for D^{*+} and D^0 (refer to Section 2.2). The comparison between the ALICE measurements and the prediction shows consistency within the measurement uncertainties. This means that the same $F_{MS}(p_T)$ can be applied for D^{*+}, and thus $\tilde{f}_{D^{*+}}$ is defined as

$$\tilde{f}_{D^{*+}}(p_T) = f_{D^{*+}}^{uni} F_{MS}(p_T), \tag{6.20}$$

which is shown by the green points in Figure 6.5.

In Figure 6.7, the FONLL prediction derived with the Kartvelishvili function (Eq.(2.26)) also is shown by the red curve. The α_K values were determined for each D^{*+} and D^0 by comparing the FONLL predictions derived using the Kartvelishvili function to the ones derived with the BCFY functions. As a result, $\alpha_K = 9.5$ (D^{*+}) and 6.1 (D^0) were used in Figure 6.7, which give again consistent results compared to the ALICE measurements. The derivation of these α_K values can be found in Appendix J. The extrapolation which will be introduced in the next

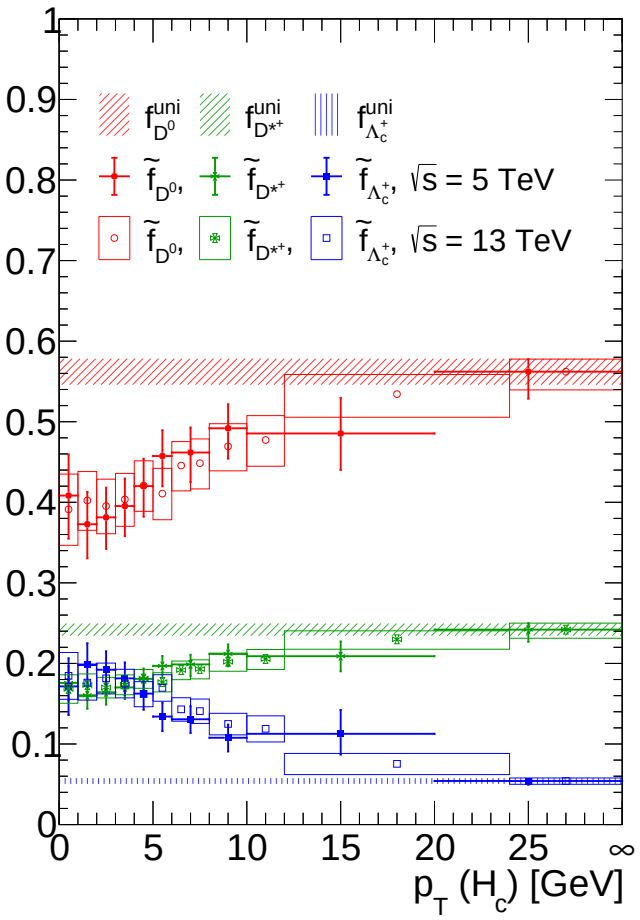
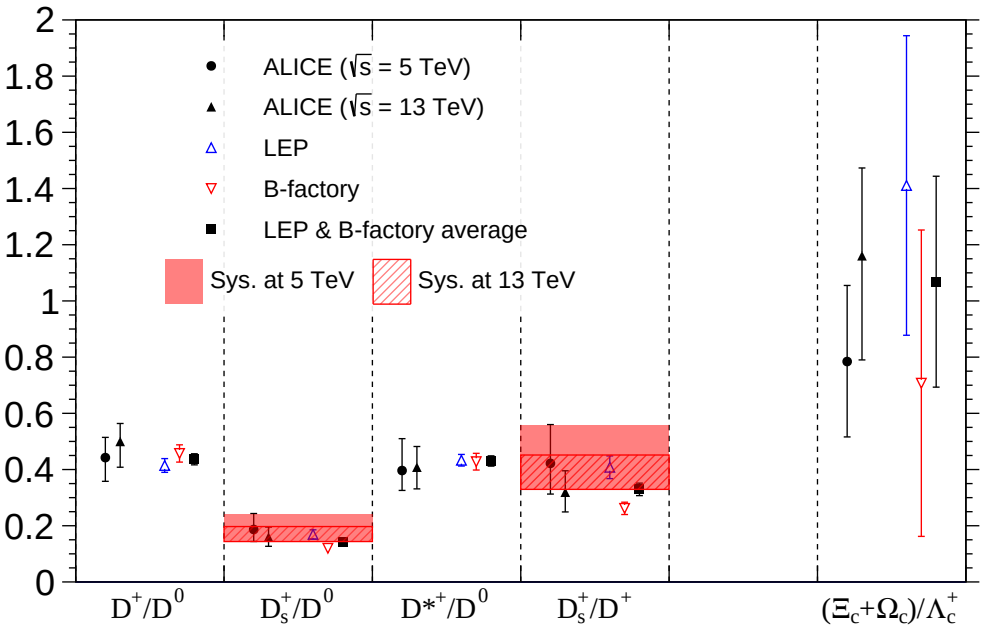
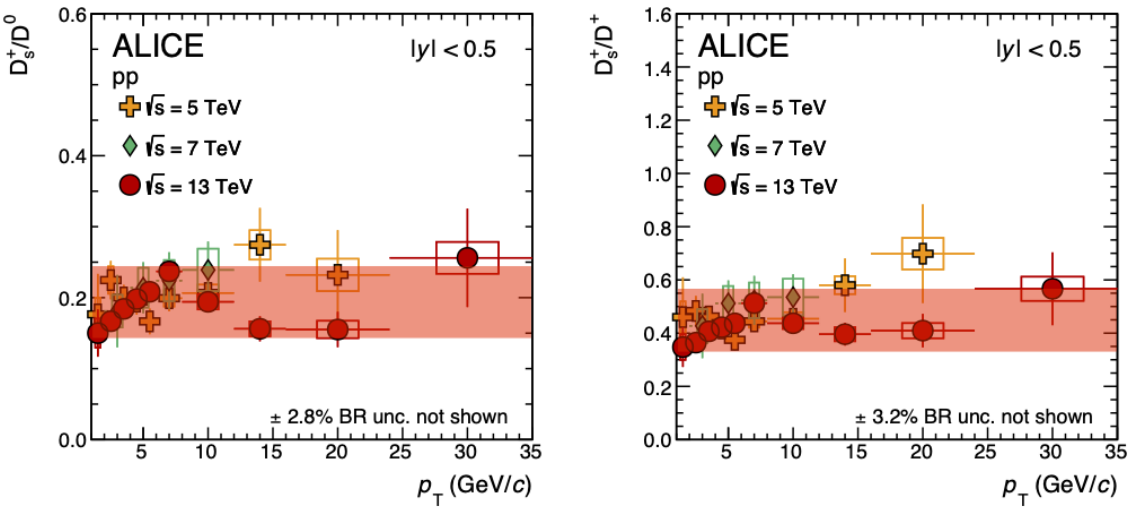


Figure 6.5: $\tilde{f}_{D^0}(p_T)$, $\tilde{f}_{\Lambda_c^+}(p_T)$ and $\tilde{f}_{D^{*+}}(p_T)$. The red, blue and green band are the D^0, Λ_c^+ and D^{*+} fragmentation fractions with the uncertainties, respectively, measured from the e^+e^- collisions.



(a) Meson-to-D^0 and baryon-to-Λ_c^+ ratios of the fragmentation fractions.



(b) Ratios of D_s^+ to D^0 (left) and D^+ (right), with figures adapted from [15]. The red bands refer to the systematic uncertainties assigned at $\sqrt{s} = 5$ TeV.

Figure 6.6: Additional uncertainty (red bands) assigned to account for a possible p_T dependence of D_s^+/D^0 and D_s^+/D^+.

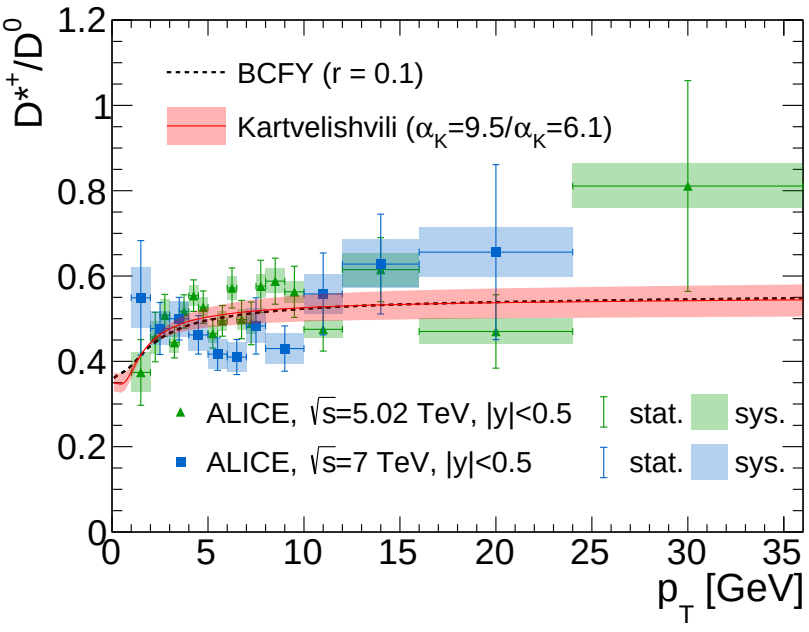


Figure 6.7: D^*/D^0 comparison between ALICE measurements [7] at $\sqrt{s} = 5$ (the green triangle points) and 7 (the blue square points) TeV and FONLL predictions (the black dashed and red solid line).

section was provided with the Kartvelishvili function instead of the BCFY one. Then, α_K was introduced as free parameter for the χ^2 scan in the extrapolation (see later). Varying α_K deals with a possible p_T dependent ratio of D^{*+} to D^0, which is expected even in the universality case, as demonstrated in Figure 6.7.

6.2.2 Data-driven FONLL

For the total charm cross section, charm hadron measurements of a constrained kinematic range should be extrapolated (and interpolated, depending on available measurements) to the full kinematic range. In this thesis, measurements were extrapolated by taking FONLL as perturbative theory. To account for the non-universal charm fragmentation, the p_T dependent production fraction ($\tilde{f}$) was introduced instead of the fragmentation fraction in the FONLL framework (Eq.(2.59)):

$$\Delta\sigma_{H_c}^{\text{FONLL with } \tilde{f}} \equiv \tilde{f}_{H_c} \cdot \left(\Delta\sigma_c \otimes D_{c\to H_c}^{NP} \right), \tag{6.21}$$

where $\Delta\sigma$s are integrated cross section over the bins. In order to reduce the FONLL theory uncertainty, which is very large compared to measurement uncertainties given to date, a χ^2 *scan* was introduced for some of QCD parameters; the two theory scales (μ_f and μ_r), the charm mass (m_c) and the α_K. In other words, the parameters describing data best are determined by χ^2 calculation defined by

$$\sum_{\text{data bins}} \frac{(\text{FONLL} - \text{data})^2}{\text{statistical unc.}^2 + \text{systematic unc.}^2}. \tag{6.22}$$

Then, the so-called *data-driven FONLL (ddFONLL)* was defined with the best parameters.

To determine the best μ_f and μ_r, new variables x_f and x_r were introduced for the theory scales as

$$\mu_f \equiv 2^{x_f} \times \mu_0 \text{ and } \mu_r \equiv 2^{x_r} \times \mu_0 \tag{6.23}$$

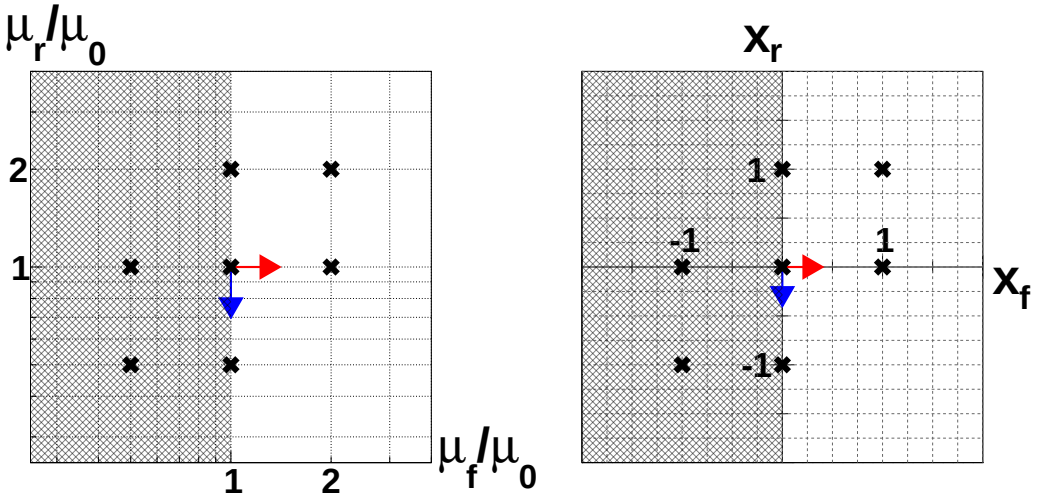


Figure 6.8: The coordinates (μ_f, μ_r) (left) and (x_f, x_r) (right). The black cross marks indicate the conventional seven scale sets listed in Eq.(2.61). The cross sections increase in the directions of the arrows. The shadowed area refers to where CTEQ6.6 PDF is not fully validated with $\mu_0 = \sqrt{m_c^2 + p_T^2}$ and $m_c \geq 1.3$ GeV, which was excluded for the χ^2 scan.

to vary the scales by the power of 2. Figure 6.8 illustrates the coordinates (μ_f, μ_r) in the left panel and the coordinates (x_f, x_r) in the right panel. The origin of the coordinates is defined by the conventional central scales, $\mu_f = \mu_0$ and $\mu_r = \mu_0$, and all the black points represent the conventional seven scale sets listed in Eq.(2.61). For the extrapolation, the CTEQ6.6 PDF was taken as a proxy of the PROSA_VFNS PDF again. However in the application of PDF sets to this extrapolation via LHAPDF [131], it was observed that the numerical results are instable at scales lower than Q_{min} of each PDF set. For example, Q_{min} is defined to be 1.3 GeV [132] for the CTEQ6.6 PDF set, which means that I could not fully trust the results in the range $\mu_f < 1.3$ GeV due to observed numerical instability. Thus for this thesis, I considered only $\mu_f \geq 1.3$ GeV where the application is fully validated down to $p_T = 0$ GeV with $m_c = 1.3$ GeV (the minimum m_c considered in the extrapolation):

$$\mu_f = 2^{x_f} \sqrt{(m_c = 1.3\ [\mathrm{GeV}])^2 + (p_T = 0\ [\mathrm{GeV}])^2} \geq 1.3\ [\mathrm{GeV}]. \tag{6.24}$$

In other words, the best scales were determined in $\mu_f \geq 1.3$ GeV.

The first study [28] was introduced to D^0 measurements at $\sqrt{s} = 5$ TeV in pp collisions. The D^0 measurements at $\sqrt{s} = 5$ TeV are available from the ALICE [9], CMS [20] and LHCb [16] experiments, of which the kinematic ranges are listed in Table 6.4. However the ALICE and CMS measurements have overlapping cross-sections within $|y| < 1$. For the integrated cross section over $|y| < 1$, the CMS measurement covers $\sim 40\%$ while the ALICE measurement covers $\sim 50\%$ with much better precision. Furthermore, the contribution of the CMS measurement in the range $36 < p_T < 100$ GeV is negligible for the total charm cross section. In other words, the single ALICE measurement covers already the maximum of the cross section in the range $|y| < 1$. Therefore, the ALICE and LHCb measurements only were considered in this thesis.

The D^0 measurements were extrapolated using the modified FONLL (Eq.(6.21)) where the p_T dependent production fraction is given by $\tilde{f}_{D^0}$ (Eq.(6.2)). However, since this was performed as the first study for this thesis, I extrapolated the measurements also using the original FONLL (Eq.(2.59)) which is fully based on the universality assumption of the charm fragmentation. This could provide a direct comparison to the previous LHC measurements which were still based on the universality assumption and detach a cross check of the extrapolation scheme from

ALICE	$\|y\| < 0.5$	$0 < p_T < 36$ GeV
CMS	$\|y\| < 1.0$	$2 < p_T < 100$ GeV
LHCb	$2.0 < y < 2.5$	$0 < p_T < 10$ GeV
	$2.5 < y < 3.0$	$0 < p_T < 10$ GeV
	$3.0 < y < 3.5$	$0 < p_T < 10$ GeV
	$3.5 < y < 4.0$	$0 < p_T < 9$ GeV
	$4.0 < y < 4.5$	$0 < p_T < 6$ GeV

Table 6.4: The kinematic ranges covered by the ALICE [9], CMS [20] and LHCb [16] experiments for D^0 measurements at $\sqrt{s} = 5$ TeV.

introducing the non-universality fragmentation (see the next section). For an application of the universality assumption, $f_{D^0}^{uni} = 0.562 \pm 0.016$ (refer to Table 6.1) was used as the fragmentation fraction in the original FONLL (Eq.(2.59)).

The 2-dimensional χ^2 tables were produced with 5 different fixed $x_f = 0$, 0.25, 0.5, 0.75 and 1, and 4 different fixed $m_c = 1.3$, 1.5, 1.7 and 1.9 GeV. Additional constraints were introduced to the scales and α_K such that $\frac{1}{5} \lesssim \mu_f/\mu_r \lesssim 5$[‡] and $\alpha_K \lesssim 30$[§]. As a result, 17(18) χ^2 tables in total were produced with the $\tilde{f}_{D^0}(f_{D^0}^{uni})$, which can be found in Appendix G. As a summary, the best parameters and uncertainty parameters are shown in Figure 6.9, where the least χ^2 results of 3-dimensional fits with μ_f, μ_r and α_K (i.e., the *local least* χ^2 results) were projected onto 2-dimensional coordinates (μ_f, μ_r) with the fixed m_c. Then, the best parameters, which were determined by the least χ^2 of 4-dimensional fits (i.e., the *global least* χ^2), are marked by a star. The *uncertainty parameters*, which were determined by $\Delta\chi^2 \sim 1\sigma$, are marked by an additional outer circle[¶]. For the purpose of this study, the uncertainties were determined only for the non-universal fragmentation applied.

After the first study with the D^0 measurements at $\sqrt{s} = 5$ TeV, similar studies were done also with D^0 measurements at $\sqrt{s} = 13$ TeV from ALICE [15] and LHCb [18], D^{*+} measurements at $\sqrt{s} = 7$ TeV from CMS (the previous chapter) and LHCb [17], and D^{*+} measurements at $\sqrt{s} = 0.9$ TeV from CMS (again the previous chapter[‖]). The kinematic ranges covered by each experiment are listed in Table 6.5 for the 13 TeV data and Table 6.6 for the 7 TeV data. To extrapolate the 7 and 0.9 TeV data, $\tilde{f}_{D^{*+}}$ determined at $\sqrt{s} = 5$ TeV was used.

In analogy with the 5 TeV D^0 measurements, 2-dimensional χ^2 tables were produced with 5 different fixed $x_f = 0$, 0.25, 0.5, 0.75 and 1, and 4 different fixed $m_c = 1.3$, 1.5, 1.7 and 1.9 GeV, but only with the non-universal fragmentation applied. Additional constraints for the scales and α_K were introduced as done in the 5 TeV study. With these constraints, 16 tables in total were provided with the 7 TeV data. In the case of 13 TeV, 5 more tables were produced with $m_c = 2.1$ GeV and one more table with $x_f = 0.6$ to determine the boundary of the uncertainty parameters. In total 22 tables were produced with the 13 TeV data in the end. The corresponding χ^2 results are summarized in Appendix G. A summary for the best and uncertainty parameters are shown again in Figure 6.9.

[‡]Outside this range, it was observed that the numerical results are instable.

[§]The maximum of α_K was chosen rather arbitrarily, where the χ^2 results start to be strongly uncorrelated with α_K variation, and so that it is not too large compared to what has been measured from experiments.

[¶]Here, $x_f = 0.5$ and 1 were taken as the uncertainty boundary with $m_c = 1.7$ GeV even though the χ^2s are outside $\Delta\chi^2$, since its total charm cross sections are within the final χ^2 uncertainty in the end.

[‖]Here, the 0.9 TeV CMS data used for the extrapolation are already outdated with respect to what is shown in the previous chapter, but the difference is not significant

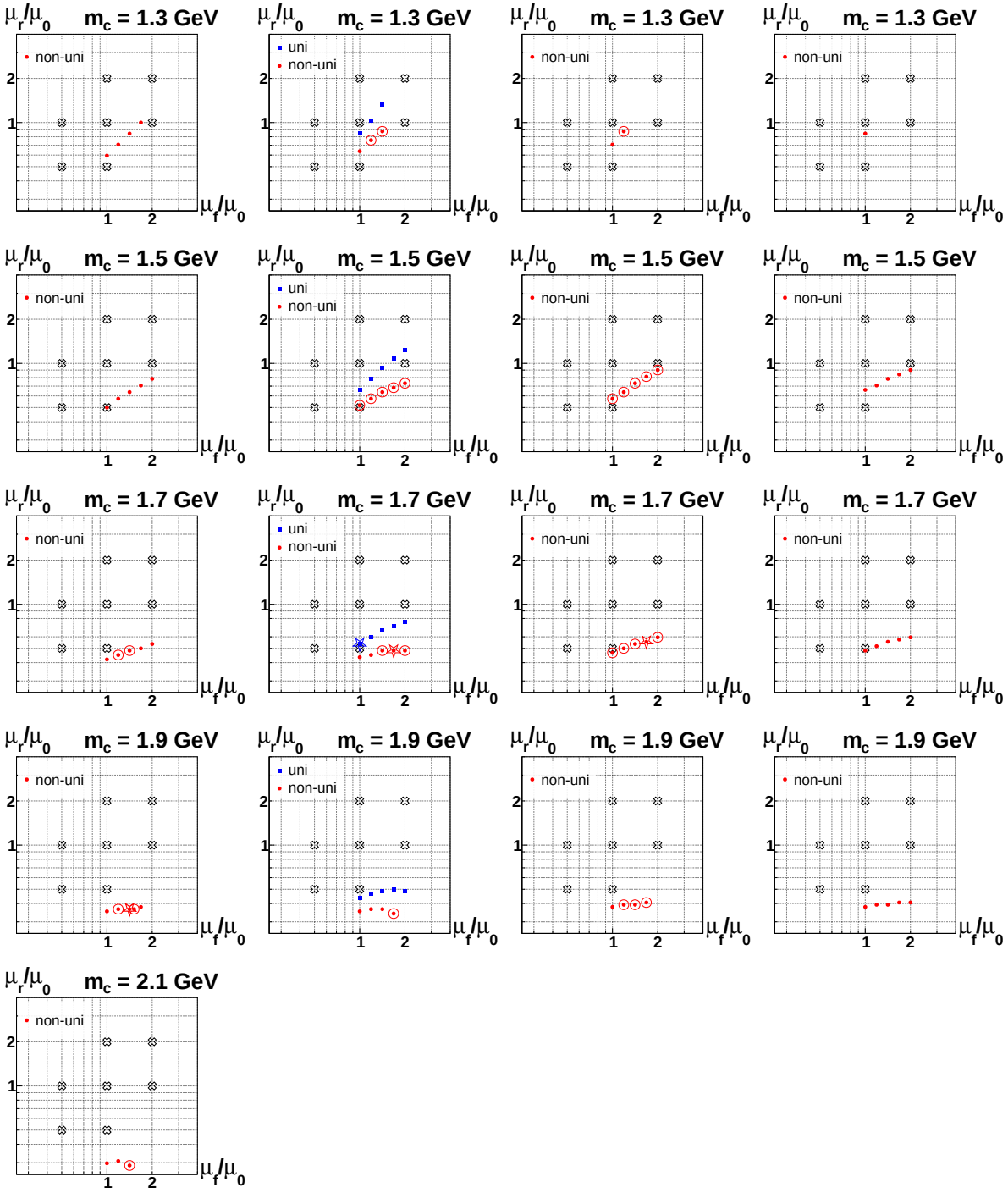


Figure 6.9: The best scales with the D^0 measurements at $\sqrt{s} = 13$ TeV (the first column) and 5 TeV (the second column), and the D^{*+} measurements at $\sqrt{s} = 7$ TeV (the third column) and 0.9 TeV (the fourth column), respectively from the left to right. The local least χ^2 results are shown by the circle/square points, while the global least χ^2 one for each $\sqrt{s}$ is marked by the additional star. The uncertainty scales are marked by the additional outer circles. The cross marks indicate the conventional scale sets for theory.

ALICE	$\|y\| < 0.5$	$p_T > 0$ GeV
LHCb	$2.0 < y < 2.5$	$0 < p_T < 15$ GeV
	$2.5 < y < 3.0$	$0 < p_T < 15$ GeV
	$3.0 < y < 3.5$	$0 < p_T < 15$ GeV
	$3.5 < y < 4.0$	$0 < p_T < 11$ GeV
	$4.0 < y < 4.5$	$0 < p_T < 7$ GeV

Table 6.5: The kinematic ranges covered by ALICE [15] and LHCb [18] for D^0 measurements at $\sqrt{s} = 13$ TeV.

CMS	$0.0 < \|y\| < 0.5$	$p_T > 1$ GeV
	$0.5 < \|y\| < 1.0$	$p_T > 1$ GeV
	$1.0 < \|y\| < 1.5$	$p_T > 1$ GeV
	$1.5 < \|y\| < 2.0$	$p_T > 1$ GeV
	$2.0 < \|y\| < 2.5$	$p_T > 2$ GeV
LHCb	$2.0 < y < 2.5$	$3 < p_T < 8$ GeV
	$2.5 < y < 3.0$	$1 < p_T < 8$ GeV
	$3.0 < y < 3.5$	$0 < p_T < 7$ GeV
	$3.5 < y < 4.0$	$0 < p_T < 7$ GeV
	$4.0 < y < 4.5$	$0 < p_T < 5$ GeV

Table 6.6: The kinematic ranges covered by CMS (provided in the previous chapter) and LHCb [17] for D^{*+} measurements at $\sqrt{s} = 7$ TeV.

	$\sqrt{s} = 0.9$ TeV	$\sqrt{s} = 5$ TeV	$\sqrt{s} = 7$ TeV	$\sqrt{s} = 13$ TeV
μ_f/μ_0	1.68 (1.00 - 2.00)	1.68 (1.00 - 2.00)	1.68 (1.00 - 2.00)	1.41 (1.19 - 1.52)
μ_r/μ_0	0.57 (0.38 - 0.90)	0.48 (0.34 - 0.93)	0.55 (0.37 - 1.00)	0.37 (0.29 - 0.48)
m_c [GeV]	1.7 (1.3 - 1.9)	1.7 (1.3 - 1.9)	1.7 (1.3 - 1.9)	1.9 (1.7 - 2.1)
α_K	12 (8 - 25)	9 (6 - 28)	10 (6 - 30)	6 (5 - 9)

Table 6.7: The best parameters used for the ddFONLL parametrization. The parentheses indicate the ranges of the uncertainty parameters, which were used to calculate so-called χ^2 uncertainties of ddFONLL. The parameters at $\sqrt{s} = 7$ TeV were slightly updated with respect to the ones shown in [32], while the total cross section remains the same.

Furthermore at $\sqrt{s} = 0.9$ TeV, 16 tables in total were determined with the constraints, but unfortunately the global least χ^2 cannot be determined within the given ranges of the parameters (overall almost the same χ^2 observed in the 2-dimensional χ^2 tables). This could be due to the large statistical uncertainties and/or since the $p_T < 1$ GeV information is missing. It might be also related with part of PDF effects; μ_f (and m_c since $\mu_f = \sqrt{m_c^2 + p_T^2}$) is the parameter of PDFs and if the CTEQ6.6 PDF (or more generally any PDF set) has less sensitivity to μ_f at lower center-of-mass energy, the convergence of χ^2 with the same μ_f variation with the ones introduced at higher center-of-mass-energies might be relatively slow. But, all these are hypotheses yet, and no further studies are provided in this thesis. Nevertheless, it turns out that the 2-dimensional χ^2 scan results are mutually consistent for the 5, 7 and 13 TeV results (similar patterns are observed in Figure 6.9), and the 0.9 TeV results are given in this thesis by taking the same μ_f and m_c parameters as those determined from the 5 TeV results. Then, all the least χ^2 values of the 2-dimensional χ^2 scan were taken to determine the χ^2 scan uncertainty for the total charm cross section at $\sqrt{s} = 0.9$ TeV.

All the best parameters including the uncertainty ranges are summarized in Table 6.7. The uncertainties turned out to cover well the conventional scales and m_c for QCD theory and the reference α_Ks based on e^+e^- data (refer again to Appendix J). Furthermore, it was observed from this study that there are some correlations between the two theory scales especially with lower m_c (see Figure 6.9).

The ddFONLL parametrization is then defined by $d\sigma_{H_c}^{\text{FONLL with } \tilde{f}}(\mu_f^b, \mu_r^b, m_c^b, \alpha_K^b)$, where μ_f^b, μ_r^b, m_c^b and α_K^b are the best parameters. The ddFONLL parametrization and the measurements are compared as a function of p_T and $|y|$ in Appendix H with uncertainties of $\tilde{f}$, χ^2 scan (i.e., μ_f, μ_r, m_c, and α_K uncertainties), and the PDFs. The results with the total uncertainties are shown in Figure 6.10 and 6.11 for the 5 TeV D^0, Figure 6.12 and 6.13 for the 7 TeV D^{*+}, Figure 6.14 and 6.15 for the 13 TeV D^0, and Figure 6.16 and 6.17 for the 0.9 TeV D^{*+}. Note that ddFONLL describes the data well overall in the full phase space, which is consistent with the assumption of the fragmentation rapidity independence (Assumption 2).

As a cross check, the ddFONLL parametrization for the Λ_c^+ spectrum was compared with the ALICE measurements at $\sqrt{s} = 5$ and 13 TeV. The Λ_c^+ ddFONLL parametrization was derived with the best parameters of the extrapolation of the D^0 measurements at $\sqrt{s} = 5$ or 13 TeV, but applying $\tilde{f}_{\Lambda_c^+}$ (the blue points in Figure 6.5) instead of $\tilde{f}_{D^0}$, which is shown by the pink band in Figure 6.18 and 6.19. The original FONLL theory (the blue band) which is based on the universality assumption totally disagrees with the measurements in the Λ_c^+ comparisons, while, the ddFONLL parametrization describes both the D^0 and Λ_c^+ measurements well, as it should do by construction.

A further cross check for Assumption 1 and 2 can be given by comparing the ddFONLL

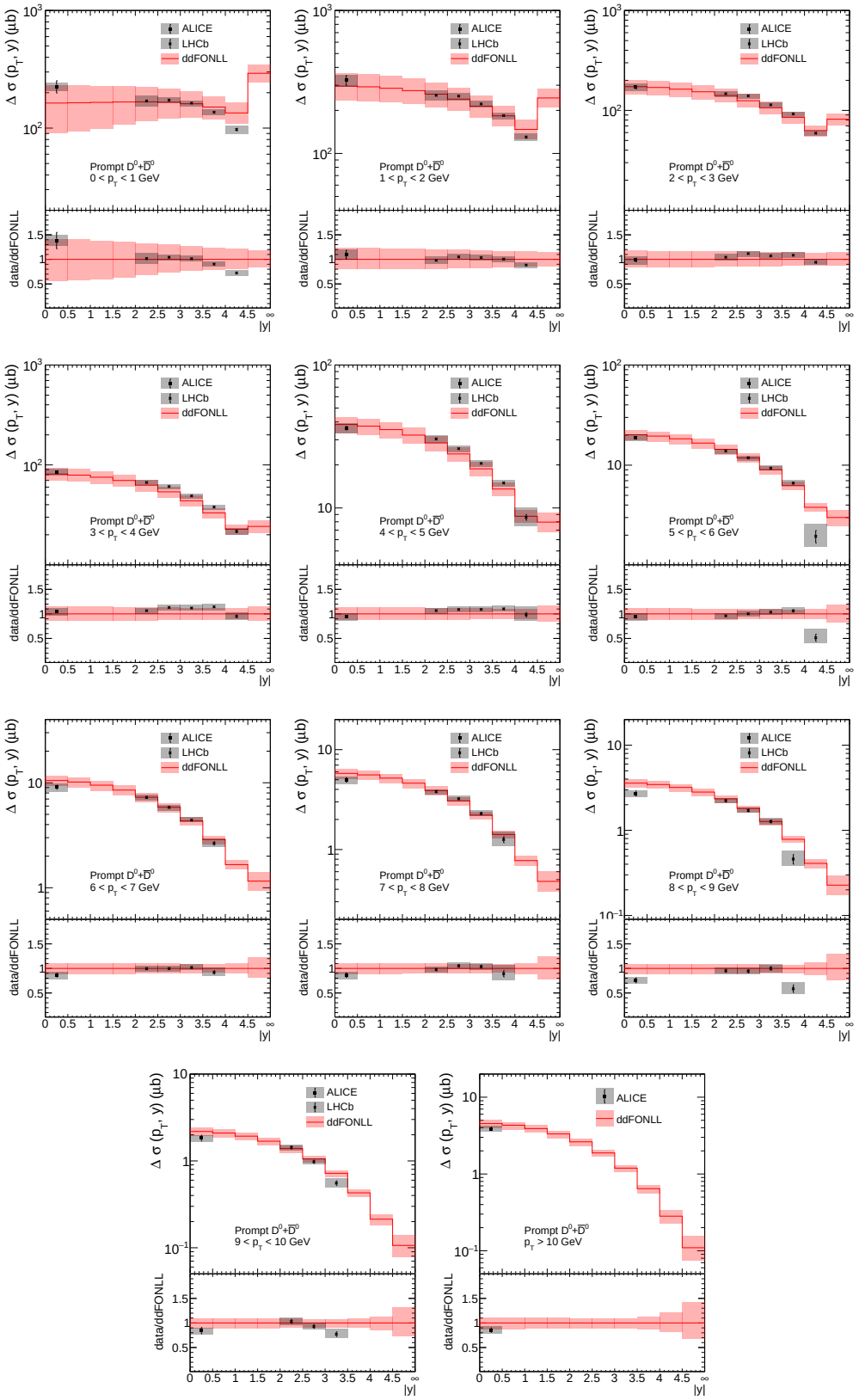


Figure 6.10: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

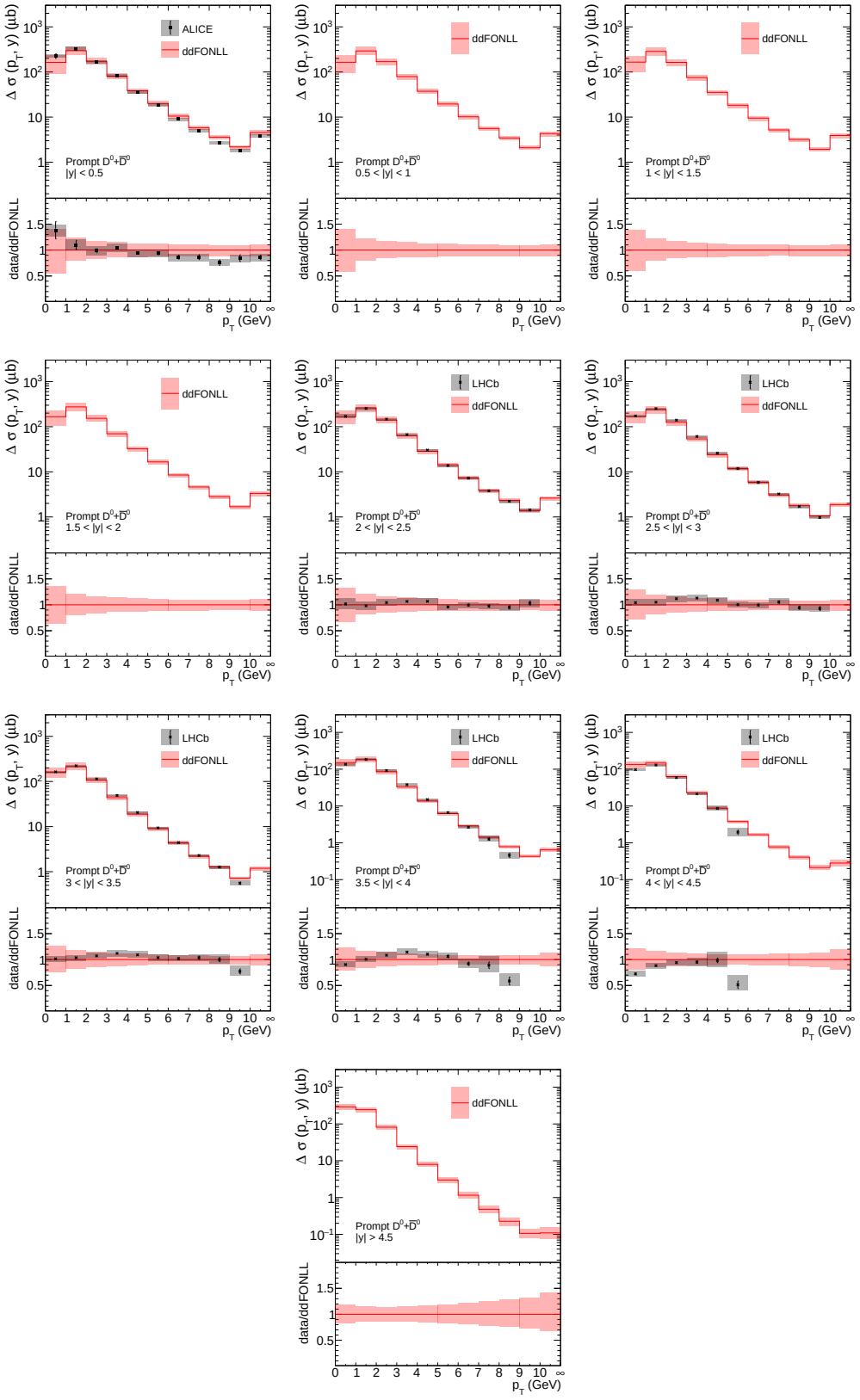


Figure 6.11: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f} \oplus \chi^2$).

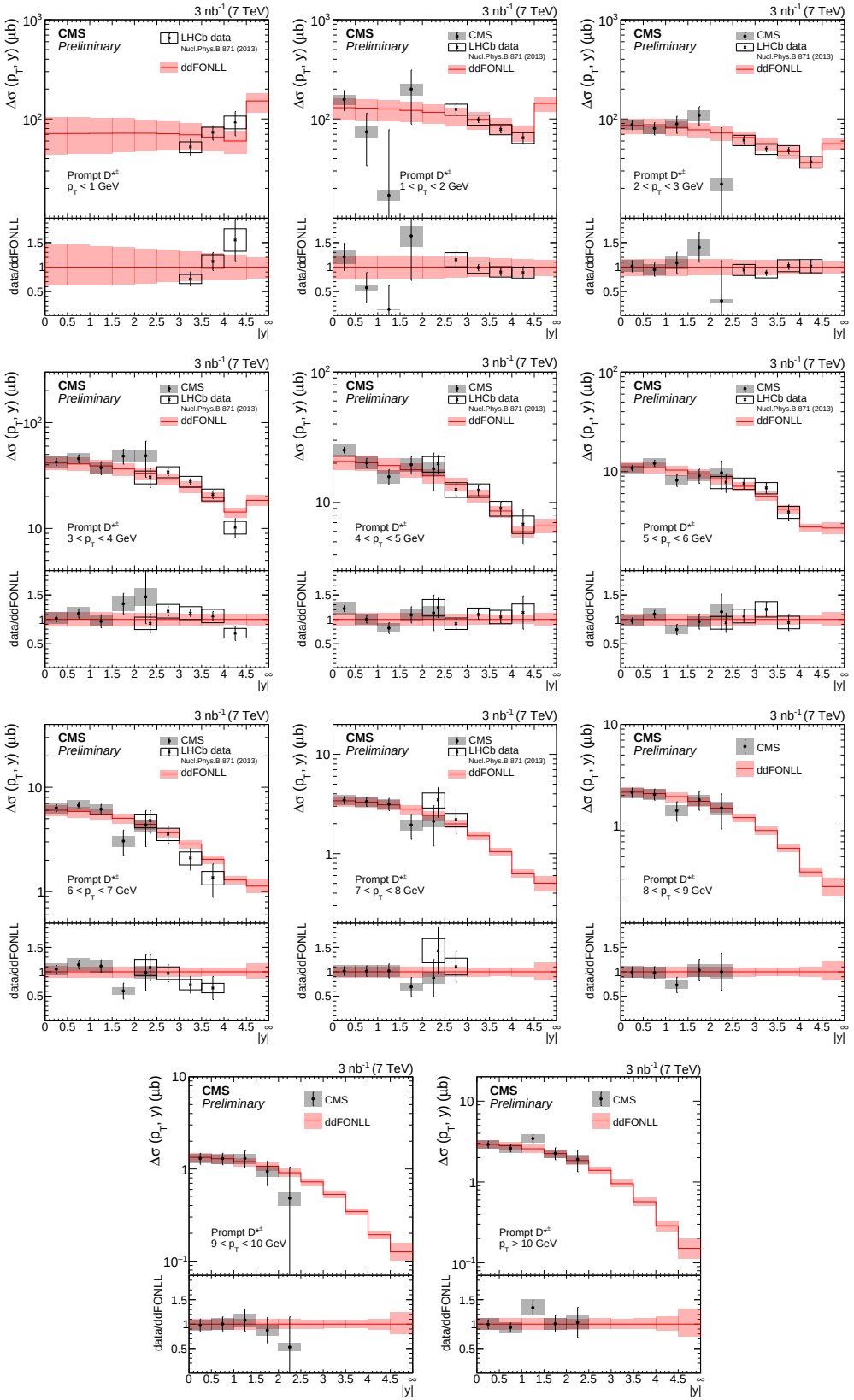


Figure 6.12: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of $|y|$. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

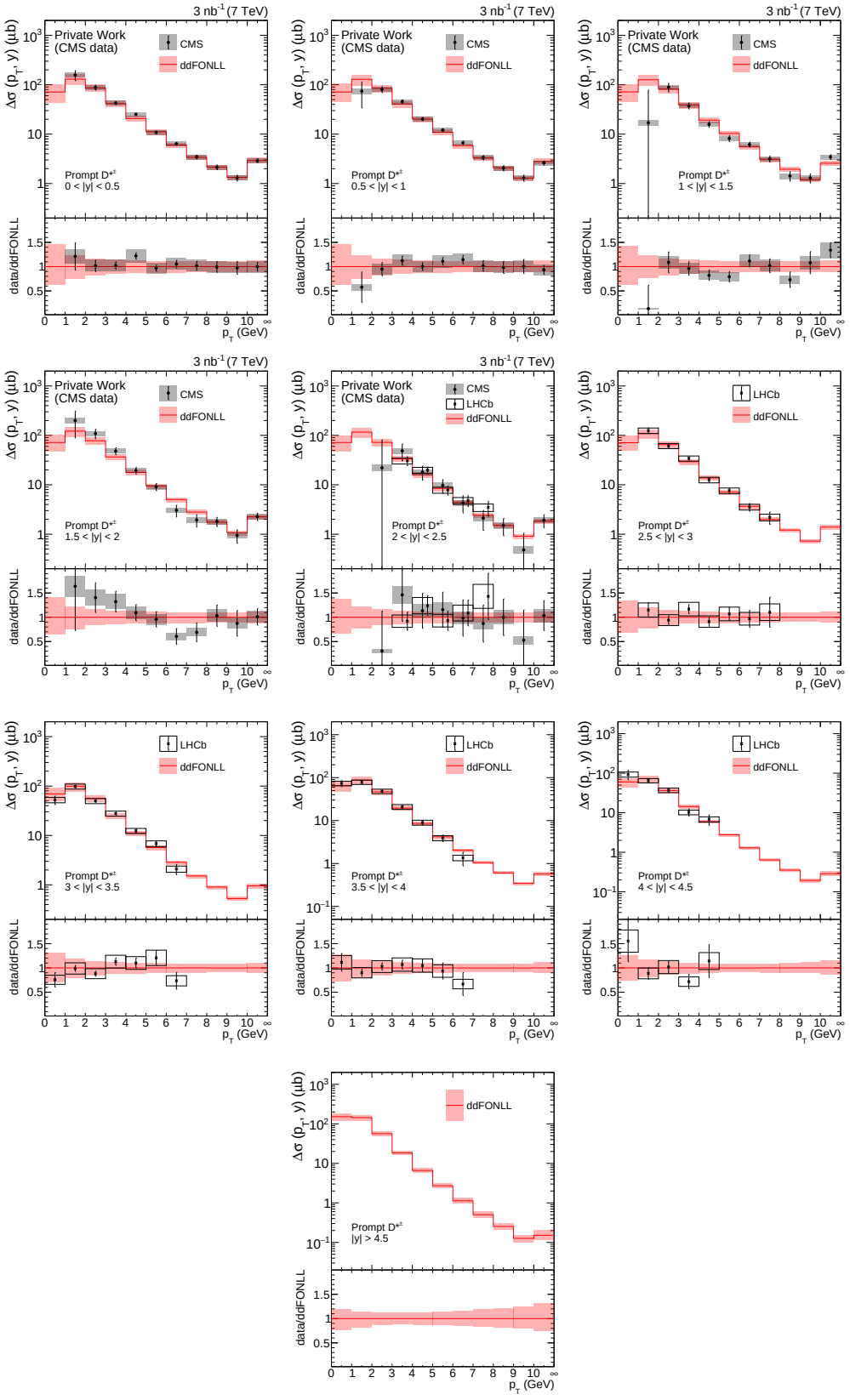


Figure 6.13: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

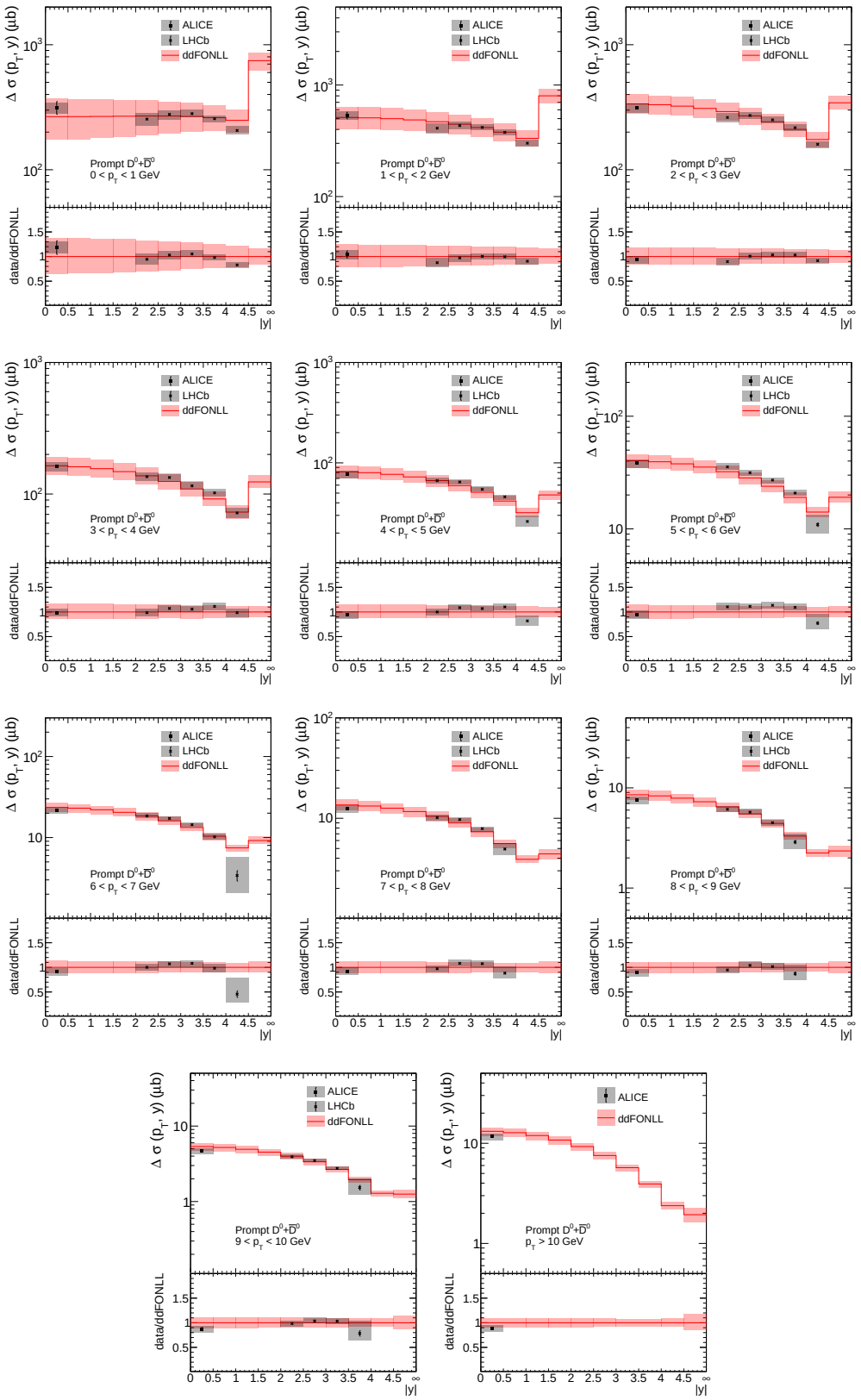


Figure 6.14: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of $|y|$. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

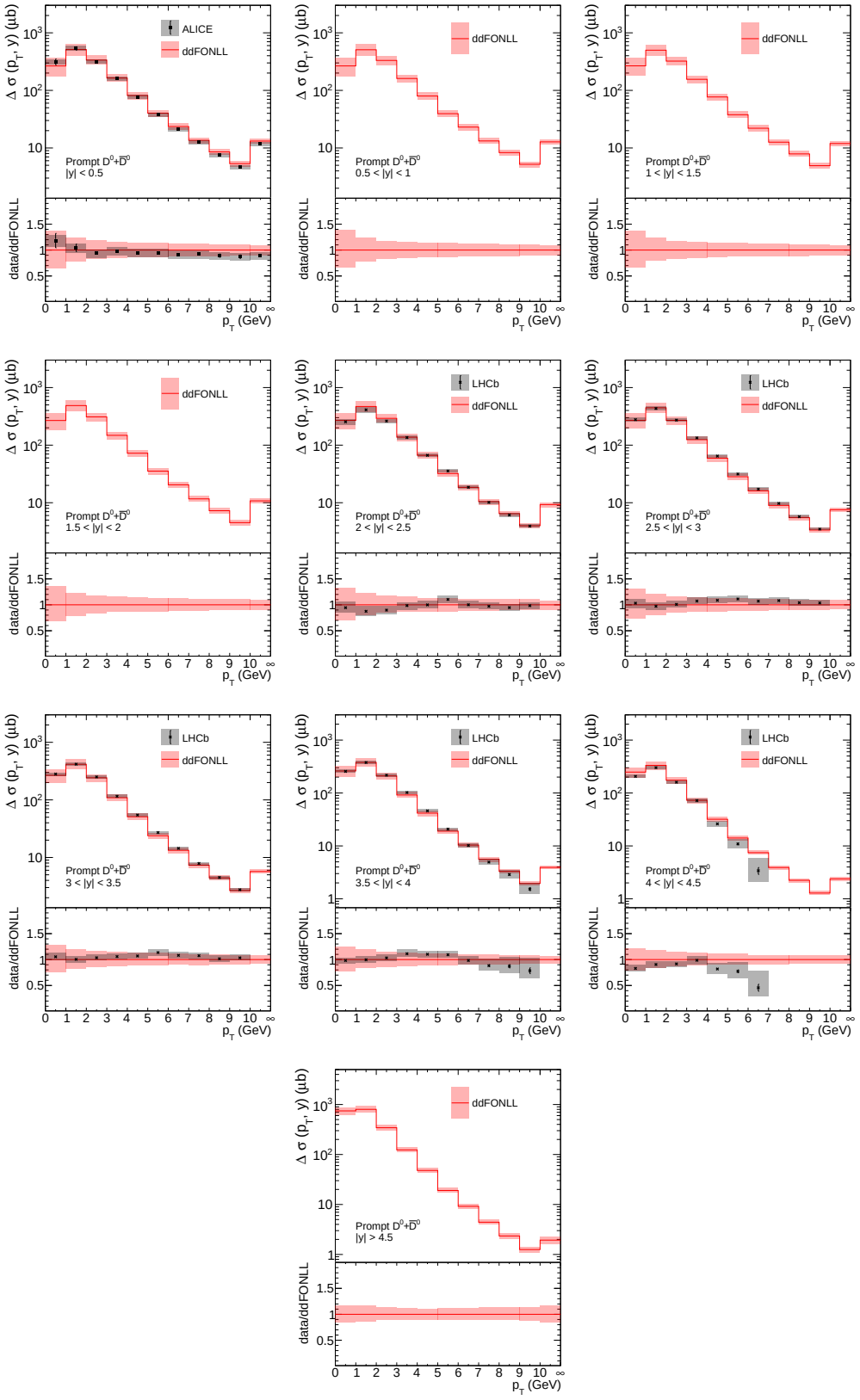


Figure 6.15: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f} \oplus \chi^2$).

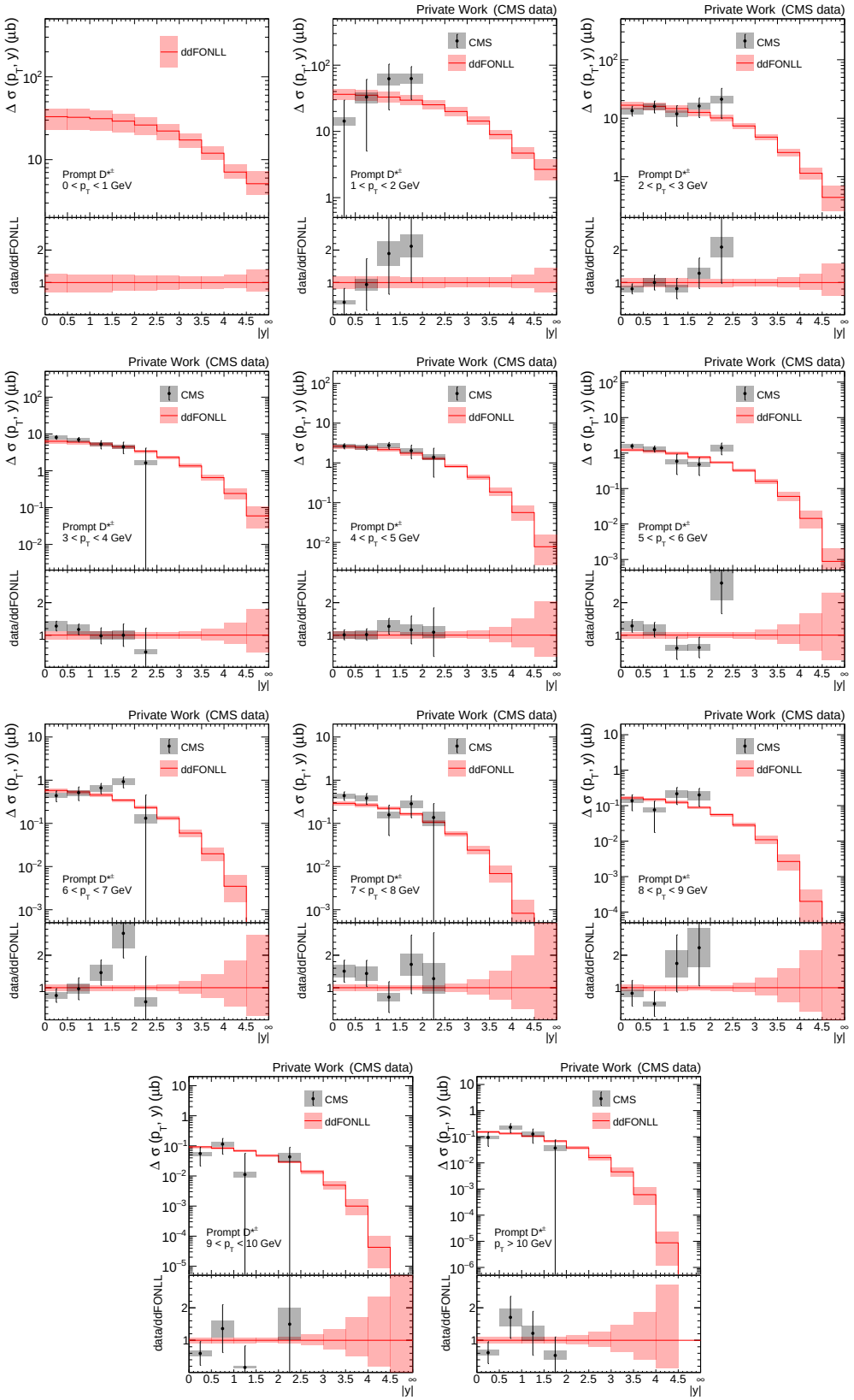


Figure 6.16: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of $|y|$. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

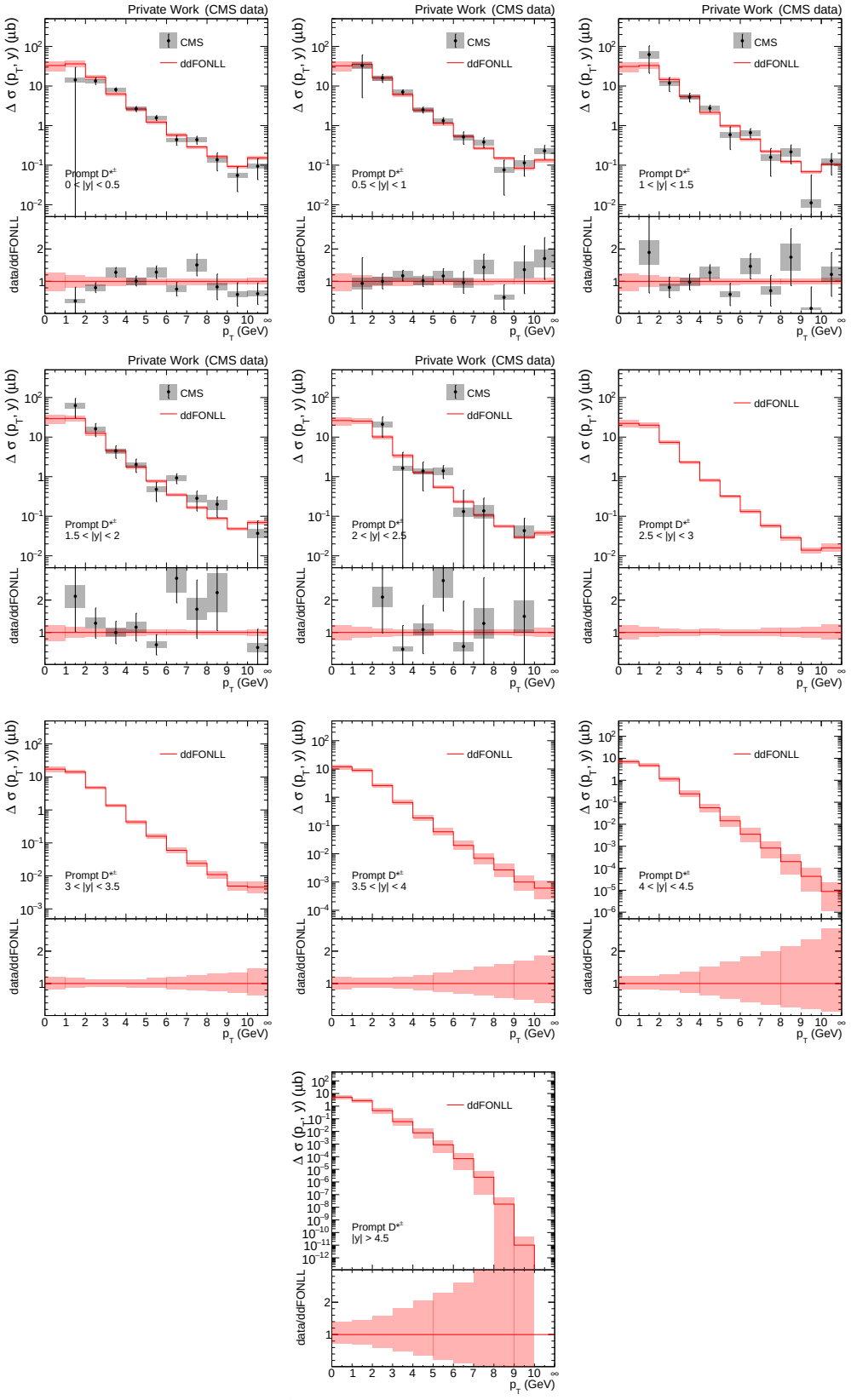


Figure 6.17: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The red bands of ddFONLL show the total uncertainty (CTEQ6.6 PDF $\oplus$ $\tilde{f}$ $\oplus$ χ^2).

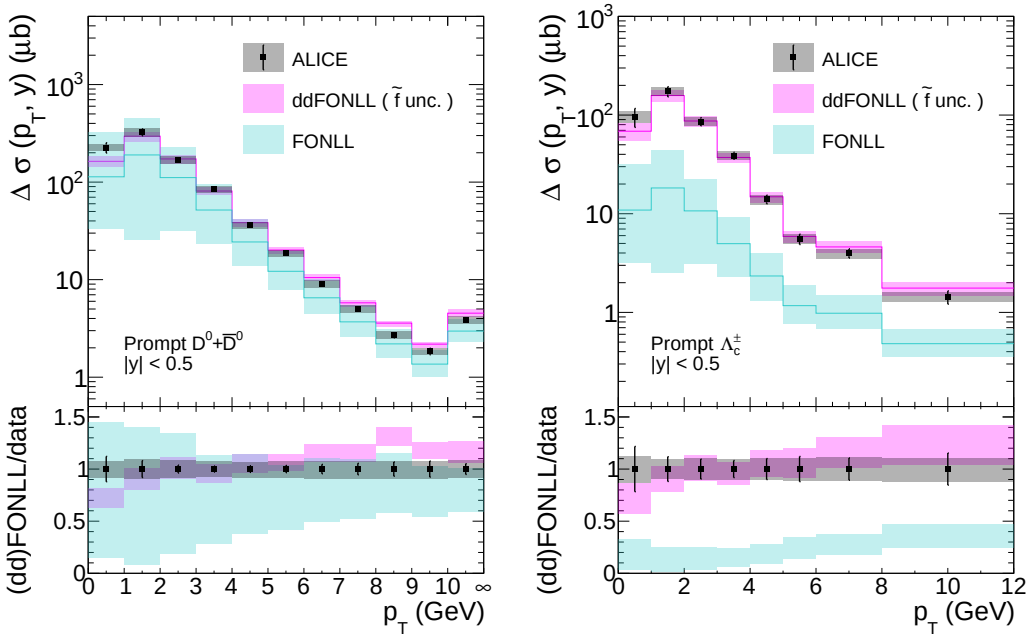


Figure 6.18: $D^0 + \overline{D}^0$ (left) and $\Lambda_c^{\pm}$ (right) cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL parametrization with $\tilde{f}$ uncertainty (the pink band) describes both the D^0 and Λ_c^+ data well.

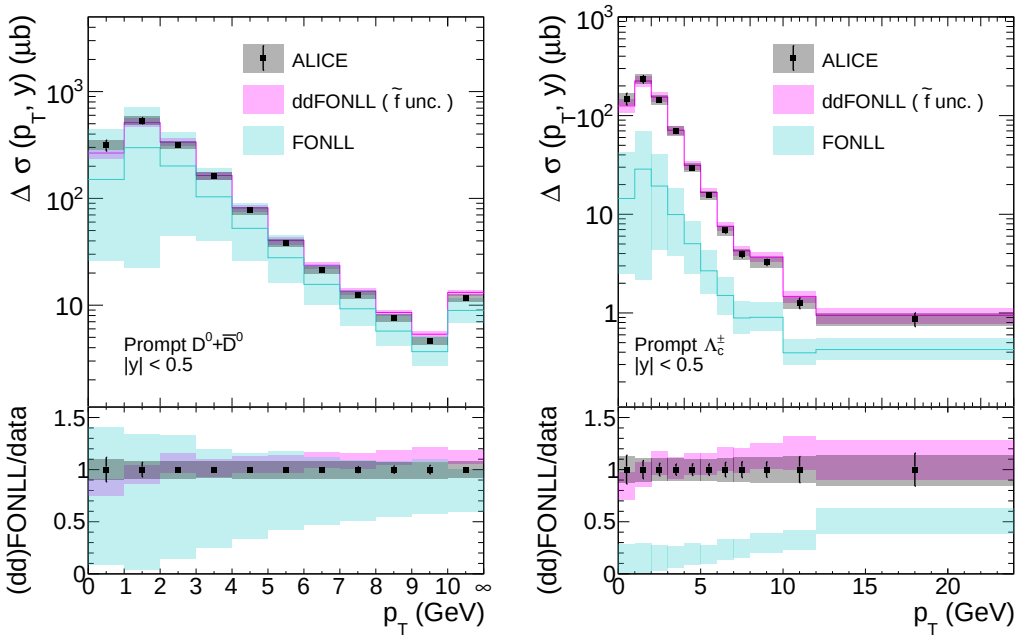


Figure 6.19: $D^0 + \overline{D}^0$ (left) and $\Lambda_c^{\pm}$ (right) cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The ddFONLL parametrization with $\tilde{f}$ uncertainty (the pink band) describes both the D^0 and Λ_c^+ data well.

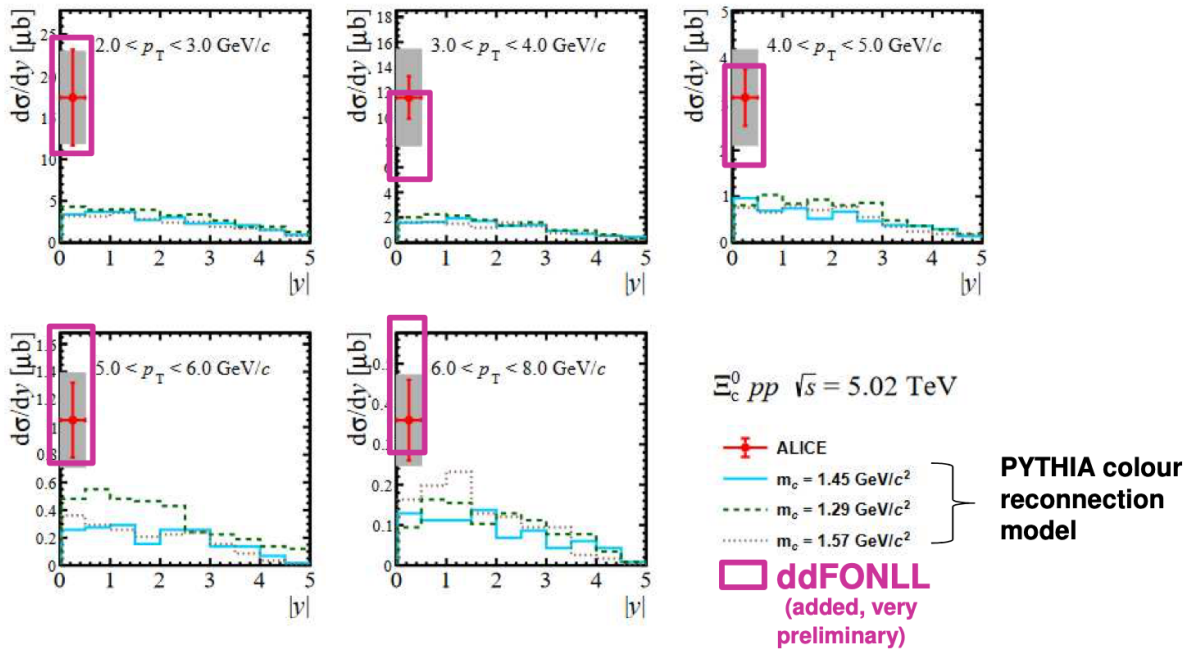


Figure 6.20: Ξ_c^0 cross sections as a function of $|y|$, with figures taken from [133]. The figures were adapted from [29].

parametrization also with the Ξ_c^0 measurement, e.g., at $\sqrt{s} = 5$ TeV, from ALICE [87]. However, since the fragmentation fraction is not measured yet for Ξ_c^0, Ξ_c^+ and Ω_c^0 from e^+e^- (or ep) collisions, $\tilde{f}_{\Xi_c^0}$ cannot be defined by taking direct measurements. Thus, a conservative ratio is introduced as

$$\frac{\Xi_c^0 + \Xi_c^+ + \Omega_c^0}{\Lambda_c^+} = 1.0 \pm 0.4 \tag{6.25}$$

which covers all points of $(\Xi_c + \Omega_c)/\Lambda_c^+$ in Figure 6.3. Then, the ddFONLL parametrization of the 5 TeV Λ_c^+ shown in Figure 6.18 was rescaled to derive the ddFONLL parametrization for Ξ_c^0, assuming strangeness suppression and isospin symmetry: $\Omega_c^0/\Xi_c^0 = \Xi_c^0/\Lambda_c^+$ and $\Xi_c^0 = \Xi_c^+$. These assumptions result in consistency with the measurements from ALICE [15]. Eventually in Figure 6.20, the Ξ_c^0 measurements were compared with ddFONLL. It shows that the ddFONLL parametrization describes the Ξ_c^0 measurement as well.

6.3 Total charm cross section

The total H_c cross section was determined by taking all measurements where available and ddFONLL was taken for the non-measured kinematic ranges:

$$\sigma_{H_c} = \Delta\sigma_{H_c}^{\text{data}}(\text{measured phase space}) + \Delta\sigma_{H_c}^{\text{ddFONLL}}(\text{unmeasured phase space}). \tag{6.26}$$

Then the total H_c cross section was divided by the fragmentation fraction of H_c measured from pp collisions (denoted by $f_{H_c}^{pp}$) to derive the total charm cross section:

$$\sigma_{c\bar{c}} = \frac{\sigma_{H_c}}{f_{H_c}^{pp}}. \tag{6.27}$$

At the moment of writing this thesis, $f_{H_c}^{pp}$ was measured only at $\sqrt{s} = 5$ and 13 TeV from ALICE, of which the latest numbers can be found all in [15].

Essentially, the fragmentation fraction should be measured from experiments and cannot be determined from theory. Although, as a cross check of the application of ddFONLL, I introduced

		[GeV]	$\Delta\sigma_{D^0+\overline{D^0}}$ [mb]
ALICE	$0.0 < \|y\| < 0.5$	$0 < p_T < 36$	$0.88^{+0.08}_{-0.08}$
LHCb	$2.0 < \|y\| < 2.5$	$0 < p_T < 10$	$0.70^{+0.06}_{-0.05}$
	$2.5 < \|y\| < 3.0$	$0 < p_T < 10$	$0.68^{+0.04}_{-0.04}$
	$3.0 < \|y\| < 3.5$	$0 < p_T < 10$	$0.59^{+0.03}_{-0.03}$
	$3.5 < \|y\| < 4.0$	$0 < p_T < 9$	$0.48^{+0.03}_{-0.03}$
	$4.0 < \|y\| < 4.5$	$0 < p_T < 6$	$0.32^{+0.02}_{-0.02}$
	Σ	Σ	$2.75^{+0.17}_{-0.17}$
ALICE+LHCb			$3.64^{+0.19}_{-0.19}$
ddFONLL			$2.95^{+0.31}_{-0.33}(\tilde{f})^{+0.52}_{-0.44}(\text{PDF})^{+0.10}_{-0.09}(\chi^2)$

$\sigma_{c\bar{c}}$ [mb]	$8.43^{+0.25}_{-0.25}(\text{data})^{+0.40}_{-0.42}(\tilde{f})^{+0.67}_{-0.56}(\text{PDF})^{+0.13}_{-0.12}(\chi^2)^{+0.65}_{-0.88}(f^{pp})$
	$8.43^{+1.05}_{-1.16}(\text{total})$

Table 6.8: The integrated $D^0 + \overline{D}^0$ cross section ($\Delta\sigma_{D^0+\overline{D^0}}$) and the total charm cross section ($\sigma_{c\bar{c}}$) at $\sqrt{s} = 5$ TeV with $f^{pp}_{D^0} = 0.391^{+0.030}_{-0.041}$.

a fragmentation fraction defined by ddFONLL as

$$f^{\text{ddFONLL}}_{H_c} = \frac{\sigma^{\text{FONLL with } \tilde{f}}_{H_c}(\mu^b_f, \mu^b_r, m^b_c, \alpha^b_K)}{\sigma^{\text{FONLL}}_{c}(\mu^b_f, \mu^b_r, m^b_c)}, \tag{6.28}$$

where σ^{FONLL}_c is the total charm cross section obtained by integrating Eq.(6.21) with $\tilde{f} = 1$ over the full phase space, but with the best parameters determined for the ddFONLL parametrization. If Assumption 1 and 2 are reasonable enough, then $f^{\text{ddFONLL}}_{H_c}$ should be consistent with the measured $f^{pp}_{H_c}$. This is not trivial since $f^{\text{ddFONLL}}_{H_c}$ is derived under the assumptions by the p_T-dependent Λ_c^+-to-D^0 ratios measured from pp collisions and the fragmentation fractions measured from e^+e^- collisions, while $f^{pp}_{H_c}$ is derived by measuring all the weakly-decaying ground states except a few like Ξ_c^+ at $\sqrt{s} = 5$ TeV and Ω_c^0 (counted in the measurement uncertainties) at $\sqrt{s} = 5$ and 13 TeV. Thus an additional backward cross checks will be provided with Eq.(6.28) compared to the measured $f^{pp}_{H_c}$ in the following.

6.3.1 Total charm cross section measurements at 5, 7, 13 and 0.9 TeV

Eventually, the $\sigma_{c\bar{c}}$s were determined at $\sqrt{s} = 5$, 7, 13 and 0.9 TeV with Eq.(6.26) and (6.27), of which results can be found in Table 6.8, 6.9, 6.10 and 6.11. In each table, the integrated cross sections of $H_c + \overline{H}_c$ for data and ddFONLL also are shown, of which the sum gives σ_{H_c}, dividing by 2 to average particle and anti-particle state. For the 5, 7 and 0.9 TeV results, the fragmentation fractions were taken from the ALICE measurements at $\sqrt{s} = 5$ TeV, which are $0.391^{+0.030}_{-0.041}$ and $0.155^{+0.043}_{-0.022}$ for D^0 and D^{*+}, respectively [25]**. However, the fragmentation fraction for D^{*+} has very large uncertainties especially for the upper band ($\sim 28\%$). Thus, since it is shown in Figure 6.7 that D^{*+}/D^0 is consistent between pp and e^+e^- collisions, $f^{uni}_{D^{*+}}/f^{uni}_{D^0}$

**Here the fragmentation fractions at $\sqrt{s} = 5$ TeV are still based on the earlier measurements [25] rather than the latest ones in [15].

		[GeV]	$\Delta\sigma_{D^{*+}+D^{*-}}$ [mb]
CMS 1	$\lvert y\rvert < 0.5$	$p_T > 1$	$0.34^{+0.05}_{-0.05}$
	$0.5 < \lvert y\rvert < 1.0$	$p_T > 1$	$0.25^{+0.05}_{-0.05}$
	$1.0 < \lvert y\rvert < 1.5$	$p_T > 1$	$0.18^{+0.07}_{-0.07}$
	$1.5 < \lvert y\rvert < 2.0$	$p_T > 1$	$0.40^{+0.12}_{-0.12}$
CMS 2	$2.0 < \lvert y\rvert < 2.5$	$2 < p_T < 3$ & $p_T > 8$	$0.03^{+0.06}_{-0.06}$
CMS 3	$2.0 < \lvert y\rvert < 2.5$	$3 < p_T < 8$	$0.08^{+0.02}_{-0.02}$
LHCb 1	$2.0 < \lvert y\rvert < 2.5$	$3 < p_T < 8$	$0.07^{+0.01}_{-0.01}$
LHCb 2	$2.5 < \lvert y\rvert < 3.0$	$1 < p_T < 8$	$0.25^{+0.04}_{-0.04}$
	$3.0 < \lvert y\rvert < 3.5$	$0 < p_T < 7$	$0.25^{+0.03}_{-0.03}$
	$3.5 < \lvert y\rvert < 4.0$	$0 < p_T < 7$	$0.24^{+0.03}_{-0.03}$
	$4.0 < \lvert y\rvert < 4.5$	$0 < p_T < 5$	$0.21^{+0.04}_{-0.04}$
CMS 1 + CMS 2			$1.19^{+0.21}_{-0.21}$
CMS 2 + CMS 3			$0.11^{+0.06}_{-0.06}$
CMS 1 + CMS 2 + CMS 3			$1.28^{+0.22}_{-0.22}$
LHCb 2			$0.94^{+0.12}_{-0.12}$
LHCb 1 + LHCb 2			$1.01^{+0.13}_{-0.13}$
CMS 1 + CMS 2 + LHCb 1 + LHCb 2 (Data 1)			$2.20^{+0.25}_{-0.25}$
CMS 1 + CMS 2 + CMS 3 + LHCb 2 (Data 2)			$2.22^{+0.25}_{-0.25}$
ddFONLL			$0.94^{+0.10}_{-0.11}(\tilde{f})^{+0.21}_{-0.17}(\text{PDF})^{+0.10}_{-0.14}(\chi^2)$

$\sigma_{c\bar{c}}$ [mb] (Data 1)	$9.34^{+0.74}_{-0.74}(\text{data})^{+0.31}_{-0.32}(\tilde{f})^{+0.64}_{-0.51}(\text{PDF})^{+0.30}_{-0.41}(\chi^2)^{+0.82}_{-1.06}(f^{pp})$ $9.34^{+1.35}_{-1.48}(\text{total})$
$\sigma_{c\bar{c}}$ [mb] (Data 2)	$9.39^{+0.74}_{-0.74}(\text{data})^{+0.31}_{-0.32}(\tilde{f})^{+0.64}_{-0.51}(\text{PDF})^{+0.30}_{-0.41}(\chi^2)^{+0.83}_{-1.07}(f^{pp})$ $9.39^{+1.35}_{-1.49}(\text{total})$

Table 6.9: The integrated $D^{*\pm}$ cross section ($\Delta\sigma_{D^{*+}+D^{*-}}$) and the total charm cross section ($\sigma_{c\bar{c}}$) at $\sqrt{s} = 7$ TeV with $f^{pp}_{D^{*+}} = 0.168^{+0.015}_{-0.019}$.

		[GeV]	$\Delta\sigma_{D^0+\overline{D^0}}$ [mb]
ALICE	$0.0 < \|y\| < 0.5$	$0 < p_T < 50$	$1.50^{+0.14}_{-0.14}$
LHCb	$2.0 < \|y\| < 2.5$	$0 < p_T < 15$	$1.20^{+0.12}_{-0.11}$
	$2.5 < \|y\| < 3.0$	$0 < p_T < 15$	$1.25^{+0.09}_{-0.09}$
	$3.0 < \|y\| < 3.5$	$0 < p_T < 15$	$1.18^{+0.07}_{-0.07}$
	$3.5 < \|y\| < 4.0$	$0 < p_T < 11$	$1.04^{+0.07}_{-0.06}$
	$4.0 < \|y\| < 4.5$	$0 < p_T < 7$	$0.78^{+0.06}_{-0.06}$
	Σ	Σ	$5.44^{+0.40}_{-0.38}$
ALICE+LHCb			$6.94^{+0.43}_{-0.41}$
ddFONLL			$6.38^{+0.52}_{-0.60}(\tilde{f})^{+1.12}_{-0.93}(\text{PDF})^{+0.18}_{-0.14}(\chi^2)$
$\sigma_{c\bar{c}}$ [mb]	$17.43^{+0.56}_{-0.53}(\text{data})^{+0.69}_{-0.78}(\tilde{f})^{+1.47}_{-1.22}(\text{PDF})^{+0.24}_{-0.18}(\chi^2)^{+1.19}_{-2.05}(f^{pp})$ $17.43^{+2.10}_{-2.57}(\text{total})$		

Table 6.10: The integrated $D^0 + \overline{D}^0$ cross section ($\Delta\sigma_{D^0+\overline{D^0}}$) and the total charm cross section ($\sigma_{c\bar{c}}$) at $\sqrt{s} = 13$ TeV with $f^{pp}_{D^0} = 0.382^{+0.026}_{-0.045}$.

		[GeV]	$\Delta\sigma_{D^{*+}+D^{*-}}$ [mb]
CMS	$0.0 < \|y\| < 0.5$	$p_T > 1$	$0.04^{+0.02}_{-0.02}$
	$0.5 < \|y\| < 1.0$	$p_T > 1$	$0.06^{+0.03}_{-0.03}$
	$1.0 < \|y\| < 1.5$	$p_T > 1$	$0.08^{+0.04}_{-0.04}$
	$1.5 < \|y\| < 2.0$	$p_T > 1$ (excluding $9 < p_T < 10$)	$0.09^{+0.04}_{-0.04}$
	$2.0 < \|y\| < 2.5$	$2 < p_T < 10$ (excluding $8 < p_T < 9$)	$0.03^{+0.01}_{-0.01}$
	Σ	Σ	$0.30^{+0.08}_{-0.08}$
ddFONLL			$0.31^{+0.04}_{-0.04}(\tilde{f})^{+0.05}_{-0.05}(\text{PDF})^{+0.02}_{-0.03}(\chi^2)$
$\sigma_{c\bar{c}}$ [mb]	$1.83^{+0.24}_{-0.24}(\text{data})^{+0.10}_{-0.11}(\tilde{f})^{+0.16}_{-0.14}(\text{PDF})^{+0.06}_{-0.08}(\chi^2)^{+0.16}_{-0.21}(f^{pp})$ $1.83^{+0.35}_{-0.37}(\text{total})$		

Table 6.11: The integrated $D^{*\pm}$ cross section ($\Delta\sigma_{D^{*+}+D^{*-}}$) and the total charm cross section ($\sigma_{c\bar{c}}$) at $\sqrt{s} = 0.9$ TeV with $f^{pp}_{D^{*+}} = 0.168^{+0.015}_{-0.019}$.

was taken to translate $f^{pp}_{D^0} = 0.391^{+0.030}_{-0.041}$ into $f^{pp}_{D^{*+}}$ instead of taking the direct measurement. Assuming the uncertainties are fully uncorrelated, it turned out to be $f^{pp}_{D^{*+}} = 0.168^{+0.015}_{-0.019}$ which was then used as the fragmentation fraction for the total charm cross section at $\sqrt{s} =$ 7 and 0.9 TeV. The fragmentation fraction of D^0 at $\sqrt{s} = 13$ TeV was taken from the ALICE measurements at $\sqrt{s} = 13$ TeV: $f^{pp}_{D^0} = 0.382^{+0.026}_{-0.045}$ [15].

In total 5 different uncertainties were determined for the total charm cross section: data, $\tilde{f}$, PDFs, χ^2 and $f^{pp}_{H_c}$ uncertainty. The data uncertainties were calculated by treating statistical uncertainties as fully uncorrelated and systematic uncertainties as fully correlated for each experiment, while both were treated as fully uncorrelated to the other experiment. Eventually the uncertainties were given as the sum of statistical and systematic uncertainties in quadrature. The $\tilde{f}$, PDFs and χ^2 uncertainties were propagated from ddFONLL. Then the total uncertainty was calculated by treating all the individual uncertainites as fully uncorrelated. All the uncertainties can be found also in Table 6.8, 6.9, 6.10 and 6.11.

Based on the results shown in Table 6.8, 6.9, 6.10 and 6.11, the extrapolation factors turn out to be 1.4, 1.8, 1.9 and 2.0 at $\sqrt{s} = 7$, 5, 13 and 0.9 TeV, respectively. The CMS measurements including the larger rapidity measurements from LHCb at $\sqrt{s} = 7$ TeV resulted in the smallest extrapolation factor ever achieved at the LHC. The 0.9 TeV results were derived from the CMS measurements alone for the time being in the absence of any other LHC measurement and thus 2.0 is already the smallest extrapolation factor at $\sqrt{s} = 0.9$ TeV. If the CMS measurements at $\sqrt{s} = 13$ TeV shown in the previous chapter are included, the extrapolation factor 1.9 at $\sqrt{s} =$ 13 TeV will be reduced much more, so that the smallest factor could be achieved again.

$f^{\text{ddFONLL}}_{H_c}$s which were calculated using Eq.(6.28), turned out to be 0.390 and 0.401 at $\sqrt{s} =$ 5 and 13 TeV ($H_c = D^0$), respectively, and 0.168 both at $\sqrt{s} = 7$ and 0.9 TeV ($H_c = D^{*+}$). All these results are indeed consistent with the fragmentation fractions measured from ALICE [15, 25].

Additionally, as mentioned above, I derived the total charm cross section at $\sqrt{s} = 5$ TeV based on the universality assumption for a direct comparison to the previous LHC measurements and to distinguish the effects of applying the non-universal charm fragmentation in the extrapolation. The determined parametrization with the universality assumption was shown as a function of p_T and $|y|$ in Appendix H with the $f^{uni}_{D^0}$ uncertainty. No dedicated uncertainties for PDFs and the χ^2 fit were provided in this case. Then the total charm cross section at $\sqrt{s} =$ 5 TeV was determined to be $5.84^{+0.17}_{-0.17}(\text{data})^{+0.17}_{-0.17}(f^{uni})$ mb. If uncertainties of PDFs and the χ^2 fit would be more or like similar to $\sim 7\%$ and $\sim 1.5\%$ of the non-universality case, respectively, then an estimated total uncertainty would be $\sim 8\%$. This indeed gives consistent results with $5.25^{+0.35}_{-0.26}$ mb from [24], which was provided by extrapolating the ALICE and LHCb measurements also based on the universality assumption. The estimated uncertainty from this study is comparable to the reference one. This might indicate that the relatively large total uncertainty in the non-universal fragmentation results comes from the relatively large uncertainty of the fragmentation fraction from *pp* collisions compared to e^+e^- collisions. On the other hand, it shows also that the total charm cross section indeed increases significantly as expected with the decreased fragmentation fraction in *pp* collisions.

6.3.2 Comparison with QCD prediction and constraints on m_c and PDF

The total charm cross sections measured by extrapolating the D^0 cross sections at $\sqrt{s} = 5$ and 13 TeV and the D^{*+} cross sections at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation treatment are compared to NNLO QCD predictions with various PDF sets in Figure 6.21. The predictions provided in [24] are based on the complete NNLO theory introduced in Section 2.4. All the measurements show good agreement with the upper bands of the theoretical uncertainties. Compared to the earlier measurements under the universality assumption of the fragmentation, which are shown by the blue bands in Figure 6.21, the total charm cross sec-

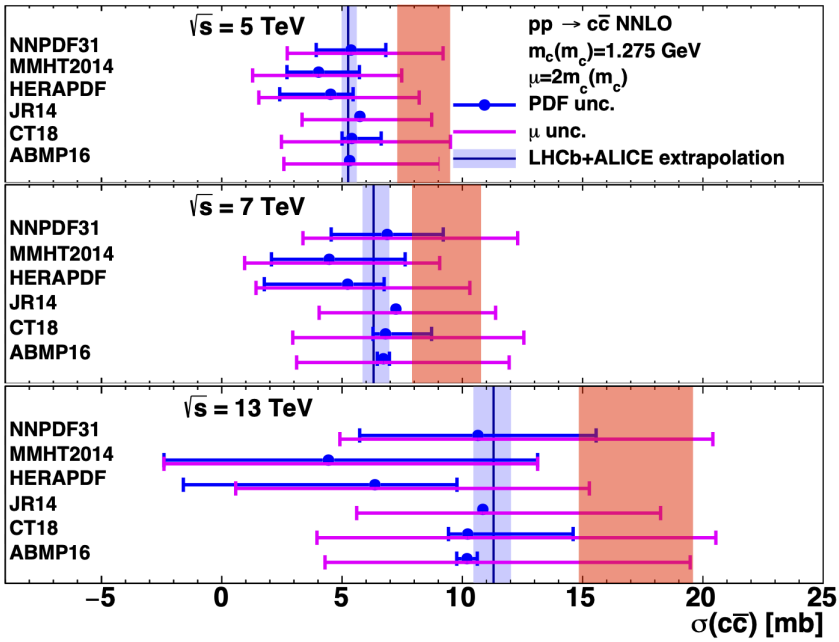


Figure 6.21: The total charm cross sections at $\sqrt{s} = 5$ (top), 7 (middle) and 13 TeV (bottom), with figures adapted from [24]. The vertical red bands are the total charm cross sections provided in this thesis.

tions at all the three center-of-mass-energies are increased significantly with the non-universal fragmentation treatment. No prediction is available in [24] for the total charm cross section at $\sqrt{s} = 0.9$ TeV.

In the meantime, the total charm cross section measurements at $\sqrt{s} = 5$, 13 and 0.9 TeV were taken to constrain the QCD parameters, specifically, the $\overline{\text{MS}}$ charm mass $m_c(m_c)$ and the PDFs at low x. The first preliminary results were kindly provided by Oleksandr Zenaiev who is one of the authors of [24]. The total charm cross sections as a function of the center-of-mass energy are shown in Figure 6.22. The data points at $\sqrt{s} = 5$ and 13 TeV were taken from Table 6.8 and 6.10. On the other hand, 1.67 ± 0.23 mb, which was determined purely based on ddFONLL and the 5 TeV extrapolation, was assigned as the 0.9 TeV point temporarily as a proxy of the measurement[††]. These three data points resulted in the charm mass to be constrained by $m_c(m_c) = 1.065(1.005)$ GeV with uncertainties up to +0.284(0.258) GeV and down to -0.383(0.086) GeV depending on the data, PDFs and theory scale uncertainties, using ABMP16_3_nnlo [134](MSHT20nnlo_nf3 [135]) PDF. Individual uncertainties can be found in Table 6.12. Examples of constraints on two PDF sets are shown in Figure 6.23. The most significant effects of including the total charm cross section measurements especially at low x were observed with the MSHT20nnlo_nf3 PDF.

6.3.3 Total charm cross section measurement as a function of $\sqrt{s}$

The total charm-pair cross section was measured as a function of $\sqrt{s}$, which is shown in Figure 6.24. The data points in the figure are from Section 6.3.1. This is the first measurement of the total charm-pair cross section as a function of $\sqrt{s}$ with the non-universal charm fragmentation, covering from $\sqrt{s} = 0.9$ TeV to 13 TeV. For a comparison to QCD predictions, the NNLO prediction with the ABMP16_3_nnlo PDF (shown in the upper panel of Figure 6.22) is also added in the figure. The measurements show good agreement with the QCD prediction up to the highest order known today.

[††]The CMS result at $\sqrt{s} = 0.9$ TeV is still internal at the moment of writing this thesis

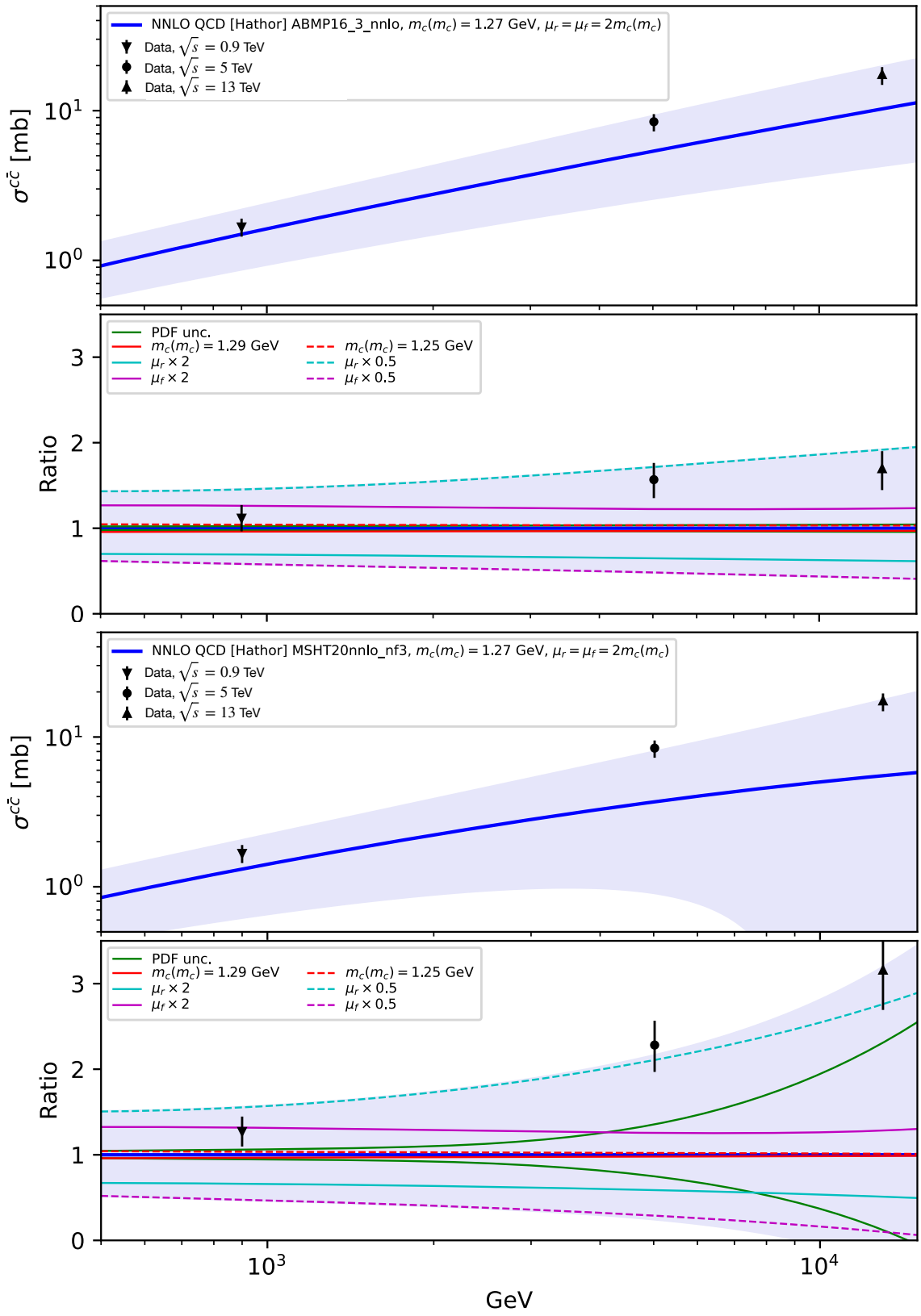


Figure 6.22: The total charm cross section as a function of the center-of-mass energy ($\sqrt{s}$). The data points at $\sqrt{s} = 5$ and 13 TeV from Table 6.8 and 6.10, while the 0.9 TeV point was determined purely based on ddFONLL and the 5 TeV extrapolation. The QCD predictions (the blue bands) were calculated with the perturbative theory up to NNLO and with ABMP16_3_nnlo (upper) and MSHT20nnlo_nf3 (lower) PDF sets.

	ABMP16_3_nnlo	MSHT20nnlo_nf3
$m_c(m_c)$	1.065	1.005
data uncertainties	$\pm$0.043	$\pm$0.023
PDF uncertainties	$\pm$0.017	$\pm$0.086
$(\mu_f/\mu_0, \mu_r/\mu_0) = (0.5, 0.5)$	+0.178	+0.250
$(\mu_f/\mu_0, \mu_r/\mu_0) = (1, 0.5)$	+0.284	+0.237
$(\mu_f/\mu_0, \mu_r/\mu_0) = (0.5, 1)$	+0.046	+0.635
$(\mu_f/\mu_0, \mu_r/\mu_0) = (2, 1)$	+0.123	-0.022
$(\mu_f/\mu_0, \mu_r/\mu_0) = (1, 2)$	-0.383	+0.258
$(\mu_f/\mu_0, \mu_r/\mu_0) = (2, 2)$	-0.021	+0.005

Table 6.12: Constraints on the charm mass. $m_c(m_c)$ indicates the charm mass determined with the central value of data and PDFs, and with the theory scales $(\mu_f/\mu_0, \mu_r/\mu_0) = (1, 1)$ where $\mu_0 = 2m_c(m_c)$. The others show the maximum and/or minimum variations of $m_c(m_c)$ within the data and PDF uncertainties or with the theory scales variations. Units are [GeV].

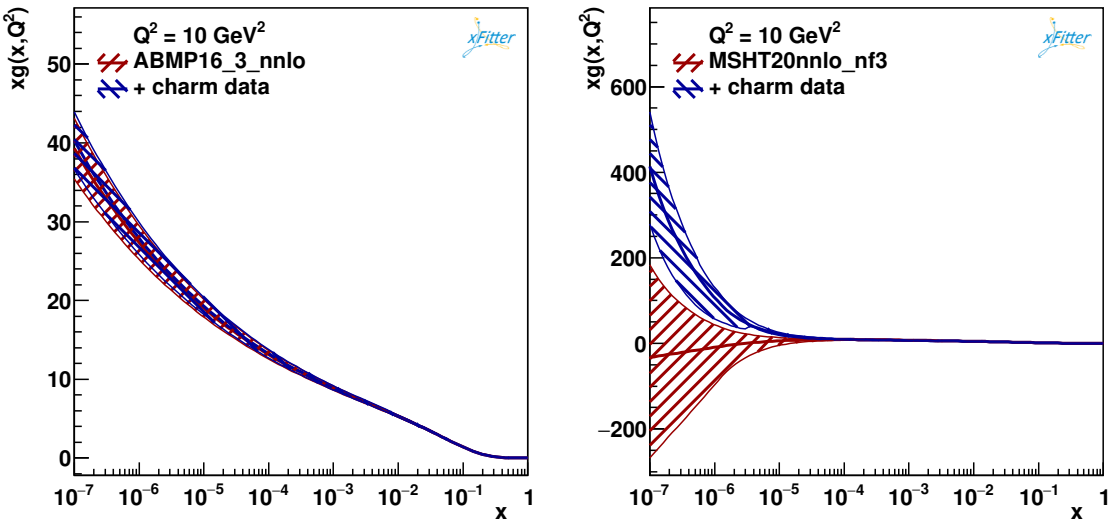


Figure 6.23: Effects of including the total charm cross section measurements (labeled by "+ charm data") to the gluon PDF of ABMP16_3_nnlo (left) and MSHT20nnlo_nf3 (right).

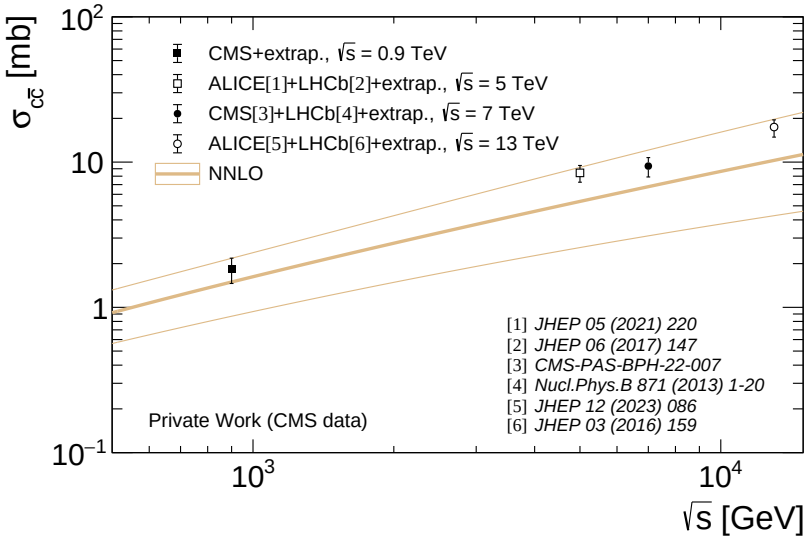


Figure 6.24: Total charm-pair cross section measurement as a function of $\sqrt{s}$.

7 | Conclusion

This thesis presented the reconstructions of $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ and $D^0 \to K^-\pi^+$ in the CMS detector at $\sqrt{s}$ = 7, 0.9 and 13 TeV, taking the largest phase space possible. The D^{*+} reconstruction at $\sqrt{s}$ = 7 TeV was performed on the MinimumBias and Next-to-MinimumBias collisions of the 2010 data with special low p_T tracking. By taking over the previous study from N. Z. Jomhari [119], further detailed studies to complete the systematics were provided in this thesis. To determine the non-prompt contamination, the theory-inspired mothod was introduced to normalize Pythia to data and extract the beauty fractions from the normalized Pythia predictions. The likelihood function of the MC-constrained fit was introduced to determine the systematics of the subtraction method, especially for the poor statistics bins where the unconstrained fit gives no reasonable results. The overall pileup systematics was determined by taking differences of the reconstruction efficiencies on the pileup and no-pileup sample, and its era dependence also was evaluated by calculating the cross sections in the 7 different subsamples of the data. As a result, the complete systematic uncertainty was presented in this thesis.

Following a similar analysis strategy, the D^{*+} reconstruction at $\sqrt{s} = 0.9$ TeV was performed on the (Zero-)Minimum-Bias collisions on the 2021/22 (Run 3) data and 2010 (Run 1) data. The very preliminary reconstruction on the early Run 3 data, which was presented as the CMS detector performance in [83], was kindly taken over by L. Della Penna to provide the cross section measurements. In this thesis, the cross section measurements on the Run 3 data were briefly introduced and the updated results by adding the Run 1 data can also be found in Appendix A.

The D^{*+} reconstruction at $\sqrt{s}$ = 13 TeV was performed on the 2018 B parking data. The Next-to-MinimumBias collisions were separated from the Trigger-Vertex collisions on the data and used for this analysis, resulting in about 1000 times larger statistics than the 7 and 0.9 TeV analysis. As introduced in the 7 TeV analysis, the subtraction method was used to extract the D^{*+} signal. The very first attempt to measure the D^0 cross sections at $\sqrt{s} = 13$ TeV also was presented. To extract the D^0 signal, the MC template fit was introduced for the backgrounds. The non-prompt contamination was temporarily determined based on the theory-inspired method. The π_s^+ tracking efficiency measurements also were introduced at $\sqrt{s} = 13$ TeV, of which the very first results can be found in Appendix C.

The prompt D^{*+} cross sections were measured in $p_T > 1$ GeV and $|y| < 2.5$ excluding $1 < p_T < 2$ GeV and $2 < |y| < 2.5$, with the complete(partial) systematic uncertainty at $\sqrt{s} =$ 7 TeV(13 TeV and 0.9 TeV). The 7 and 13 TeV cross sections show good agreement with the ALICE and LHCb measurements in $|y| < 0.5$ and $2 < |y| < 2.5$, respectively. The 0.9 TeV cross sections were presented as the first charm production measurements ever made at $\sqrt{s} = 0.9$ TeV. These measurements agree with the upper edge of the uncertainty band of the FONLL prediction which, however, is based on the universality assumption.

The total charm-pair cross sections were measured by extrapolating these fiducial cross sections into the full phase space. For the extrapolation, a phenomenological method was introduced to constrain the μ_f, μ_r, m_c and α_K parameters in the FONLL theory by the data, such that the so-called ddFONLL parametrization was derived by reducing the original FONLL theory uncertainty. To treat the non-universal charm fragmentation properly, the p_T dependent

production fractions were derived by directly taking the most precisely measured baryon-to-meson ratios represented by the Λ_c^+/D^0 as a function of p_T. The p_T dependent production fraction was then applied to modify the FONLL prediction without need to assume any particular non-universal fragmentation model. This method was used to present the total charm-pair cross section measurement for the first time with the non-universal charm fragmentation in [28].

Additionally, in this thesis, it was shown that the ddFONLL parametrization describes not only the D^{*+} and D^0 but also the Λ_c^+ and Ξ_c^0 data well, while the FONLL prediction or any other particular model cannot. For the time being, the ddFONLL parametrization is the only one which can be applied for all the weakly-decaying ground states in pp collisions.

Eventually using this ddFONLL parametrization, the D^{*+} cross section measurements at $\sqrt{s} = 7$ and 0.9 TeV from CMS in $|y| < 2.5$ were taken to measure the total charm-pair cross section at $\sqrt{s} = 7$ and 0.9 TeV, respectively. In the case of 7 TeV, the larger rapidity measurements from LHCb in $2.5 < |y| < 4.5$ also were taken, and the resulting extrapolation factor, 1.4, is the smallest factor ever achieved for charm at the LHC. The total charm-pair cross section was measured to be $9.40^{+1.35}_{-1.49}$ mb at $\sqrt{s} = 7$ TeV. The CMS measurements at $\sqrt{s} = 0.9$ TeV resulted in the extrapolation factor 2.0 which is the first one introduced at the LHC for $\sqrt{s} = 0.9$ TeV. The total charm-pair cross section turned out to be $1.83^{+0.35}_{-0.37}$ mb at $\sqrt{s} = 0.9$ TeV. This will be updated with better precision by including the Run 1 data.

This thesis presented also the total charm-pair cross section measurement as a function of the LHC center-of-mass energy. For this, the additional 5 and 13 TeV data were provided with extrapolating the D^0 cross section measurements from ALICE and LHCb, resulting in $8.43^{+1.05}_{-1.16}$ mb at $\sqrt{s} = 5$ TeV and $17.43^{+2.10}_{-2.57}$ mb at $\sqrt{s} = 13$ TeV. Comparing to the NNLO QCD prediction, the total charm-pair cross section measurement shows good agreement with the upper edge of the theoretical uncertainty band. This measurement can be used to constrain the QCD parameters. Already the very first example was kindly provided by O. Zenaiev, where the 5, 13 and 0.9 TeV data were taken to constrain the charm mass and the low-x region of ABMP16_3_nnlo and MSHT20nnlo_nf3 PDFs, of which results were briefly shown in this thesis.

All the total cross sections presented in this thesis supersede the earlier LHC measurements which were derived under the universality assumption.

A Prompt D^{*+} Cross Sections at 0.9 TeV on Run 1 and 3 Data

Prompt D^{*+} cross sections at $\sqrt{s} = 0.9$ TeV shown in Figure 5.34 were updated with combining the 2010 data, and are shown in Figure A.1. The total charm cross section at $\sqrt{s} = 0.9$ TeV presented in Chapter 6 is still based on Run 3 data only.

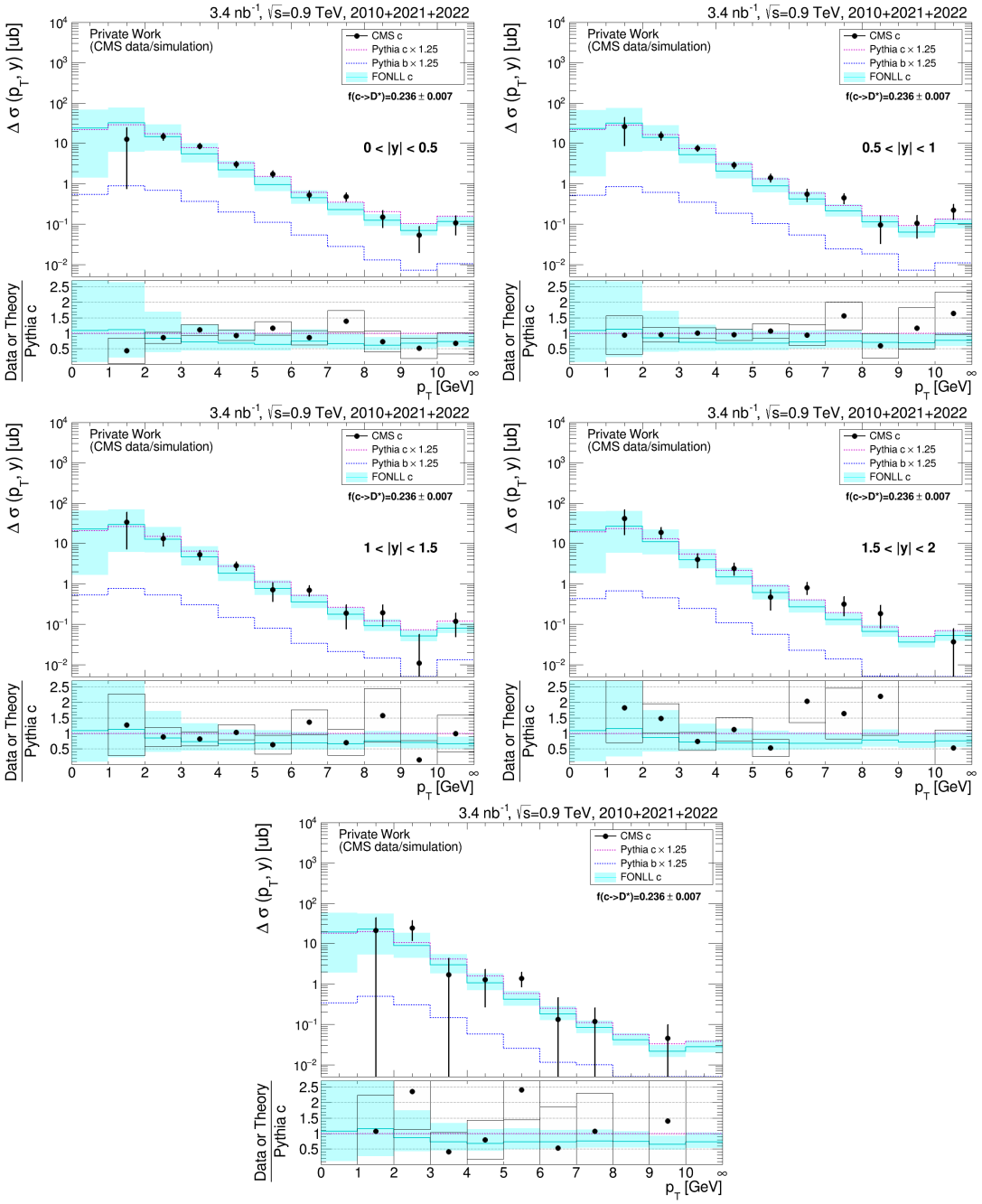


Figure A.1: Prompt $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The systematic uncertainties are incomplete. Figure are from [121].

B | Prompt D^0 Cross Sections at 13 TeV

Very preliminary prompt D^0 cross sections at $\sqrt{s} = 13$ TeV are shown in Figure B.1. Besides the effective luminosity which is not determined yet, many of studies are left to finalize this measurement.

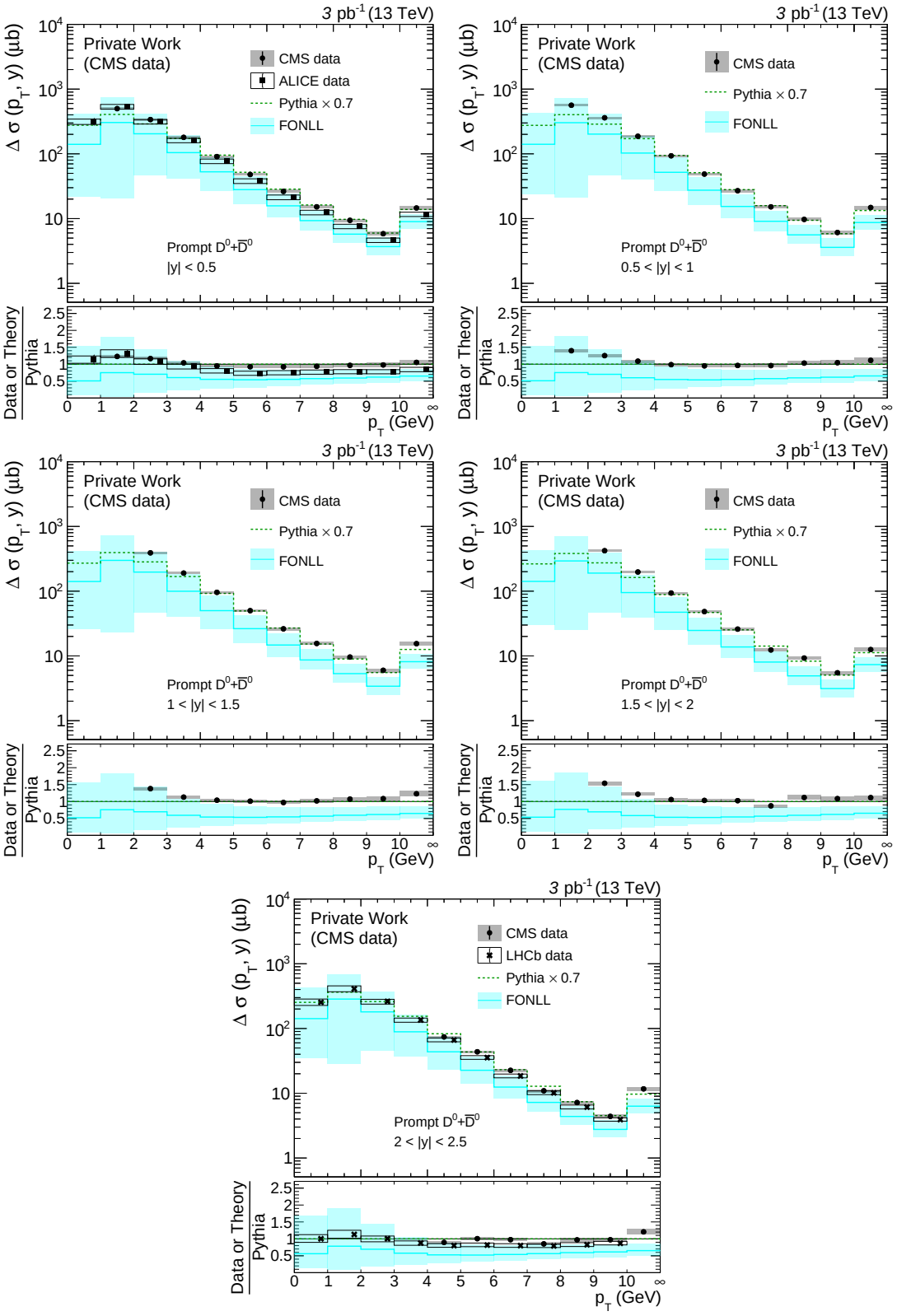


Figure B.1: Prompt D^0+$\overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The vertical error bars indicate the statistical uncertainties, and the error boxes show the systematic uncertainties. The systematic uncertainties of CMS are incomplete.

C | π_s^+ Efficiency Measurements at 13 TeV

As explained in Section 5.1, the kinematics of π_s^+ is strongly correlated to that of D^0. Therefore, if cuts applied to K^- and π^+ candidates are the same independent of whether D^0 is from D^{*+} or not, a ratio of D^{*+} to D^0 can give a measurement of π_s^+ efficiency, where all the other efficiencies are canceled by taking the ratio. A study was done to measure π_s^+ efficiency on the 2018 B parking data.

For the purpose of measuring π_s^+ efficiency, a new definition was introduced in binning. A finner binning was introduced in $|y|$ by a width 0.25 instead of 0.5 for the cross section measurements in the previous sections. Here it is assumed that π_s^+ and D^0 have the same $|y|$ distribution. By taking a traditional way of measuring tracking efficiency, $p_T^{\pi_s^+}$ was binned in a log scale with a width 0.2, where $p_T^{\pi_s^+}$ was translated by $p_T^{D^0}$ with a relative mass 0.1454 GeV/1.8648 GeV (0.1454 GeV is defined by a difference between $M_{D^{*+}}$ and M_{D^0}).

$D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ and $D^0 \to K^-\pi^+$ (including charge conjugate) were reconstructed by sharing the same preselection given for the cross section measurements, i.e., the cuts in Table 5.7. On the other hand, the final selection cuts were defined as listed in Table C.1 and were applied independent of whether D^0 is from D^{*+} or not. In the case of the $D^0 \to K^-\pi^+$ reconstruction, the cuts on M_{D^0} in the table, which are introduced for the $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ reconstruction, were applied in a M_{D^0} fit.

Table C.1: $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ and $D^0 \to K^-\pi^+$ selection applied for π_s^+ tracking efficiency measurement after preselection given by Table 5.7. The M_{D^0} cuts were defined for the $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ reconstruction, and applied in a M_{D^0} fit in the case of the $D^0 \to K^-\pi^+$ reconstruction. The last row here is only for the $D^0 \to K^-\pi^+$ selection, which was applied also for the $D^{*+} \to D^0\pi_s^+ \to K^-\pi^+\pi_s^+$ in the preselection.

$p_T^{D^0}$	$< 1.8648/0.1454 \times 10^{-0.6}$ GeV	$> 1.8648/0.1454 \times 10^{-0.6}$ GeV
M_{D^0}	$\in [1.84, 1.89]$ GeV	$\in [1.85, 1.88]$ GeV
dl_{sig}	> 3.5	> 4
$D^0 p_T^{frac}$	> 0.1	0.03
$\cos\phi$	> 0.99	
$p_T^{D^0}$	> 1.4 GeV	

The statistics of the NMB collisions on the 13 TeV B parking data covers well $p_T^{D^0}$ up to 100 GeV which is in turn $p_T^{\pi_s^+} \sim 10$ GeV, as shown in Figure C.1. To extract the number of signal events, fit methods were used for both the D^{*+} and D^0 reconstruction. Ratios of the number of the D^{*+} and D^0 signal events are summarized in Figure C.2.

To derive the actual π_s^+ efficiency, the ratio should be scaled by a relative fraction:

$$\frac{f_{D^{*+}} \times \mathrm{BR}_{D^{*+} \to D^0\pi^+}}{f_{D^0}} \tag{C.1}$$

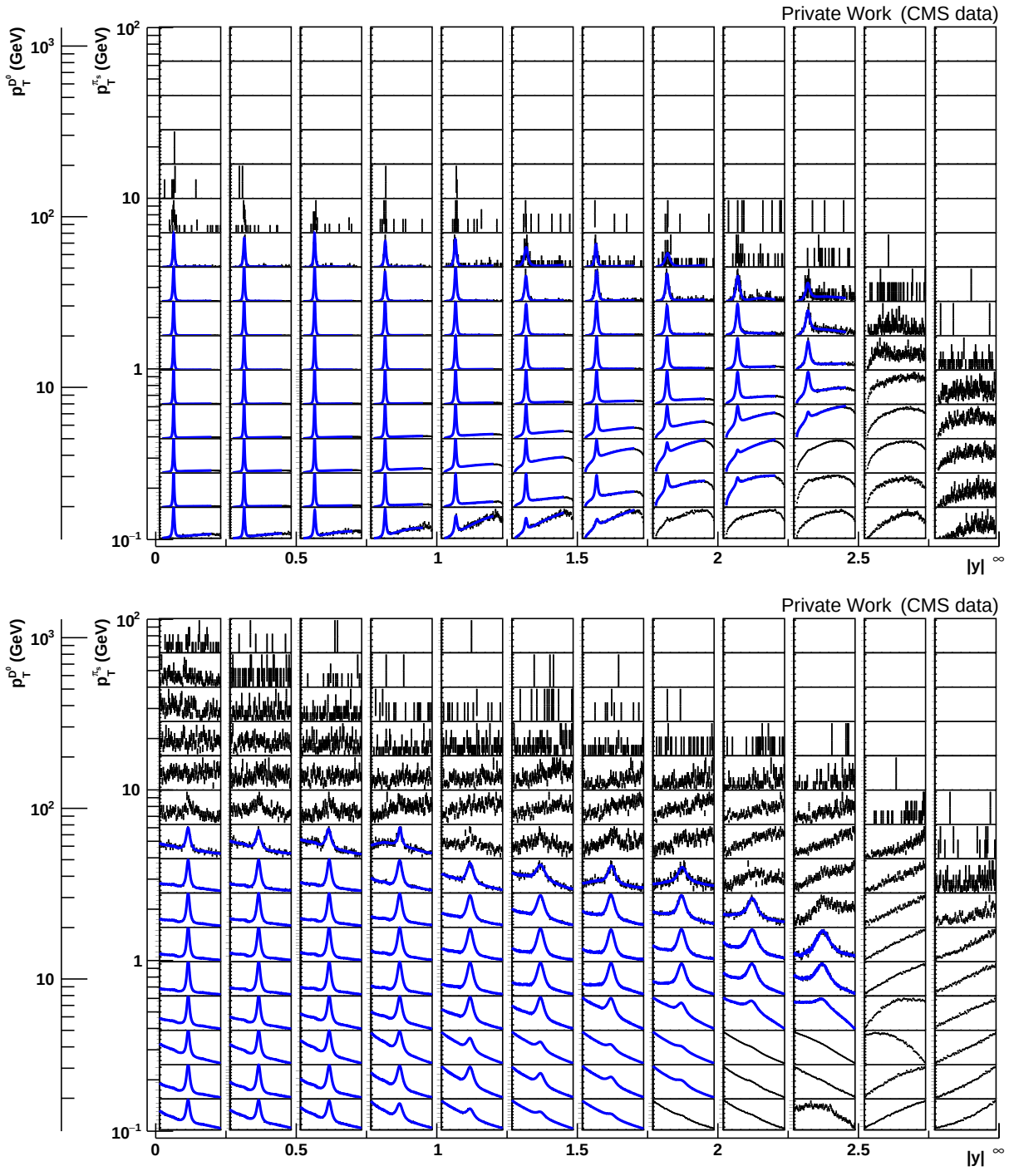


Figure C.1: D^* (top) and D^0 (bottom) reconstruction on the 13 TeV B parking data for π_s^+ efficiency measurement.

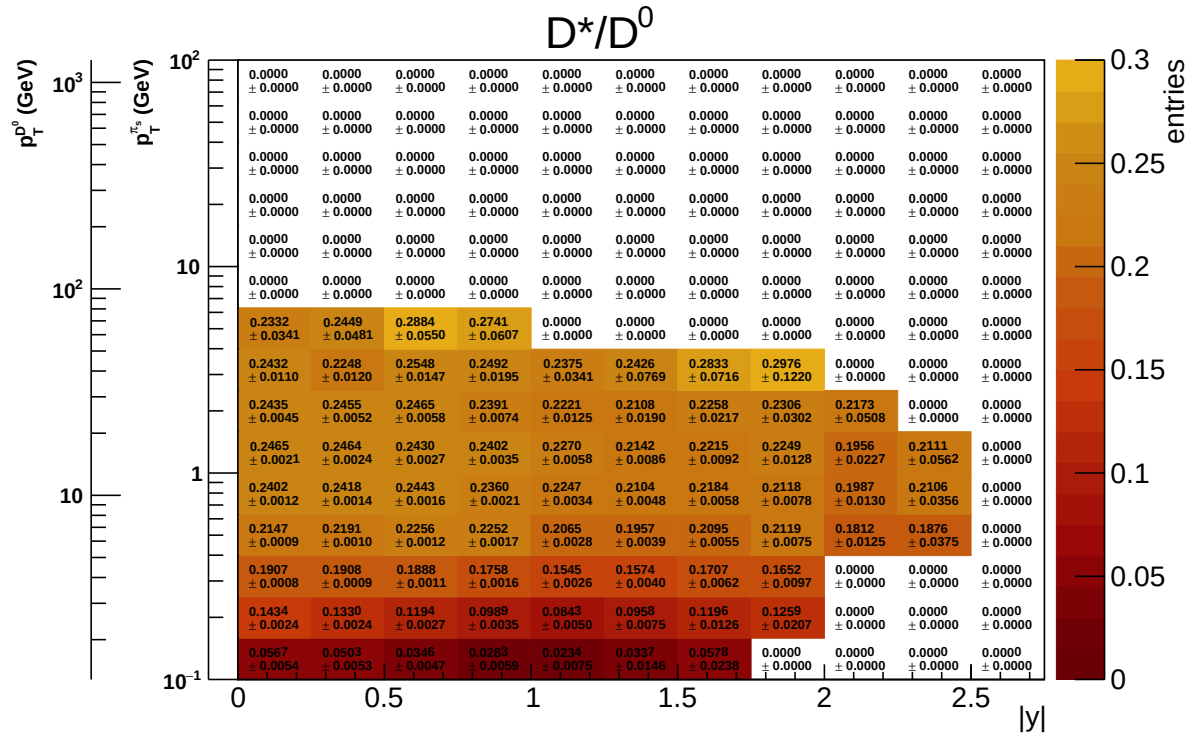


Figure C.2: Ratios of the number of the D^* and D^0 events.

where f is fragmentation fraction and BR is branching fraction. Depending on a relative contribution of prompt and non-prompt, the fraction can differ and be dependent on p_T. With fragmentation fraction 0.236(0.221) and 0.542(0.587)* of prompt(non-prompt) D^{*+} and D^0 [136], respectively, Eq.(C.1) results in minimum 0.255 (with beauty fraction 1) and maximum 0.295 (with charm fraction 1). The relative contribution of prompt and non-prompt for the π_s^+ efficiency measurement was not fully determined yet. However, it was shown in Section 5.6 that charm fractions are roughly 0.9-0.95 (refer to Figure 5.19) depending on bins up to $p_T \sim 10$ GeV of the D meson ($p_T^{\pi_s^+} \sim 1$ GeV). This then results in roughly 0.290-0.293 of Eq.(C.1). Assuming that the difference on the final cuts in Table 5.8 and 5.9 for the cross section measurements and in Table C.1 for the efficiency measurement gives almost no effect on the charm fractions, 0.290 was set as central value with +0.003 uncertainty and a very conservative uncertainty of -0.035 (with the minimum of Eq.(C.1)). Then π_s^+ efficiency can be derived with

$$\frac{N^{D^{*+}}}{N^{D^0}} \times \frac{f_{D^0}}{f_{D^{*+}} \times \mathrm{BR}_{D^{*+}\to D^0\pi^+}} \tag{C.2}$$

A result is shown in Figure C.3, where preliminary π_s^+ efficiency is shown as a function of p_T and $|y|$.

As references, the same results were derived also on the D^0 and MinimumBias MC sample, which are also shown in Figure C.3 (D^{*+} and D^0 signal and ratios between the two on the MC samples can be found in Appendix F). The same fraction 0.290 is applied for the MC samples. Comparing the two MC samples, it looks like that pileup effect on the efficiency is not significant. Meanwhile, it was observed that there are up to $\sim 10\%$ discrepancies between the MC and data at high p_T and large $|y|$.

*Here the numbers are based on LEP measurements with charm fragmentation universality. With the non-universal charm fragmentation, the individual numbers can differ but the relative ratio between D^{*+} and D^0 is observed to be universal to date. Further discussions can be found in the next chapter.

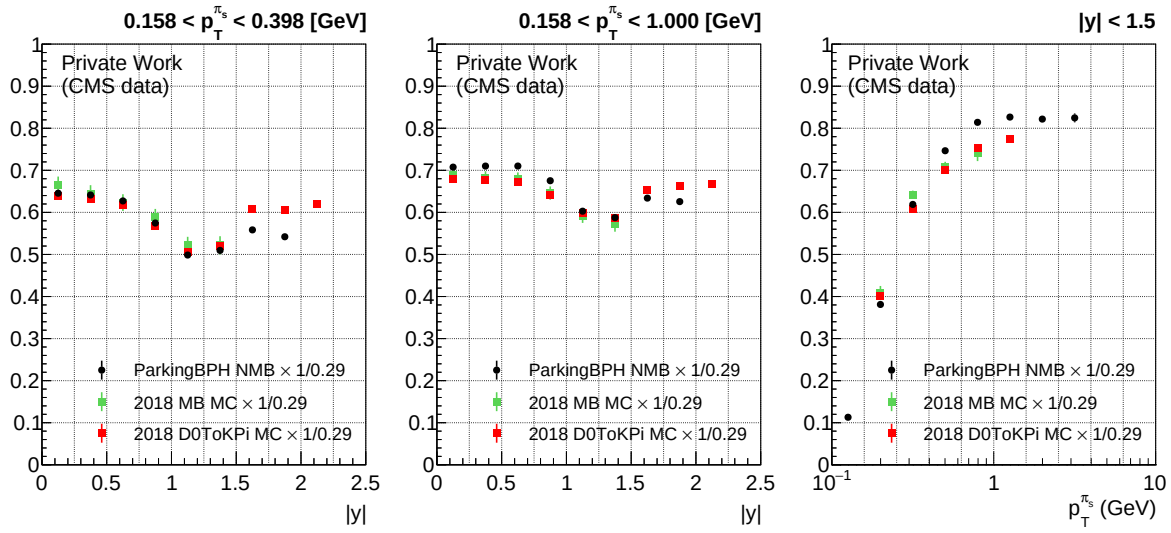


Figure C.3: Preliminary π_s^+ efficiency on the 13 TeV samples with statistical uncertainties only. The factor 0.290 was tentatively assigned with uncertainties of $^{+0.003}_{-0.035}$.

Since this is a very preliminary result and systematic uncertainties are not fully determined, no correction or systematic uncertainty was added on the cross section measurements to account for this result.

D | D^{*+} and D^0 Fit Results

D.1 7 TeV D^{*+} signal fit

In this appendix, fit results of D^{*+} signal on the 7 TeV data are shown, which the MC constrained fit was used for.

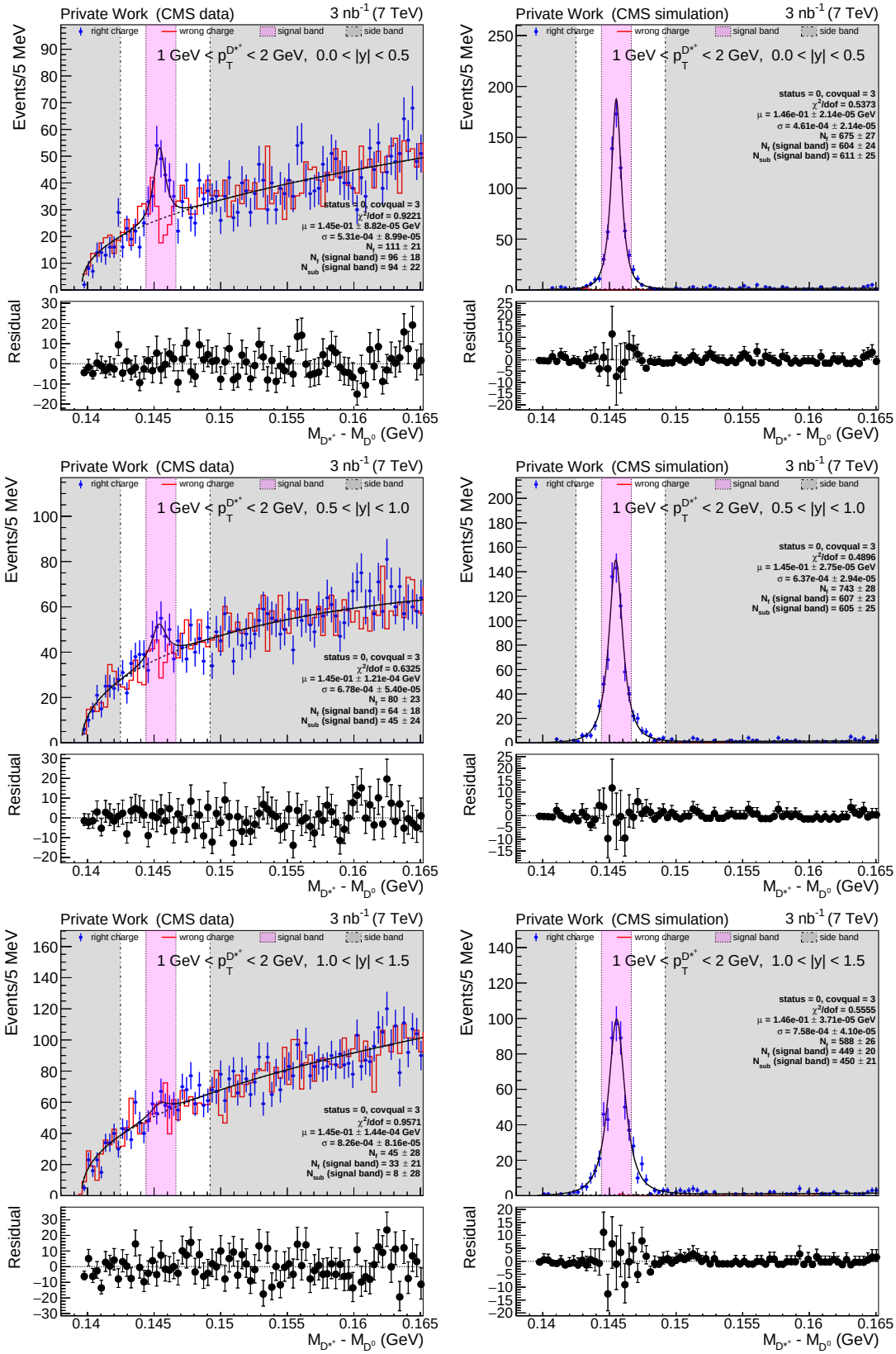


Figure D.1: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

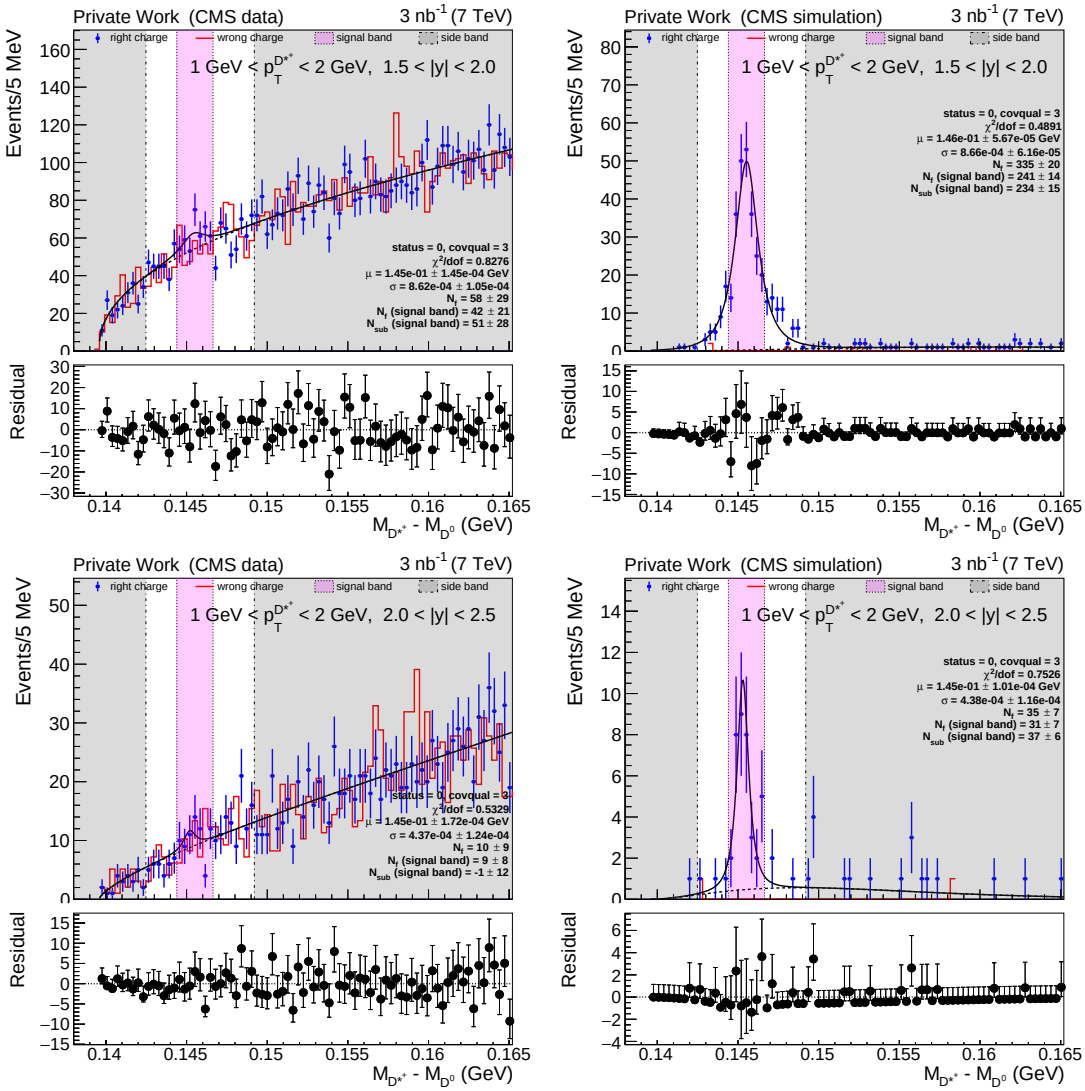


Figure D.2: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

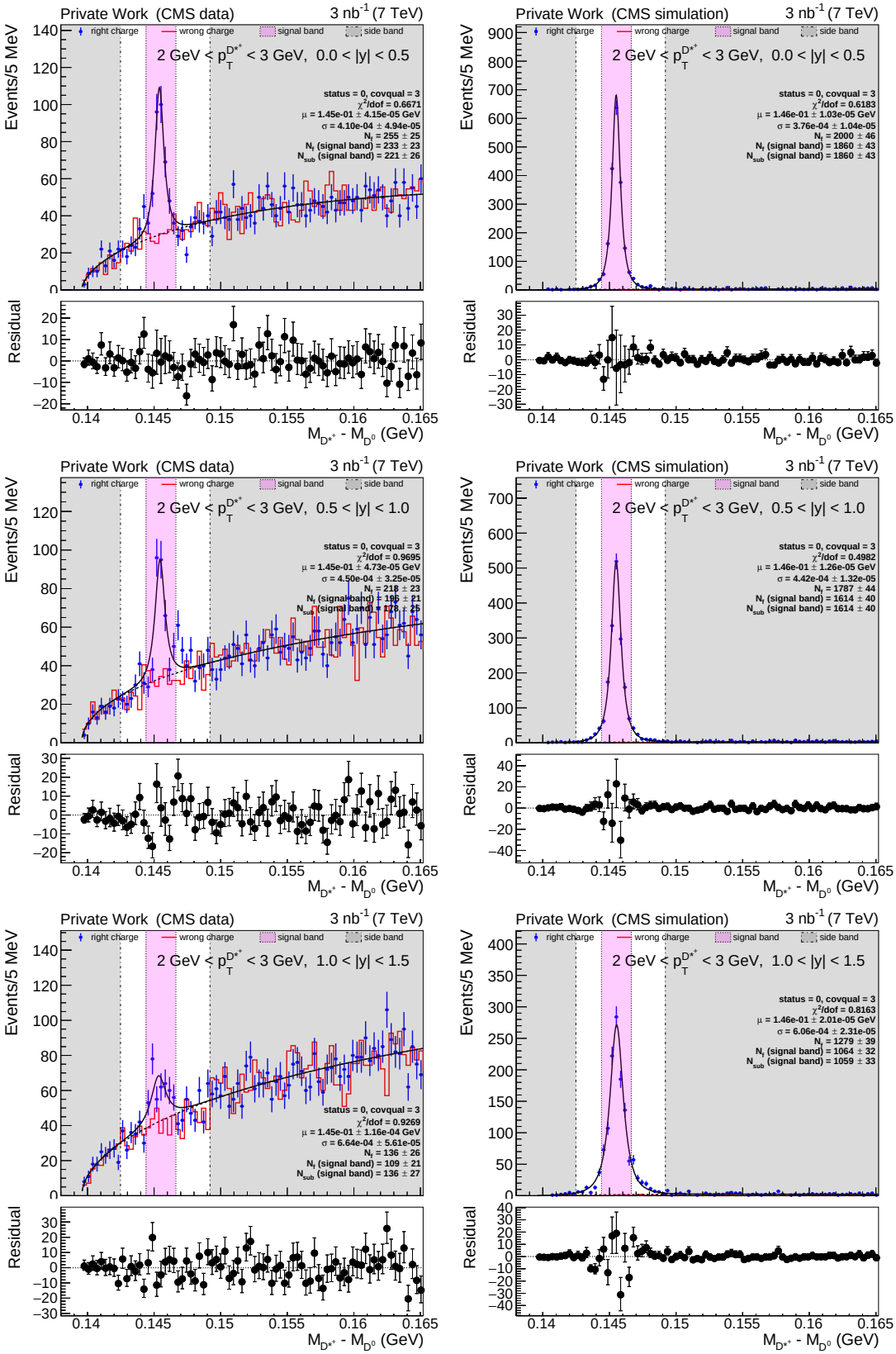


Figure D.3: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

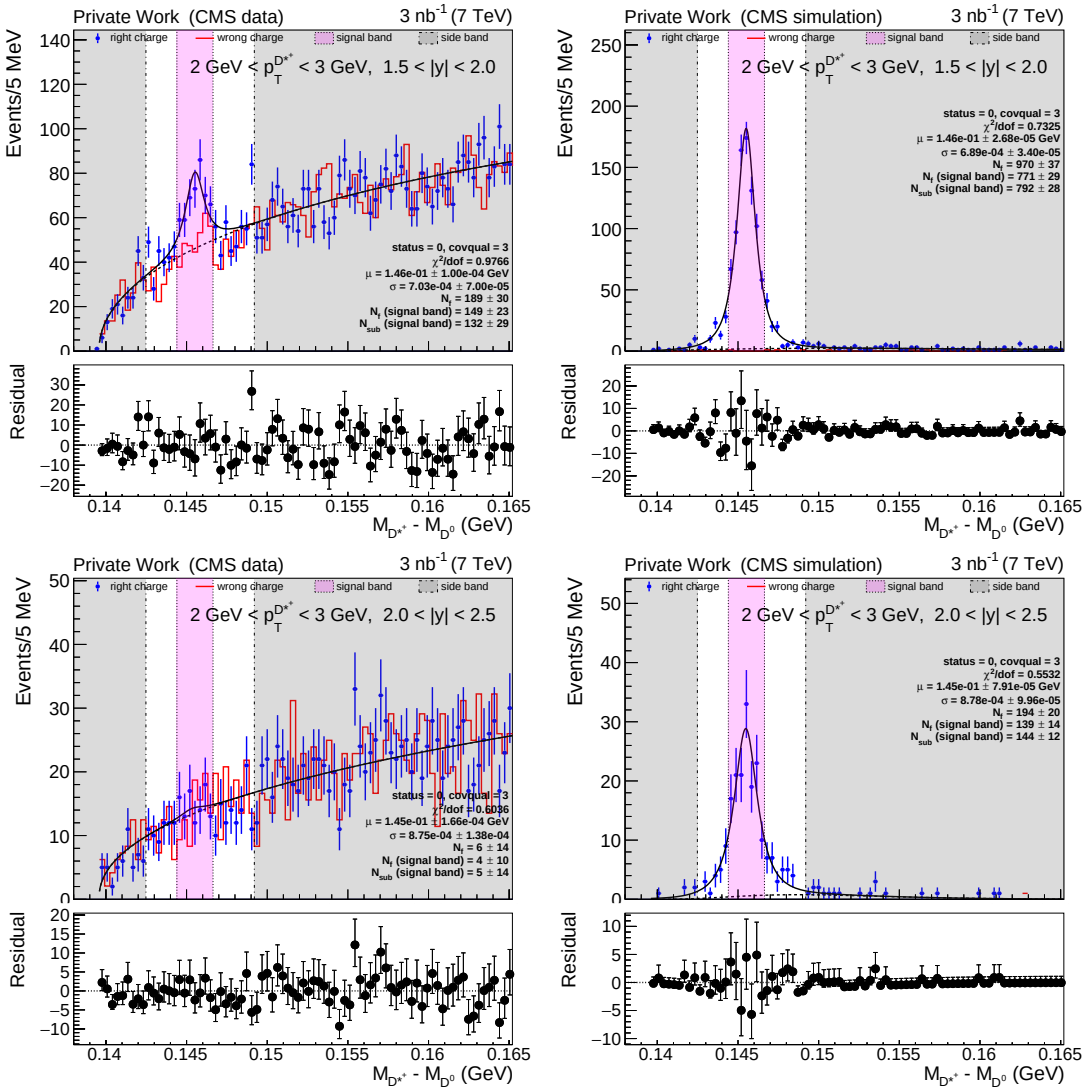


Figure D.4: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

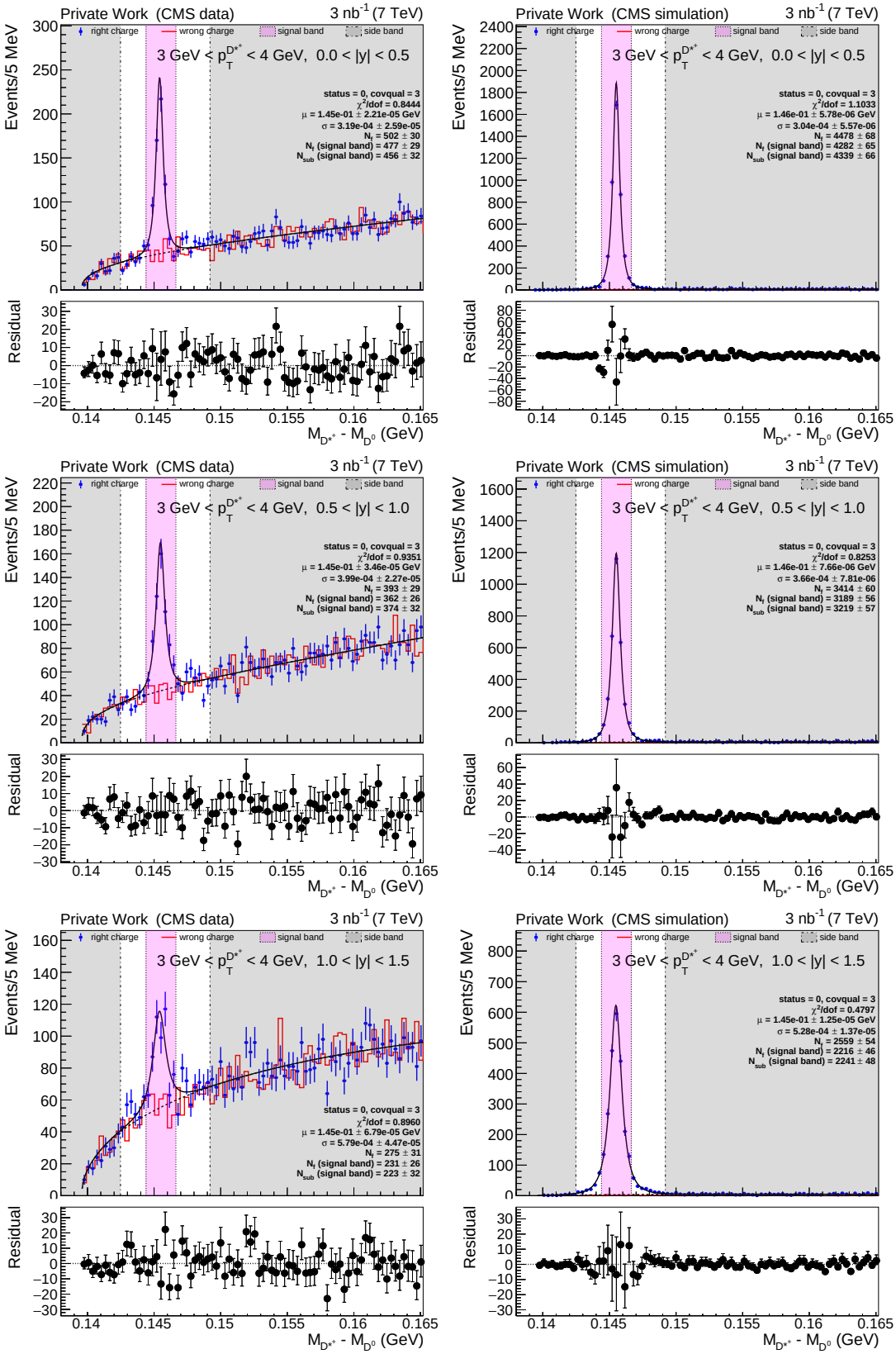


Figure D.5: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

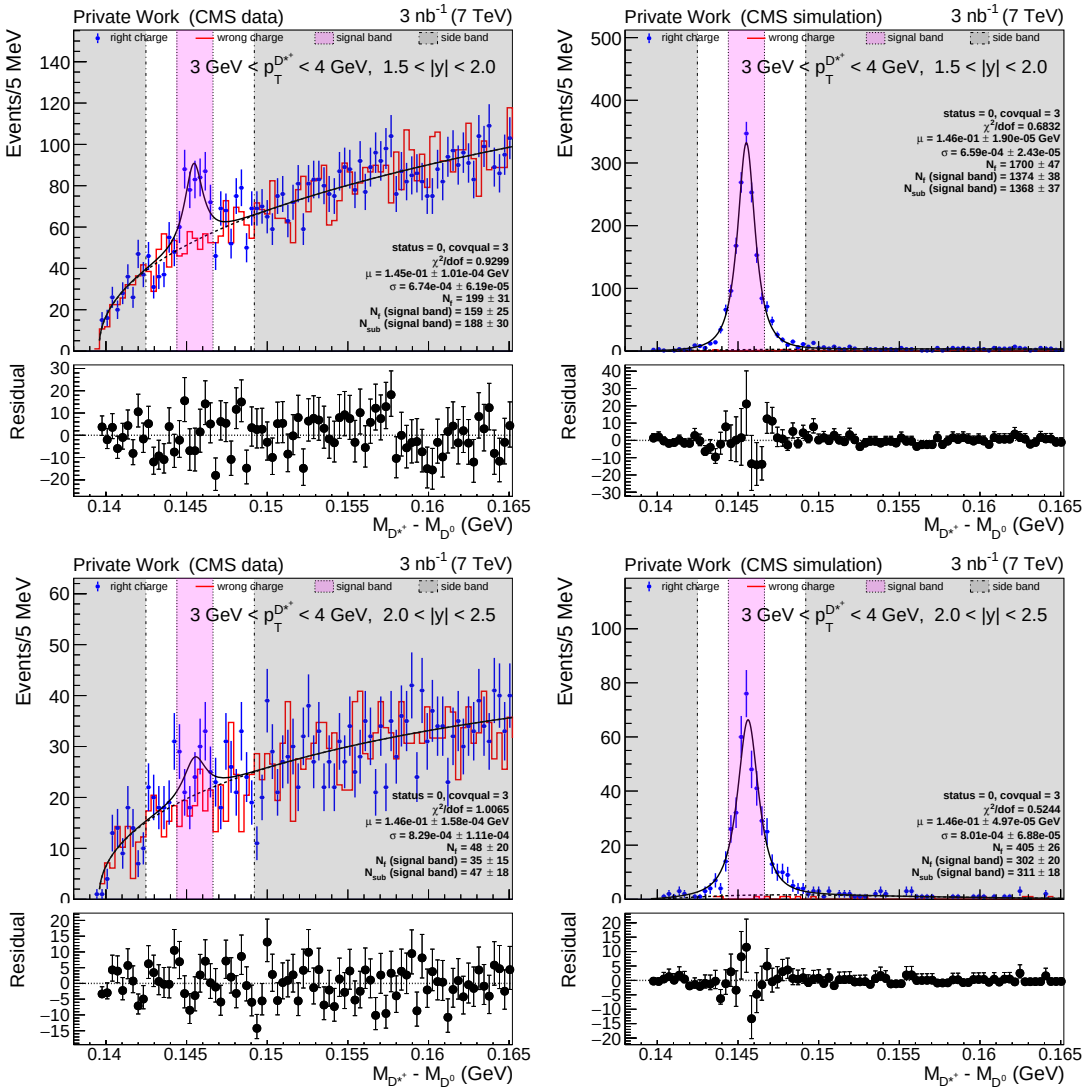


Figure D.6: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

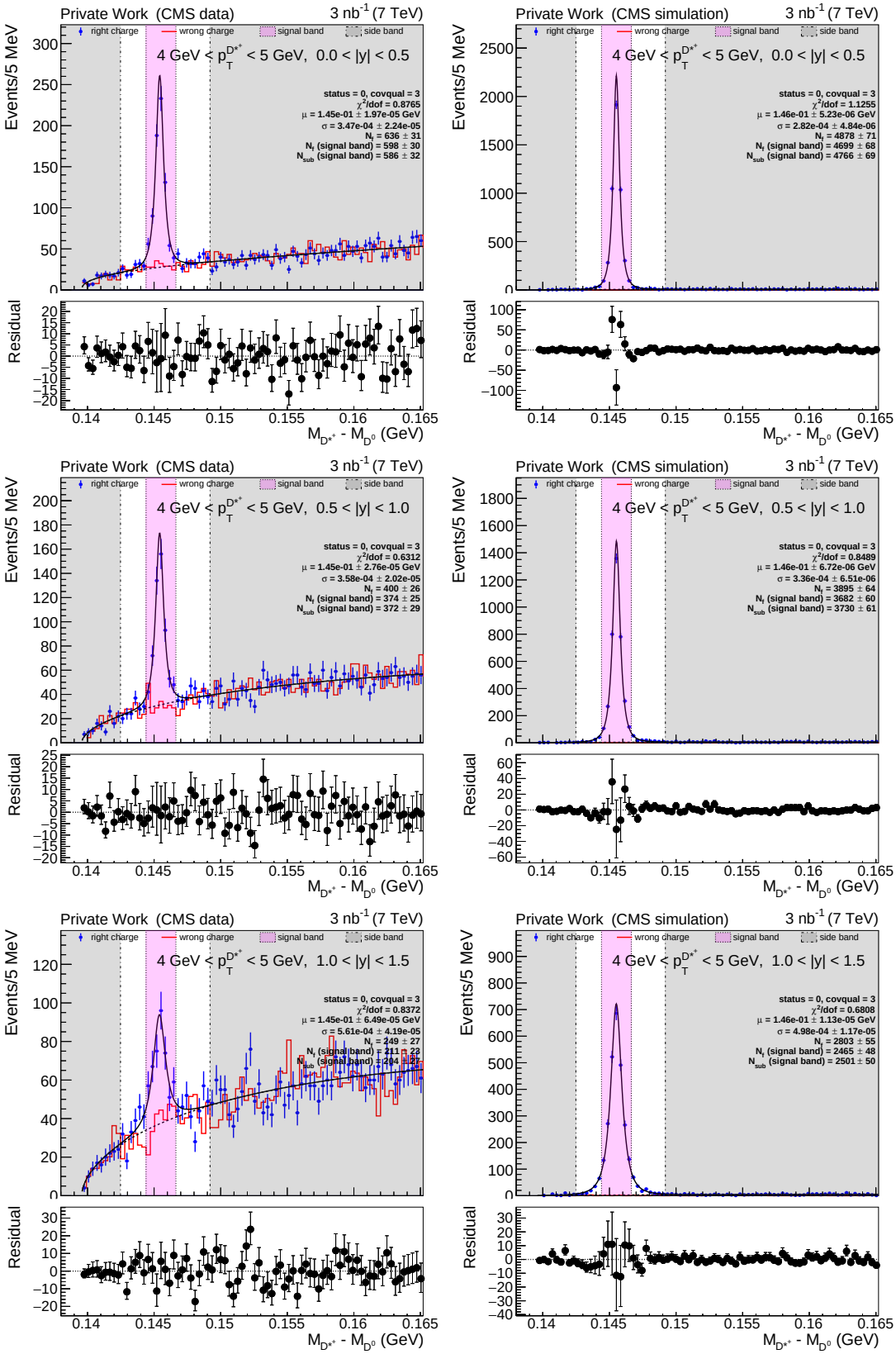


Figure D.7: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

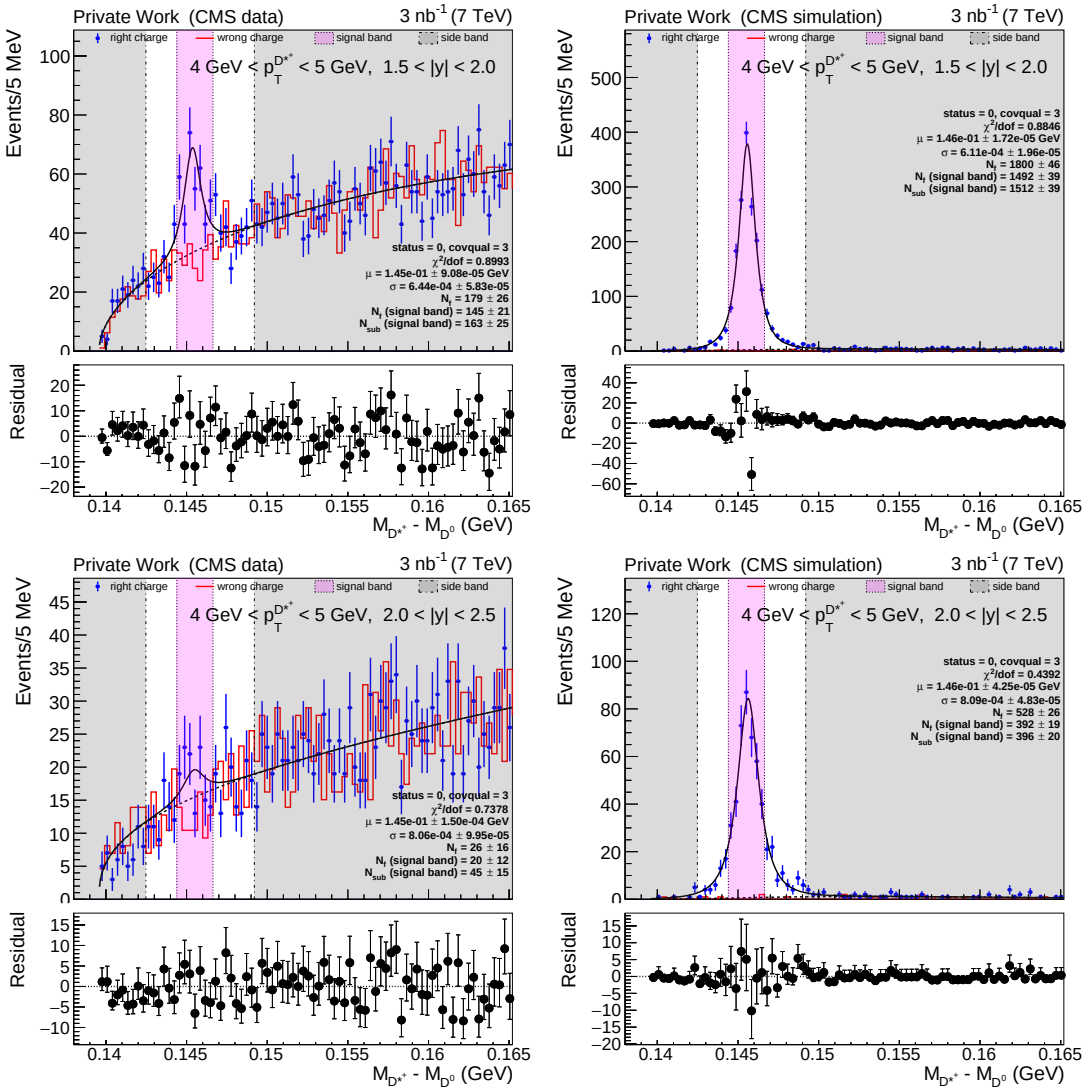


Figure D.8: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

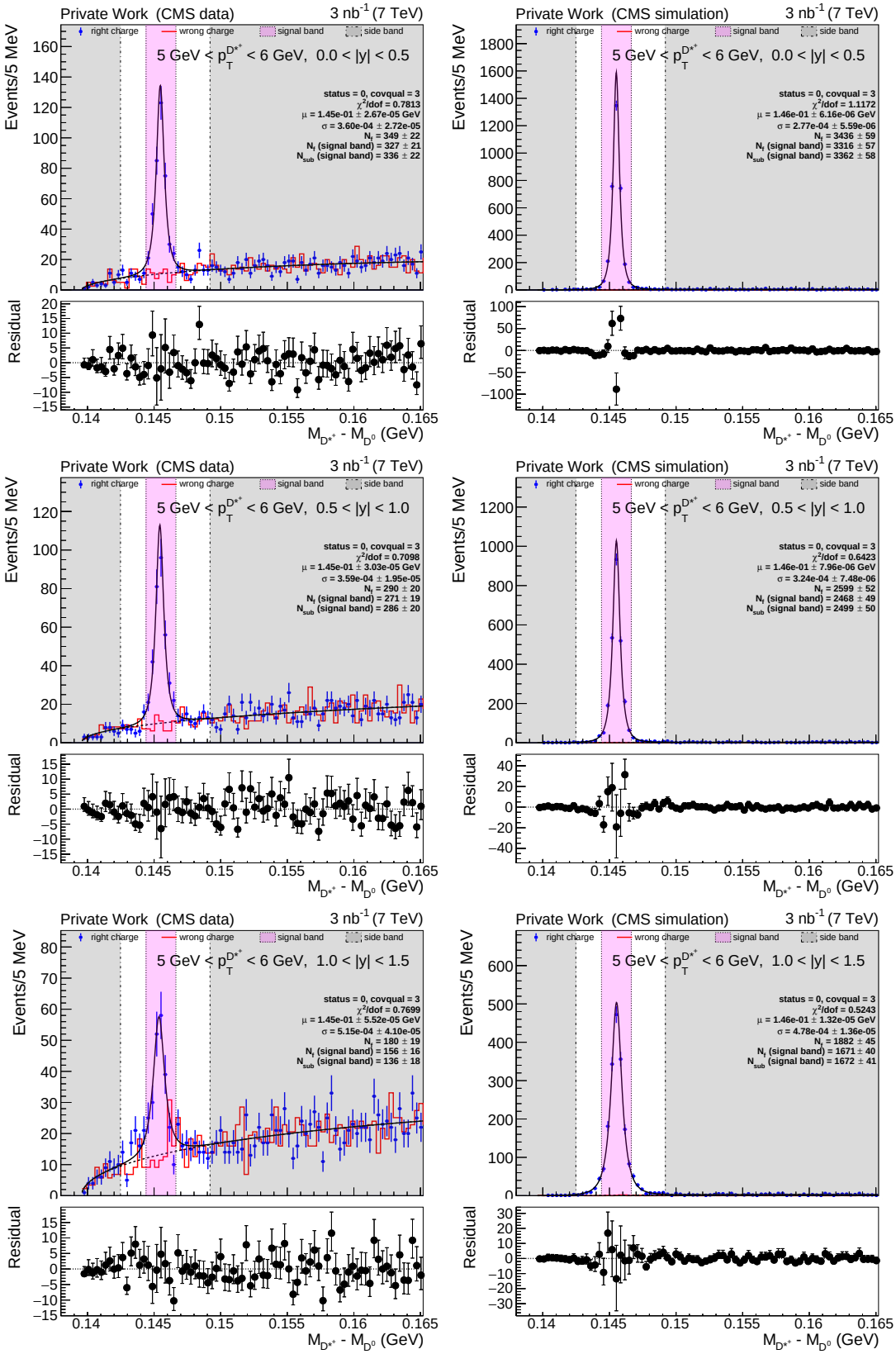


Figure D.9: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

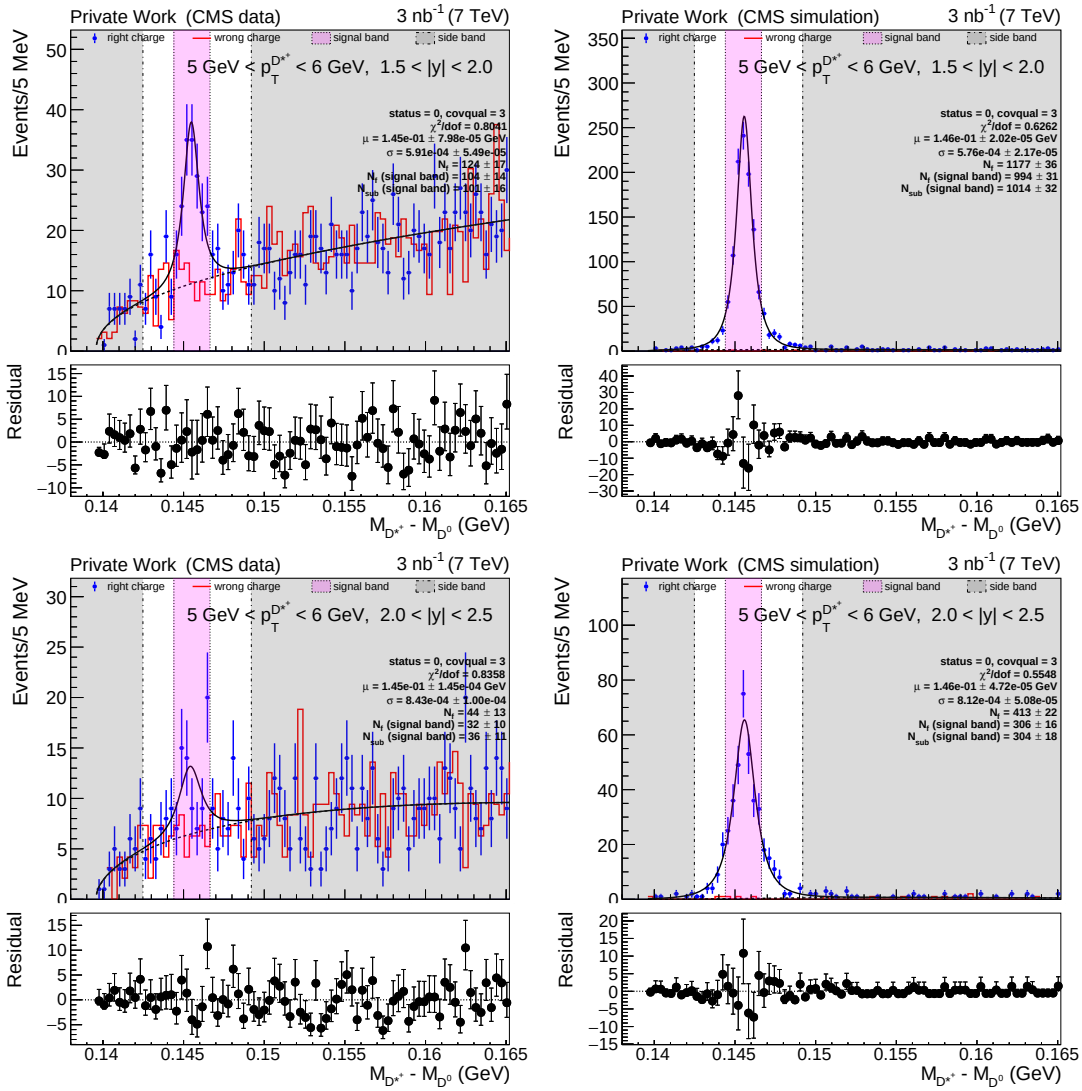


Figure D.10: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

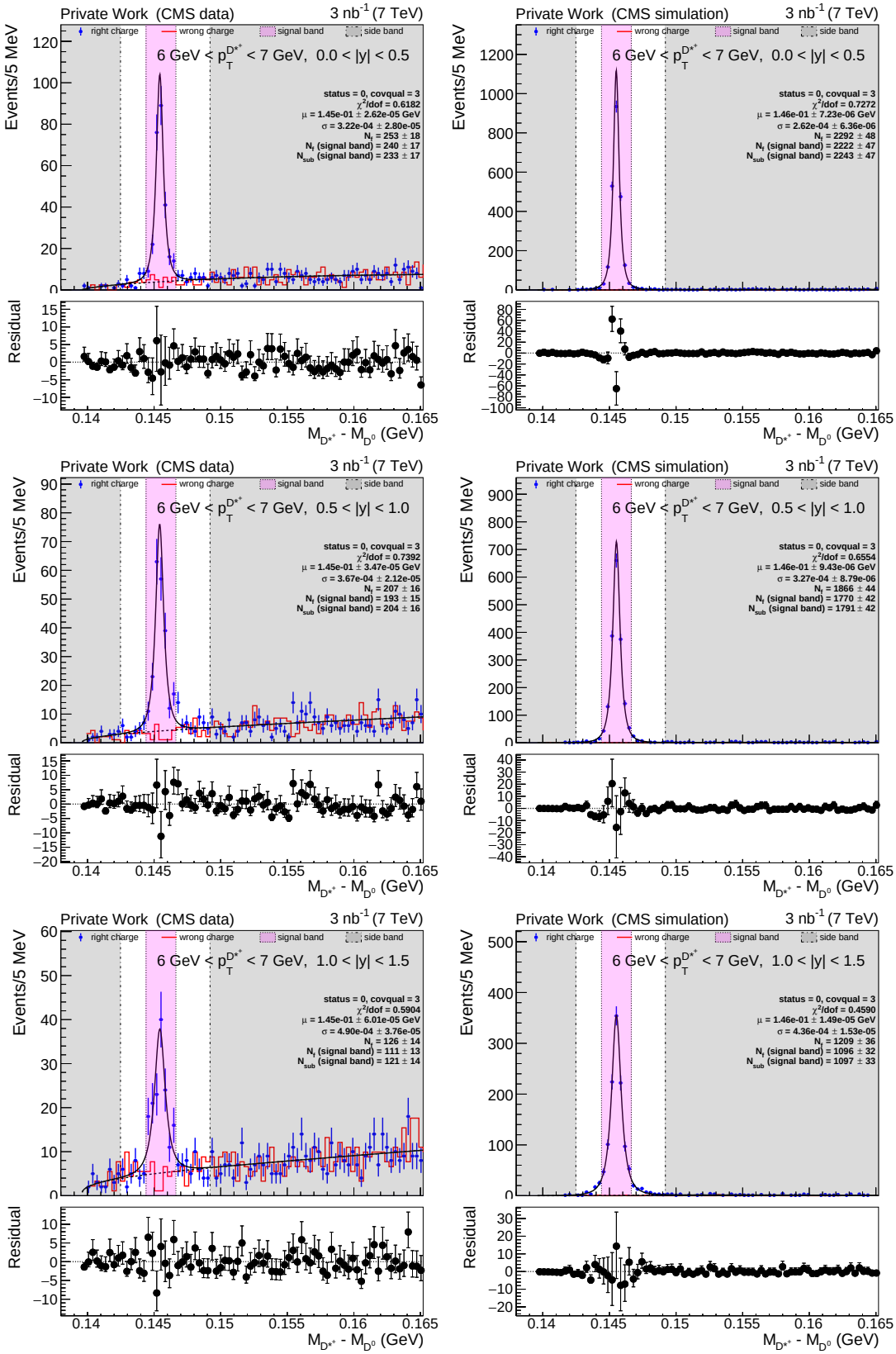


Figure D.11: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

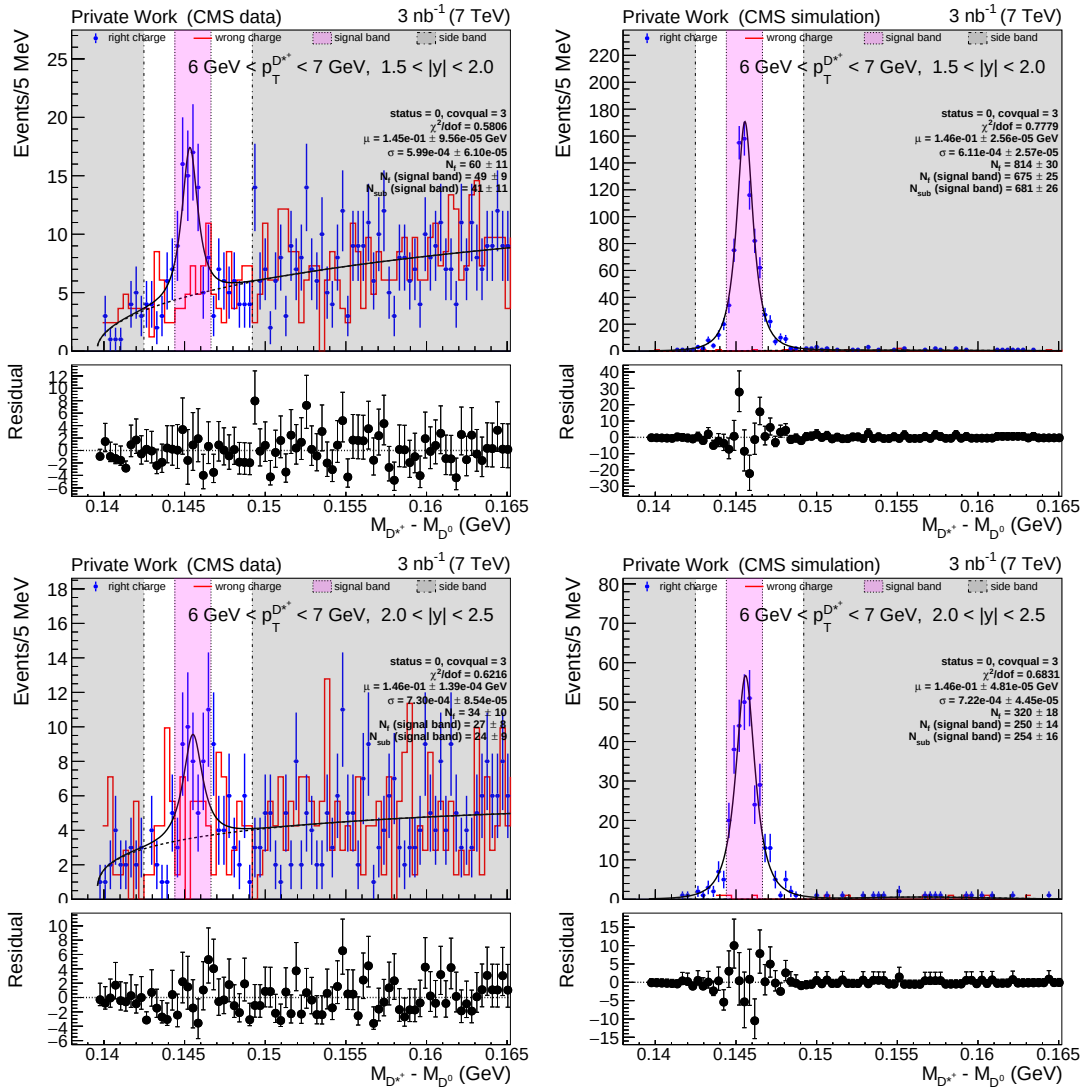


Figure D.12: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

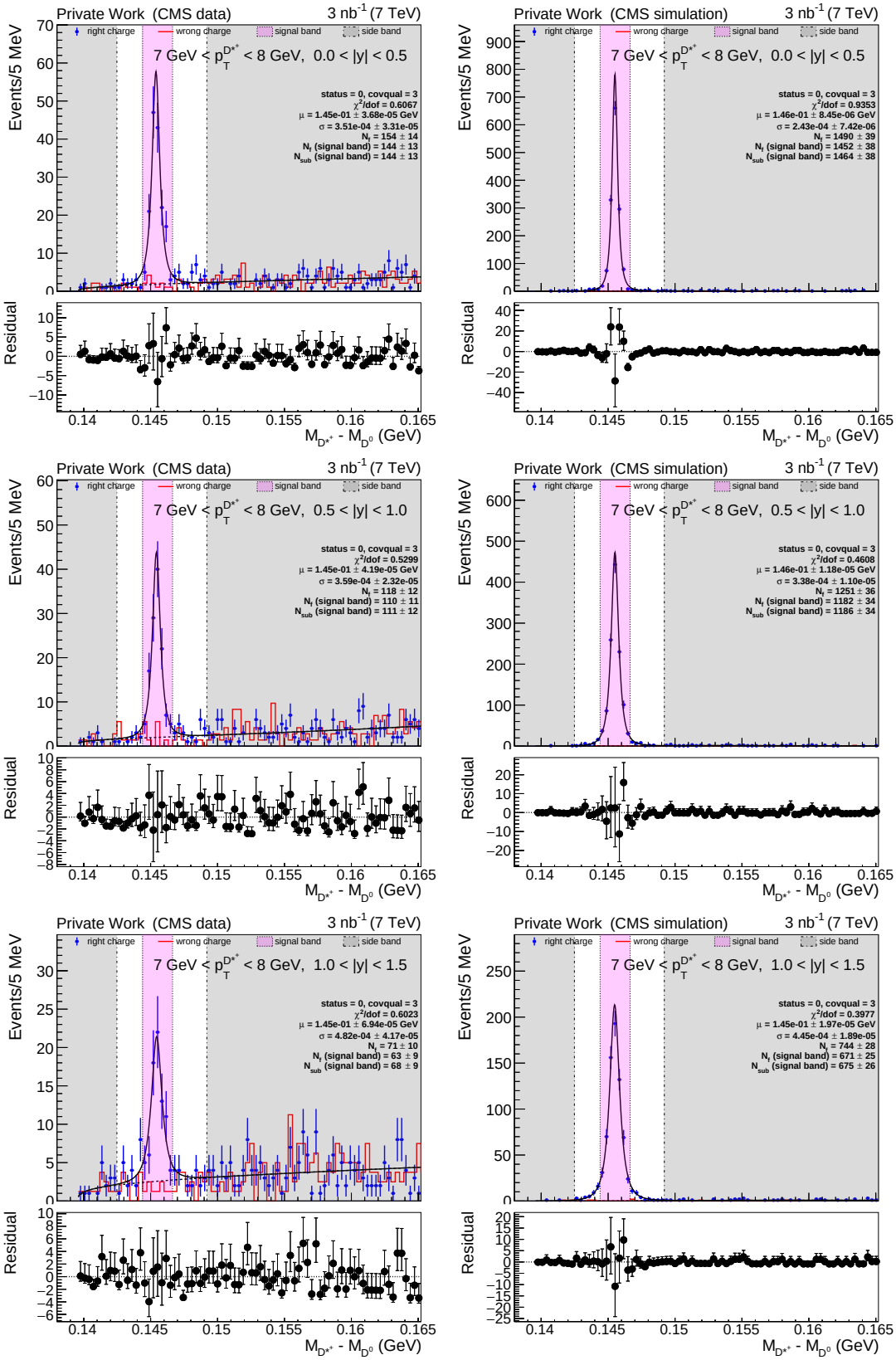


Figure D.13: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

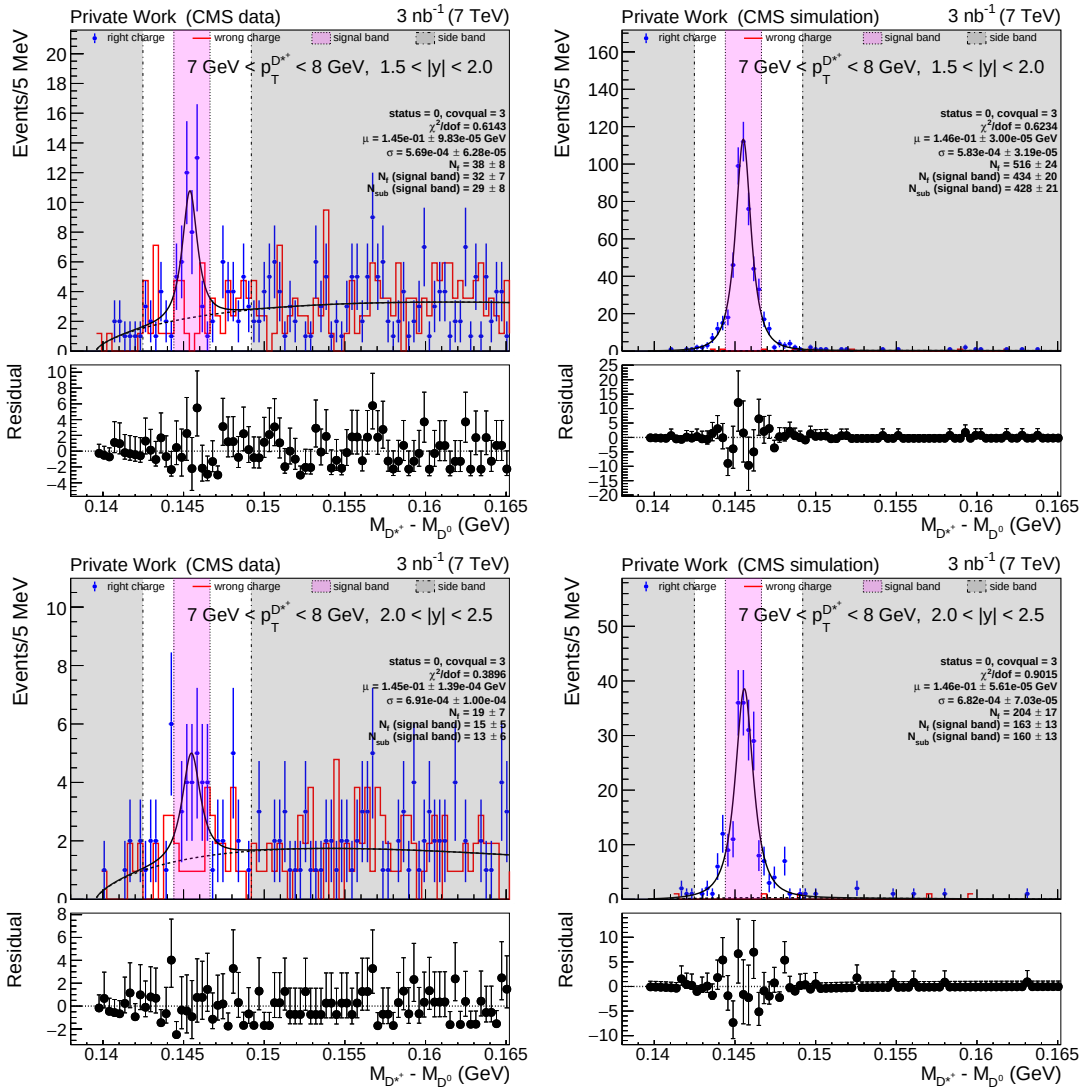


Figure D.14: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

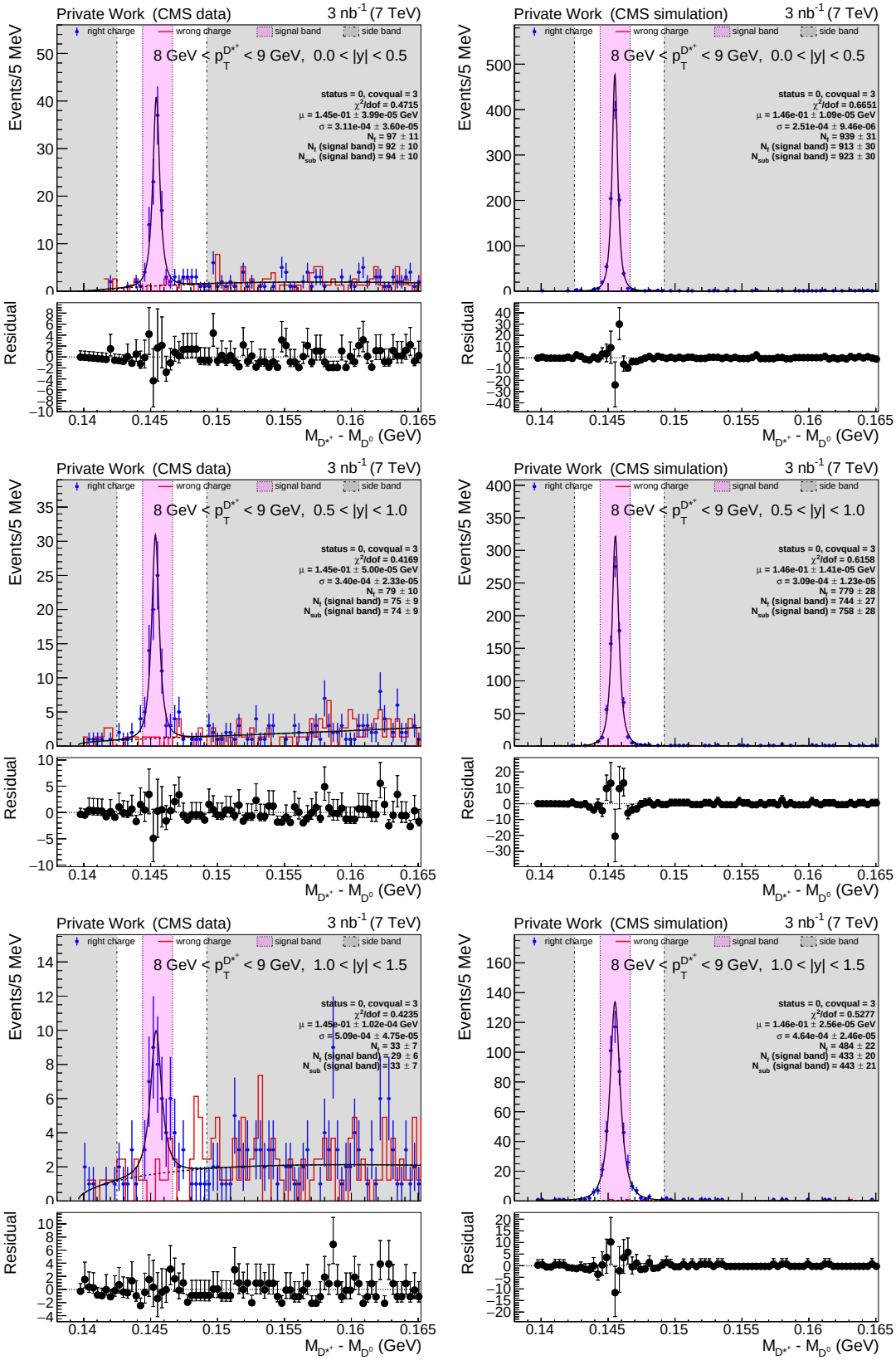


Figure D.15: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

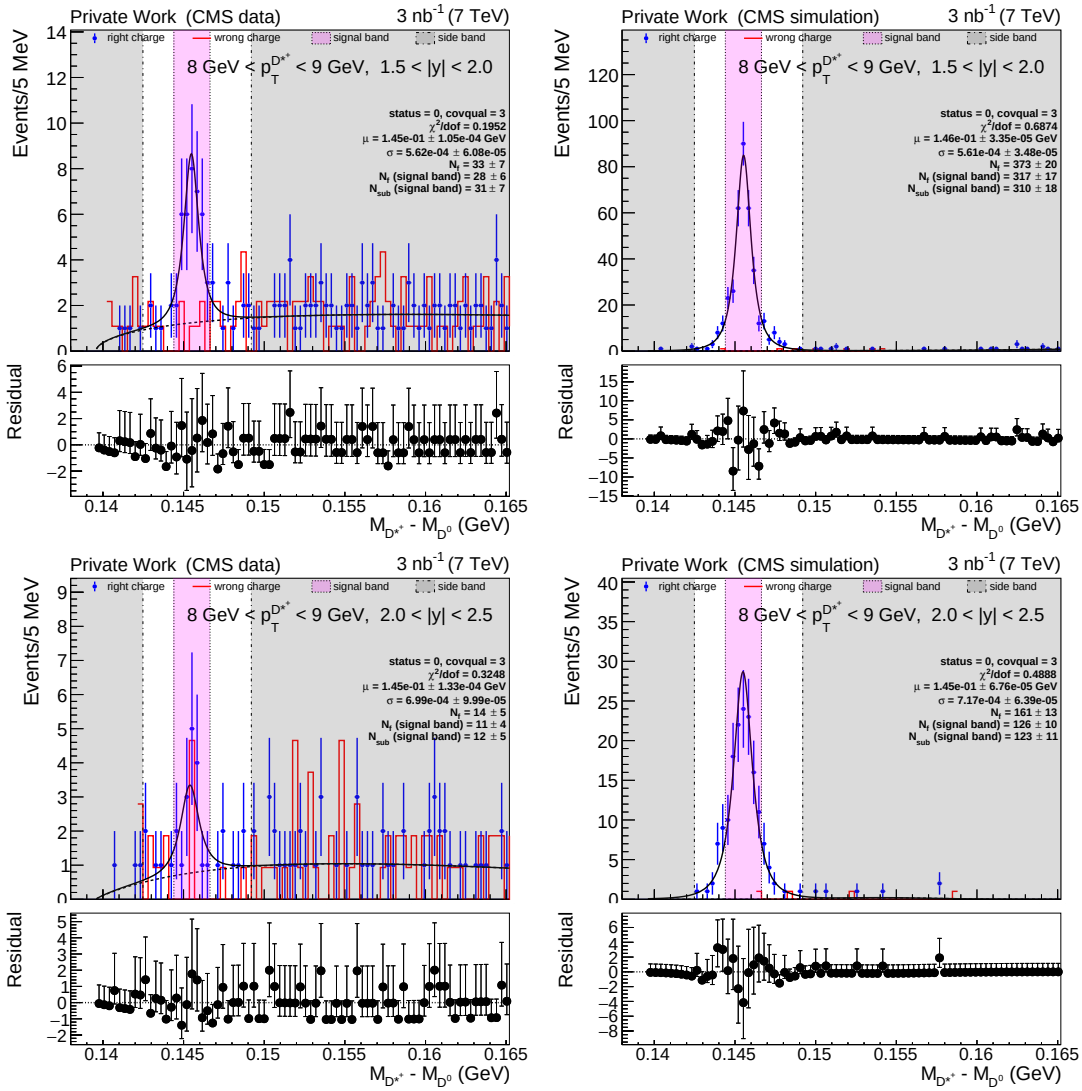


Figure D.16: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

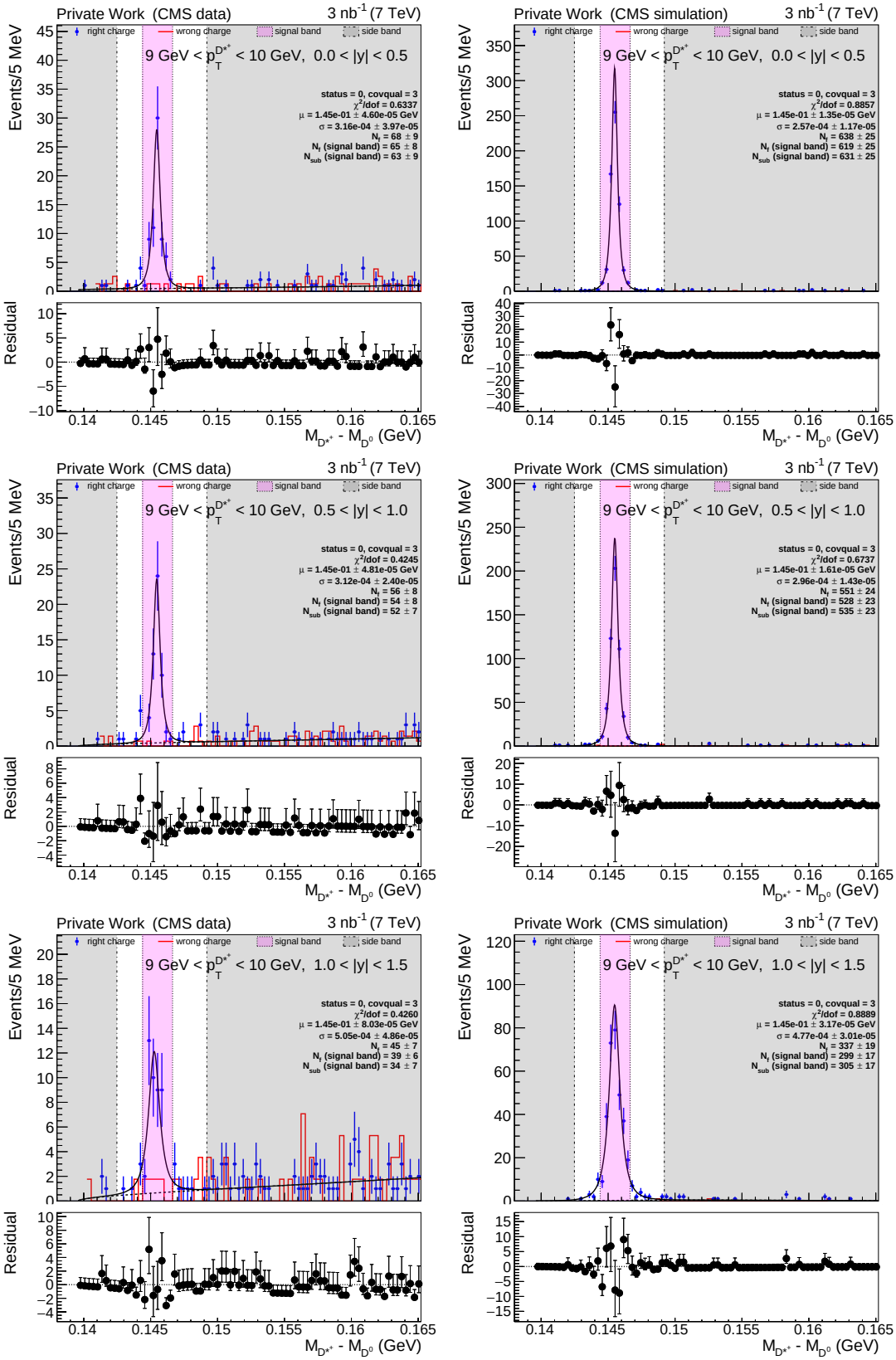


Figure D.17: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

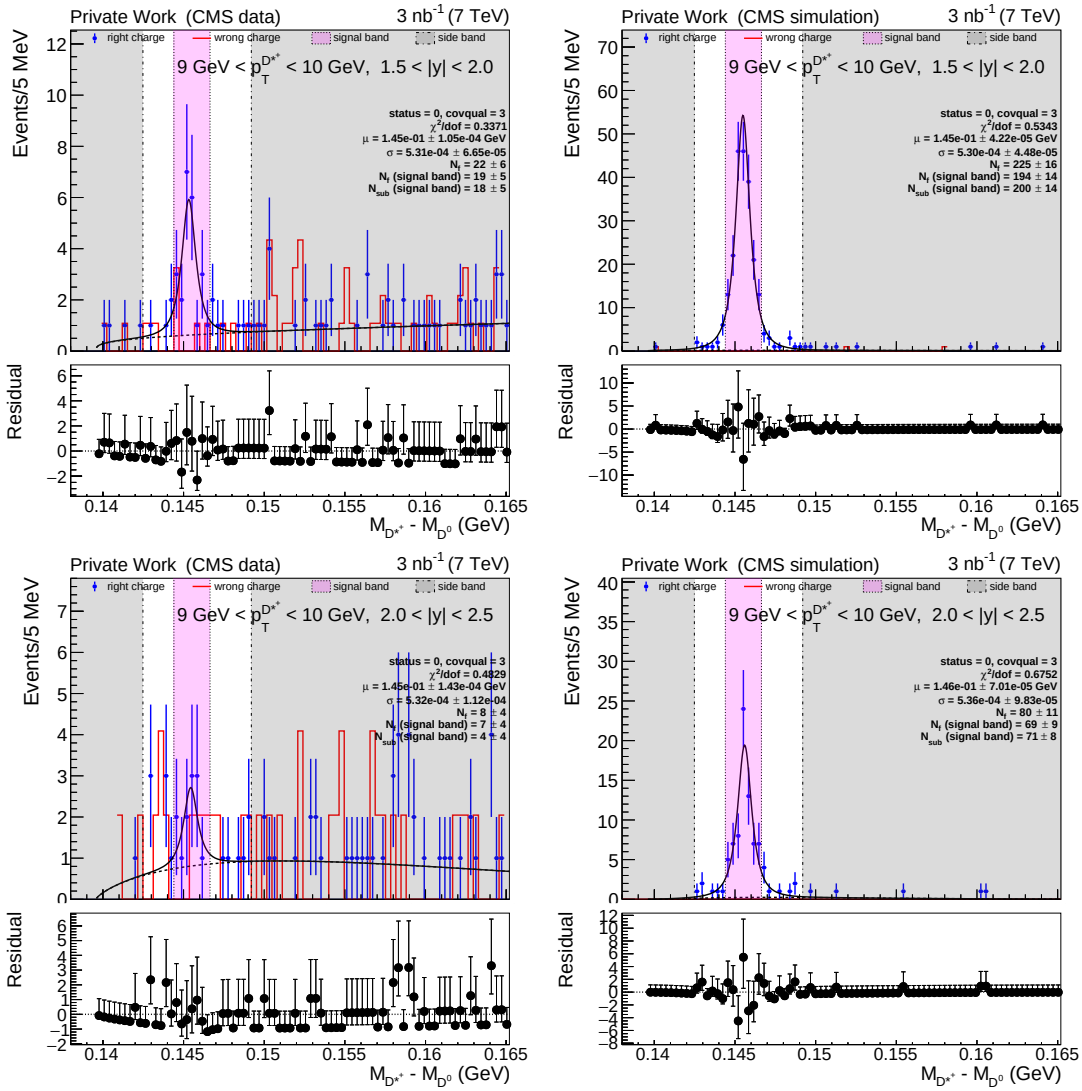


Figure D.18: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

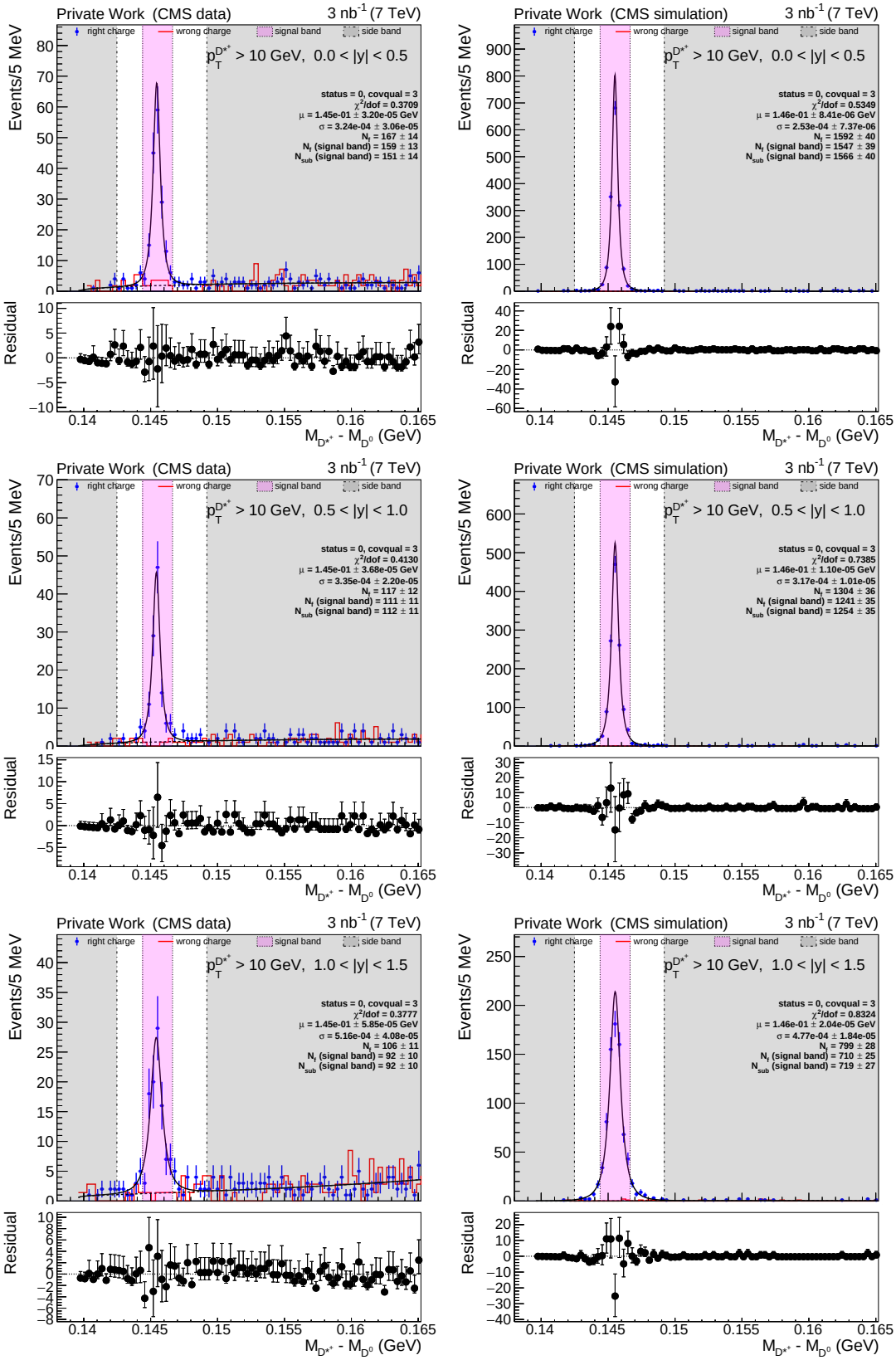


Figure D.19: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

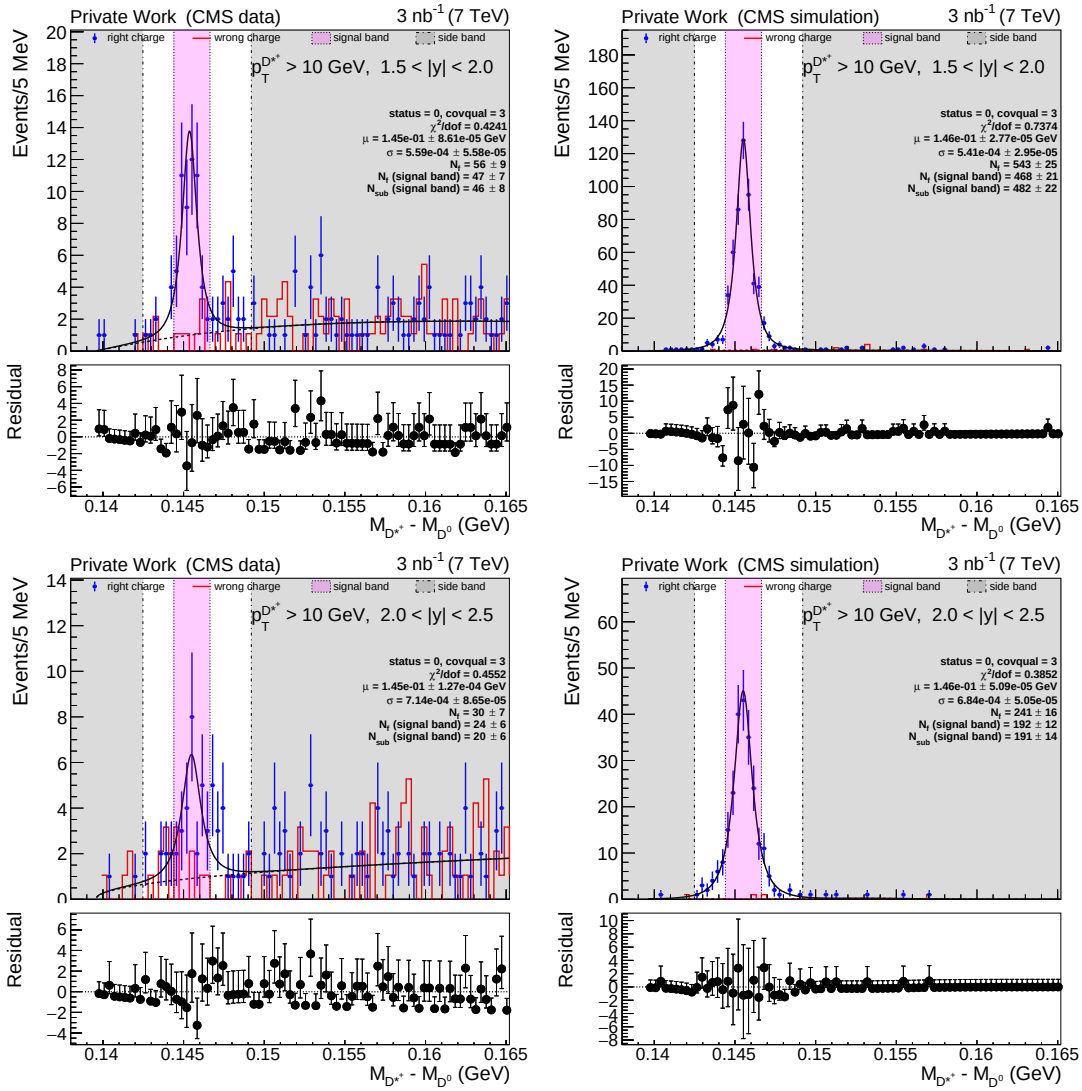


Figure D.20: $M_{D^{*+}} - M_{D^0}$ distributions and fit on data (left) and MC (right).

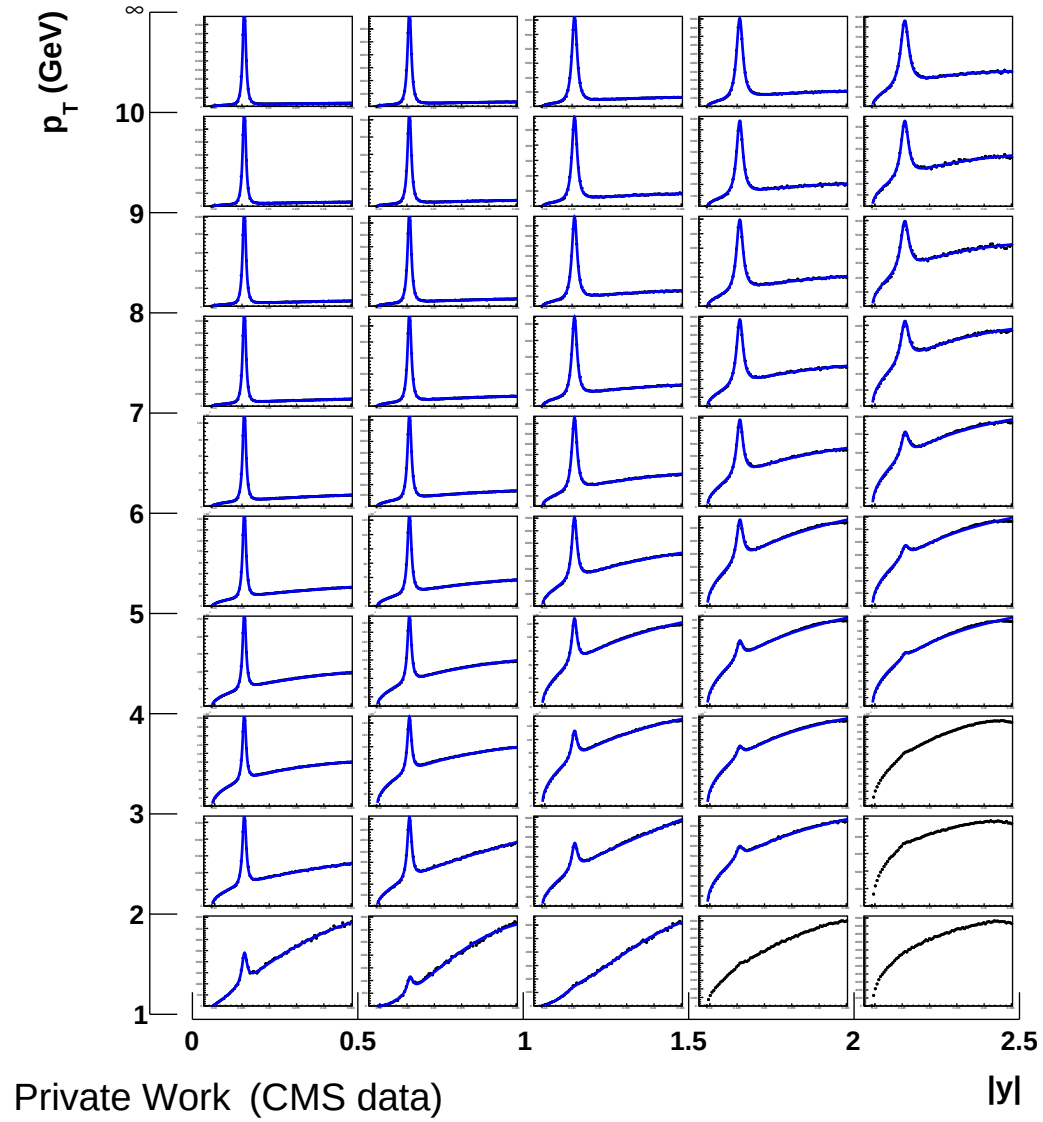


Figure D.21: D^{*+} signal on the 13 TeV data. The blue curves indicate fits to the right charge distribution.

D.2 13 TeV fit results

D.2.1 D^{*+} signal fit on data

In this appendix, D^{*+} signal on the 13 TeV data is summarized and shown for each bin of the cross section measurement. Fit results for D^{*+} are also shown as a reference to the background subtraction number, but were not used for determination of the cross section.

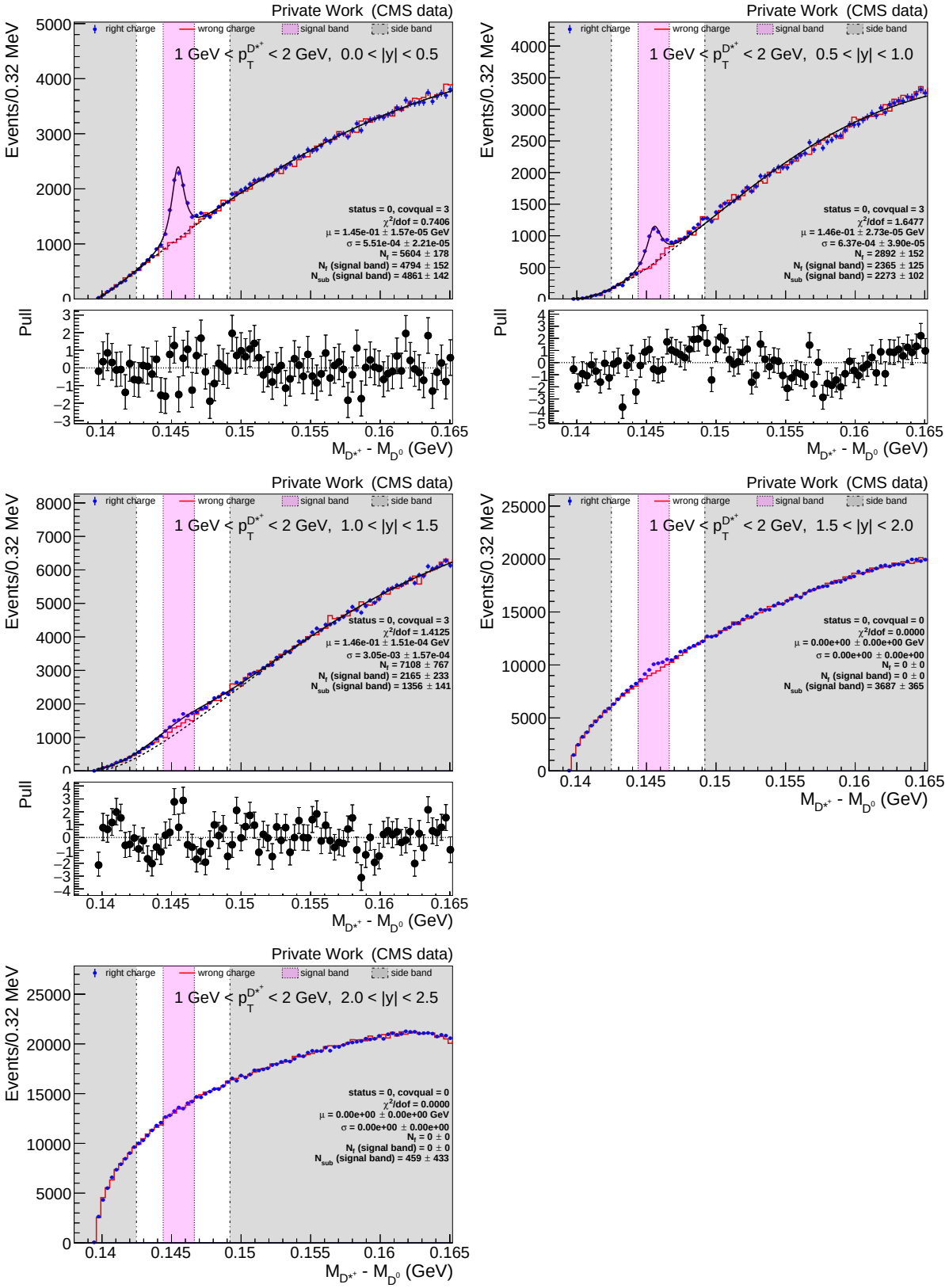


Figure D.22: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

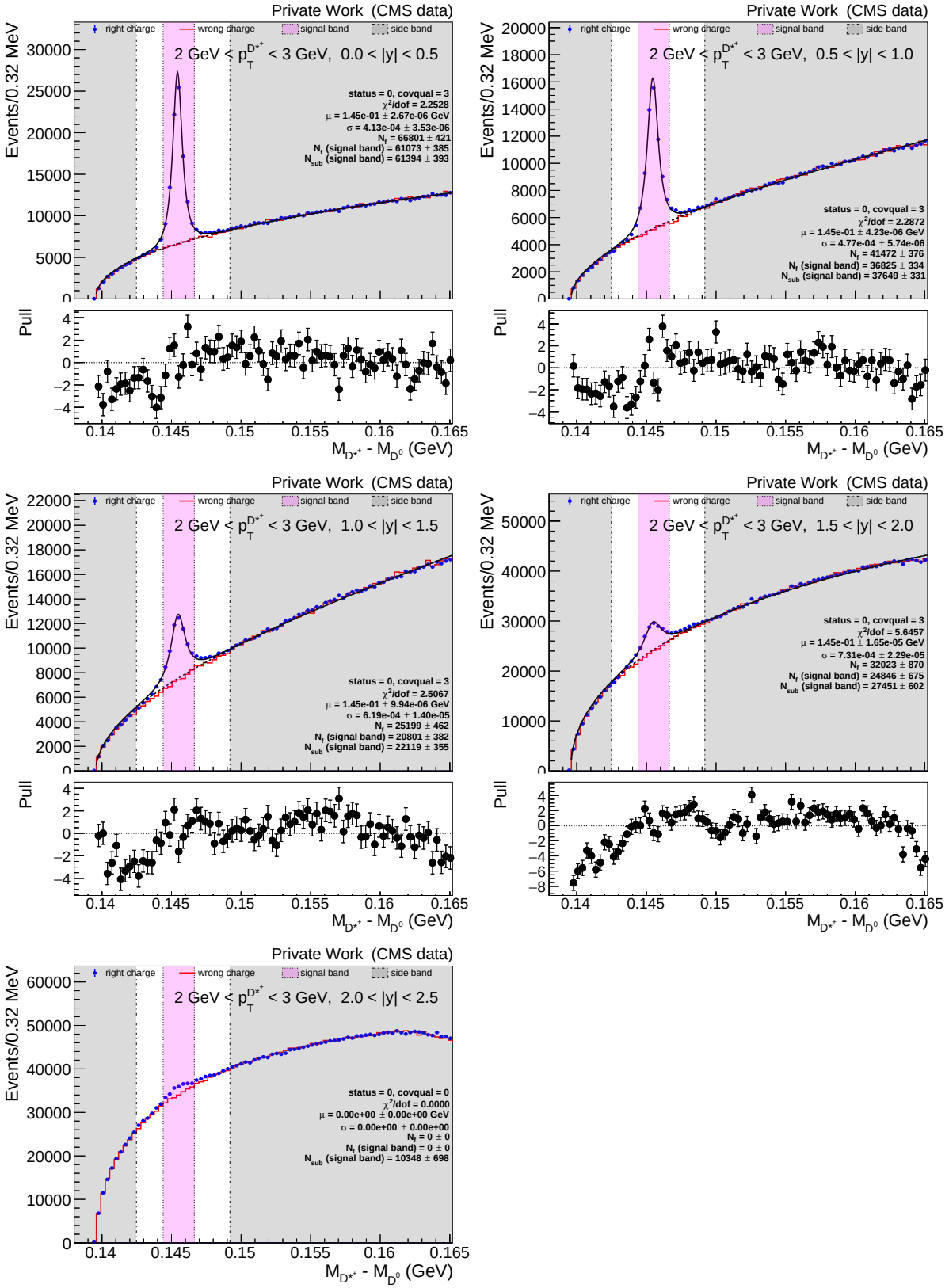


Figure D.23: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

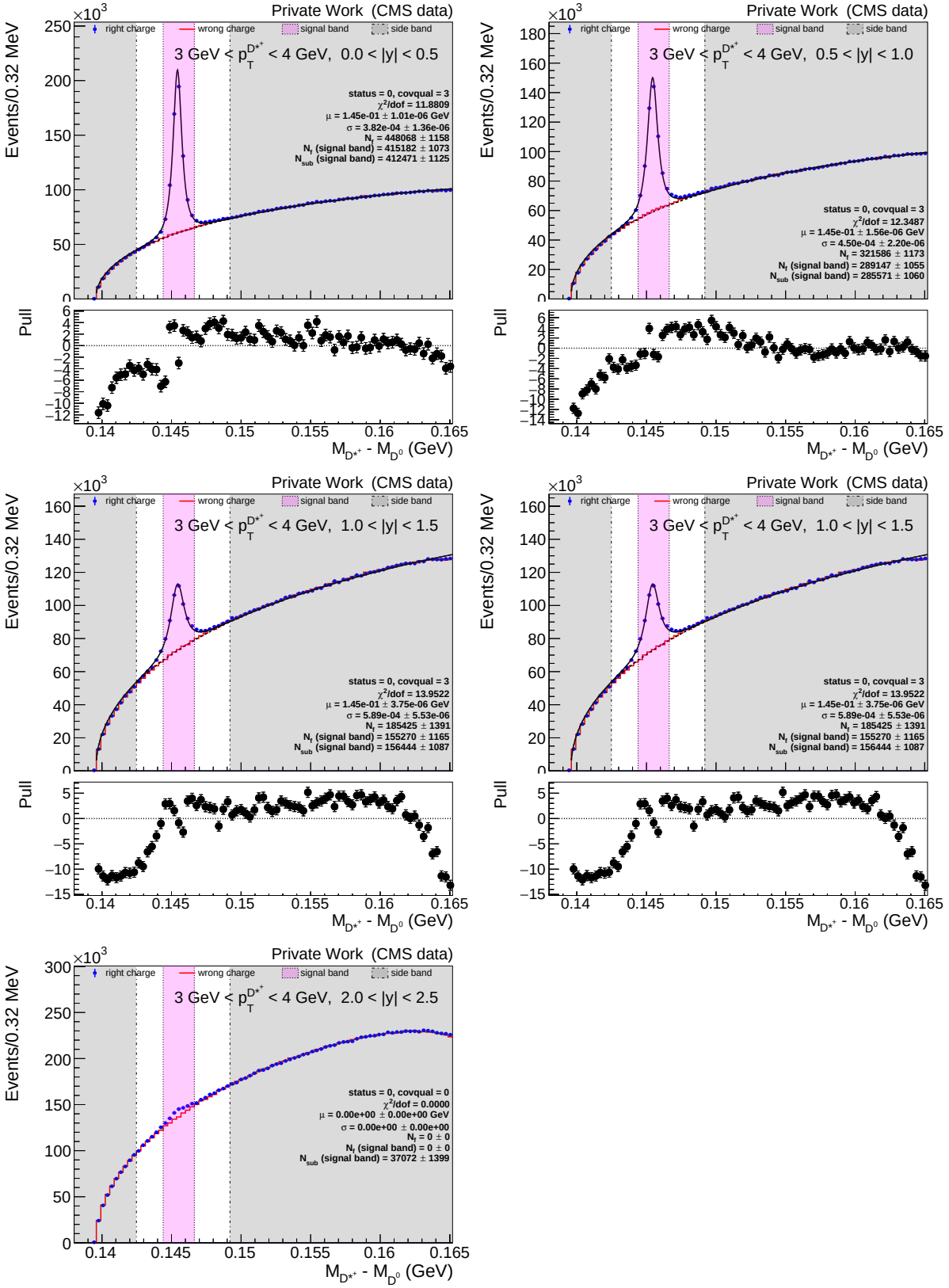


Figure D.24: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

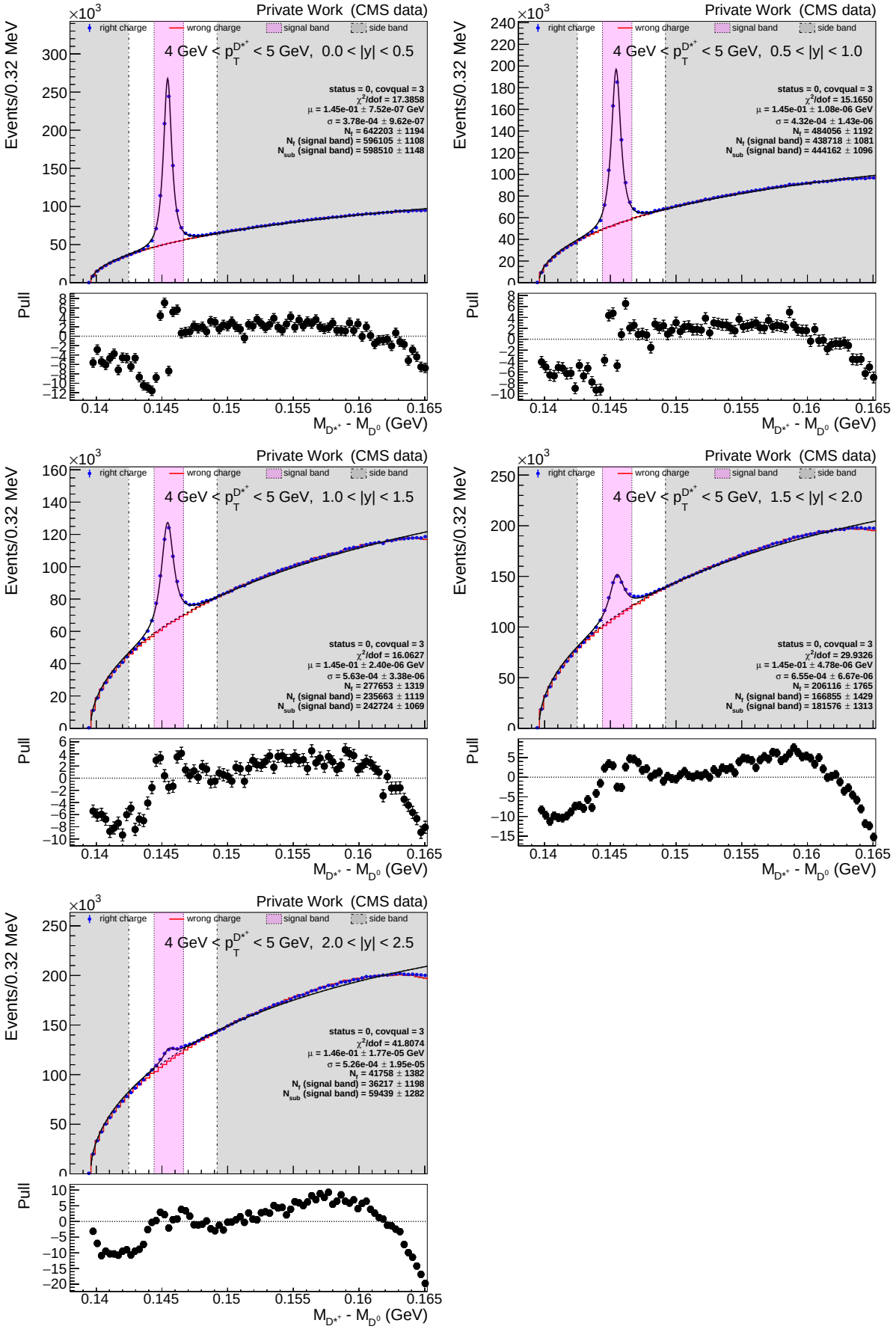


Figure D.25: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

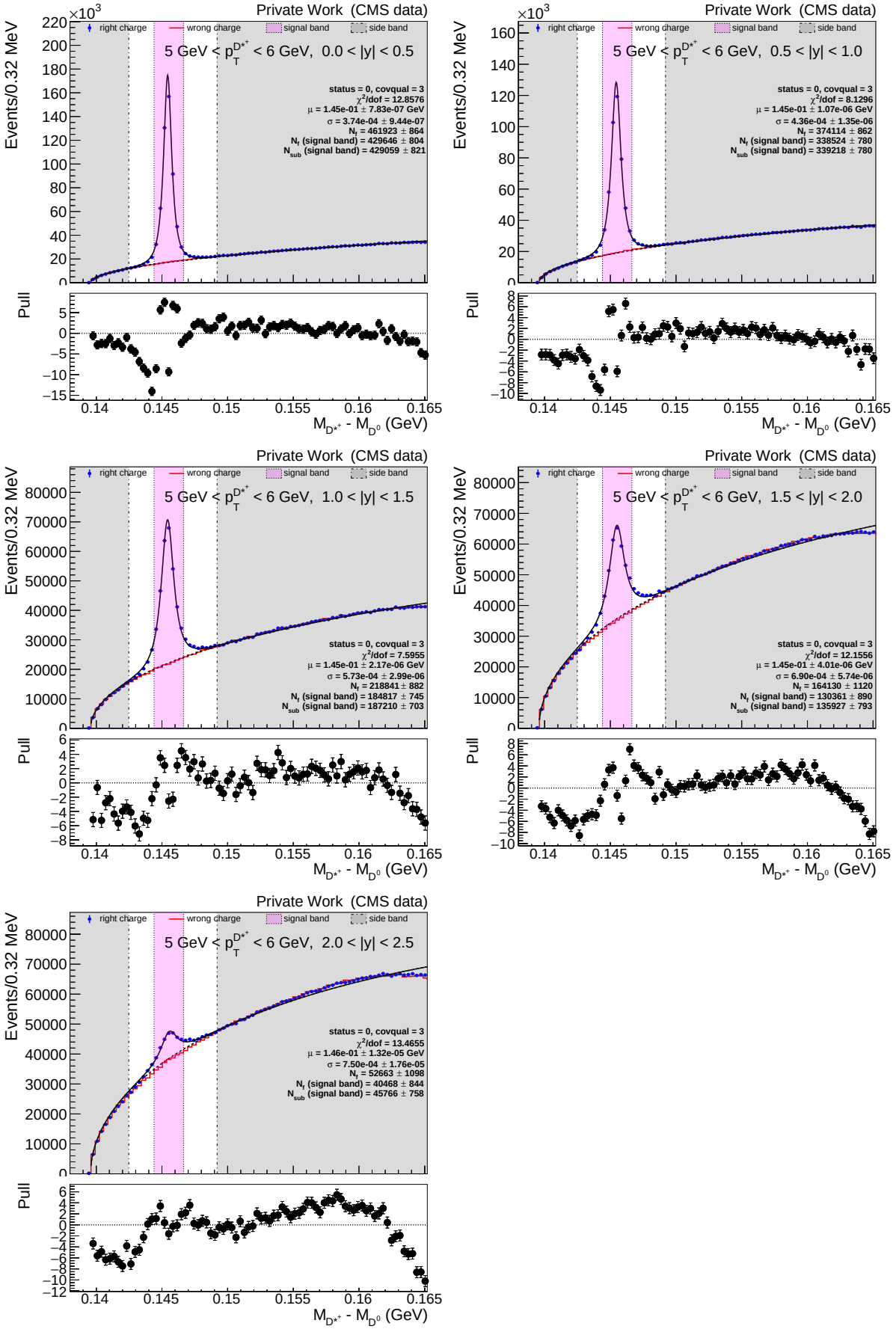


Figure D.26: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

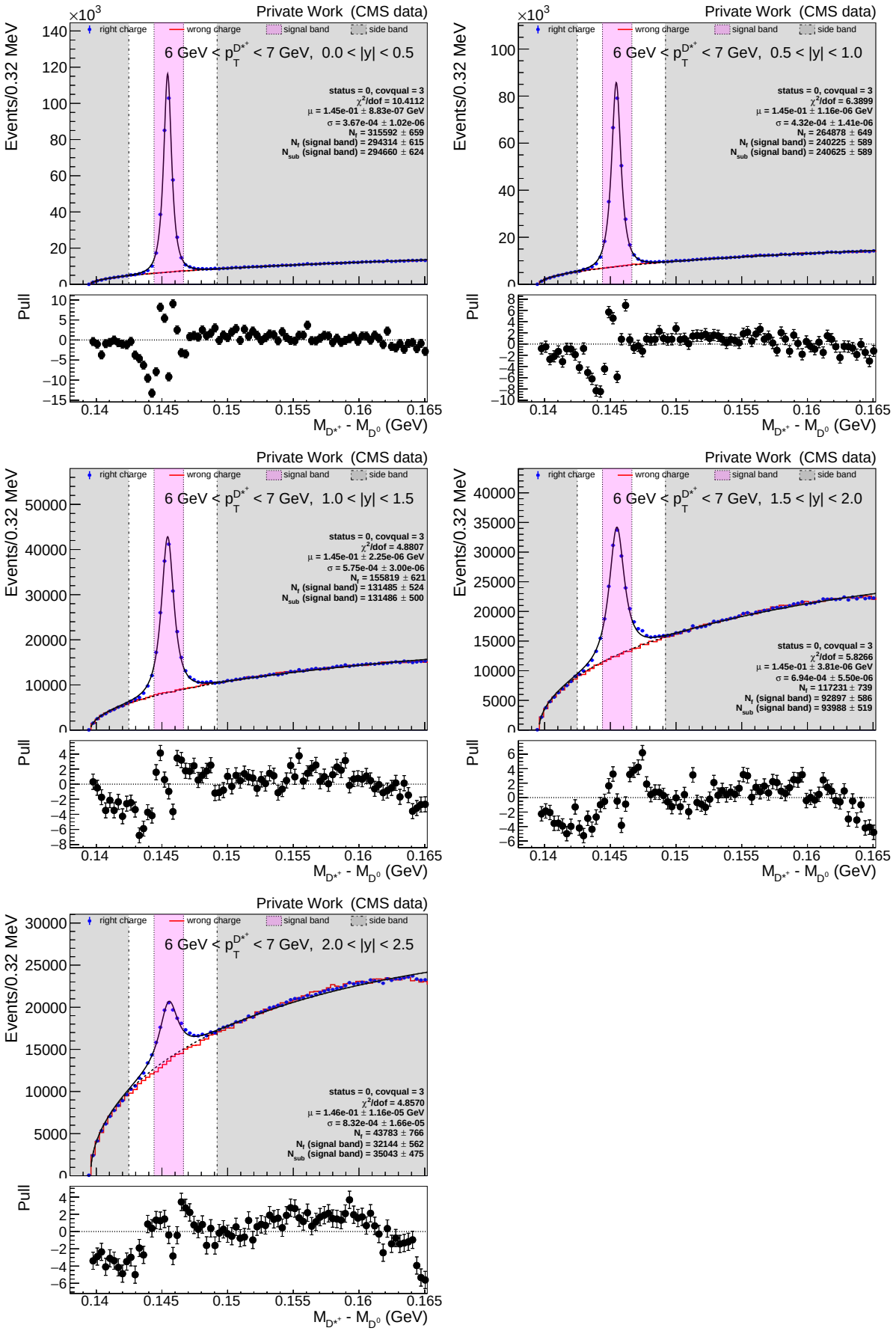


Figure D.27: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

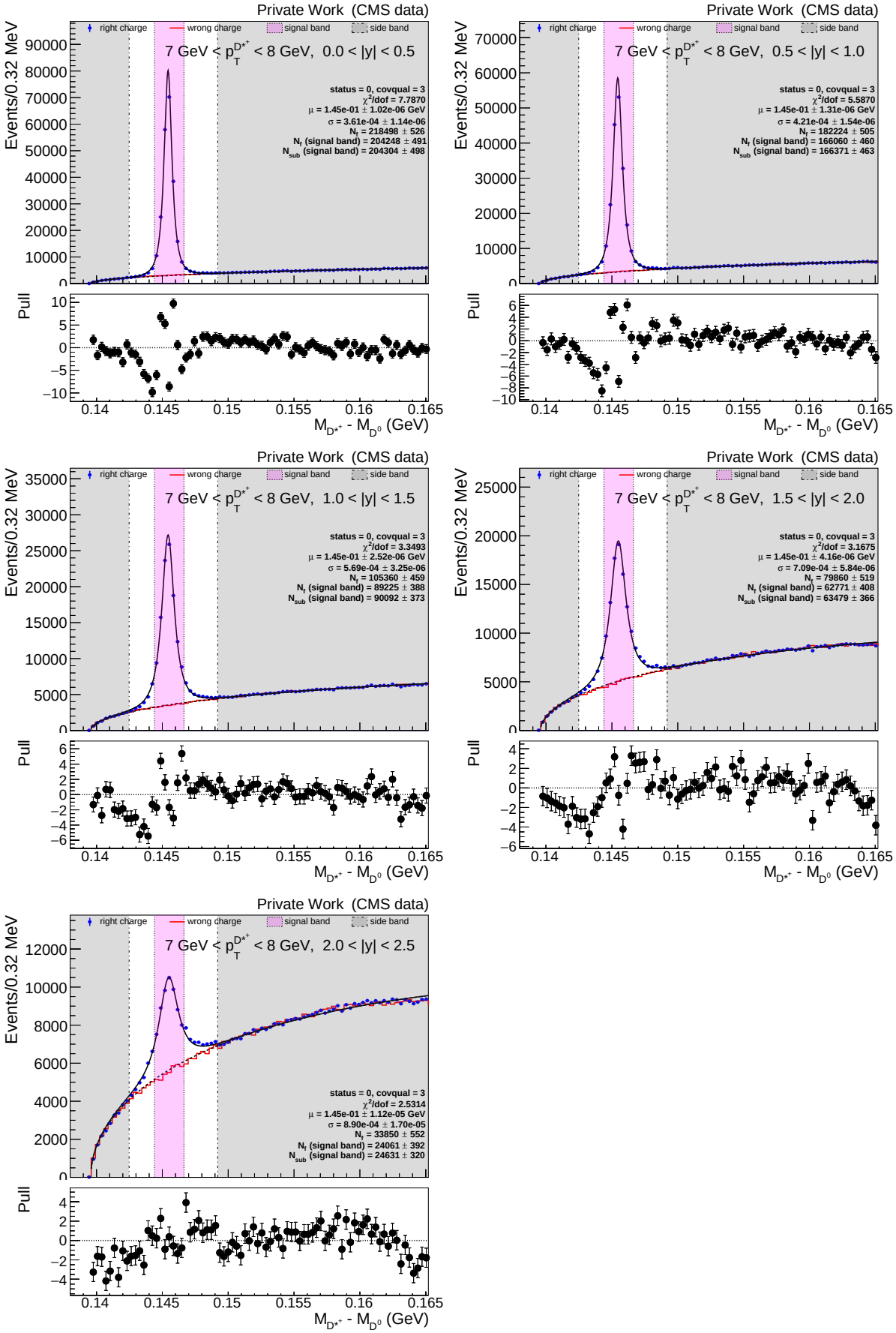


Figure D.28: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

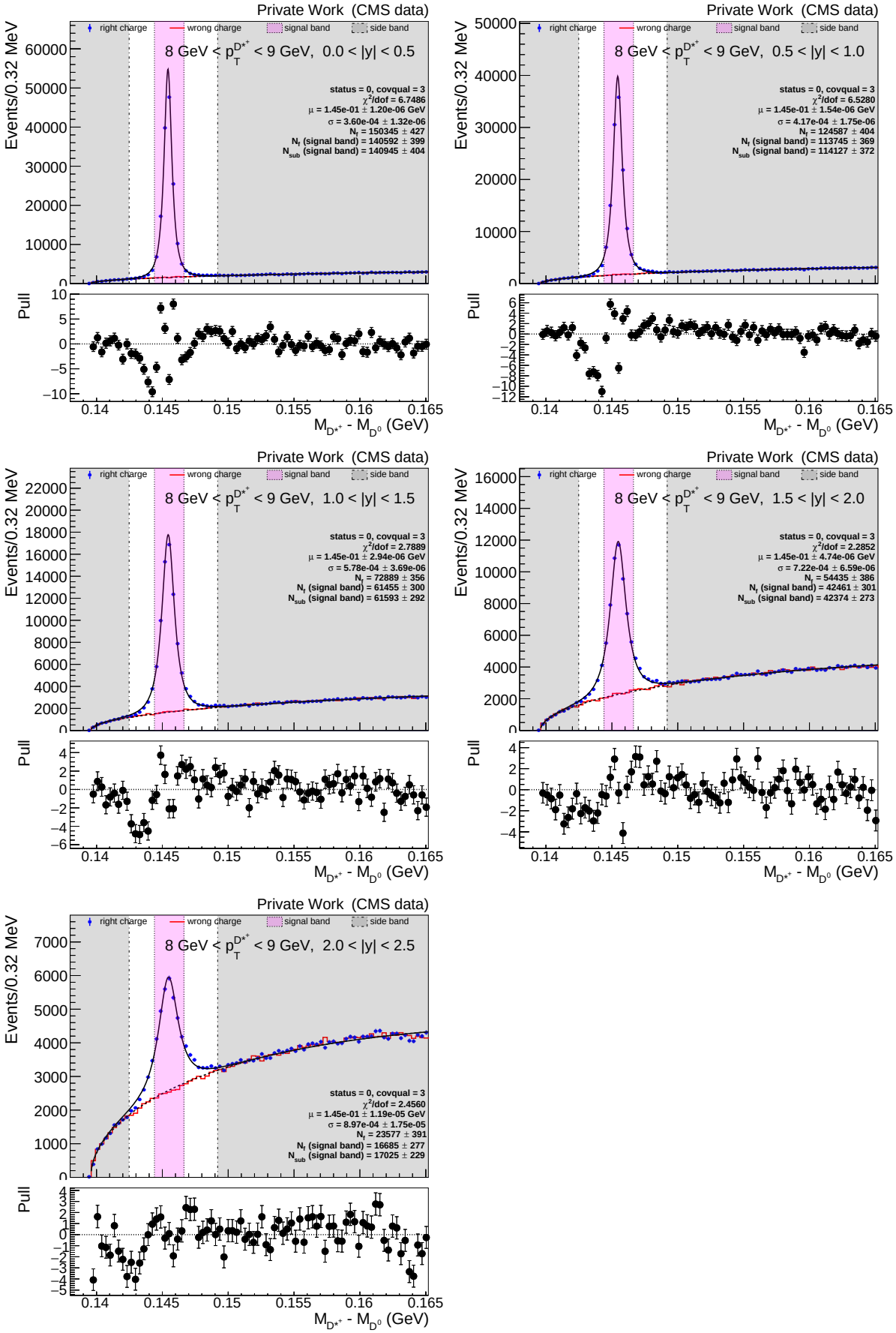


Figure D.29: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

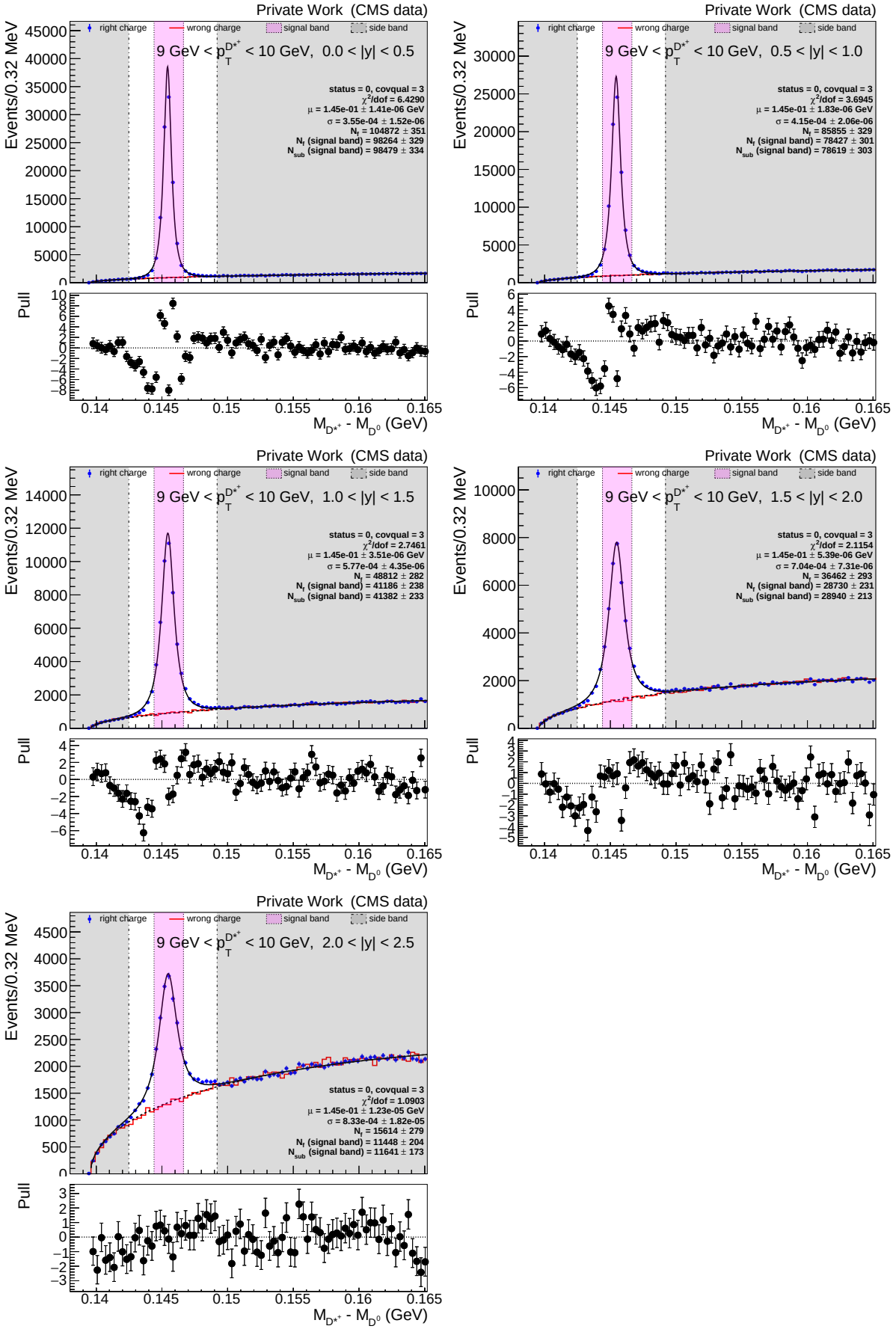


Figure D.30: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

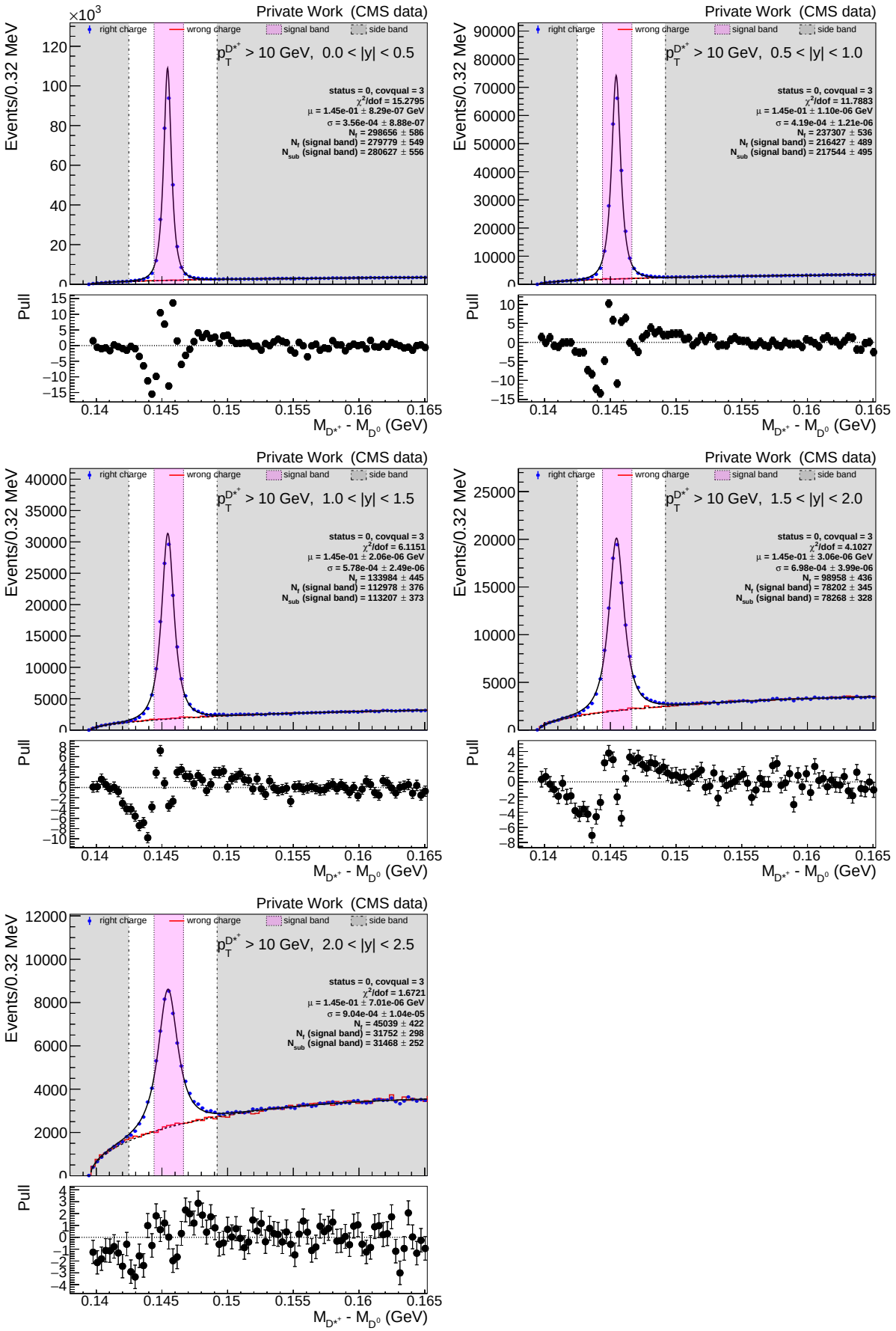


Figure D.31: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

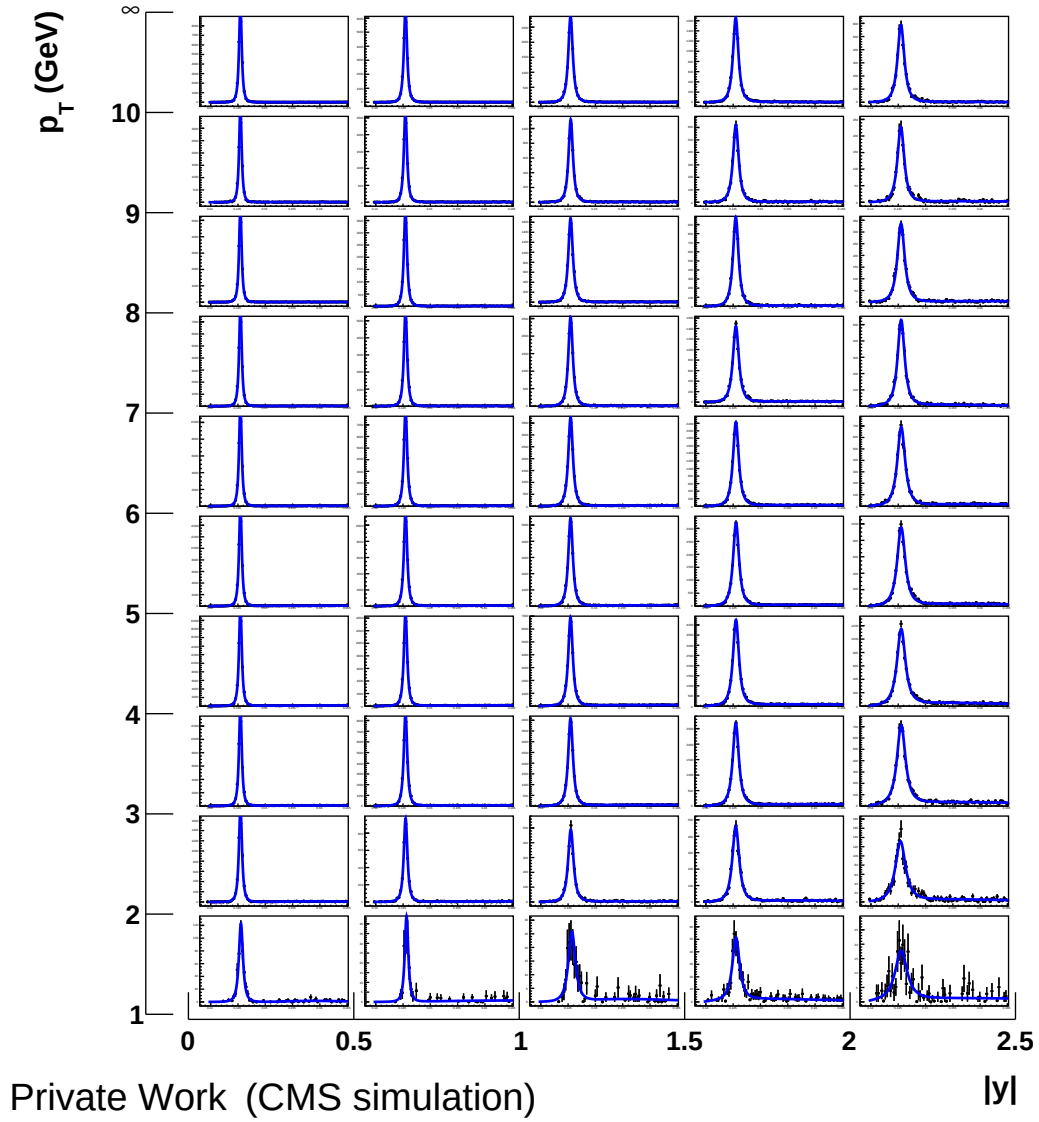


Figure D.32: D^{*+} signal on the 13 TeV D^0 MC sample. The blue curves indicate fits to the right charge distribution.

D.2.2 D^{*+} signal fit on MC

In this appendix, D^{*+} signal on the 13 TeV D^0 MC sample is summarized and shown for each bin of the cross section measurement. Fit results for D^{*+} are also shown as a reference to the background subtraction number, but were not used for determination of the cross section. In this case, χ^2/ndof results are sometimes very large as a result of reweighting events.

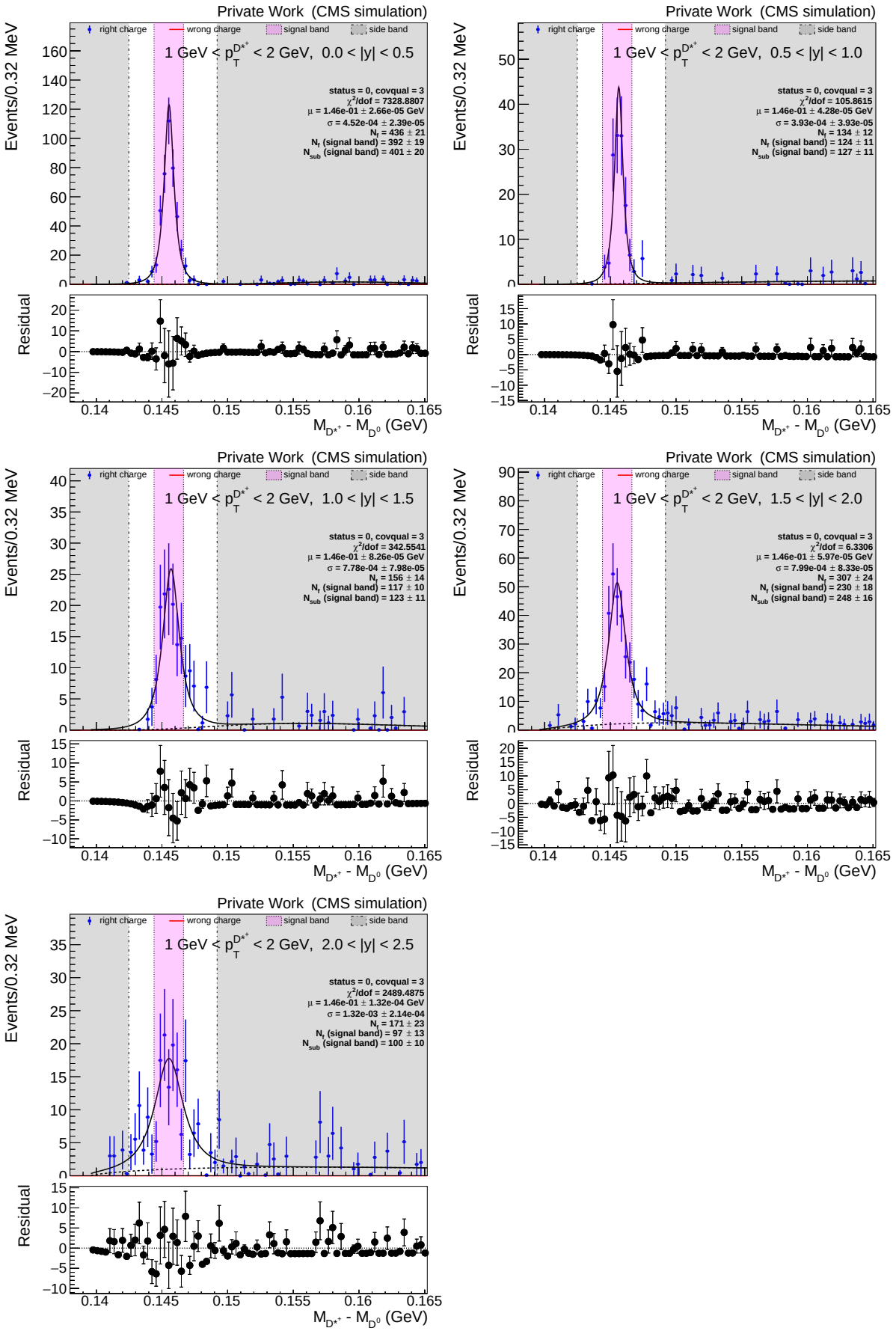


Figure D.33: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

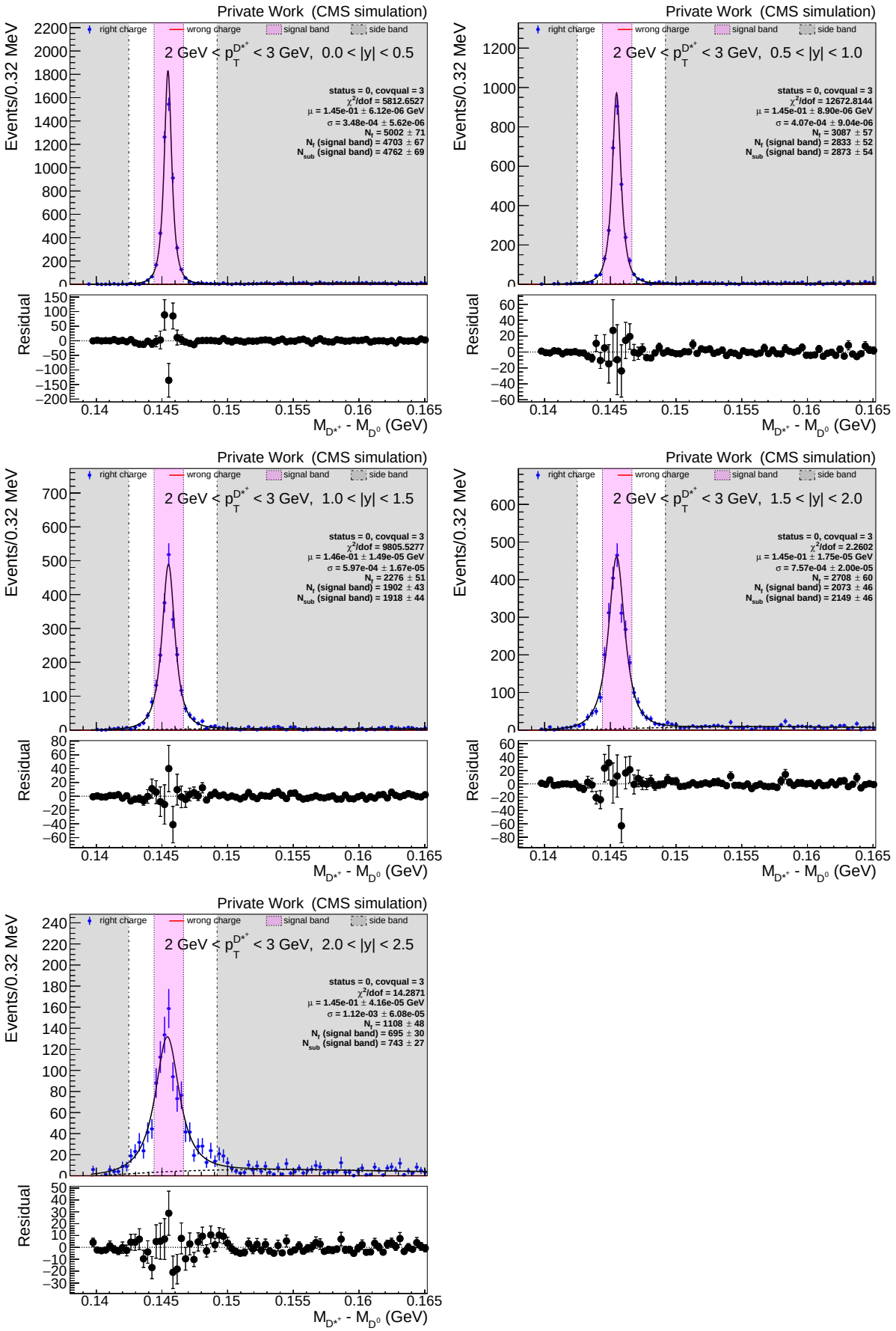


Figure D.34: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

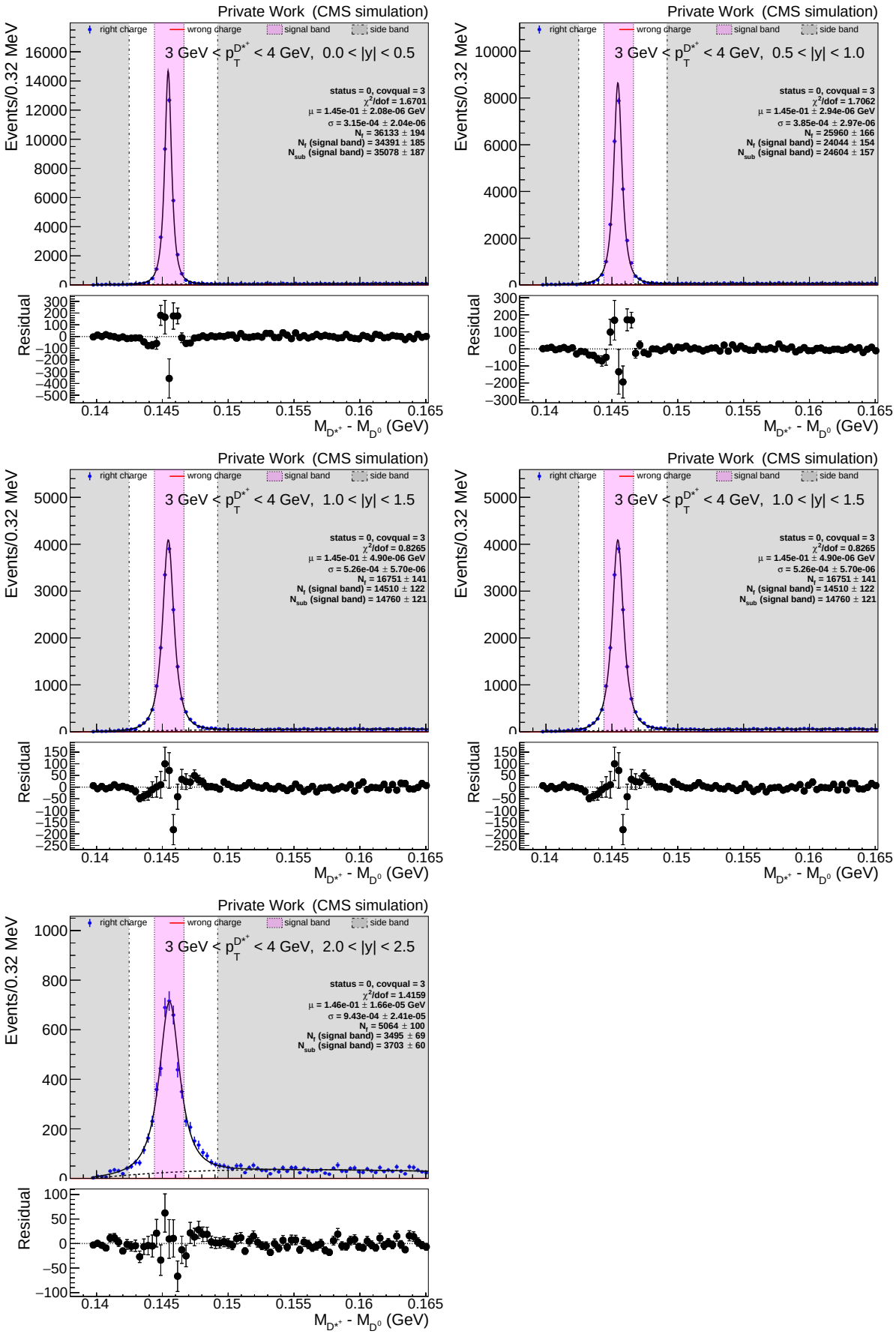


Figure D.35: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

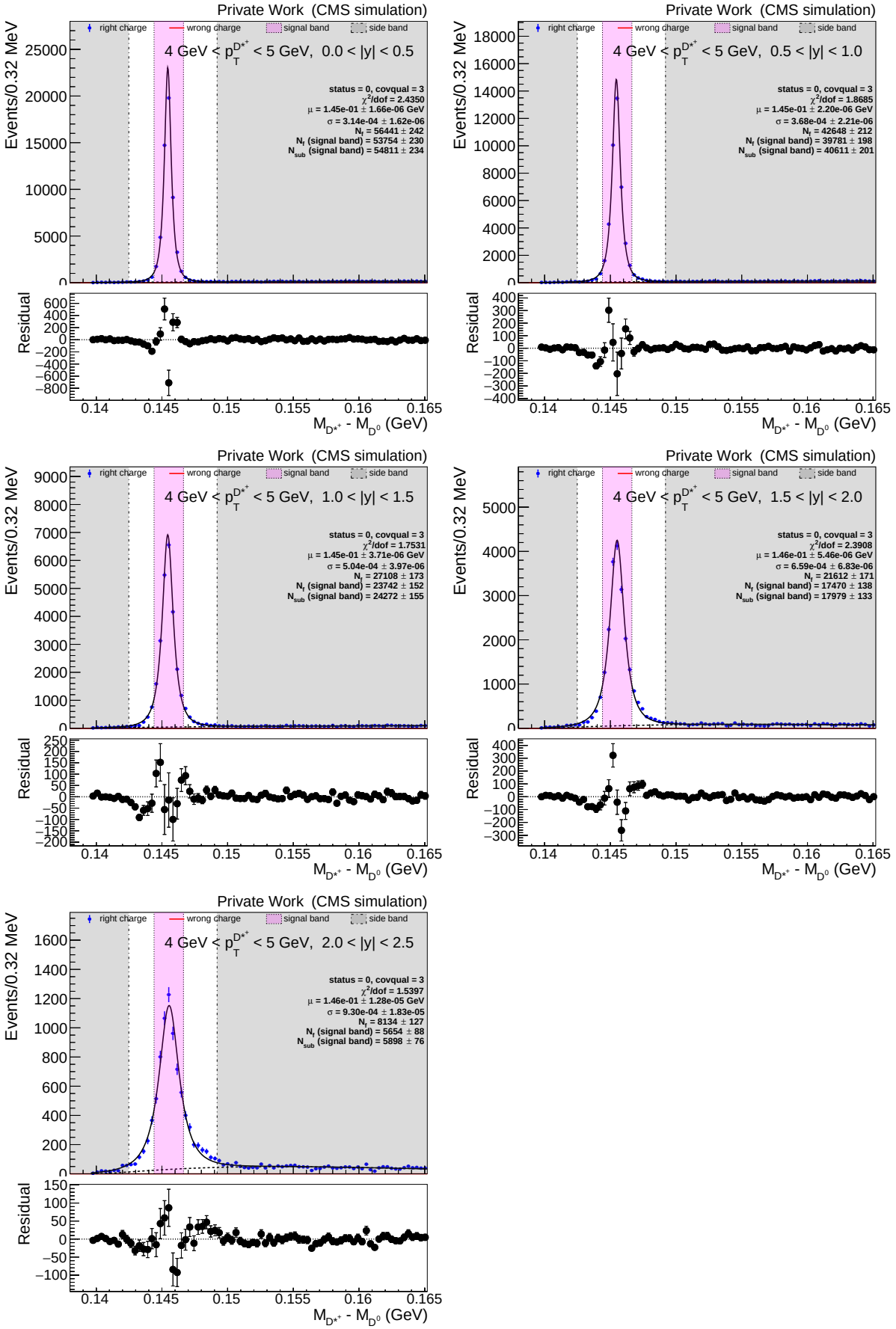


Figure D.36: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

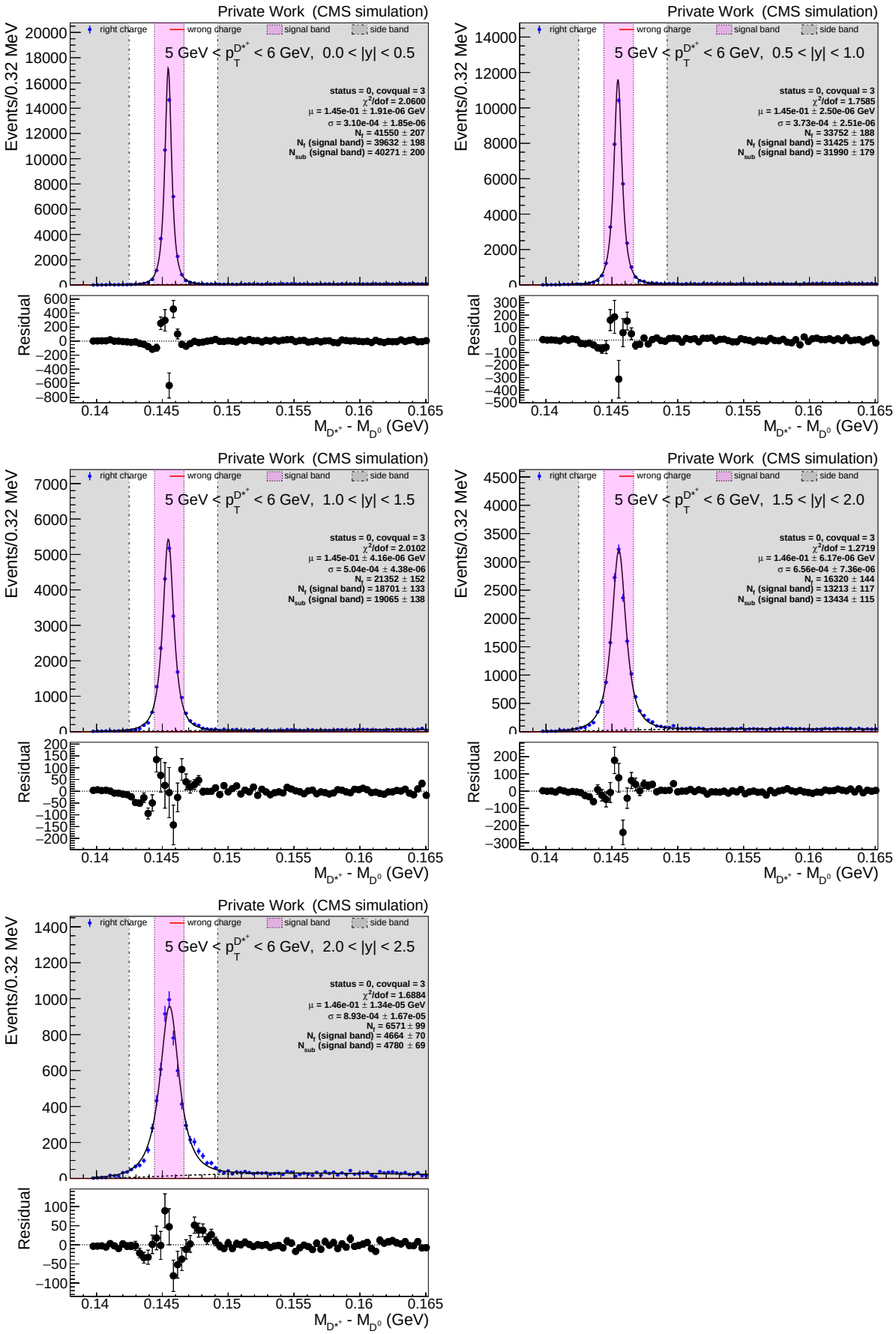


Figure D.37: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

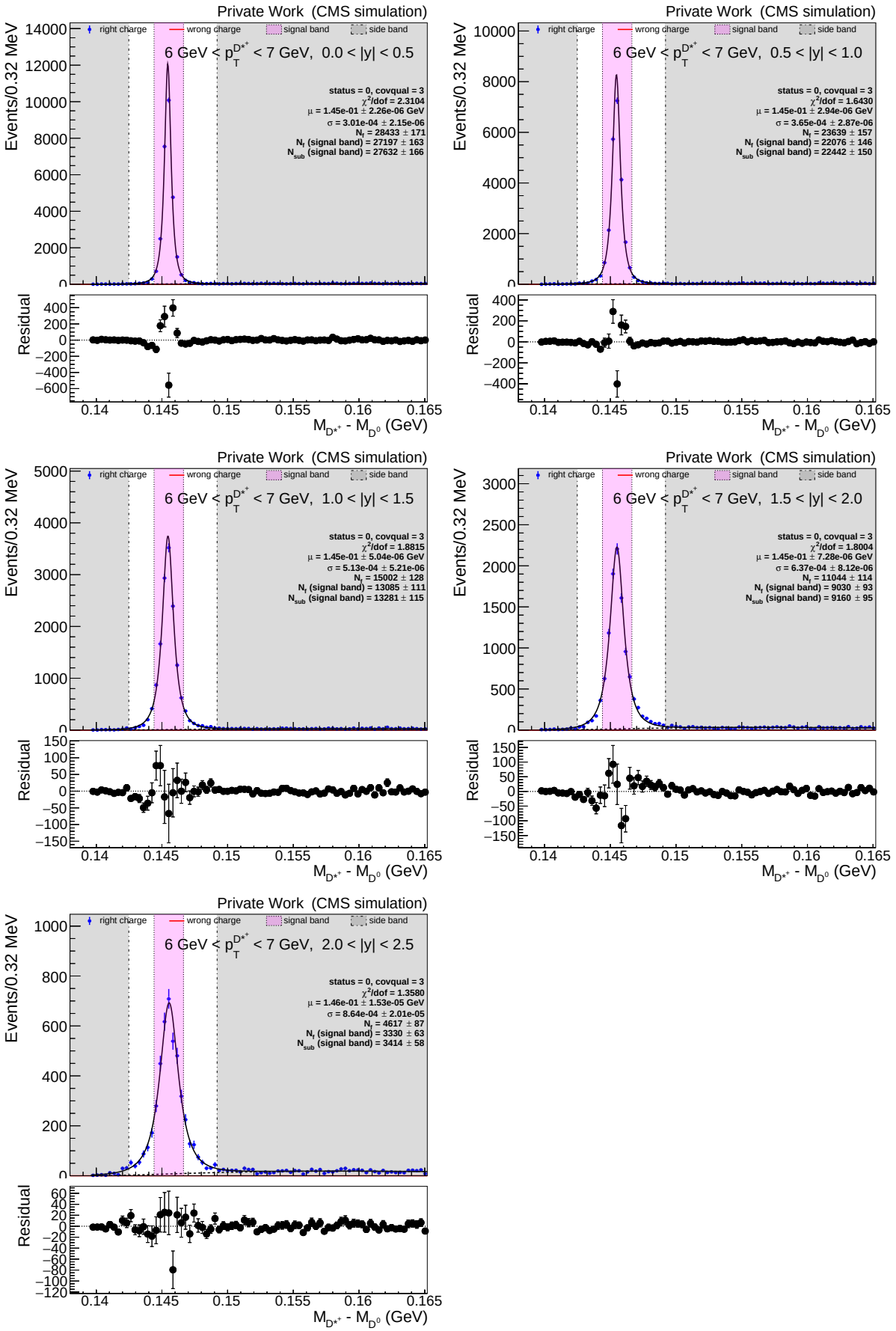


Figure D.38: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

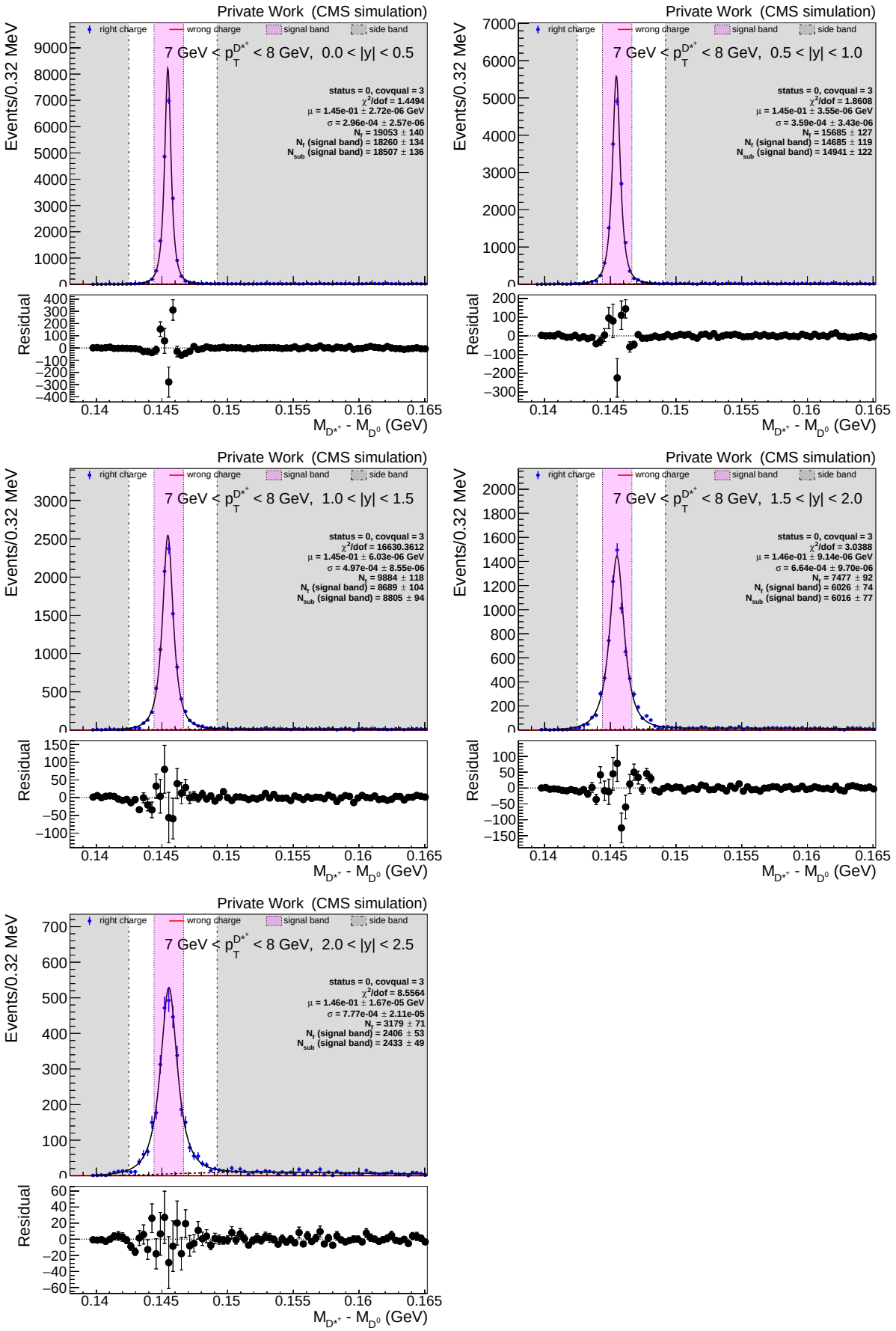


Figure D.39: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

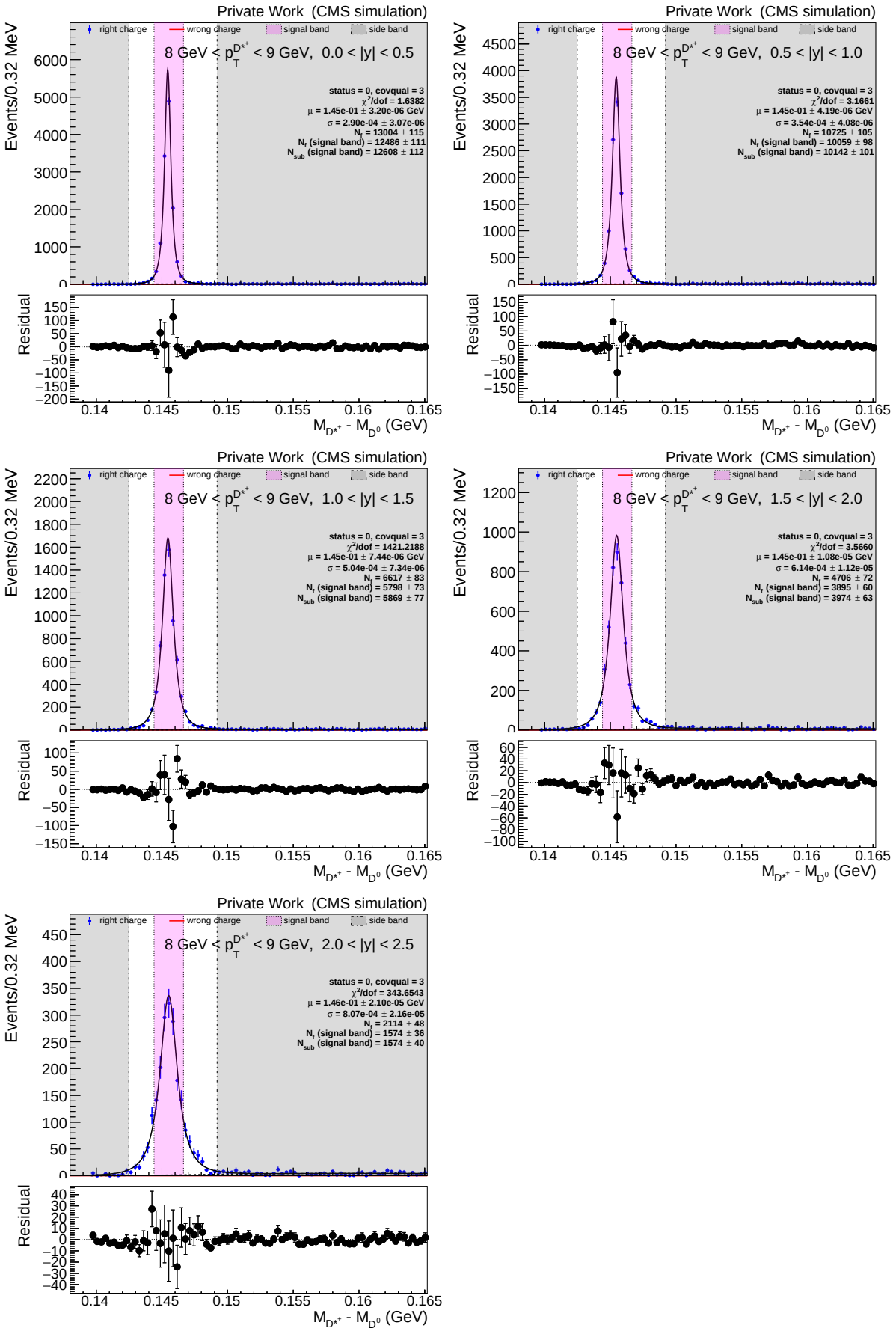


Figure D.40: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

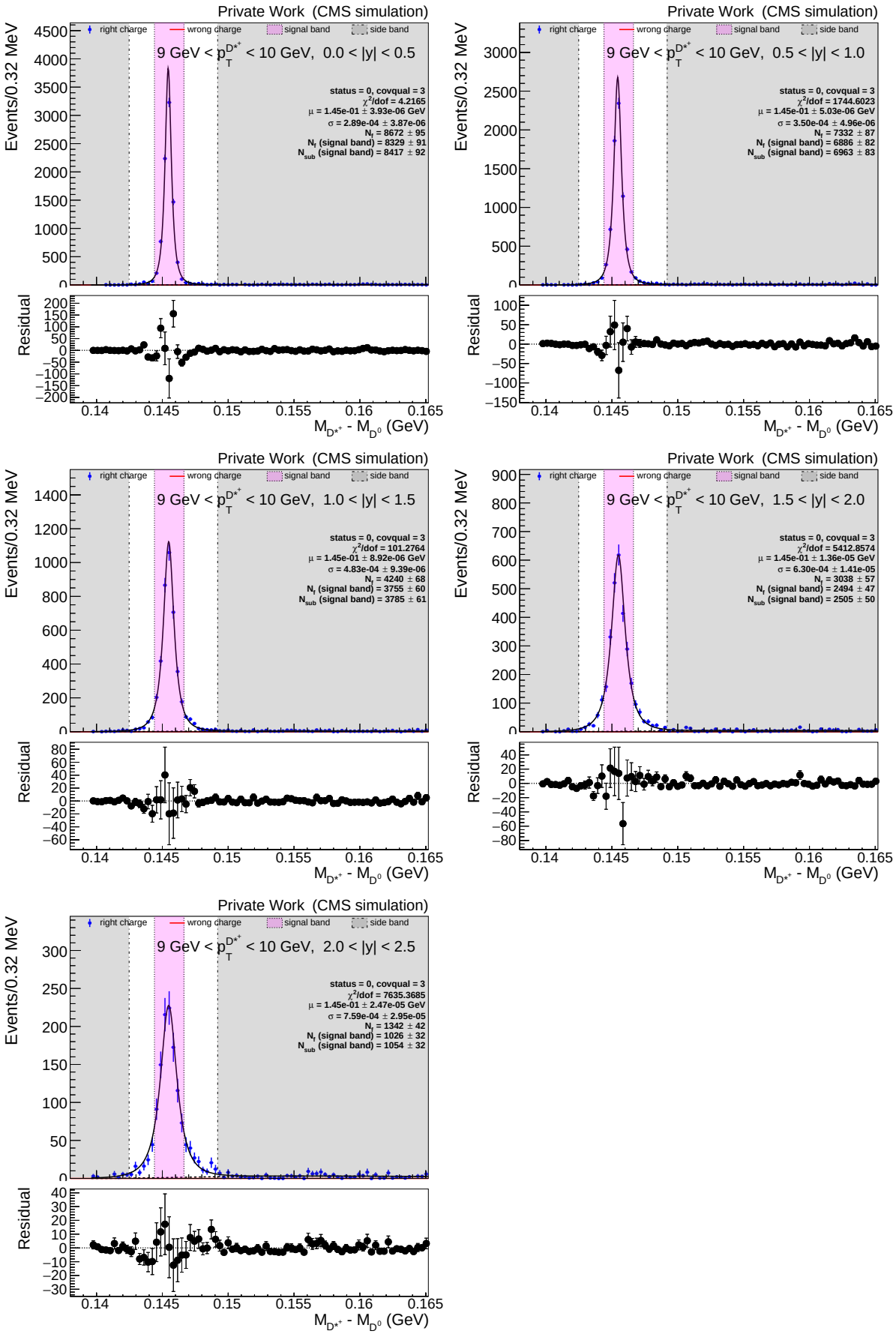


Figure D.41: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

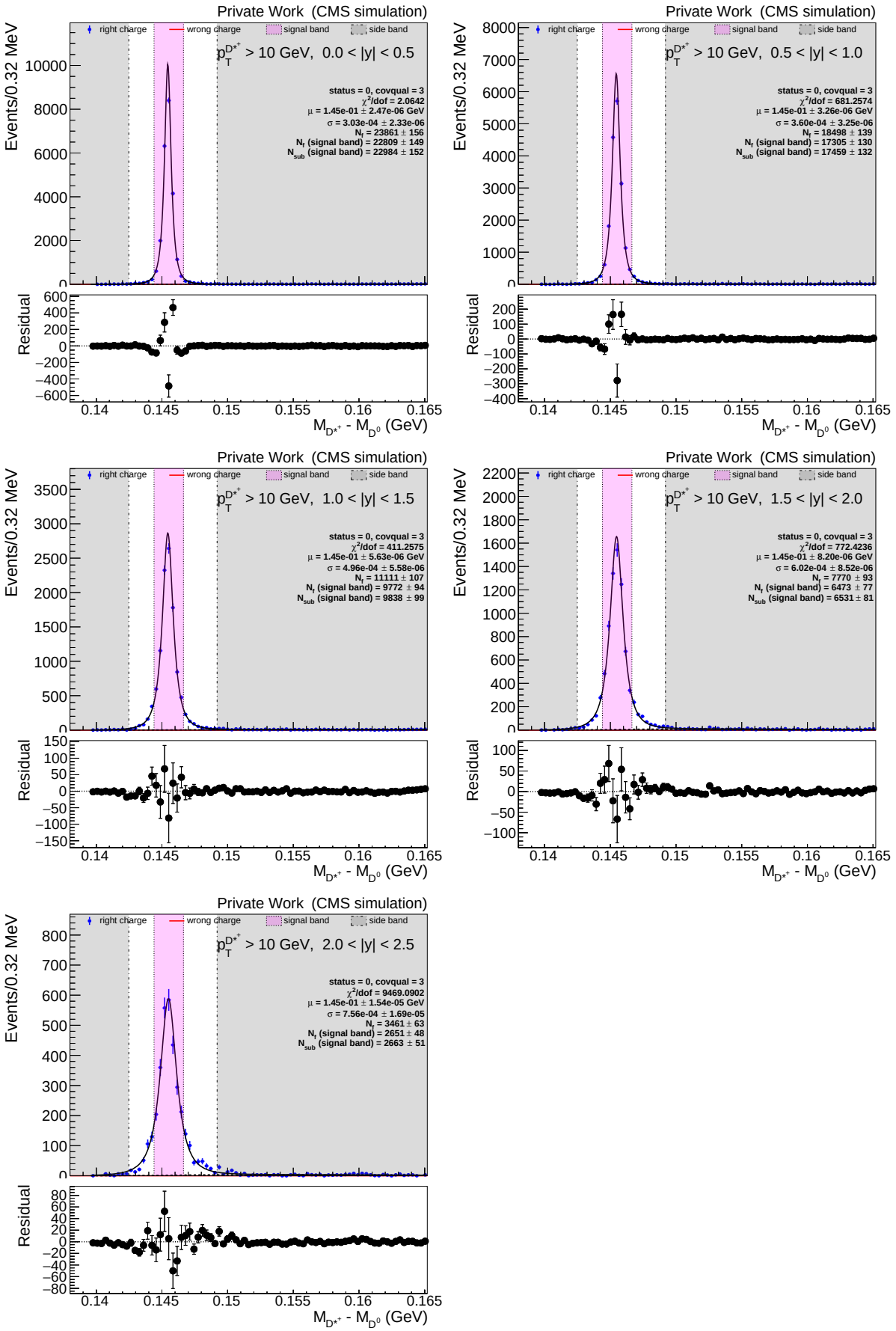


Figure D.42: $M_{D^{*+}} - M_{D^0}$ distributions and fit.

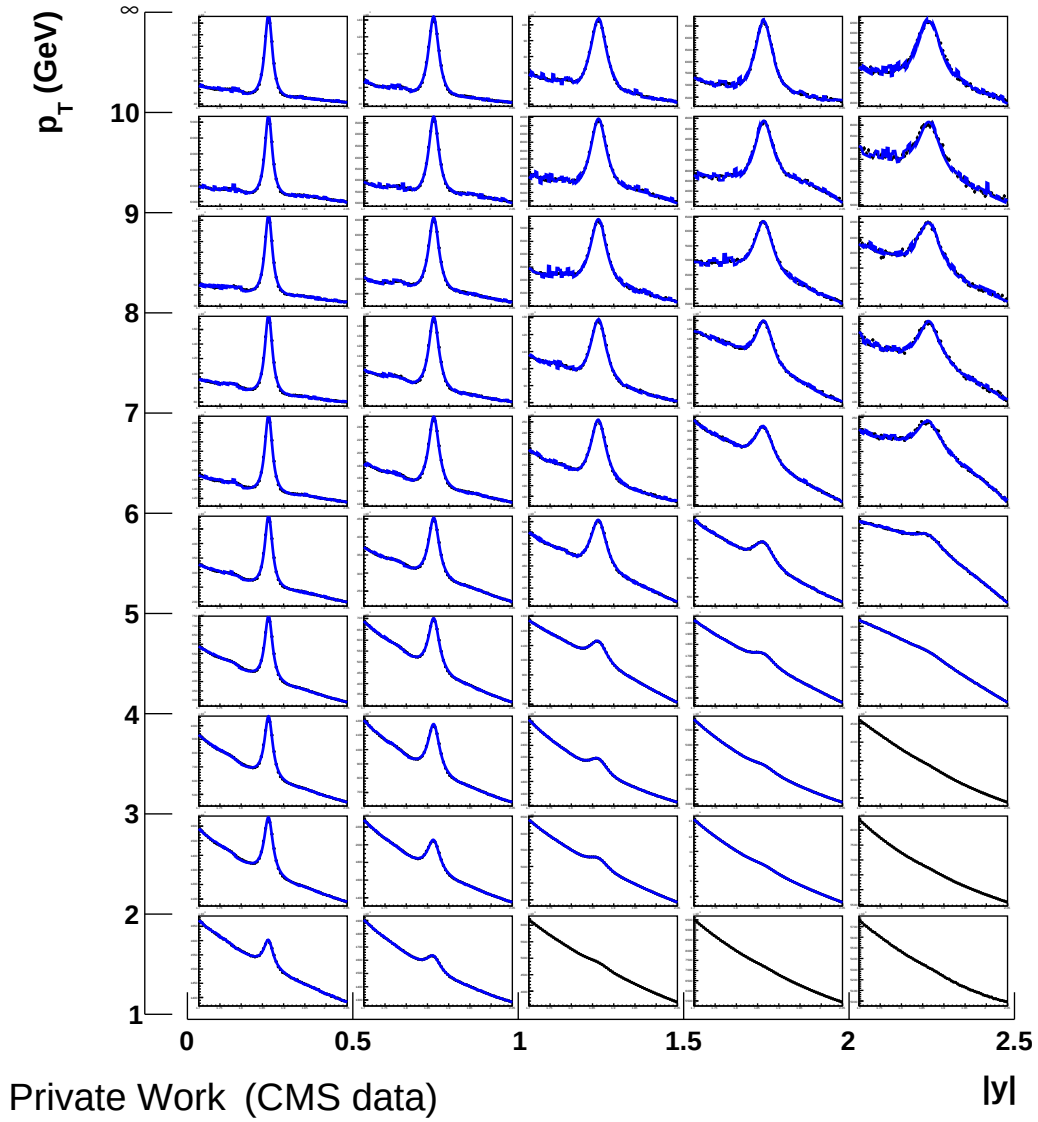


Figure D.43: D^0 signal on the 13 TeV data. The blue curves indicate fits to M_{D^0} mass distribution.

D.2.3 D^0 signal fit on data

In this appendix, D^0 signal on the 13 TeV data is summarized and shown for each bin of the cross section measurement including fit results which were used for the cross section measurement. In each figure, MC templates used for the data fit are also shown.

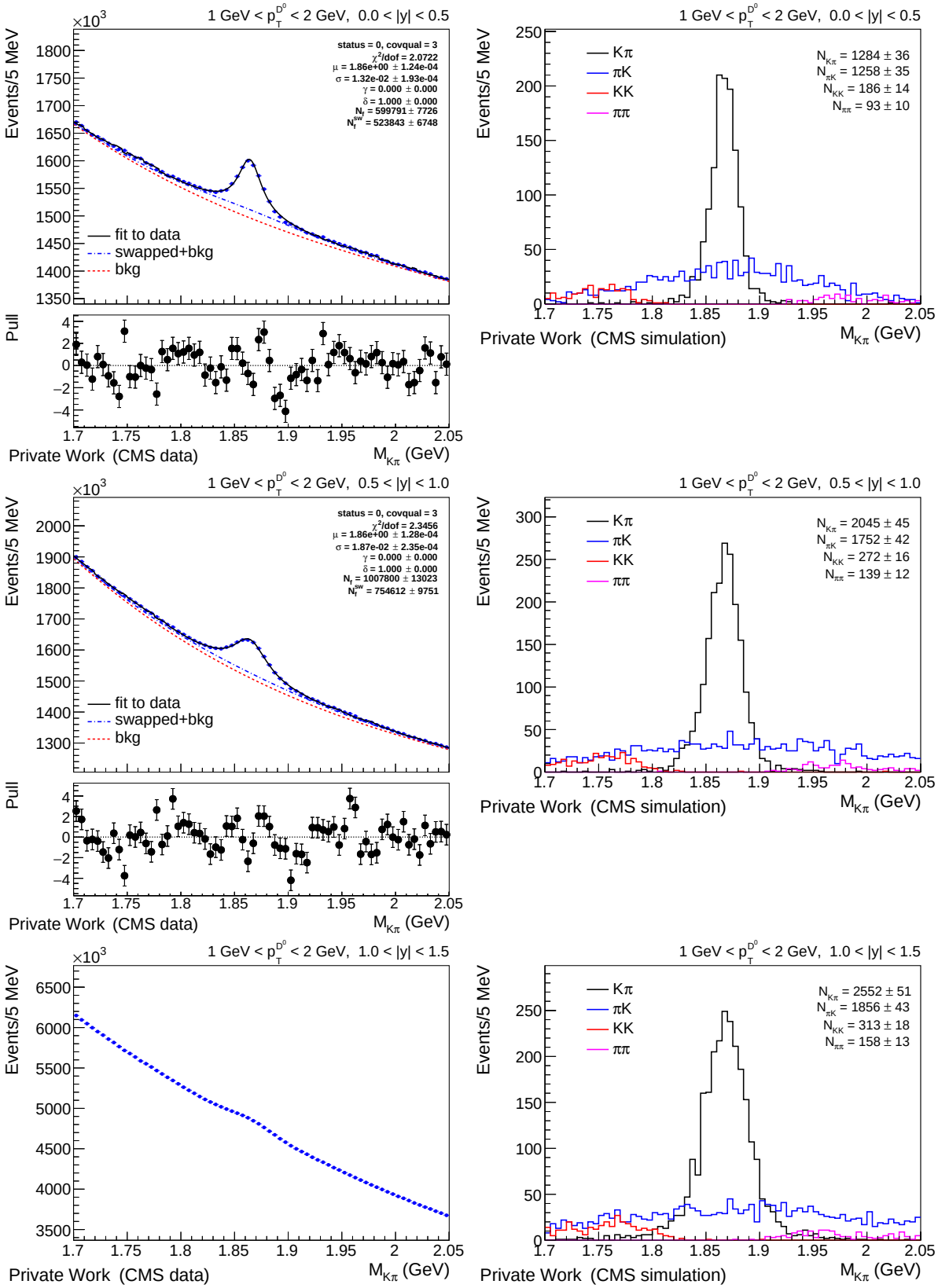


Figure D.44: M_{D^0} distributions and fit on data (left), and MC template (right).

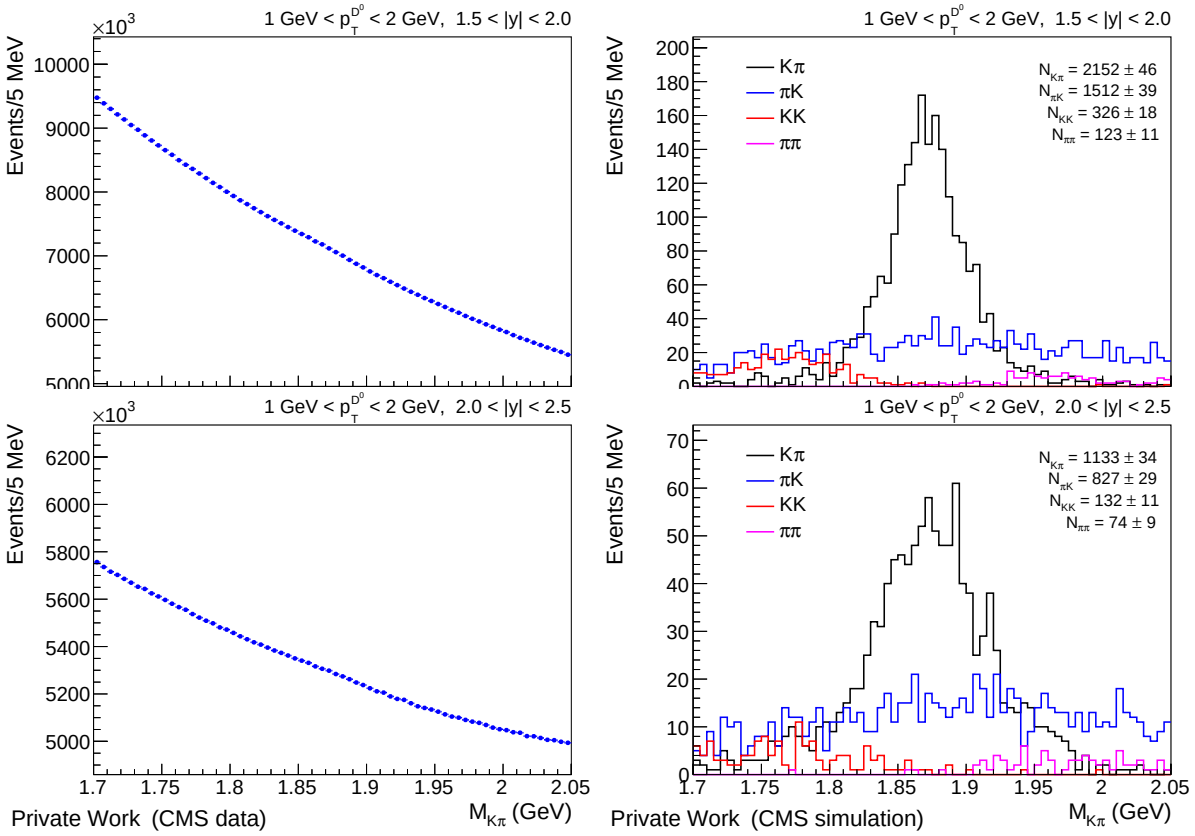


Figure D.45: M_{D^0} distributions and fit on data (left), and MC template (right).

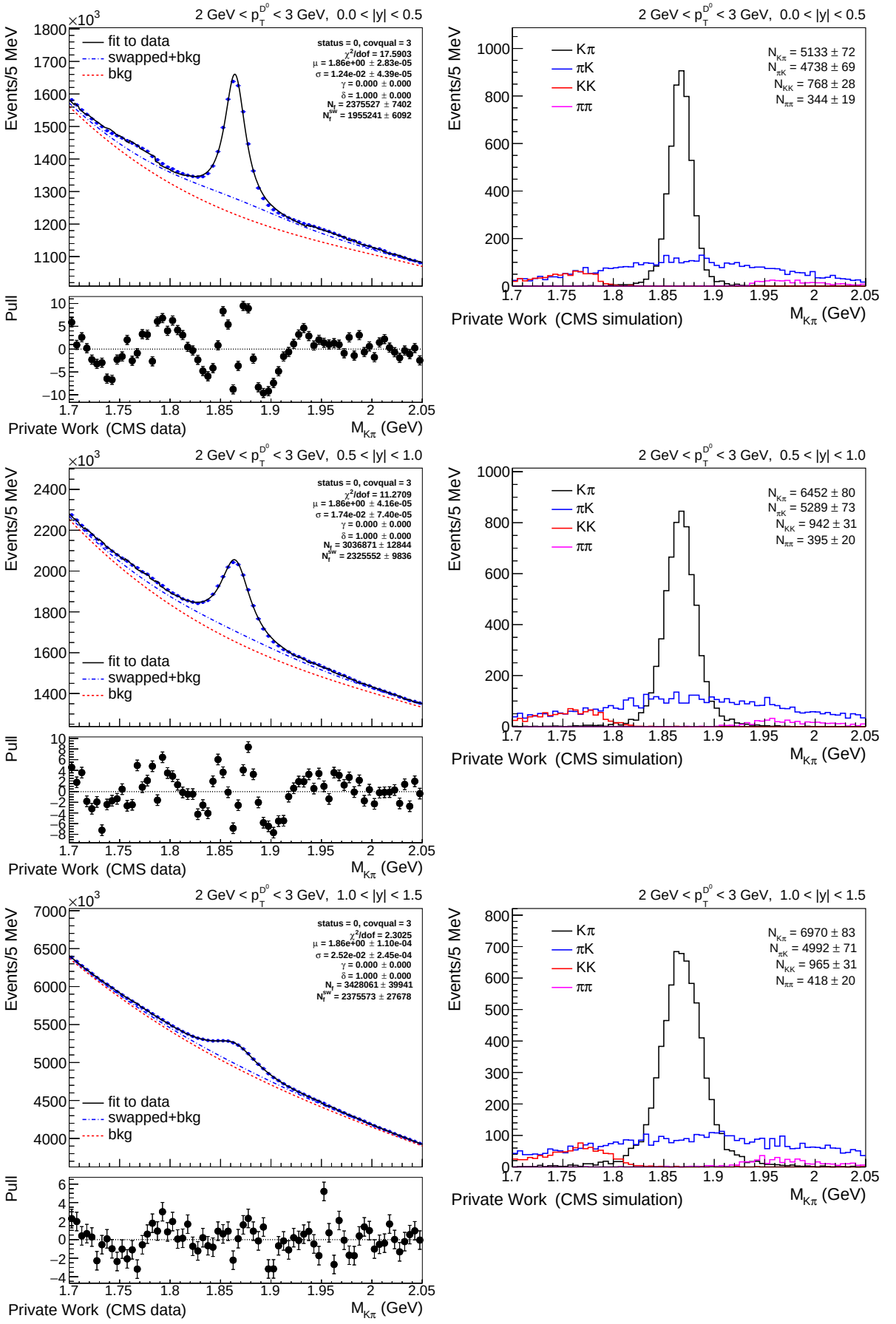


Figure D.46: M_{D^0} distributions and fit on data (left), and MC template (right).

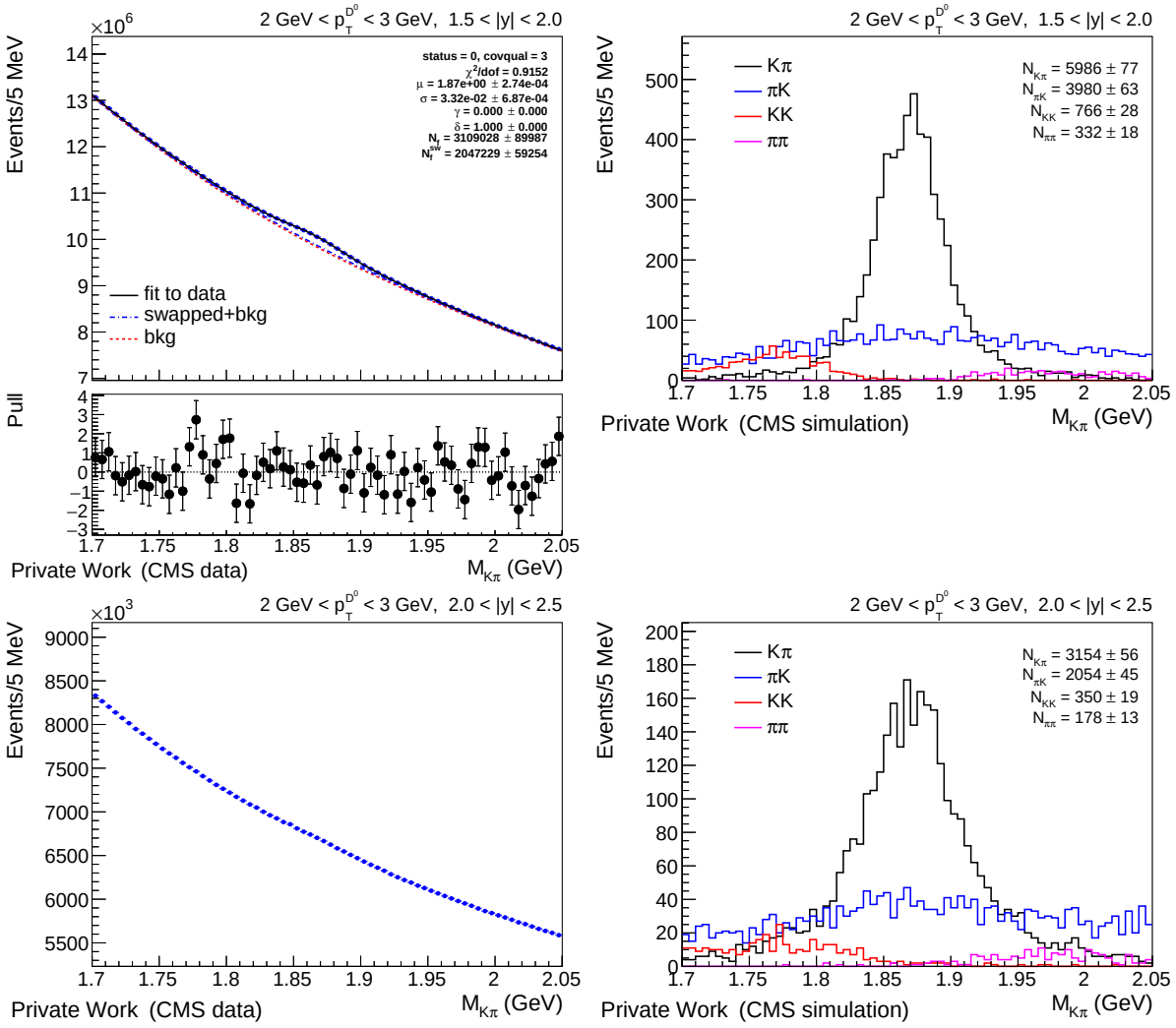


Figure D.47: M_{D^0} distributions and fit on data (left), and MC template (right).

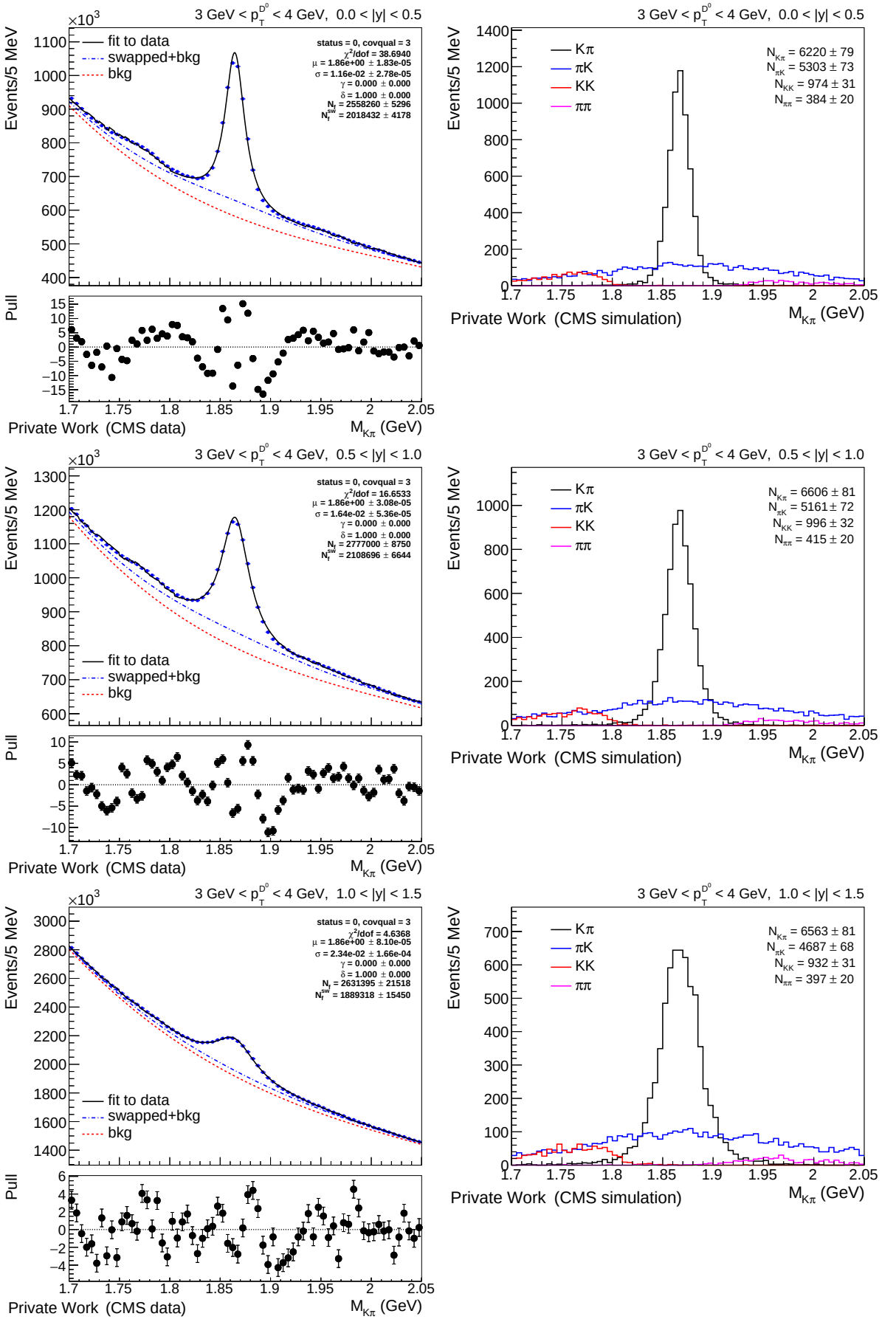


Figure D.48: M_{D^0} distributions and fit on data (left), and MC template (right).

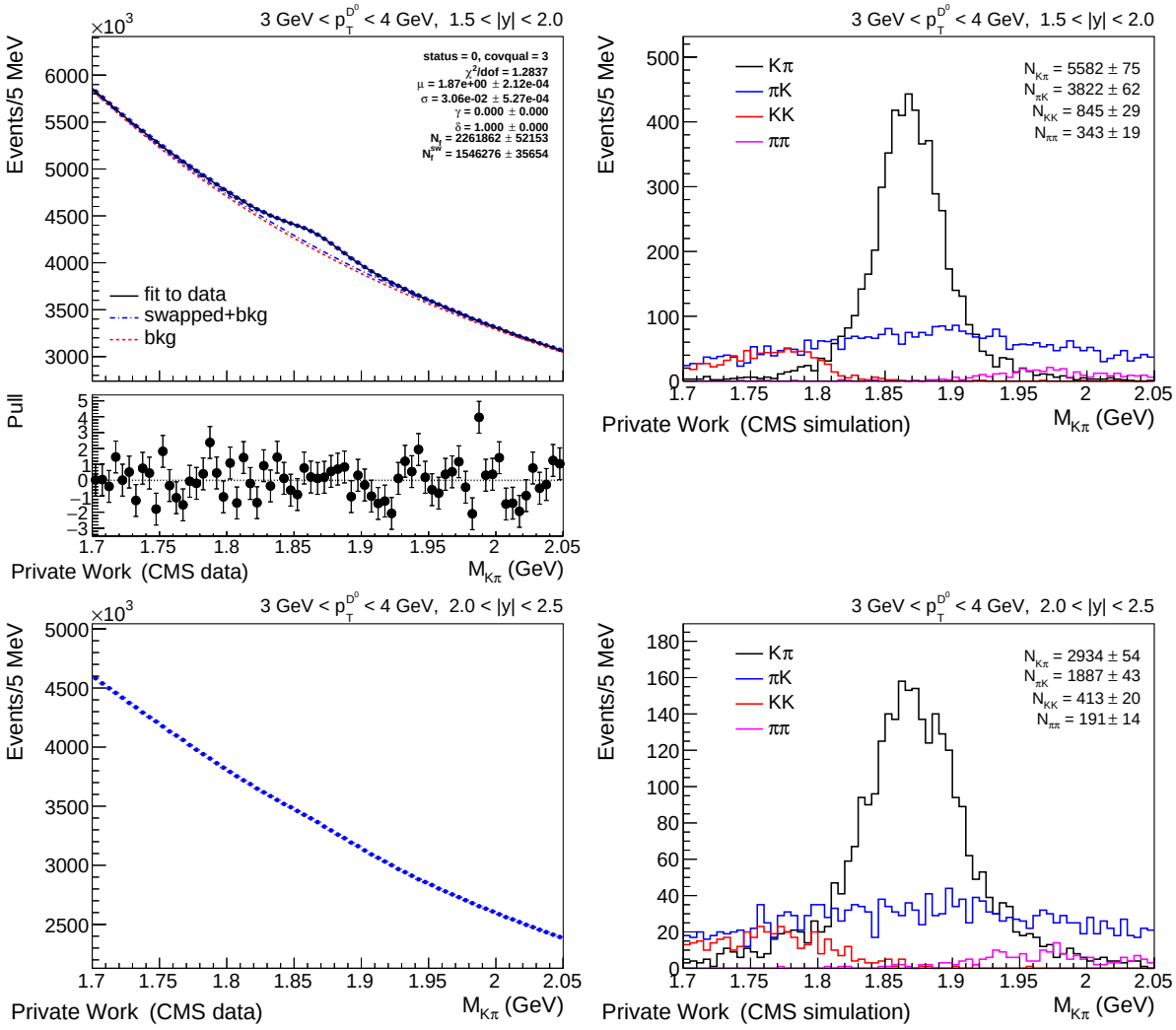


Figure D.49: M_{D^0} distributions and fit on data (left), and MC template (right).

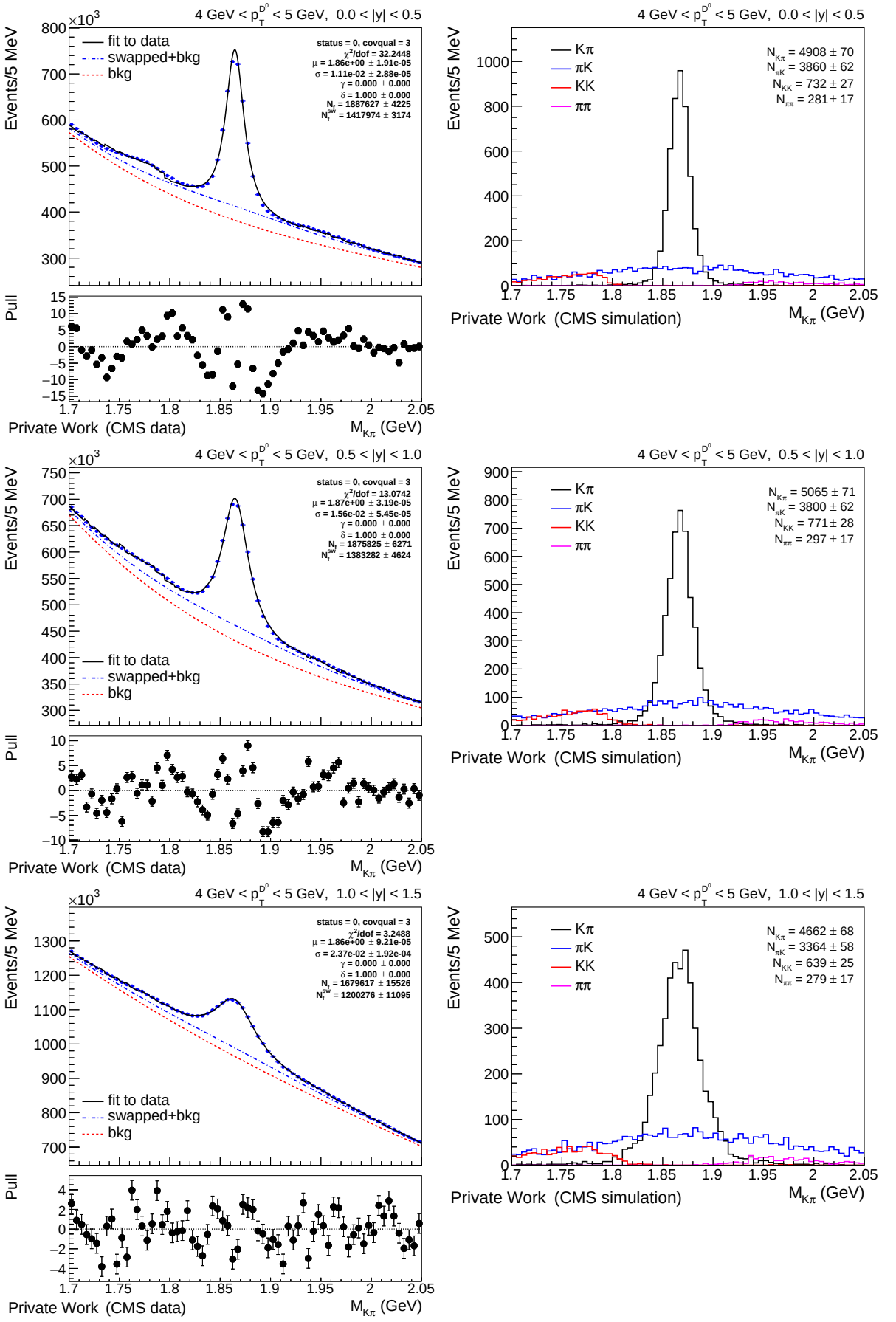


Figure D.50: M_{D^0} distributions and fit on data (left), and MC template (right).

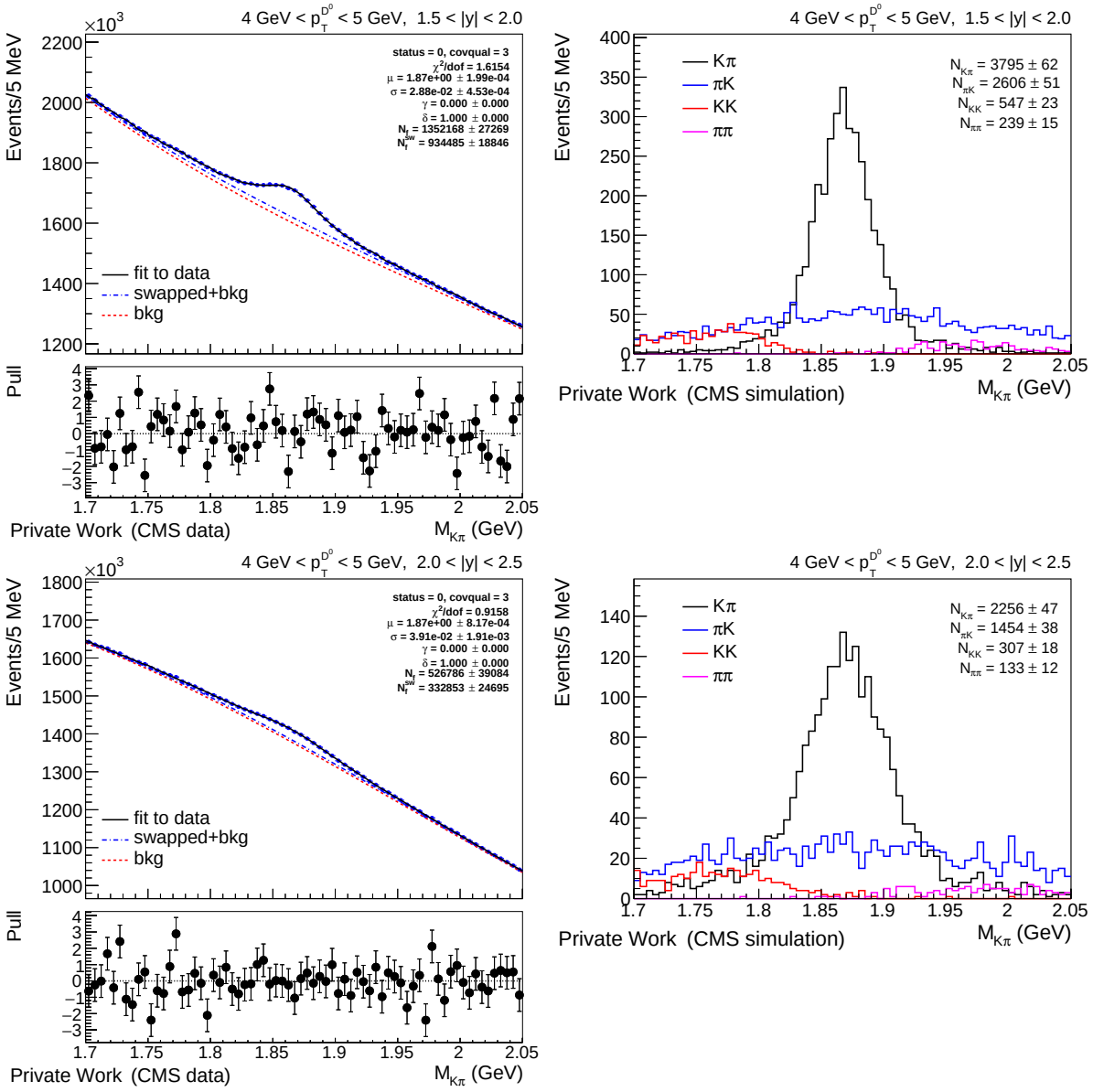


Figure D.51: M_{D^0} distributions and fit on data (left), and MC template (right).

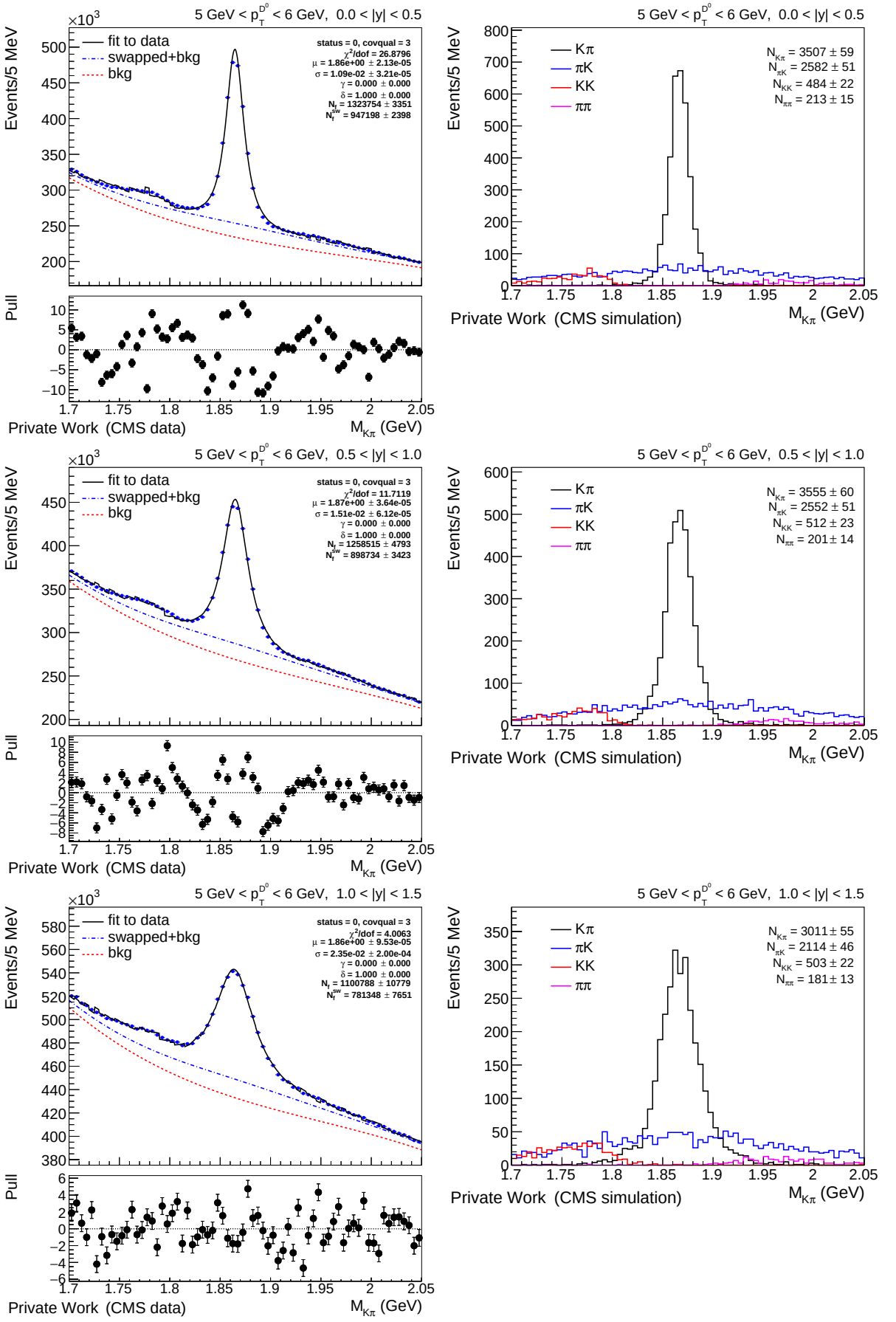


Figure D.52: M_{D^0} distributions and fit on data (left), and MC template (right).

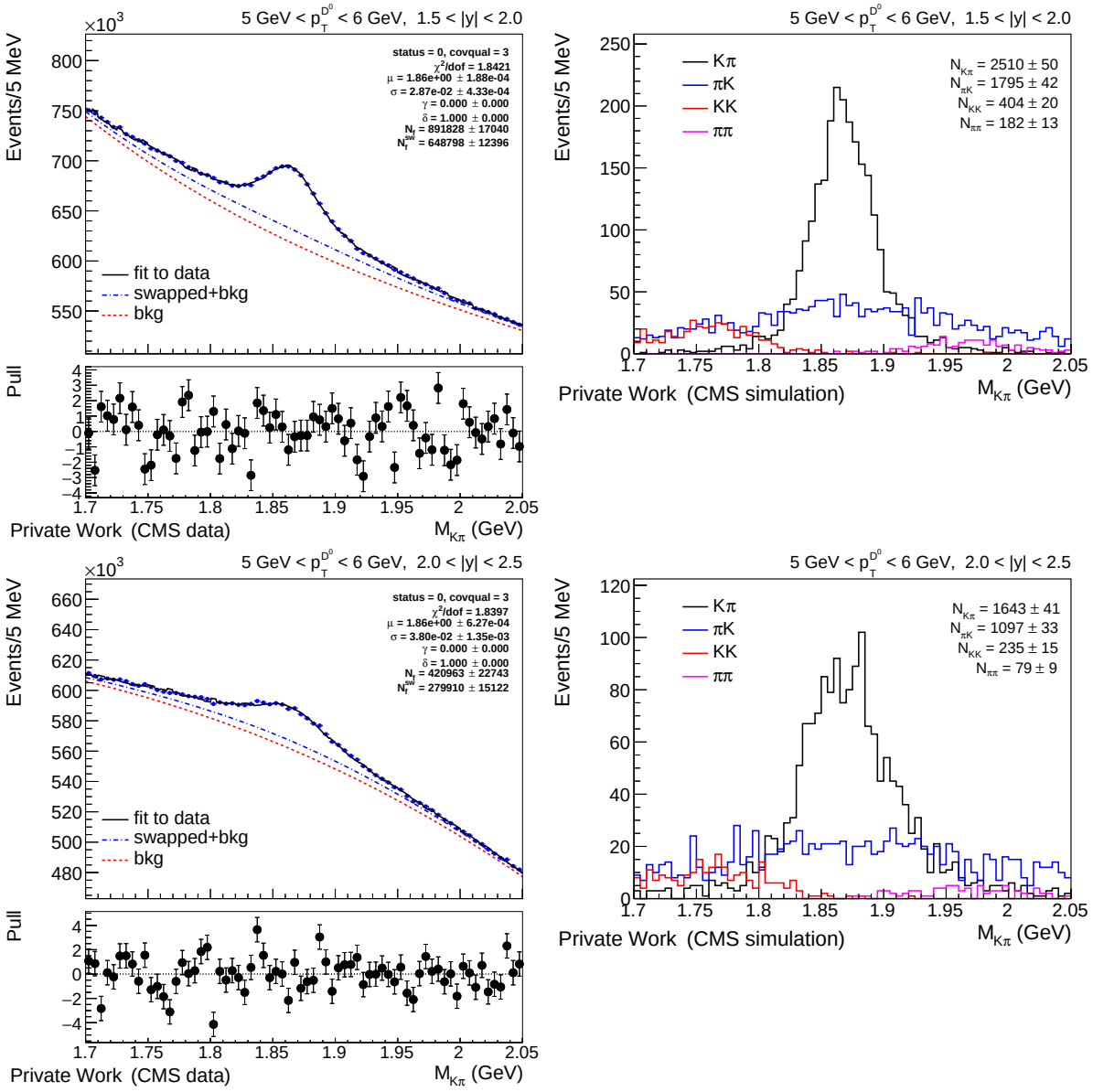


Figure D.53: M_{D^0} distributions and fit on data (left), and MC template (right).

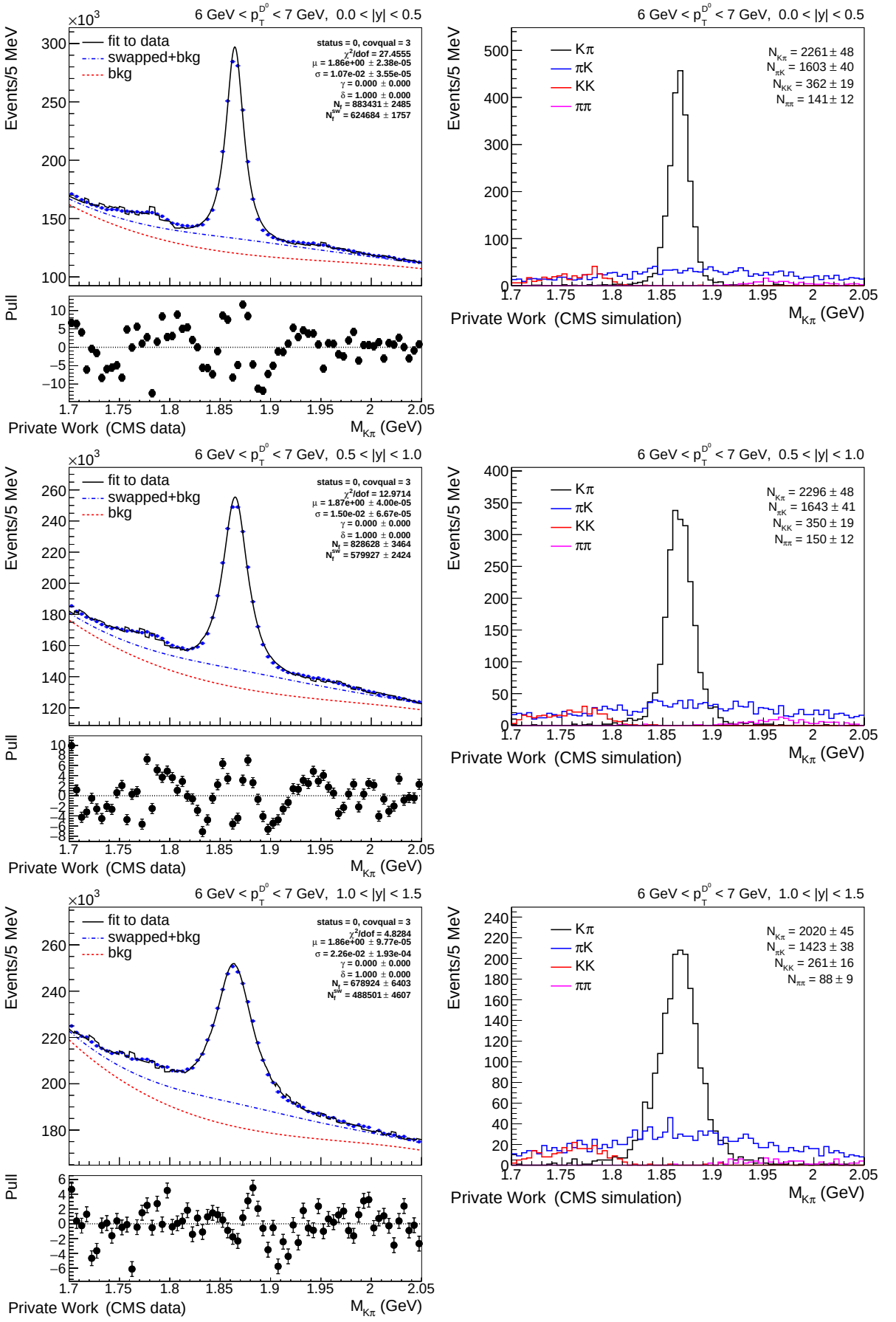


Figure D.54: M_{D^0} distributions and fit on data (left), and MC template (right).

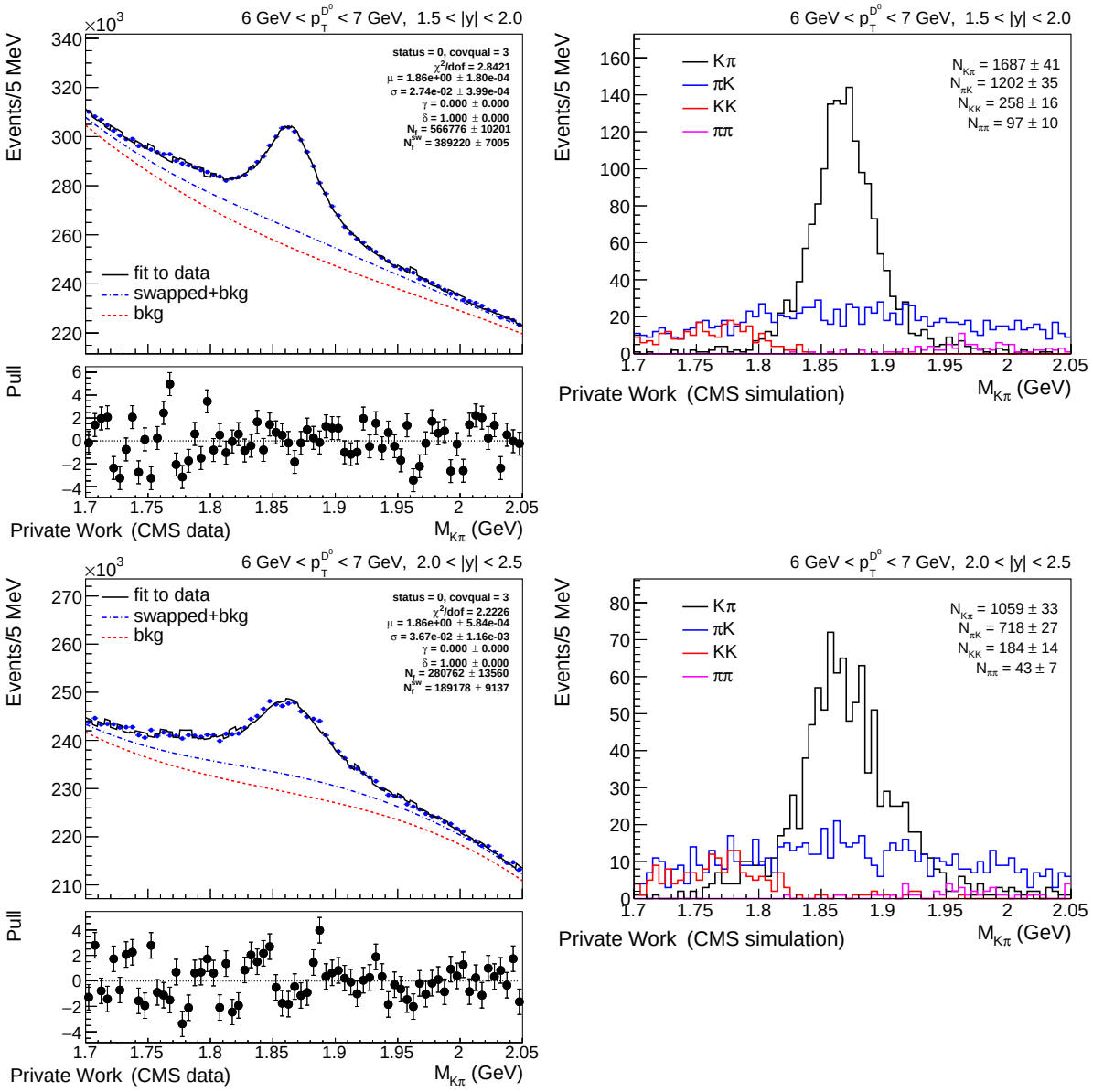


Figure D.55: M_{D^0} distributions and fit on data (left), and MC template (right).

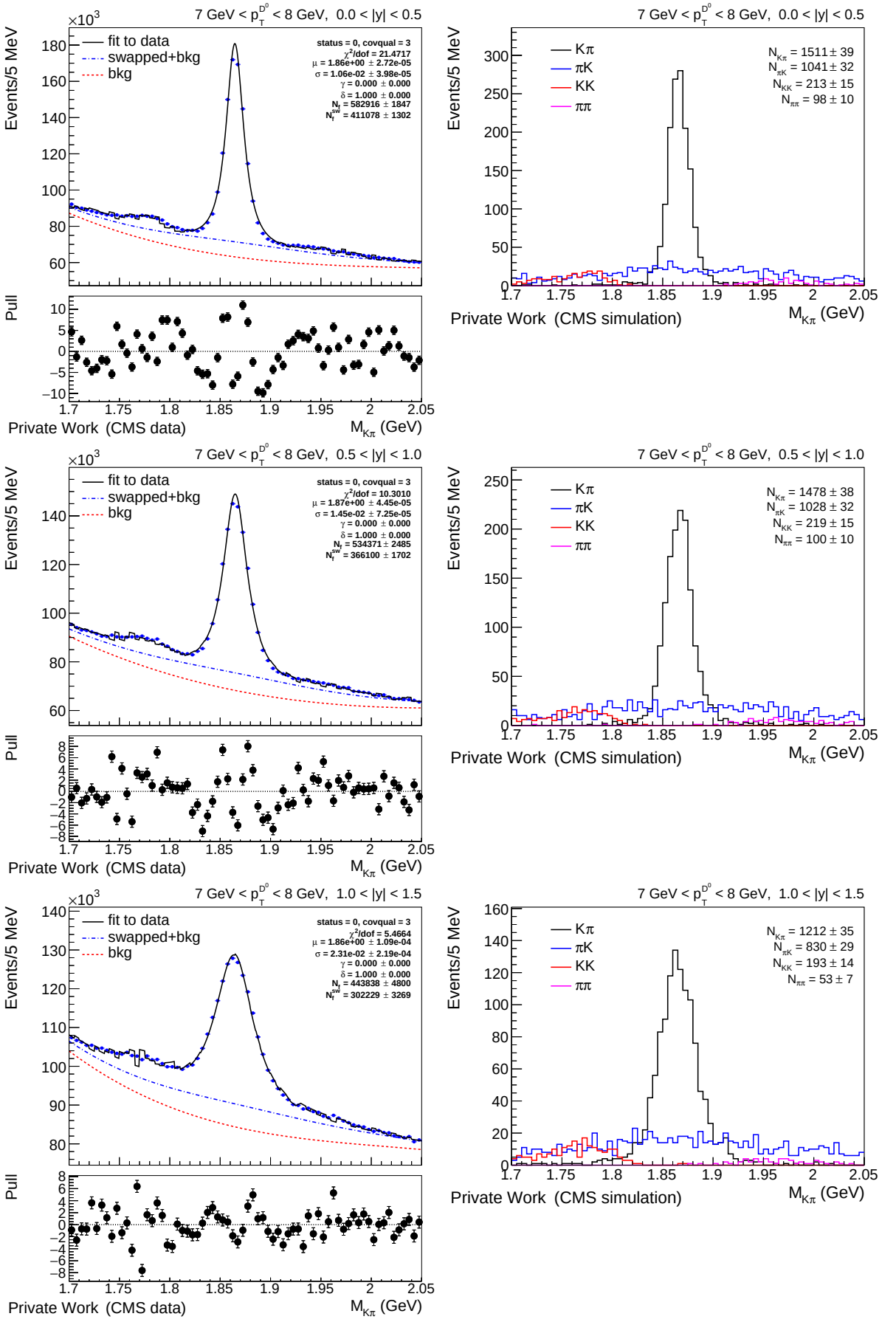


Figure D.56: M_{D^0} distributions and fit on data (left), and MC template (right).

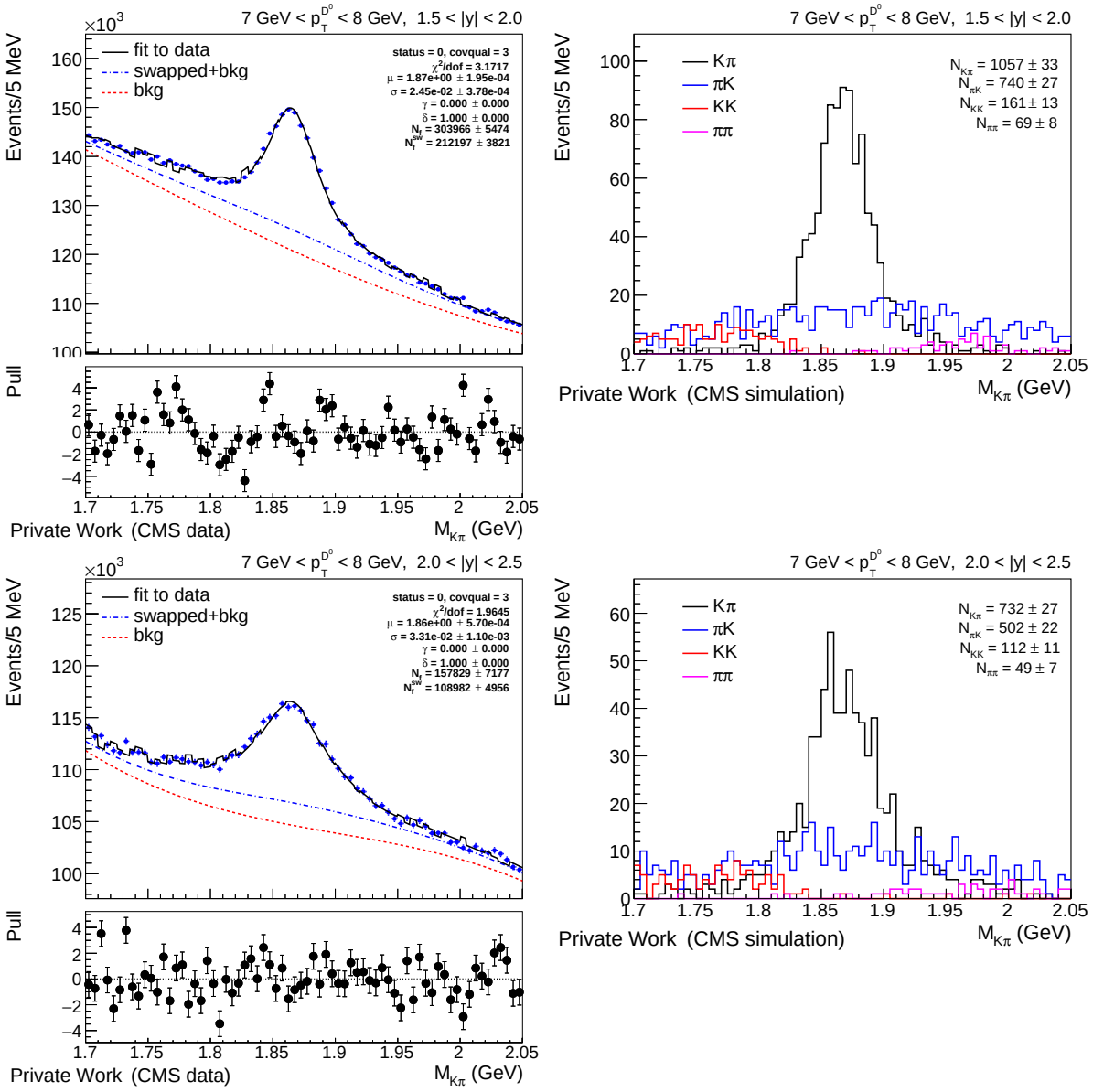


Figure D.57: M_{D^0} distributions and fit on data (left), and MC template (right).

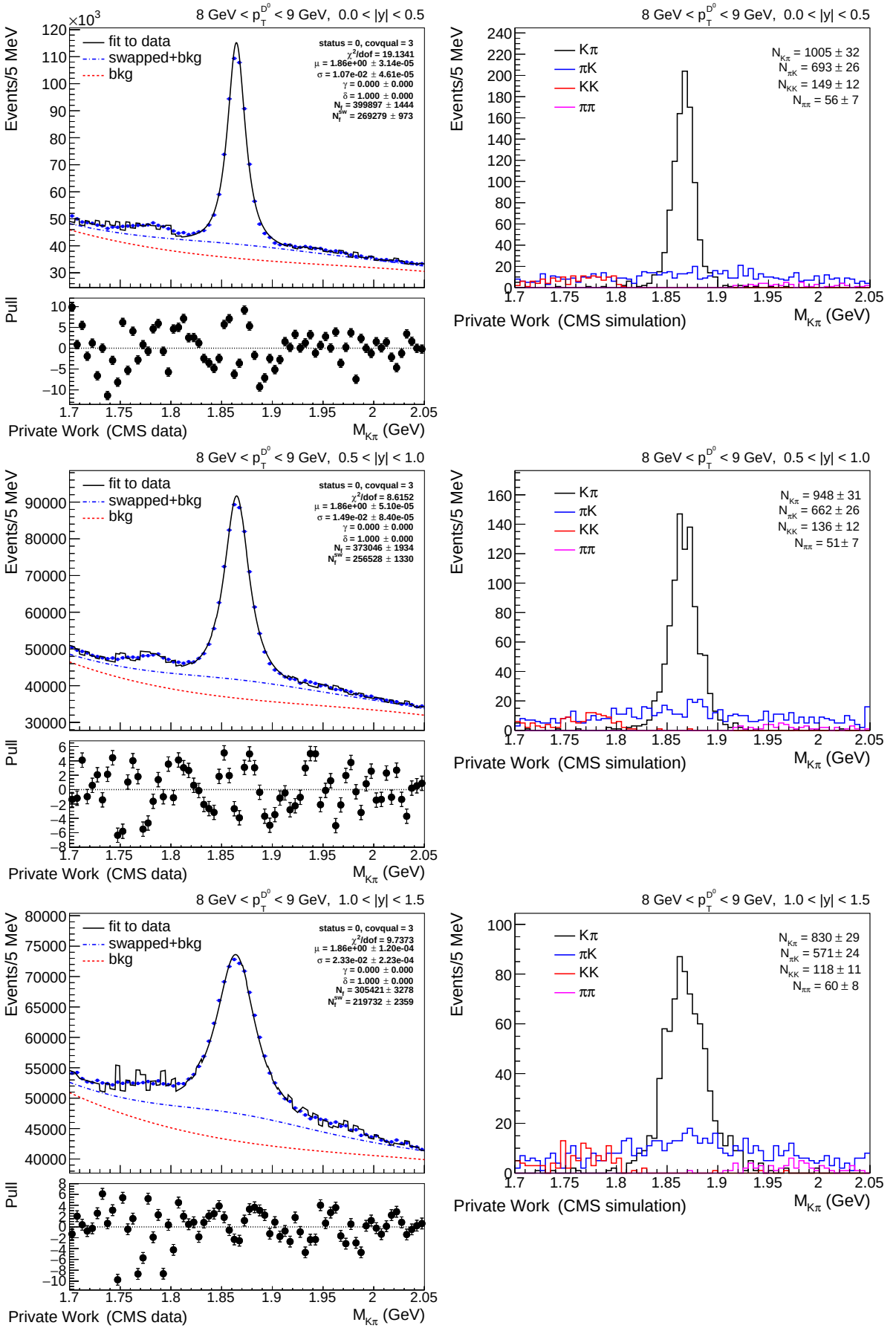


Figure D.58: M_{D^0} distributions and fit on data (left), and MC template (right).

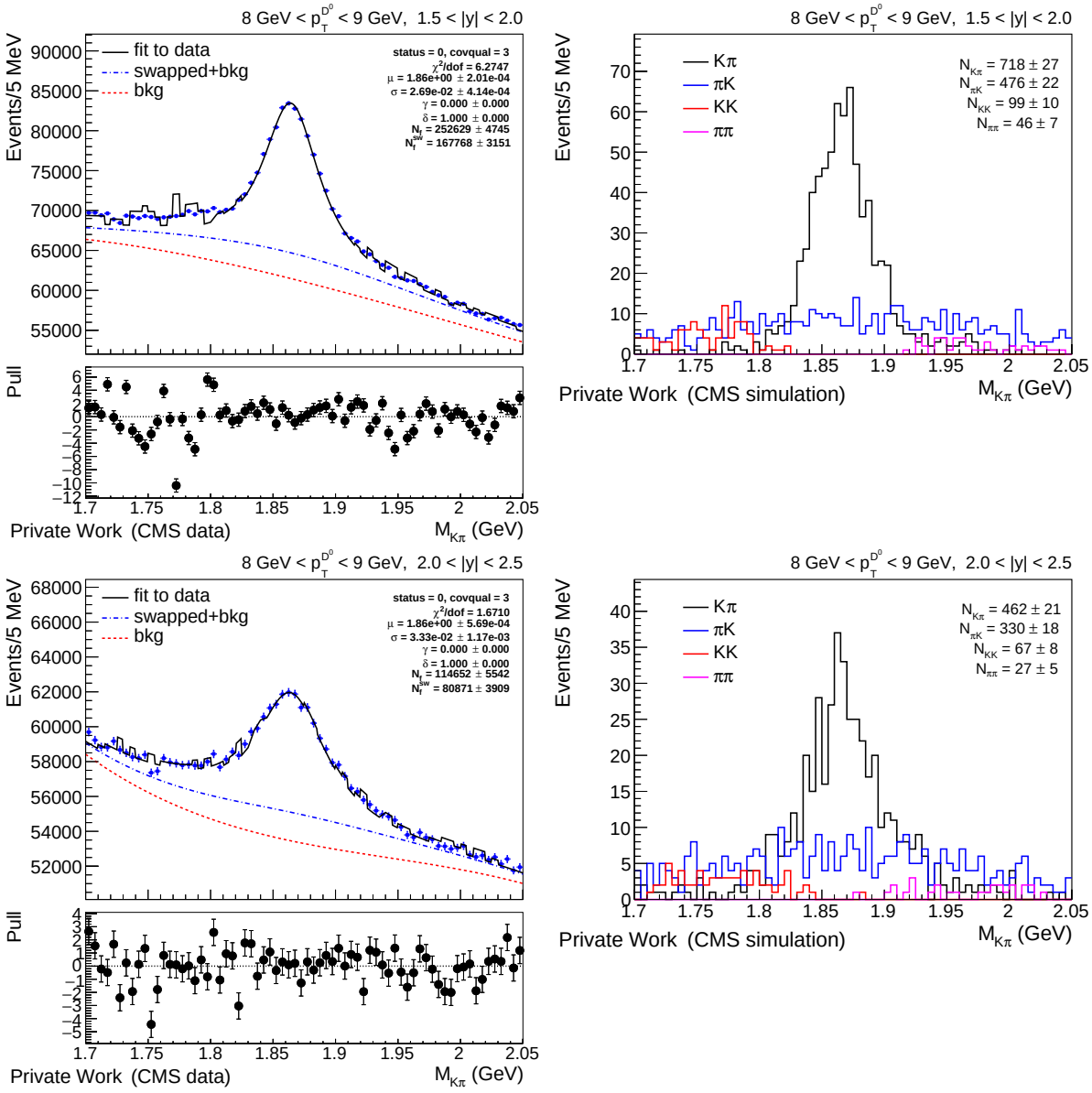


Figure D.59: M_{D^0} distributions and fit on data (left), and MC template (right).

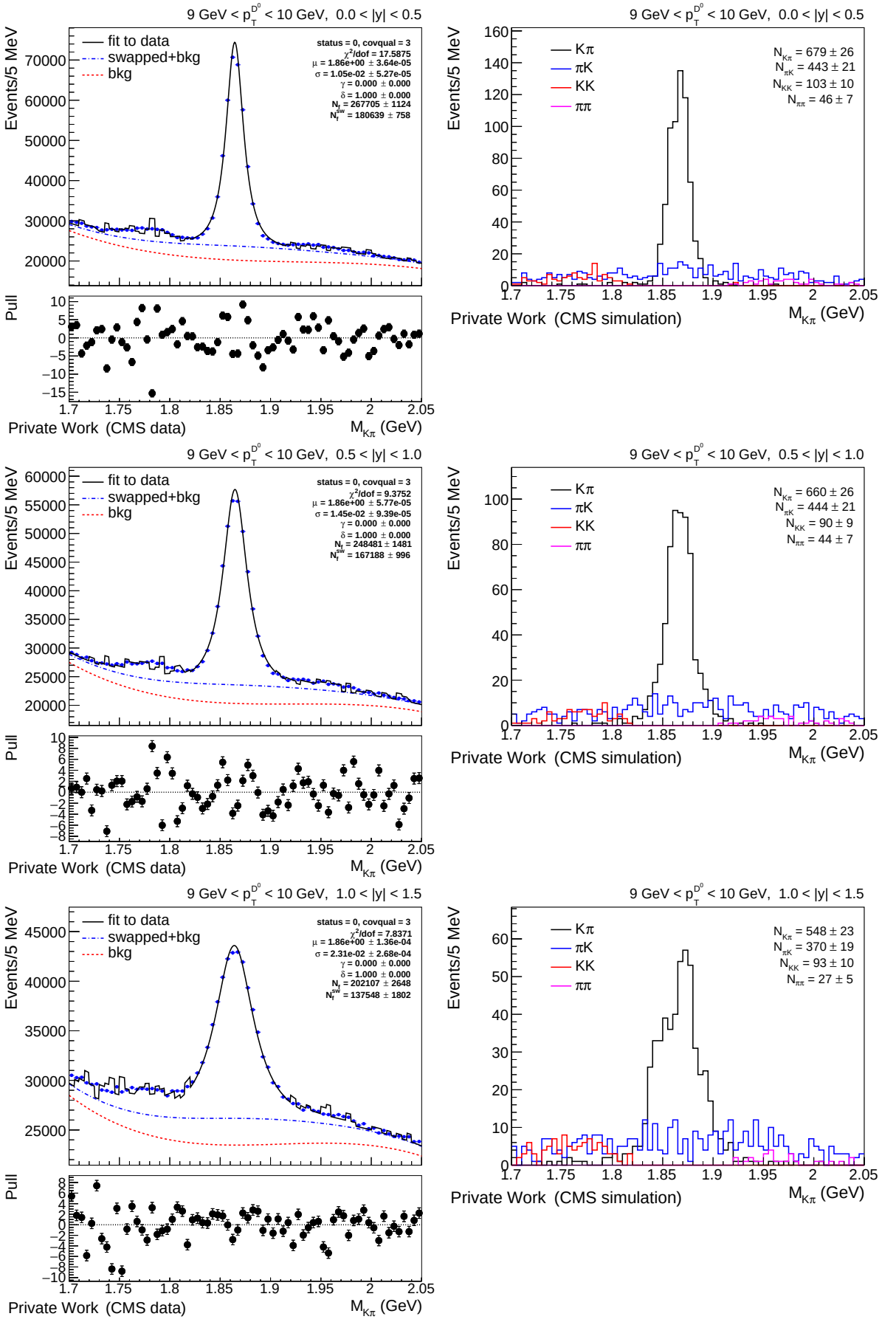


Figure D.60: M_{D^0} distributions and fit on data (left), and MC template (right).

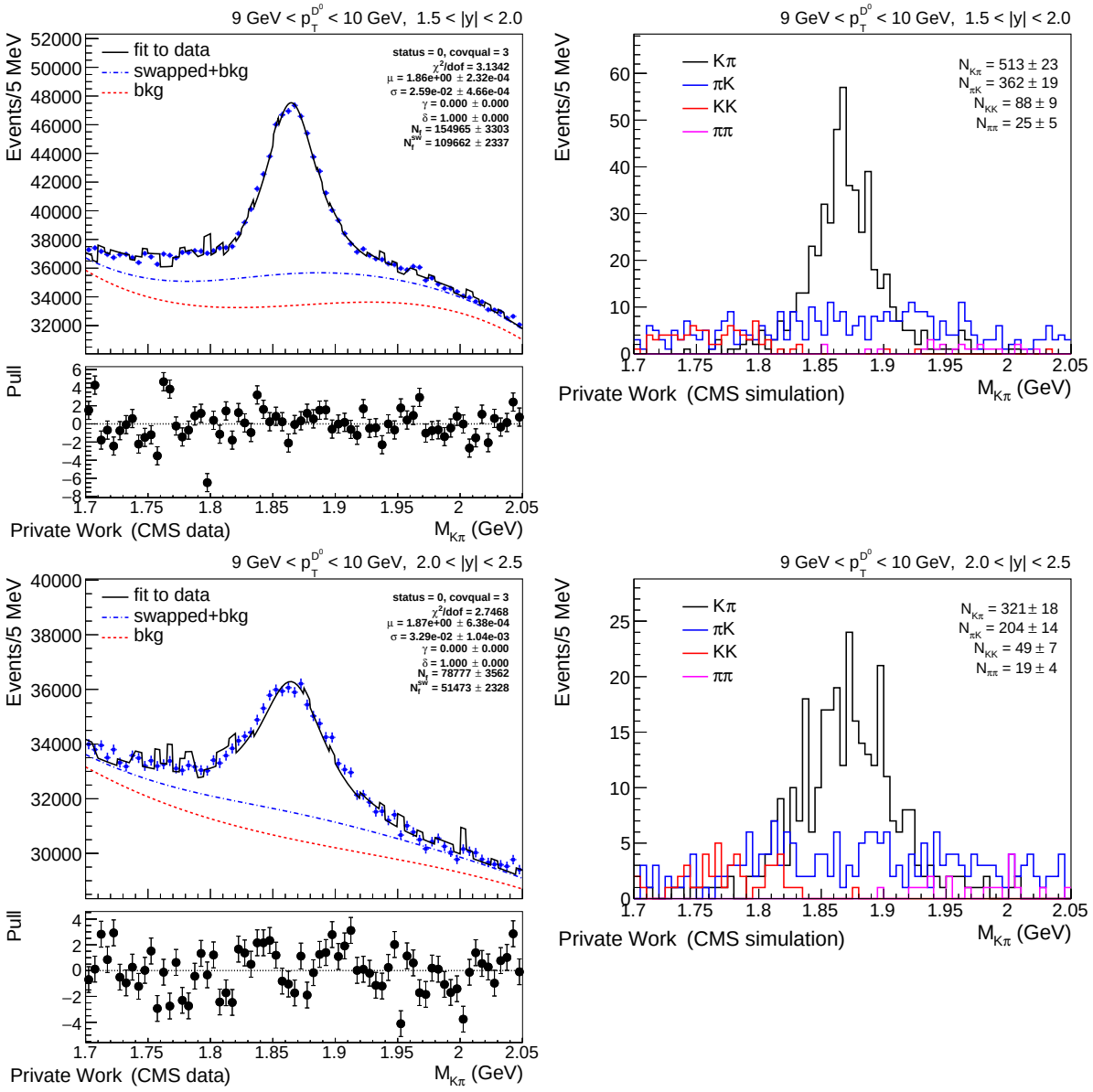


Figure D.61: M_{D^0} distributions and fit on data (left), and MC template (right).

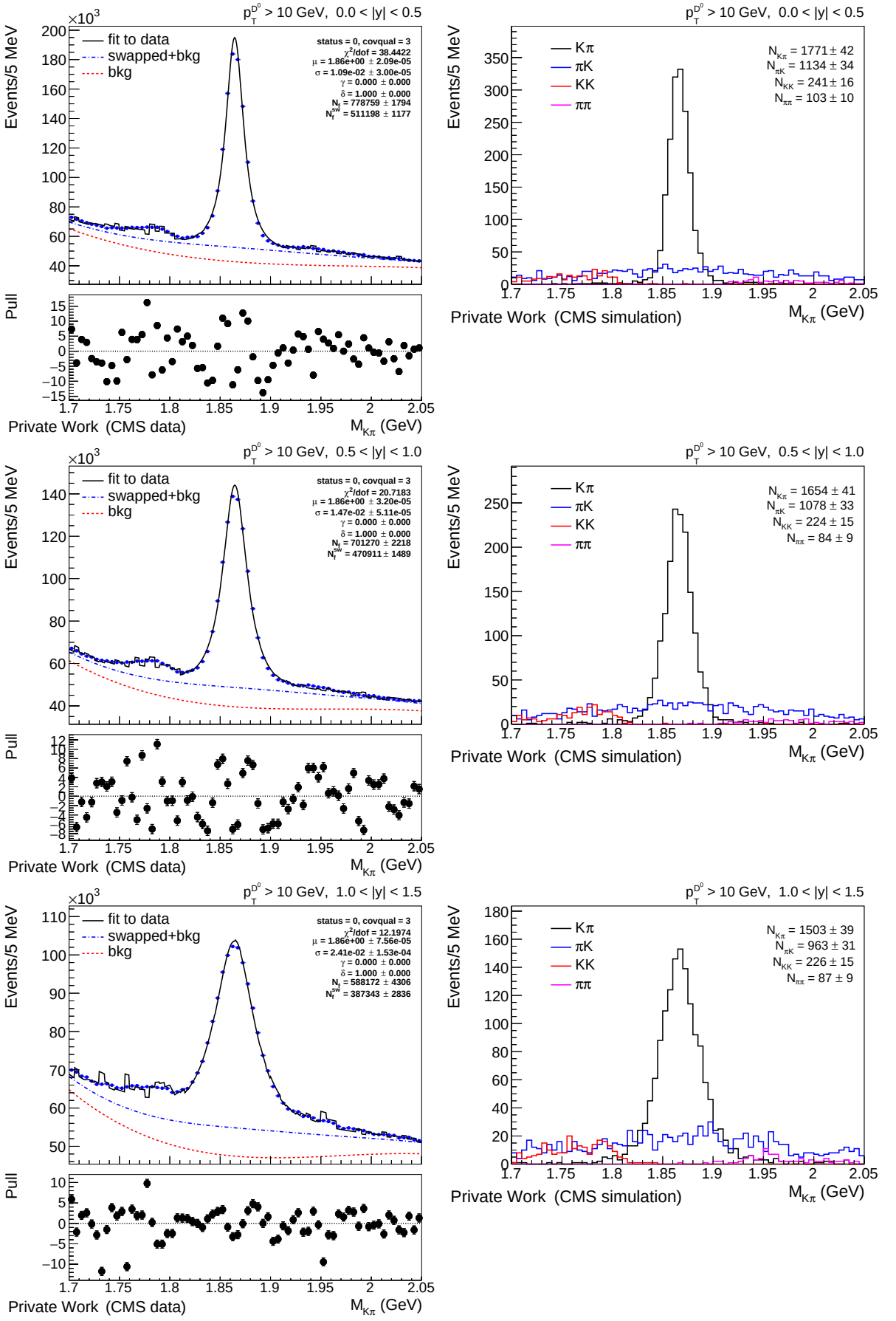


Figure D.62: M_{D^0} distributions and fit on data (left), and MC template (right).

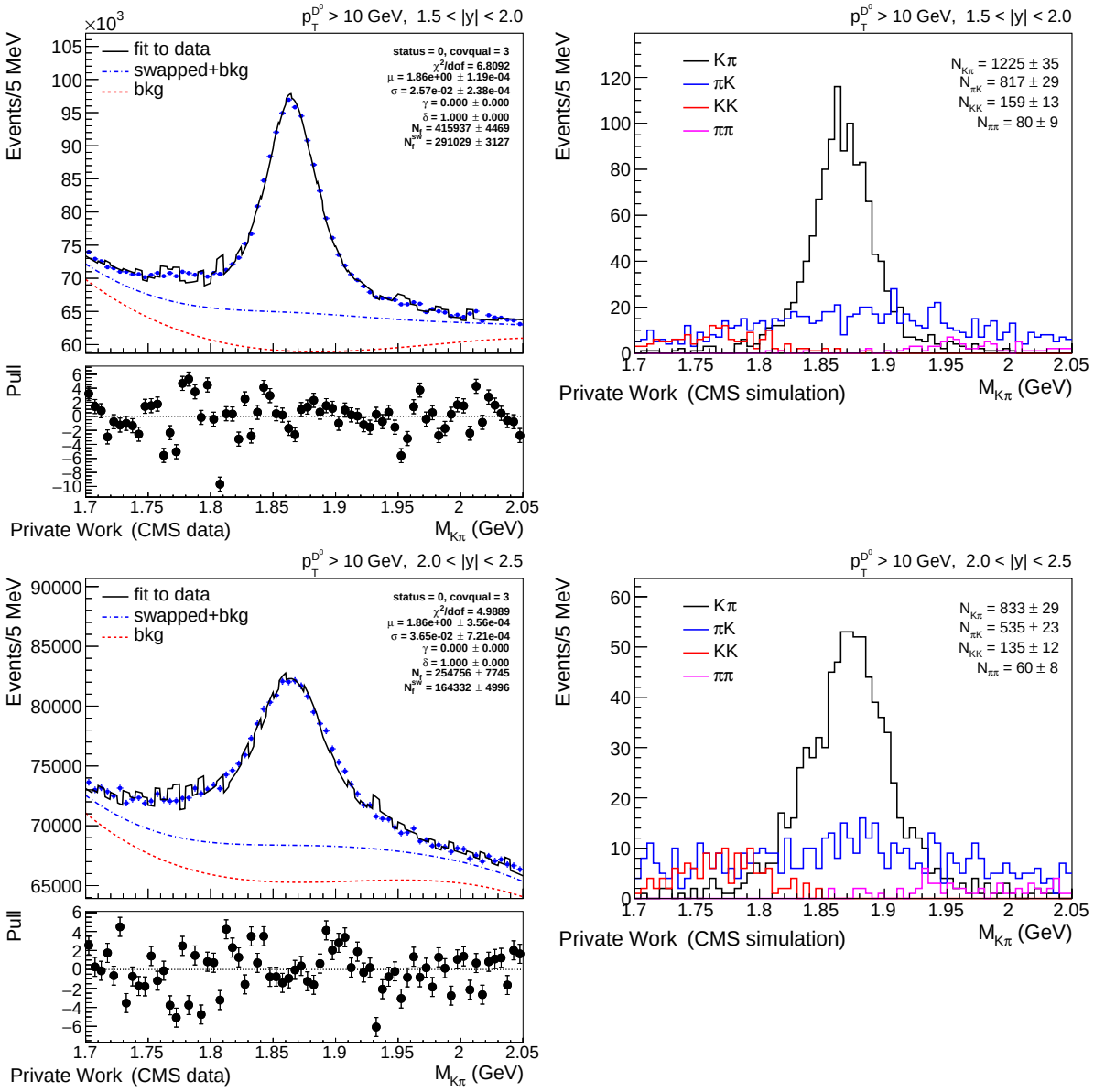


Figure D.63: M_{D^0} distributions and fit on data (left), and MC template (right).

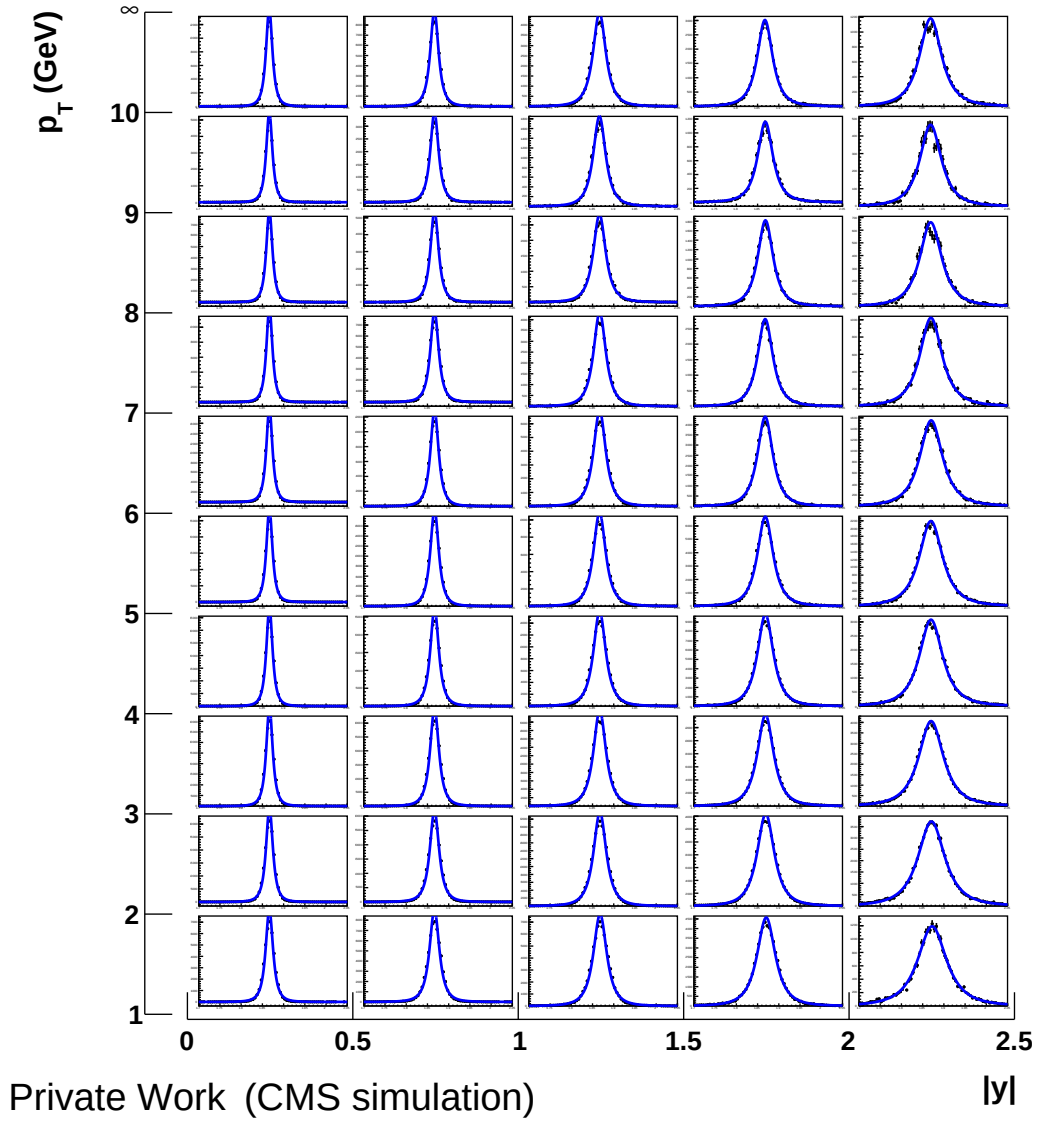


Figure D.64: D^0 signal on the 13 TeV D^0 MC sample. The blue curves indicate fits to M_{D^0} mass distribution.

D.2.4 D^0 signal fit on MC

In this appendix, D^0 signal on the 13 TeV D^0 MC sample is summarized and shown for each bin of the cross section measurement including fit results which were used for the cross section measurement. In this case, χ^2/ndof results are sometimes very large as a result of reweighting events.

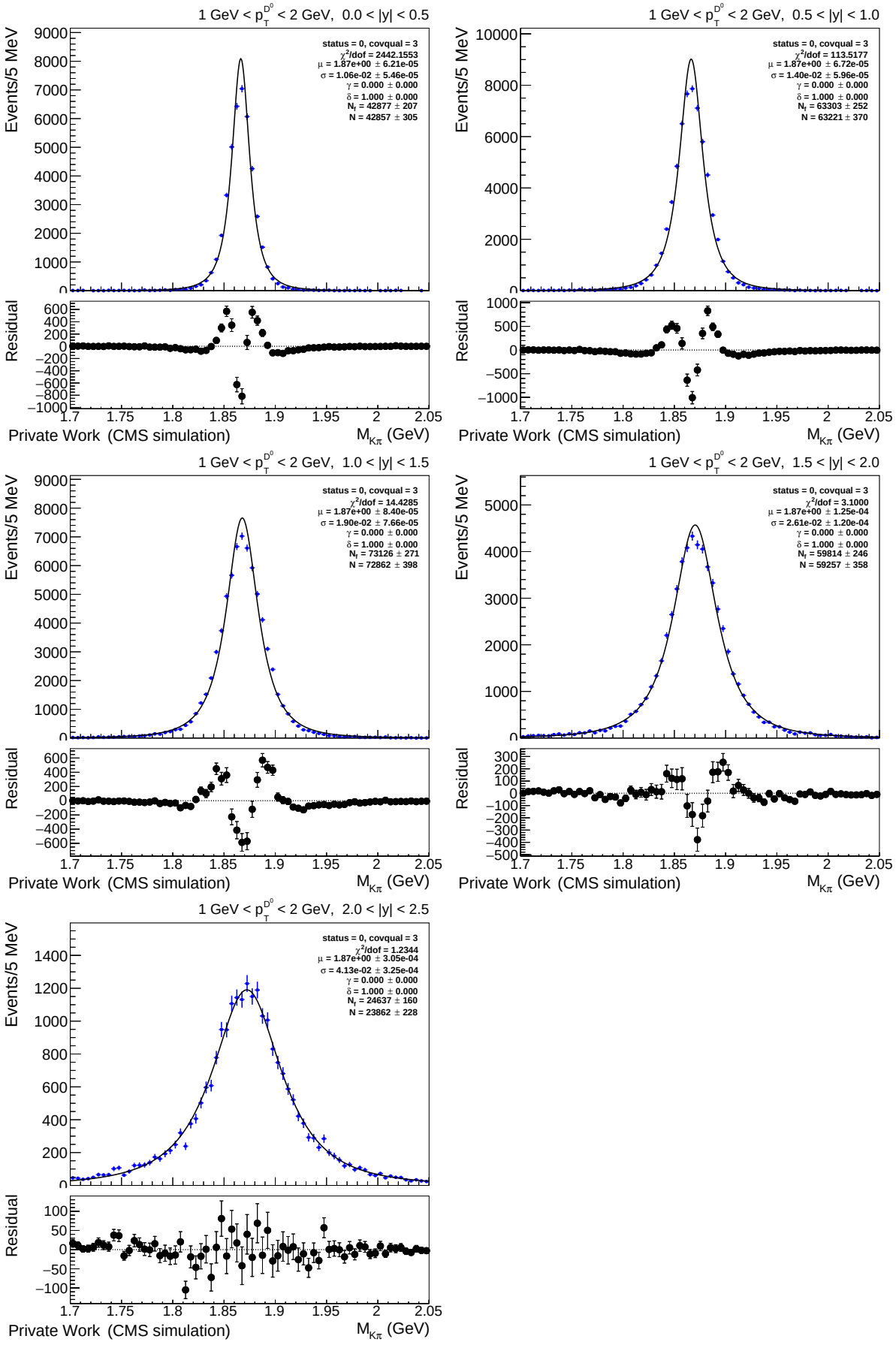


Figure D.65: M_{D^0} distributions and fit.

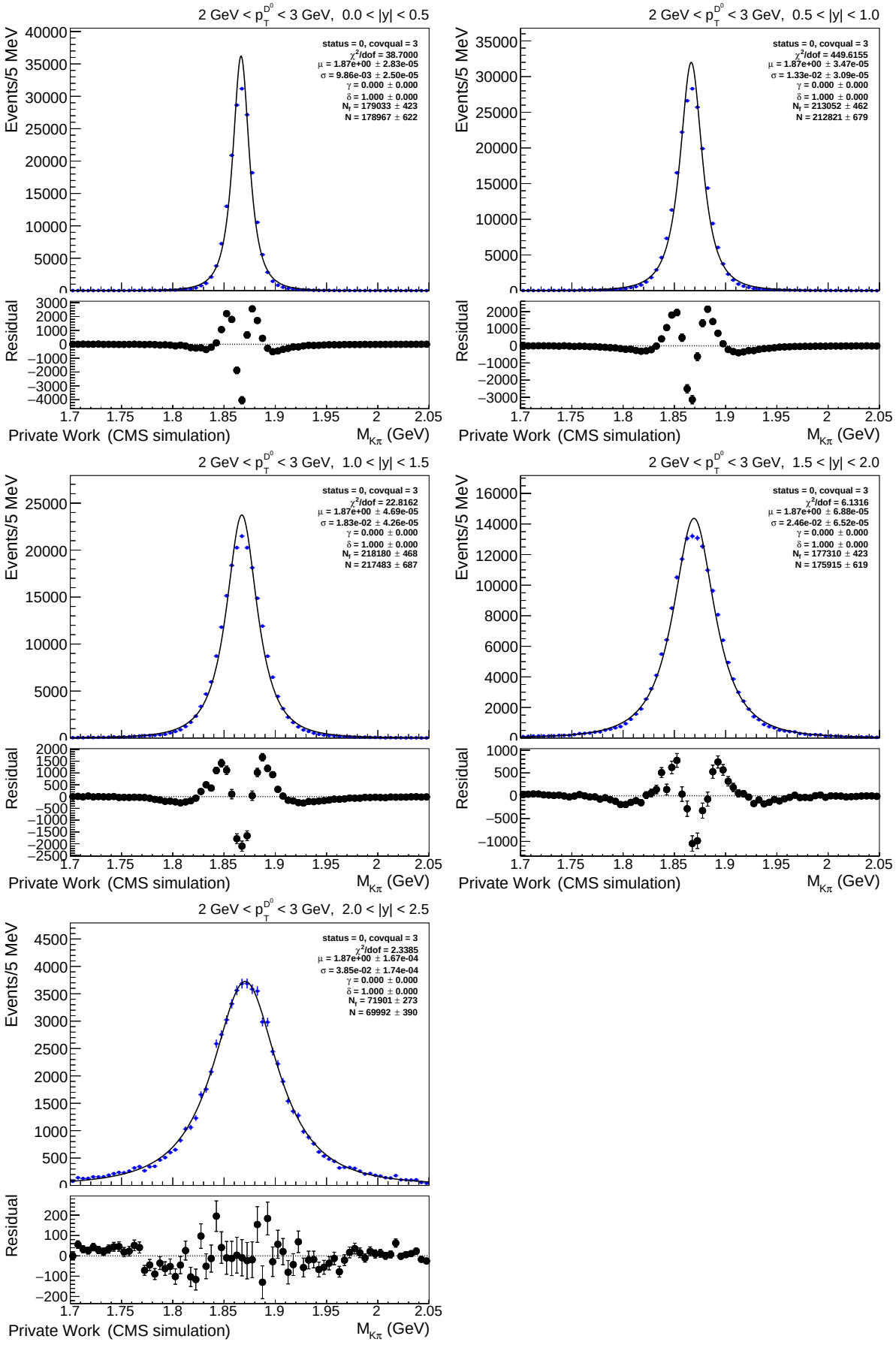


Figure D.66: M_{D^0} distributions and fit.

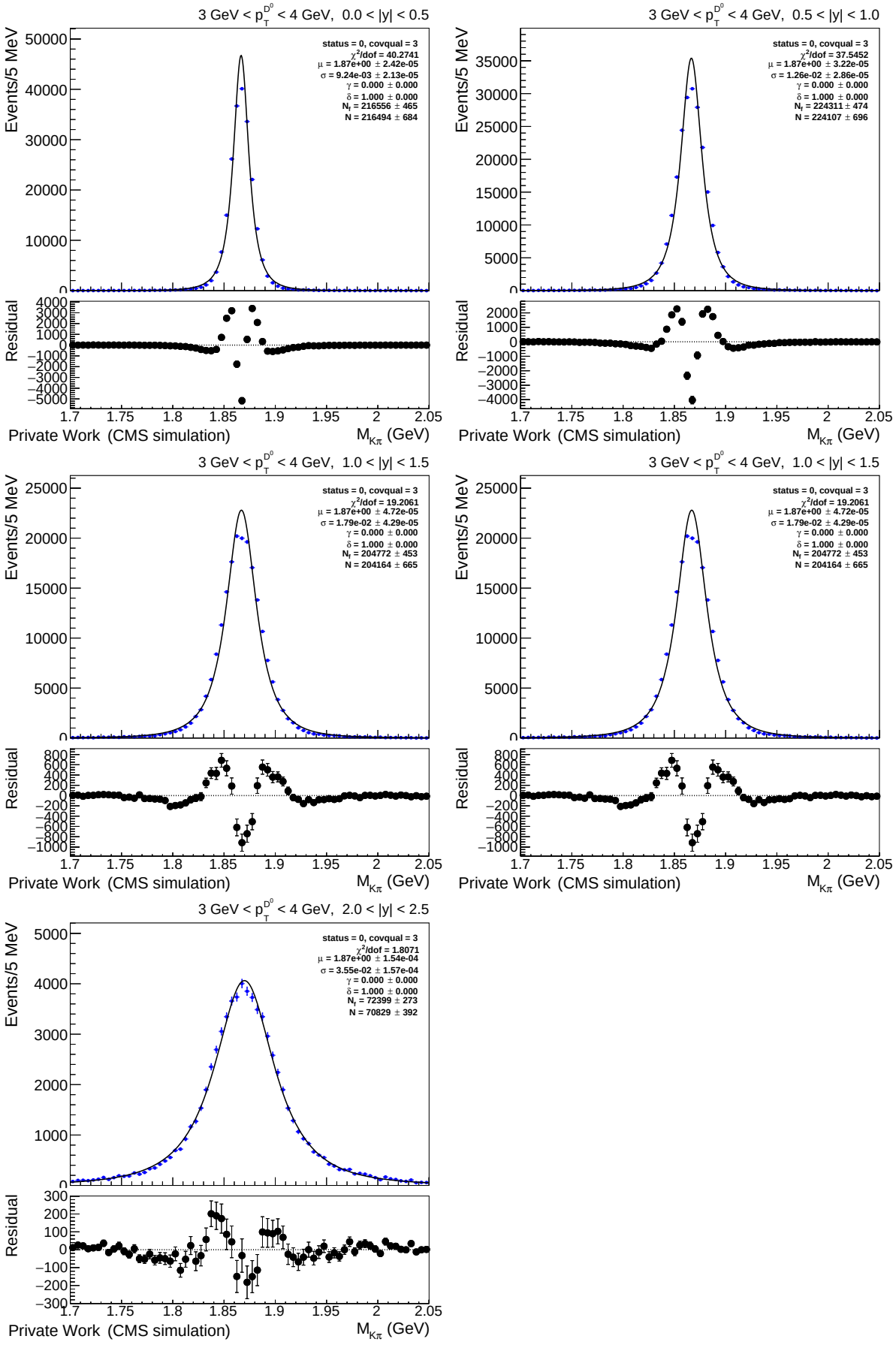


Figure D.67: M_{D^0} distributions and fit.

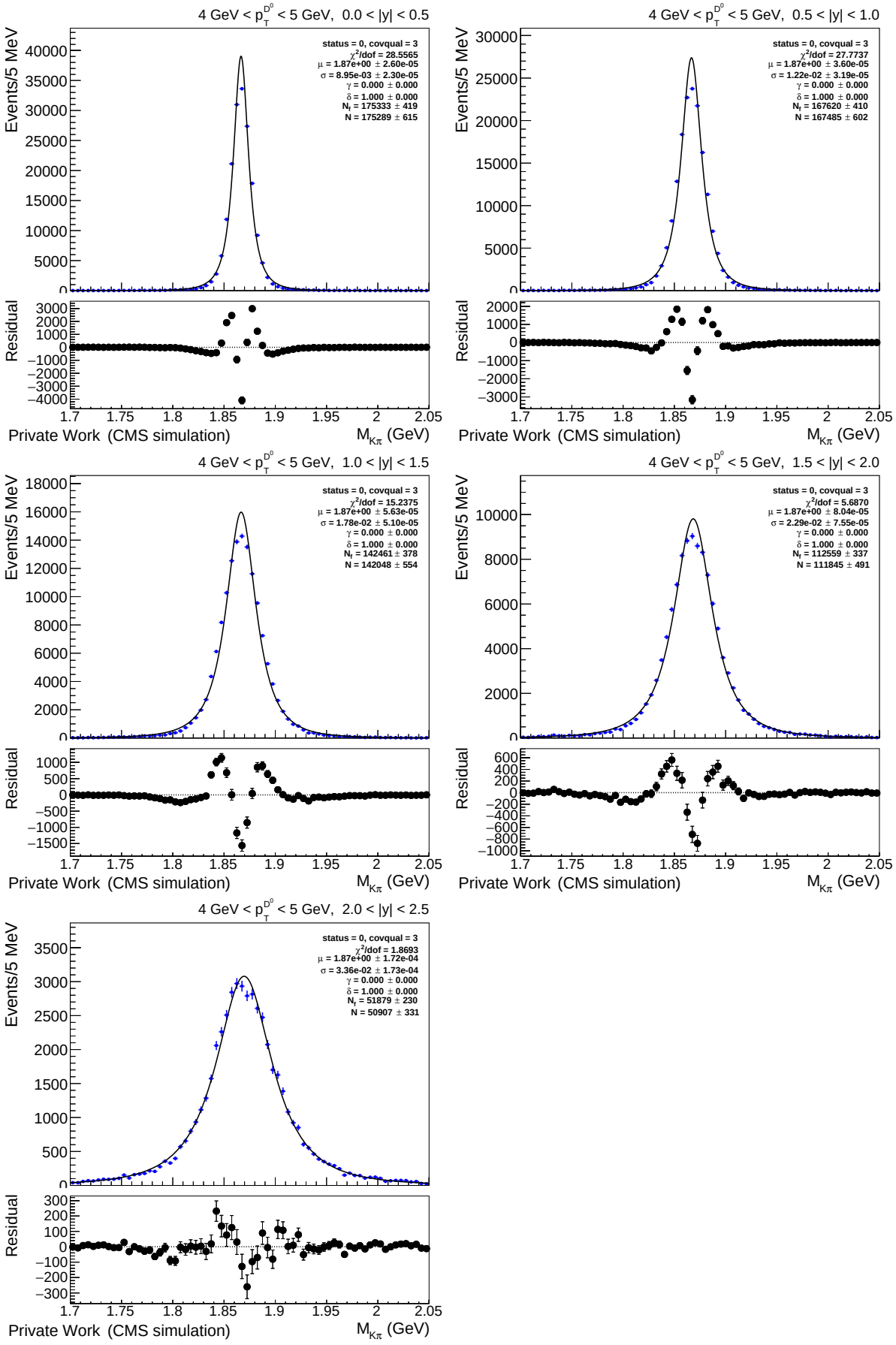


Figure D.68: M_{D^0} distributions and fit.

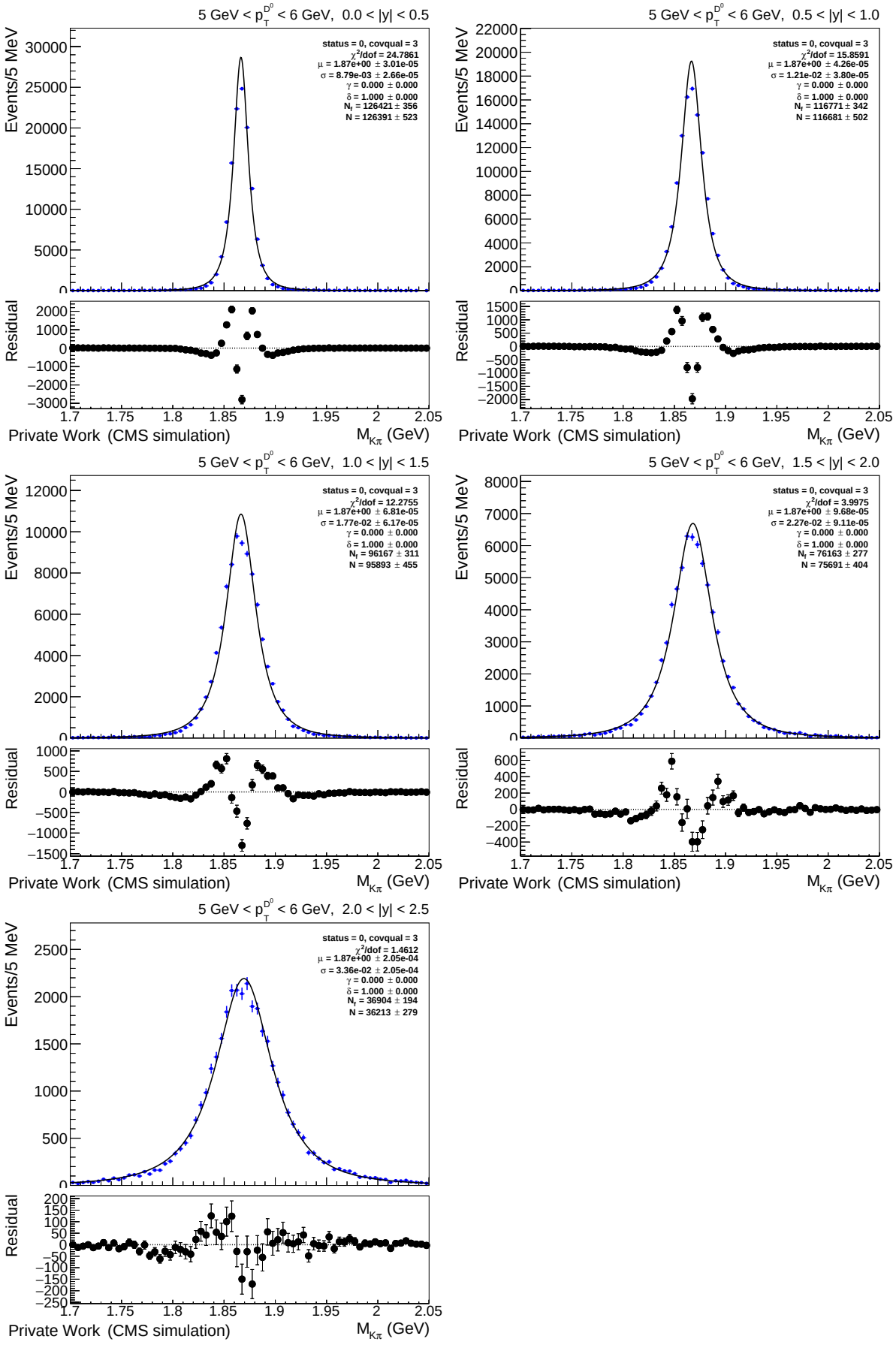


Figure D.69: M_{D^0} distributions and fit.

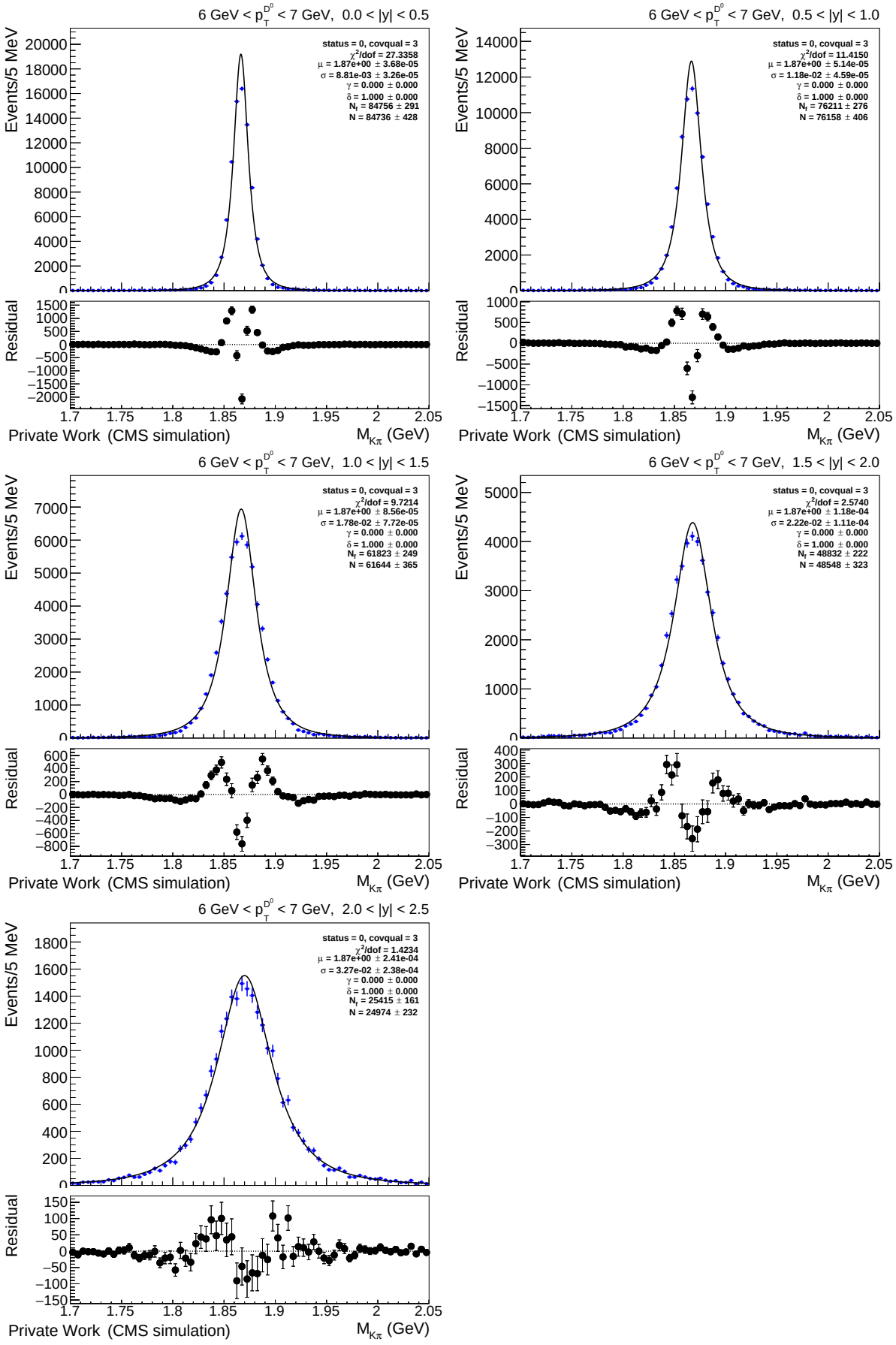


Figure D.70: M_{D^0} distributions and fit.

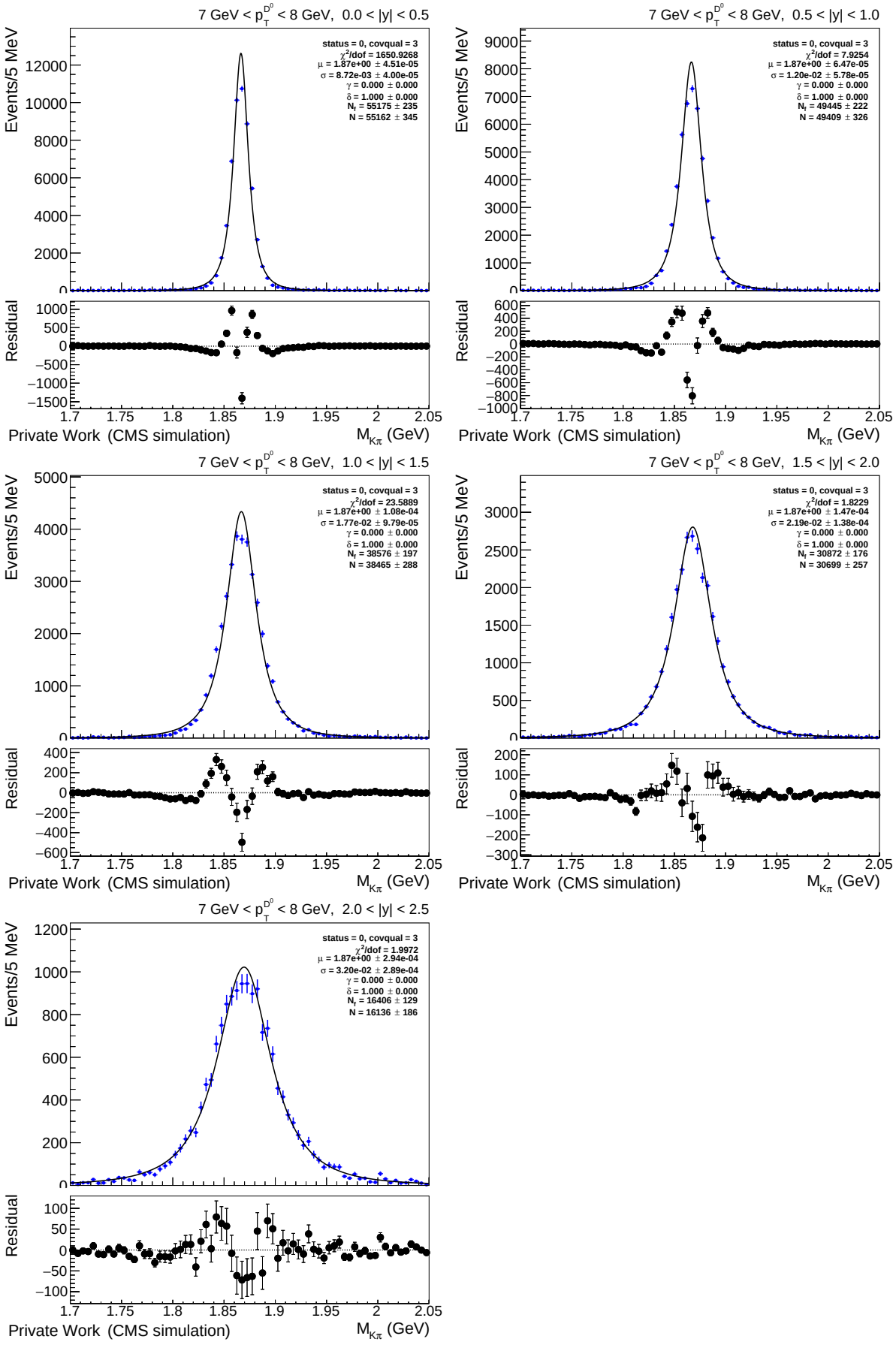


Figure D.71: M_{D^0} distributions and fit.

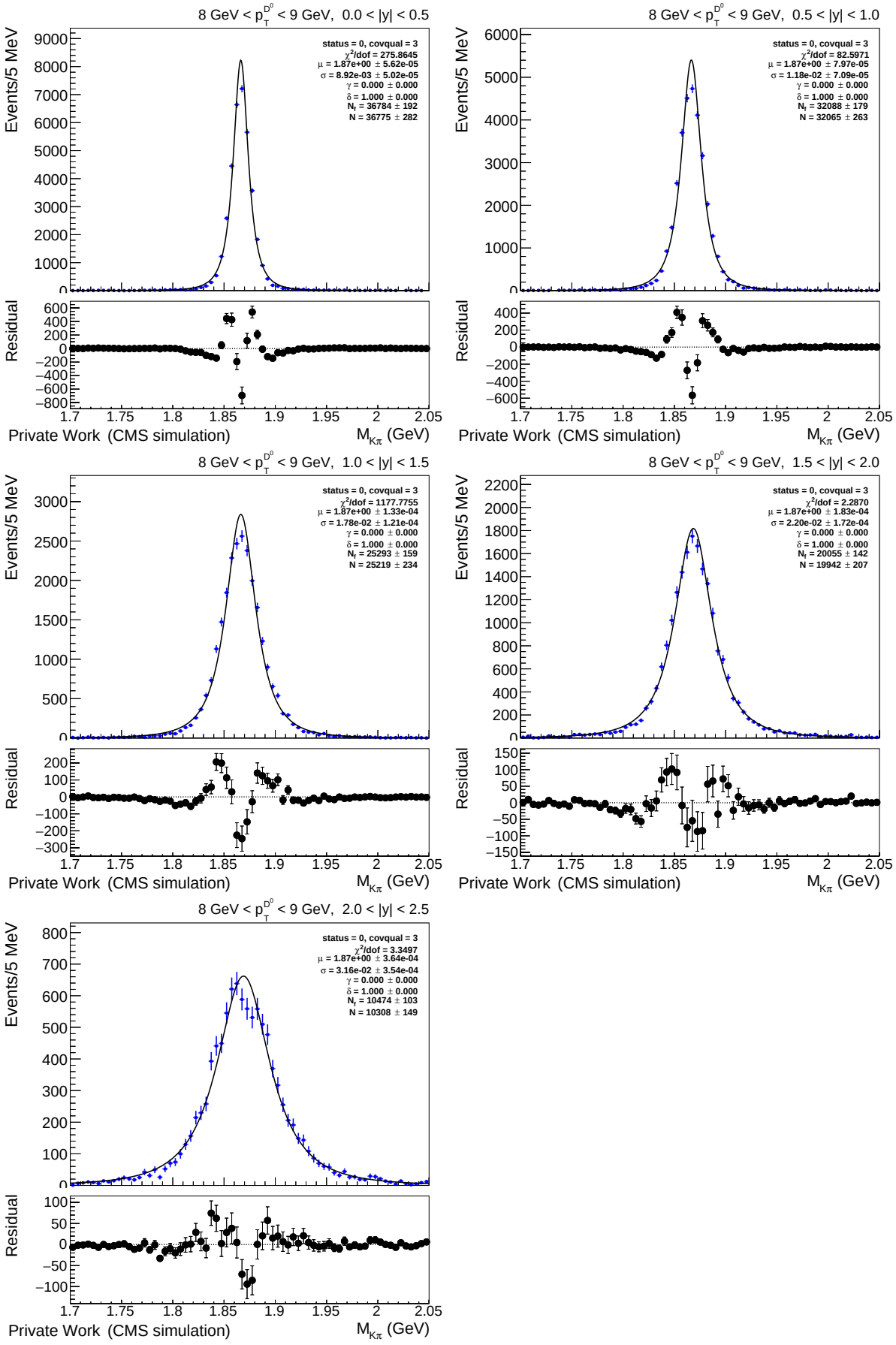


Figure D.72: M_{D^0} distributions and fit.

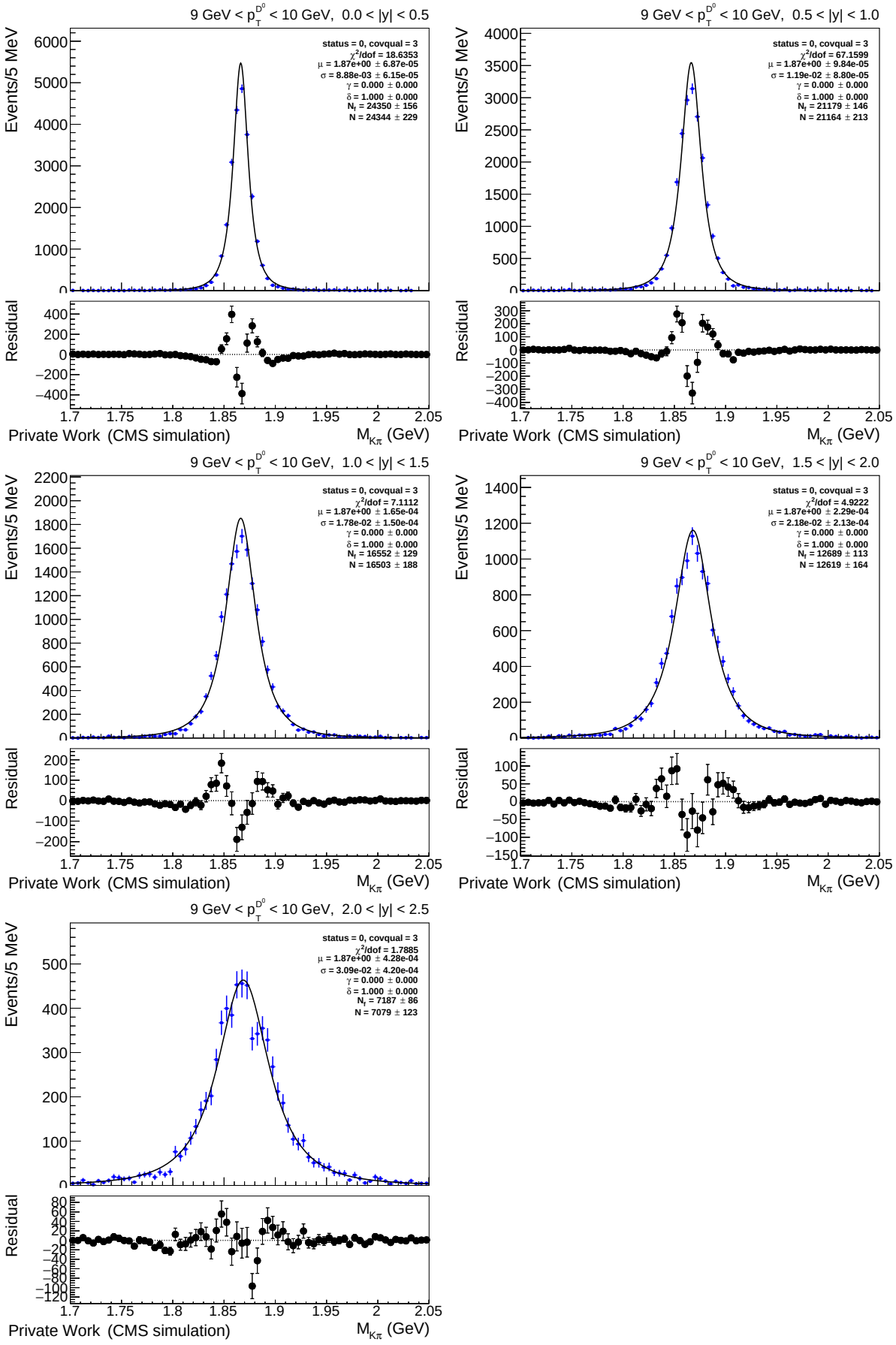


Figure D.73: M_{D^0} distributions and fit.

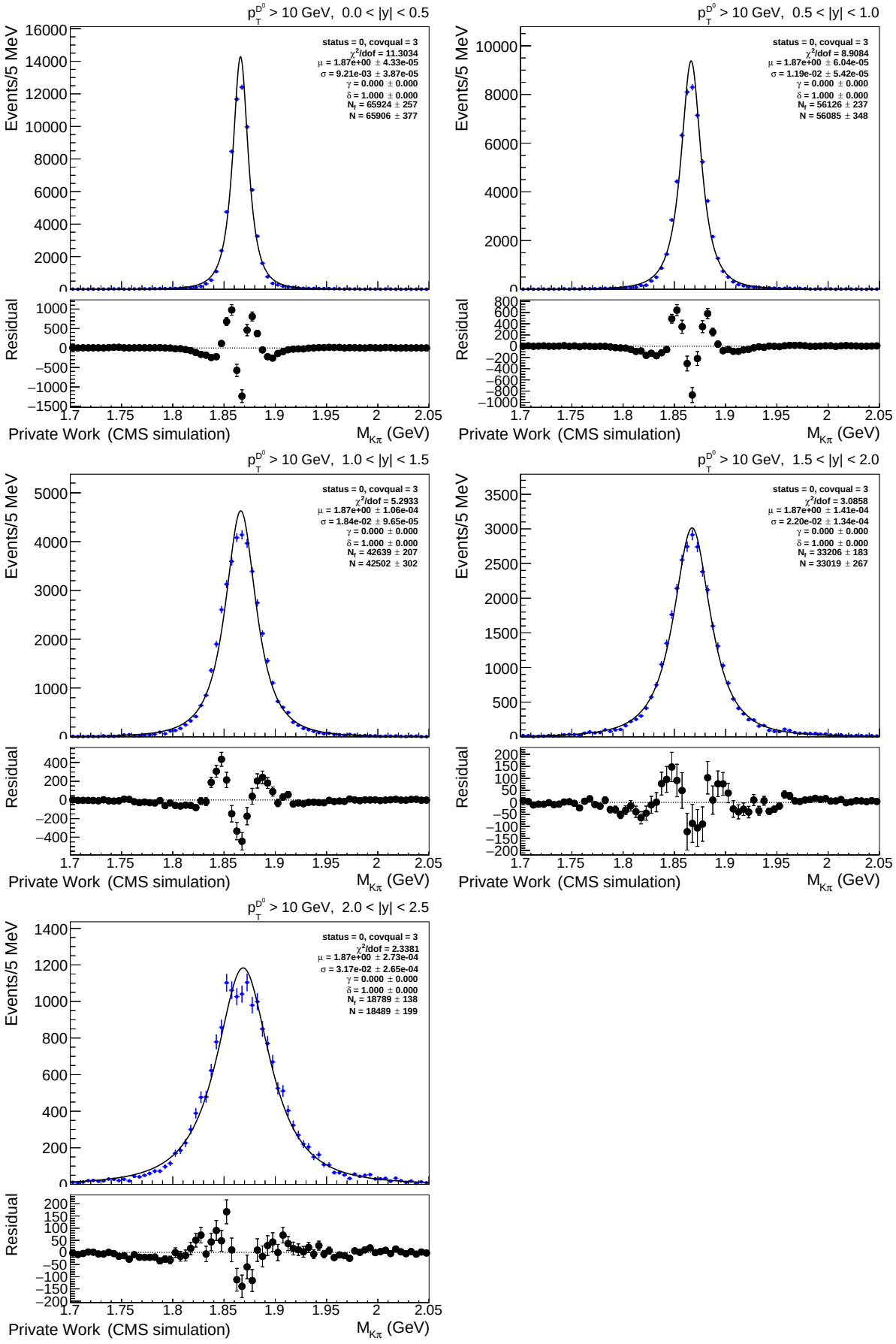


Figure D.74: M_{D^0} distributions and fit.

E | 13 TeV Data and MC Comparison

E.1 Data and MC comparison with PU reweight

In this appendix, the 13 TeV B parking data are compared to the D^0 MC sample. PU distribution (PV_npvs) of the D^0 MC was reweighted to that of the data (refer to Figure 5.4). After reweighting, the MC PU distribution of D^{*+} candidate events is shifted a bit compared to the data (refer to Figure E.14 and E.15). It looks like, although, that overall comparison shows reasonable agreement between the MC and data, while up to $\sim 10\%$ discrepancies might have effects on the cross section measurement.

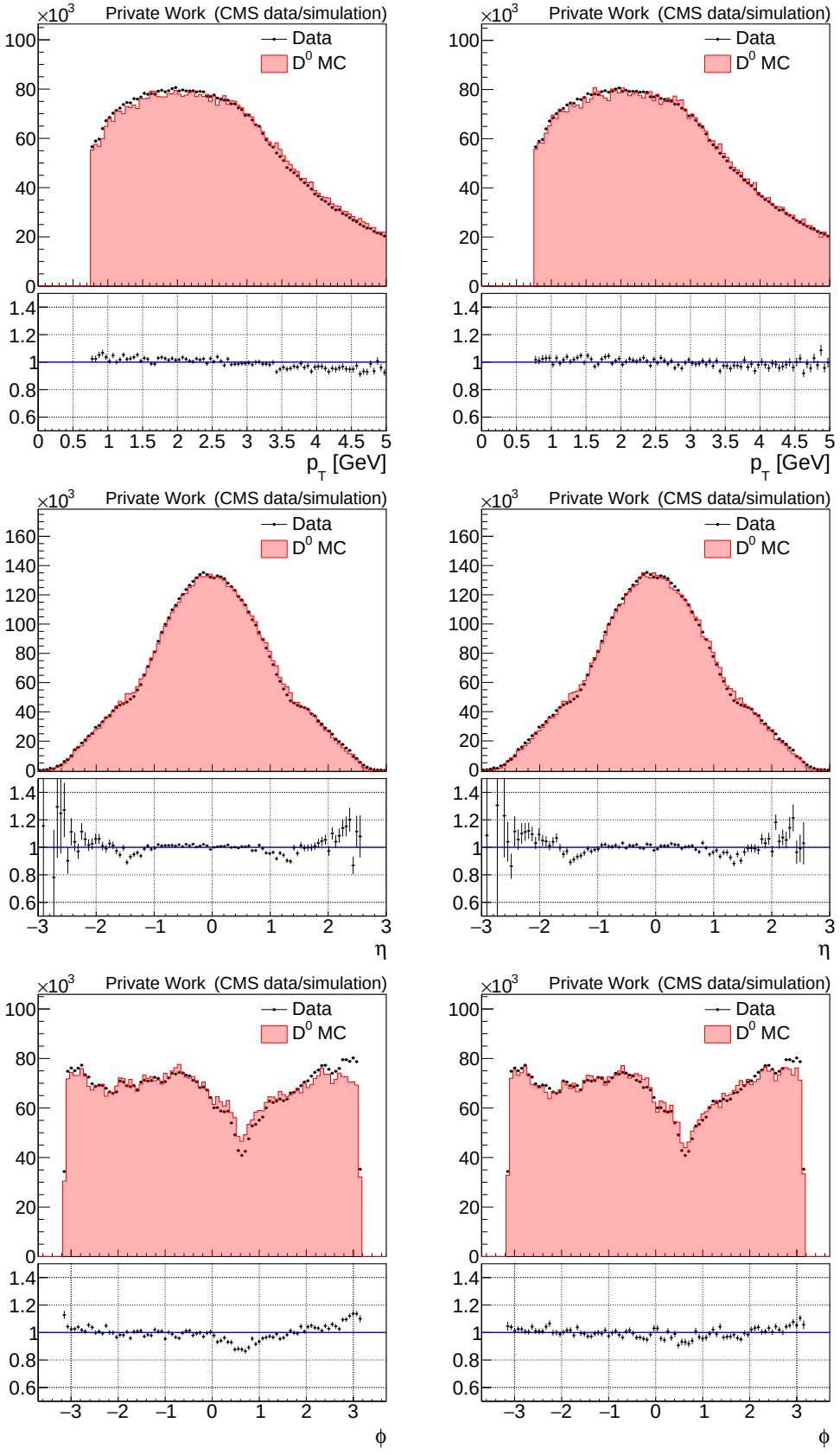


Figure E.1: K from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

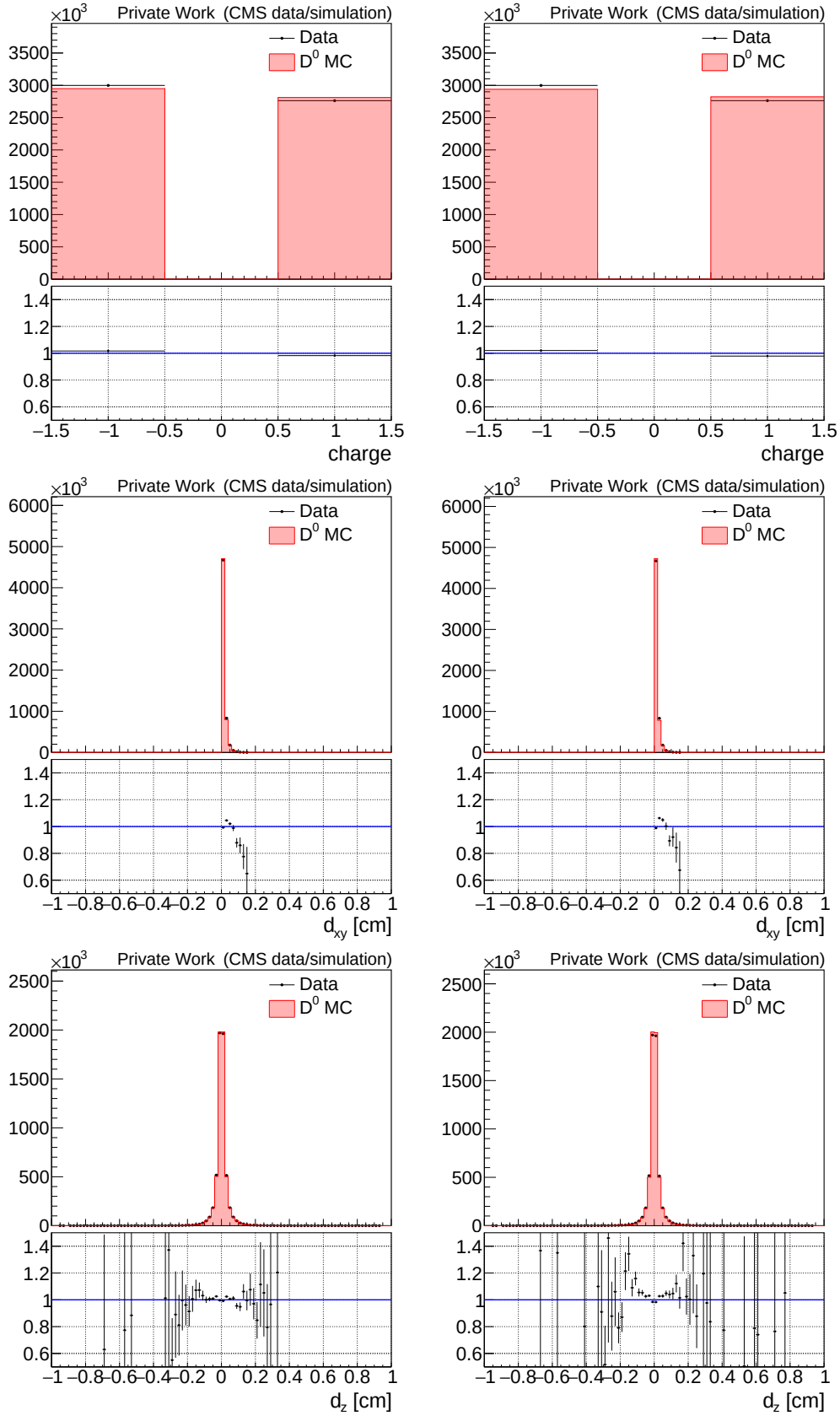


Figure E.2: K from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

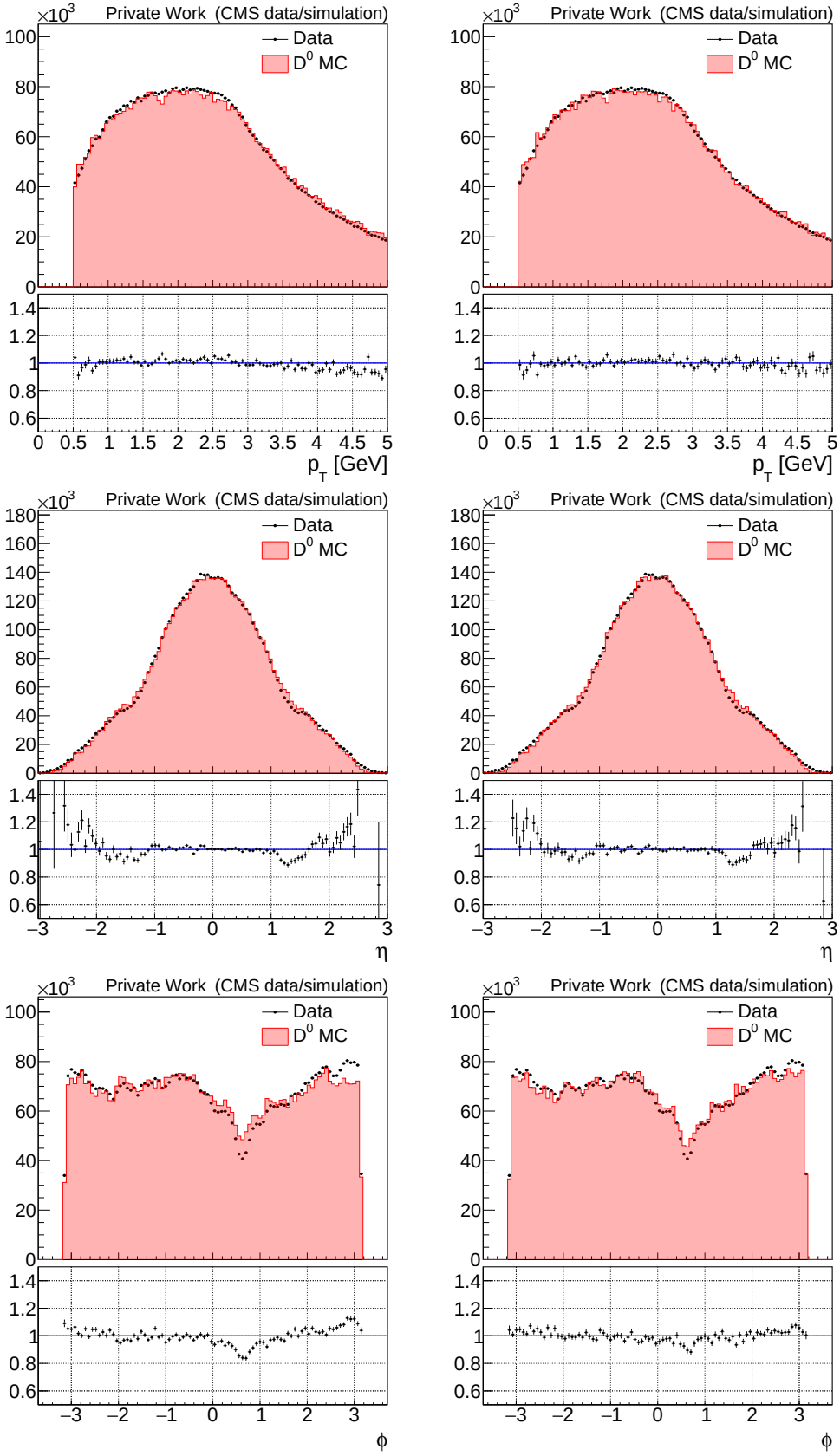


Figure E.3: π from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

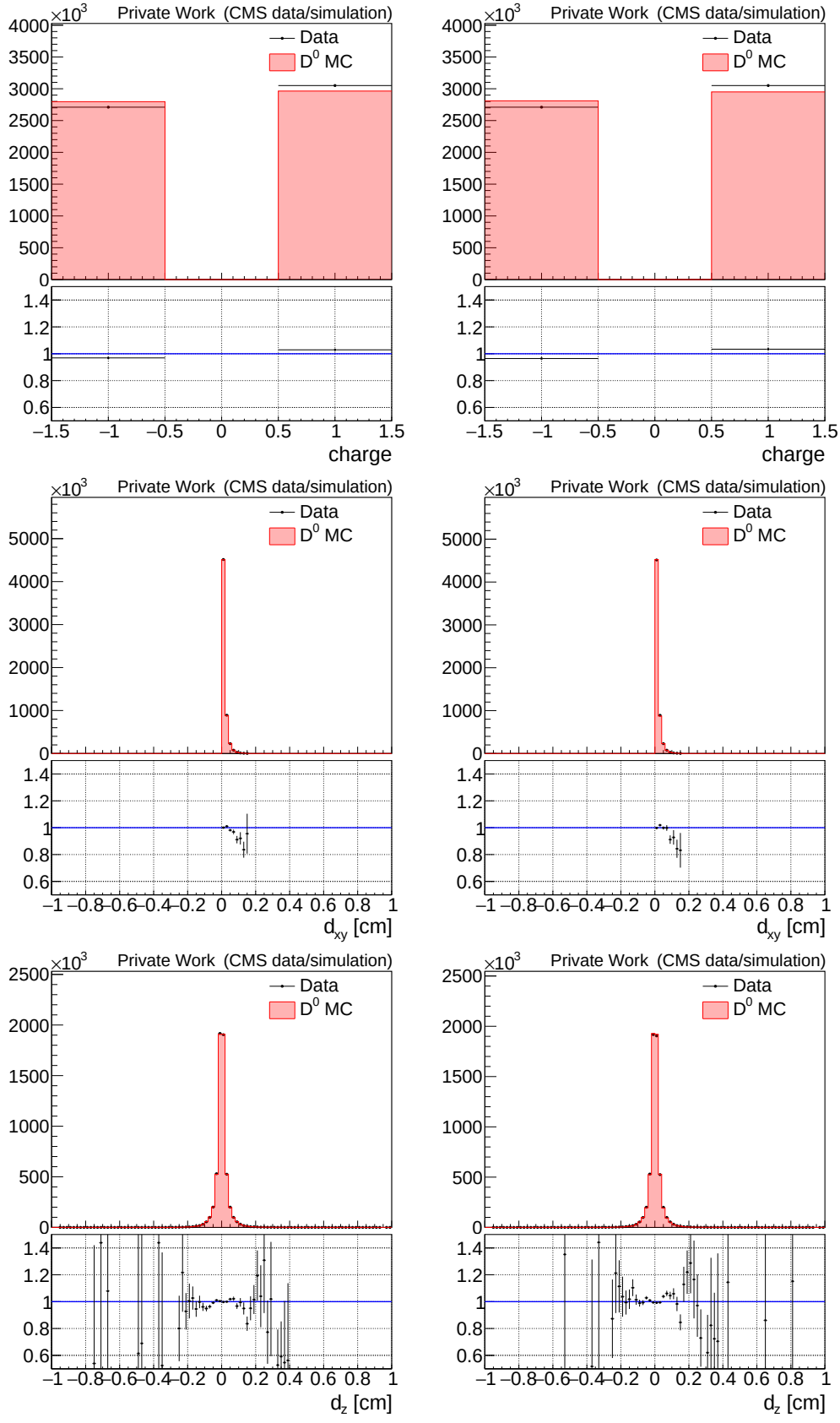


Figure E.4: π from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

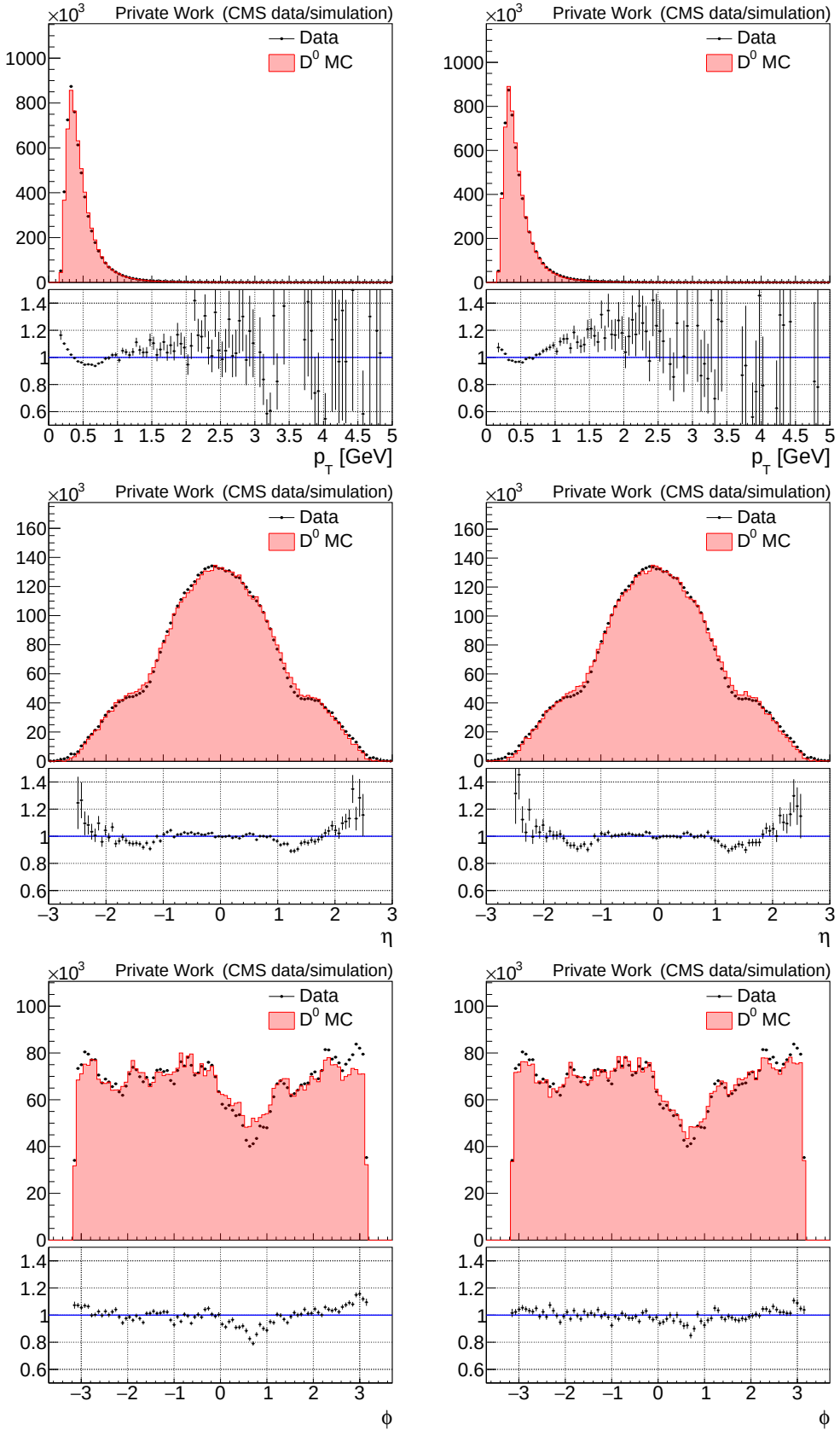


Figure E.5: π_s from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

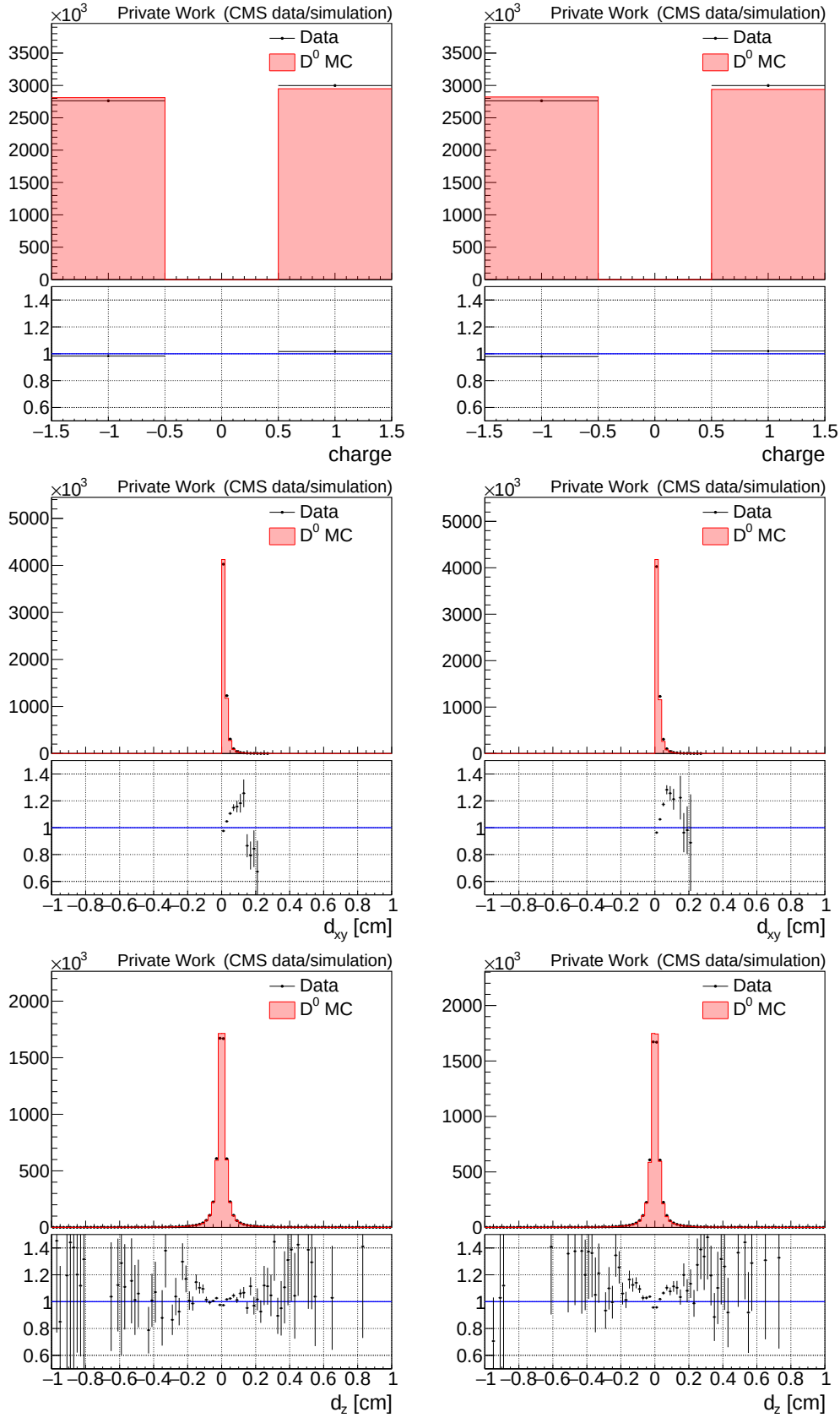


Figure E.6: π_s from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

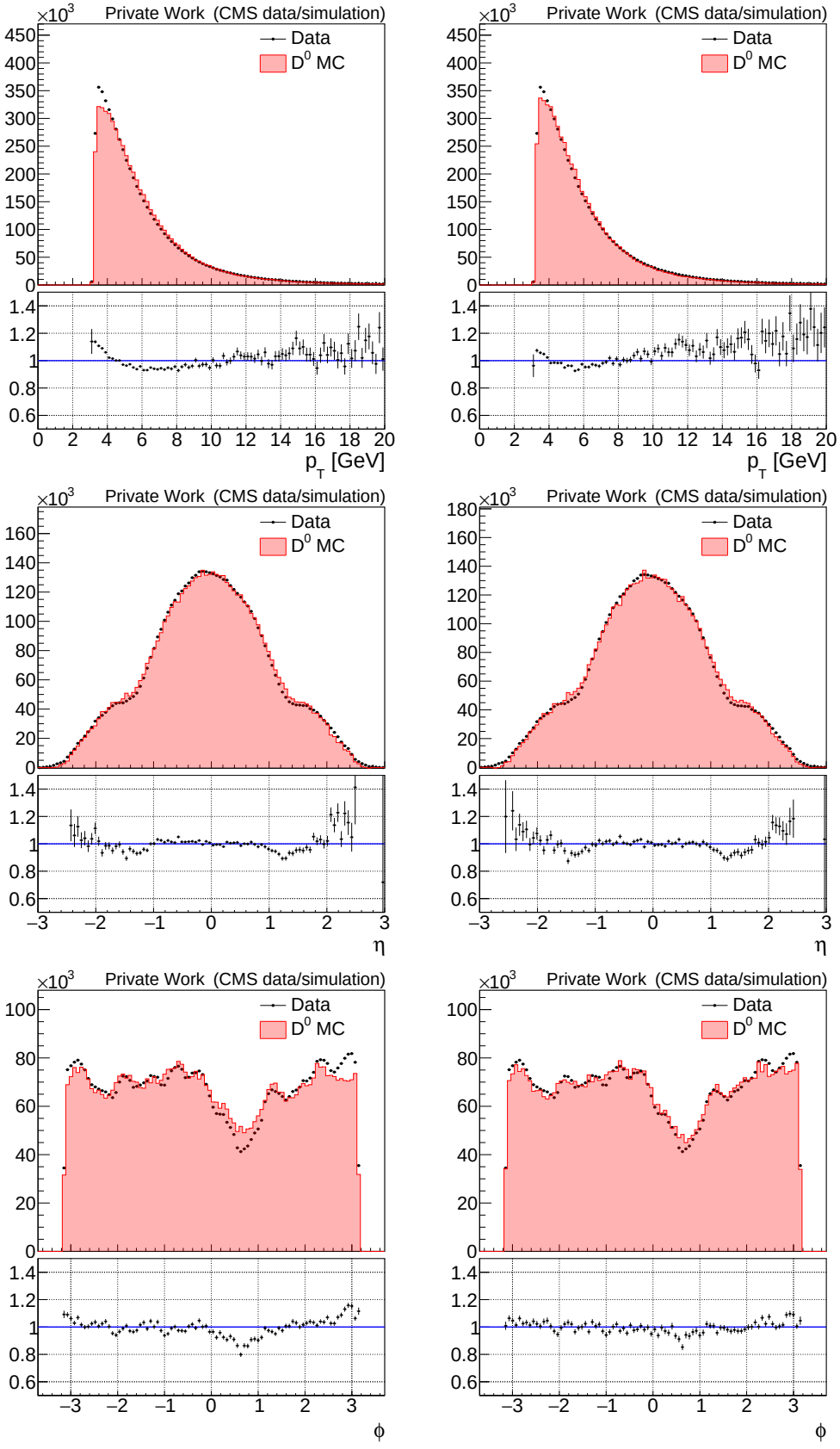


Figure E.7: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

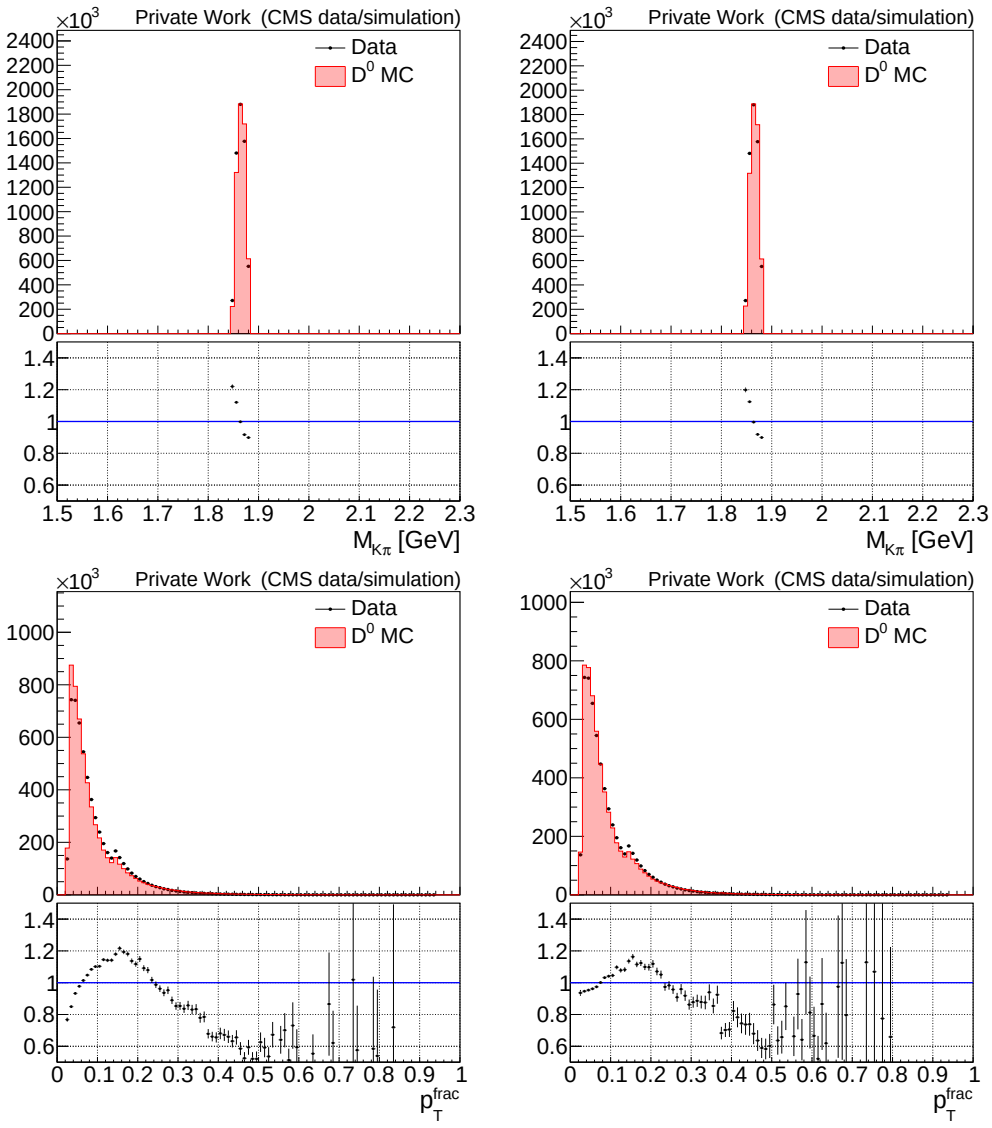


Figure E.8: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

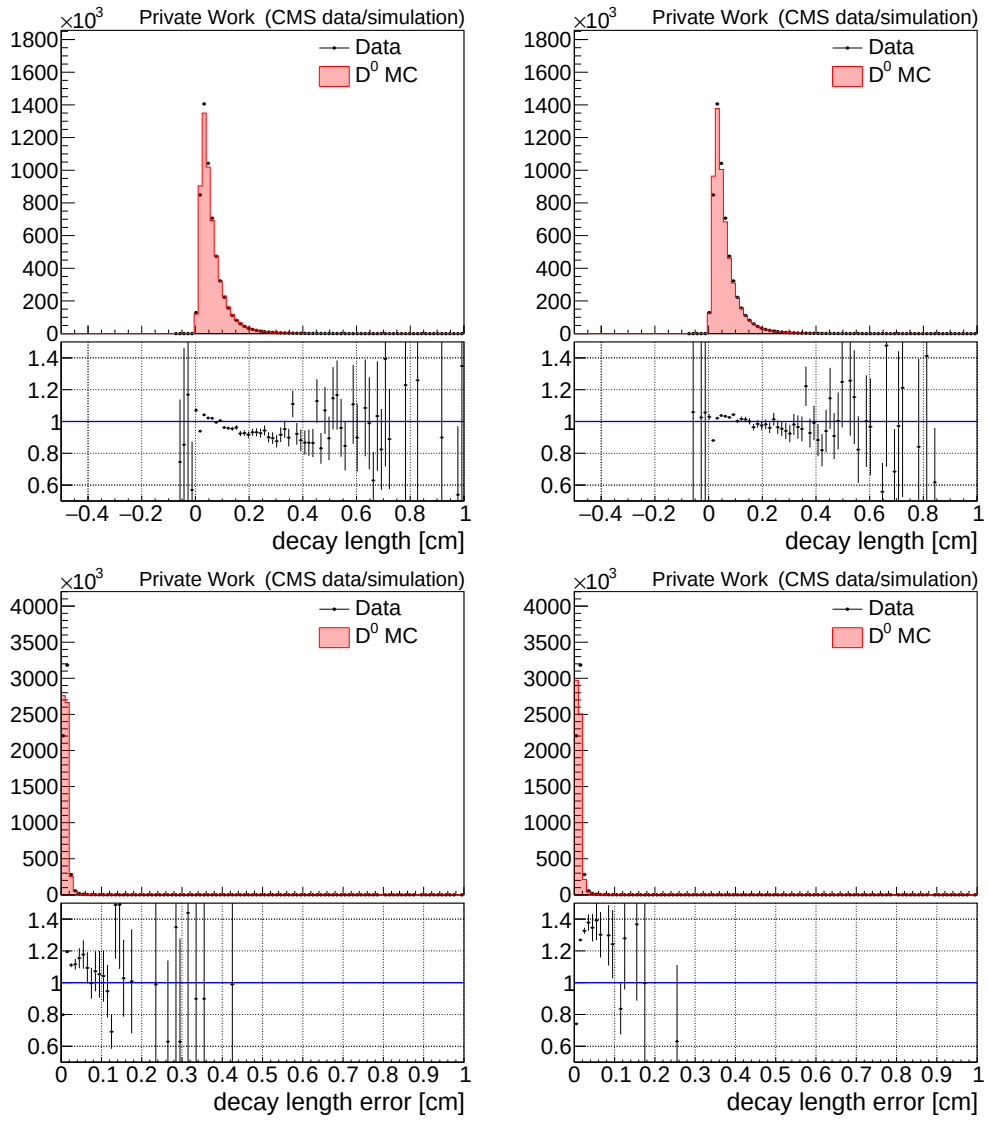


Figure E.9: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

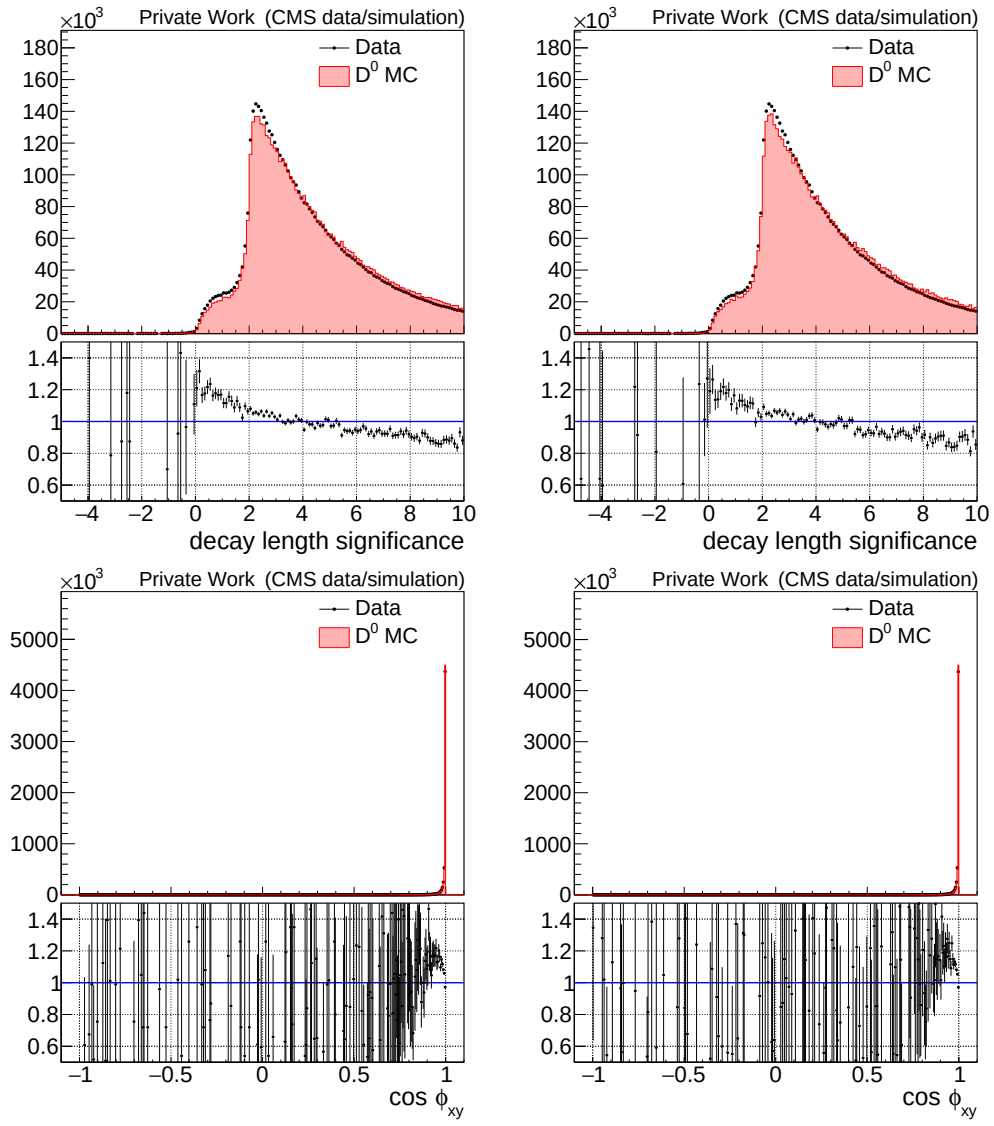


Figure E.10: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

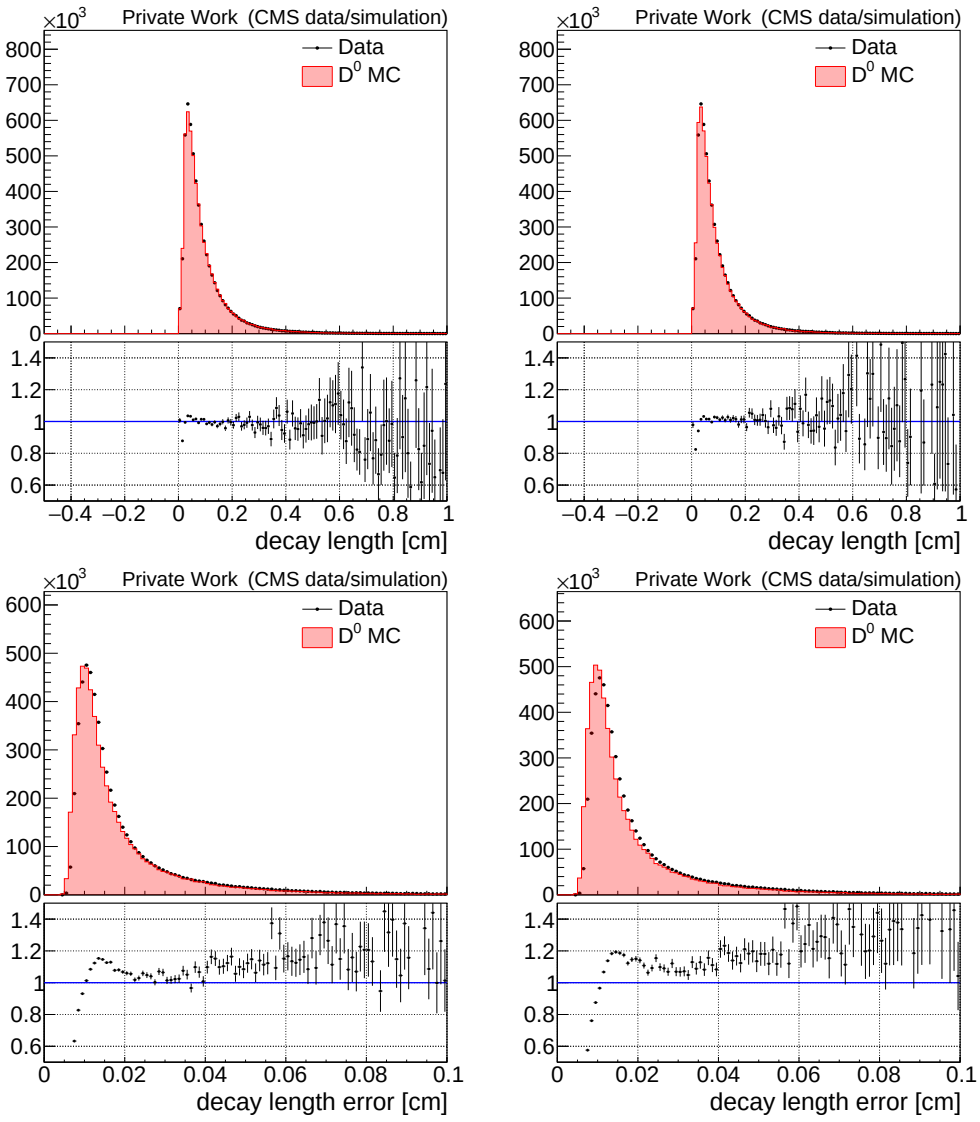


Figure E.11: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

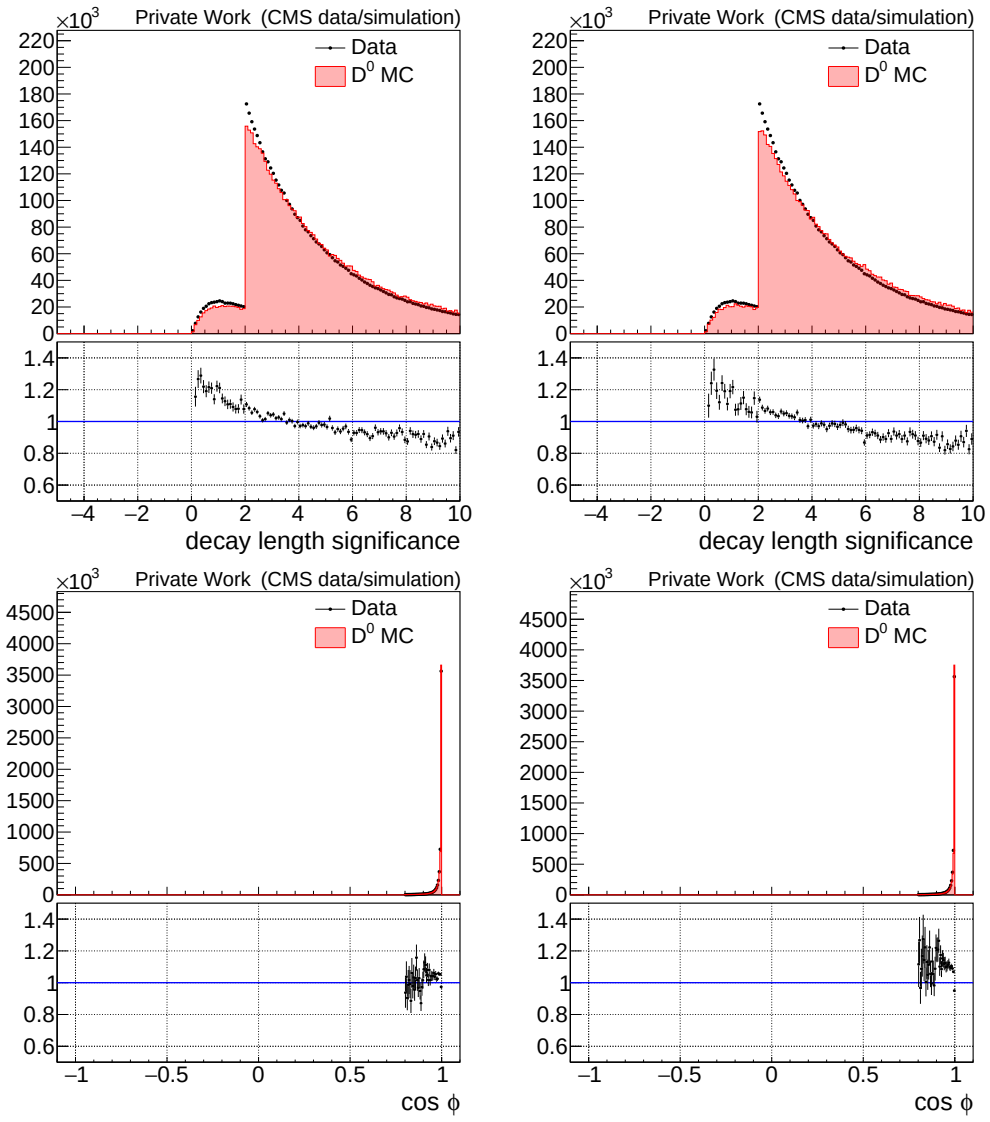


Figure E.12: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

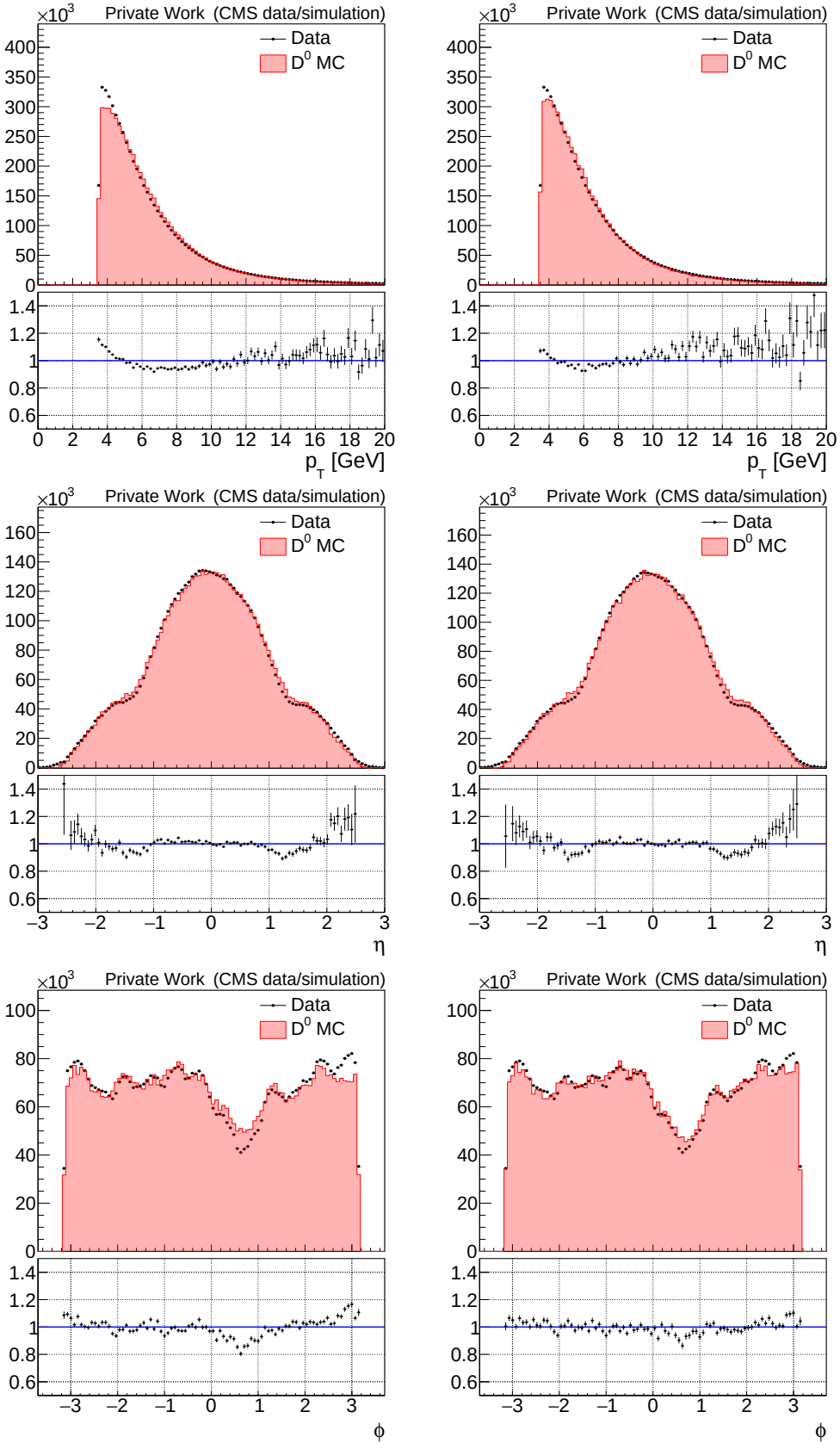


Figure E.13: D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

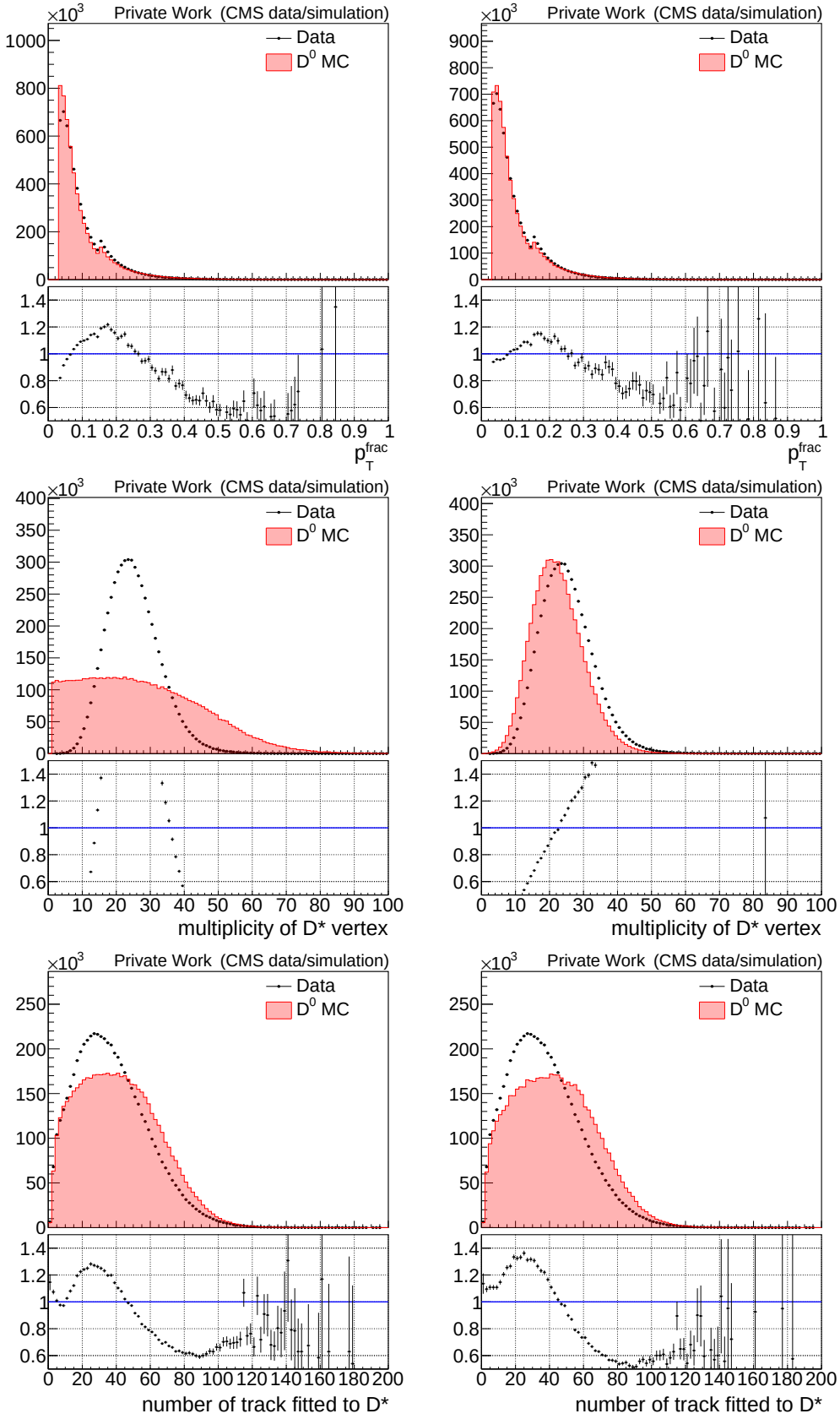


Figure E.14: D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

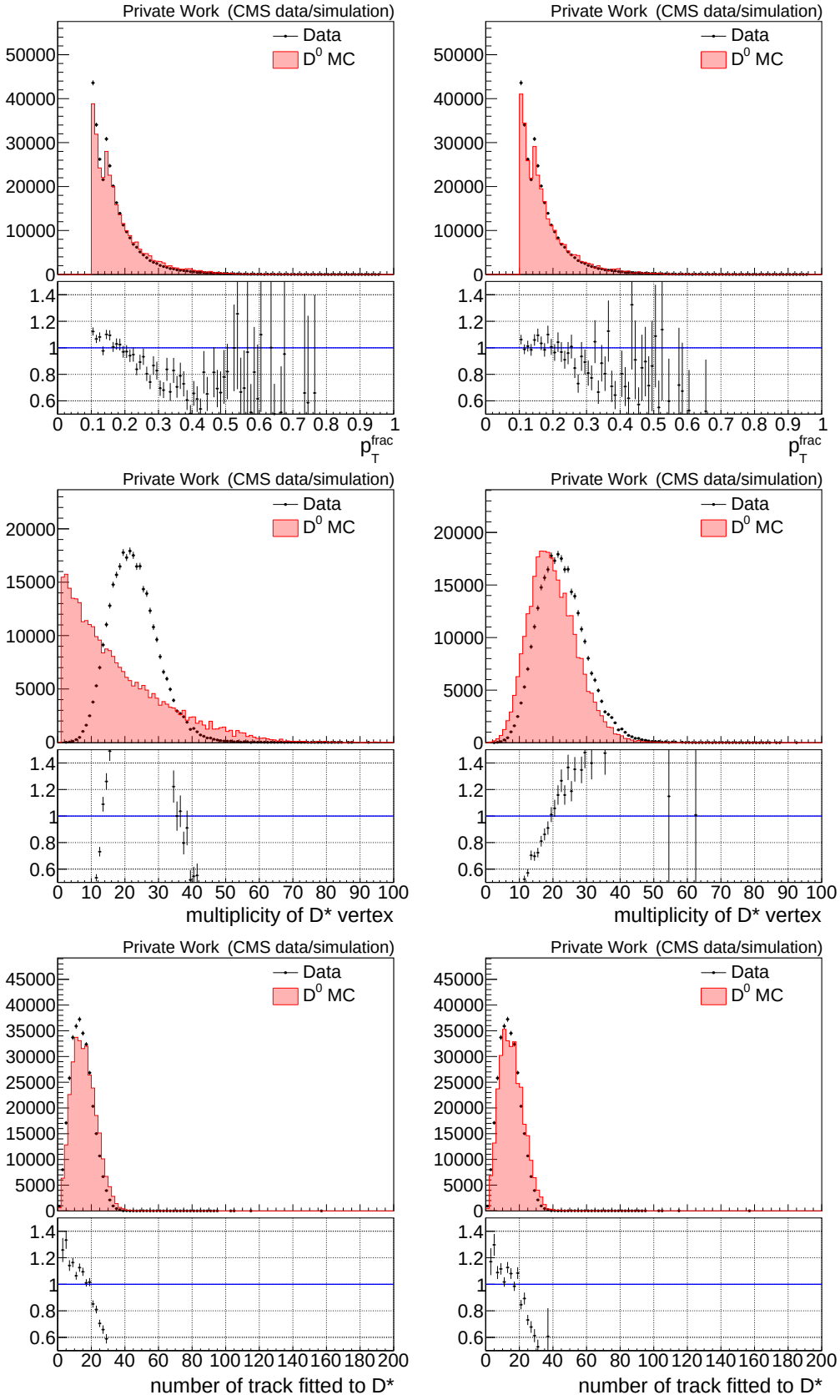


Figure E.15: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

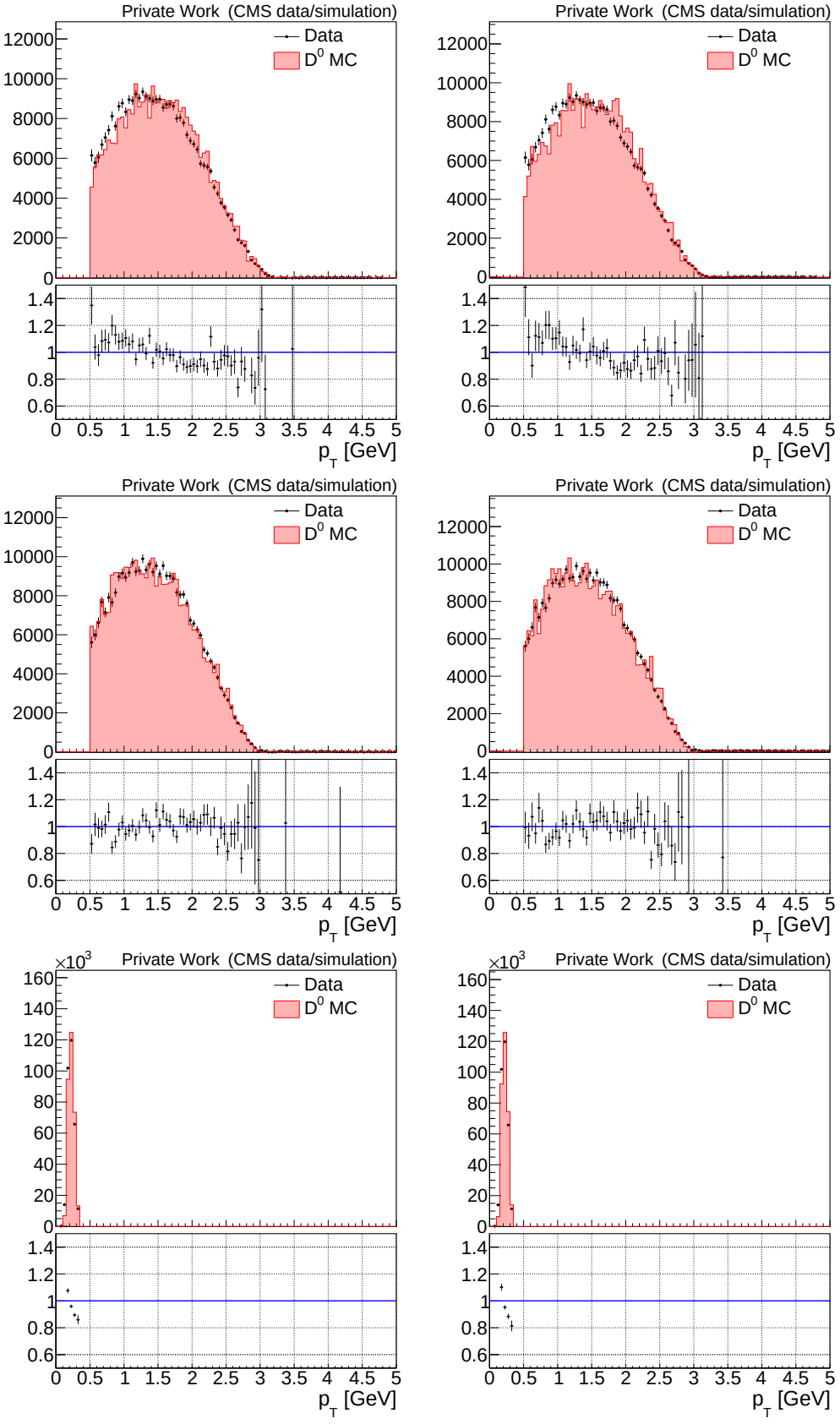


Figure E.16: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

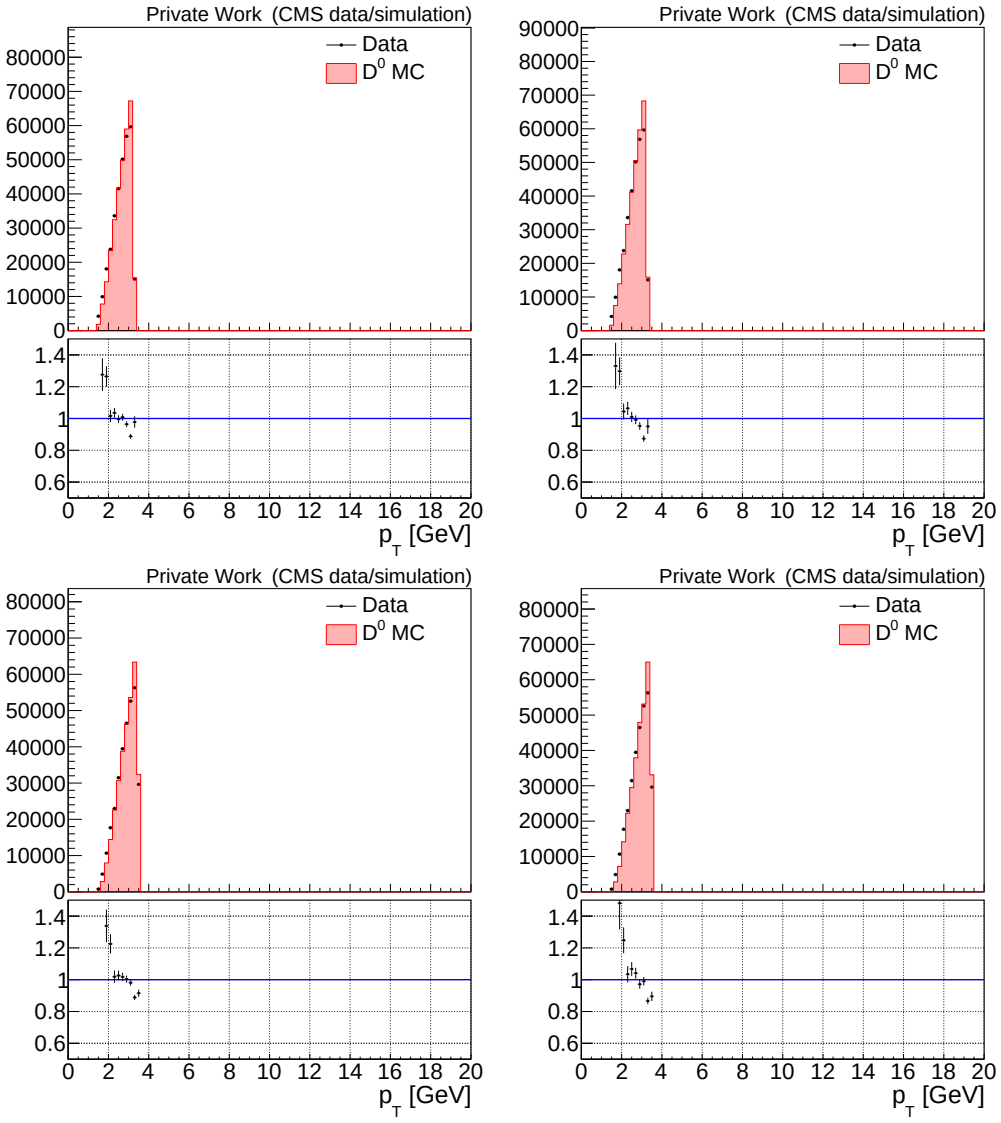


Figure E.17: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

E.2 MC comparison with PU reweight

In this appendix, the 13 TeV MinimumBias MC is compared to the D^0 MC sample. PU distribution (PV_npvs) of the D^0 MC was reweighted to that of the MinimumBias MC (refer to Figure 5.4). No significant discrepancy is observed between the MC samples after reweighting.

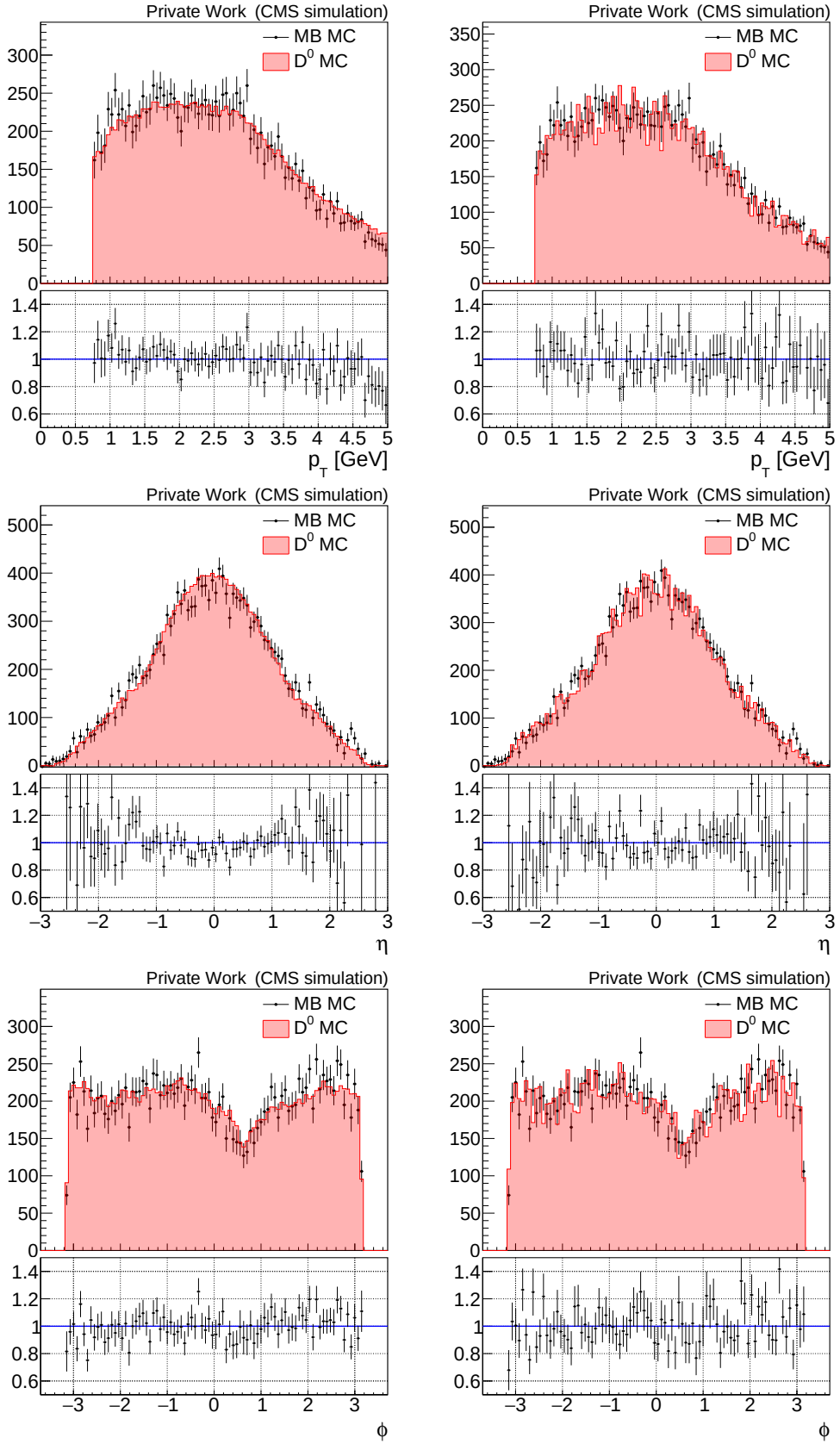


Figure E.18: K from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

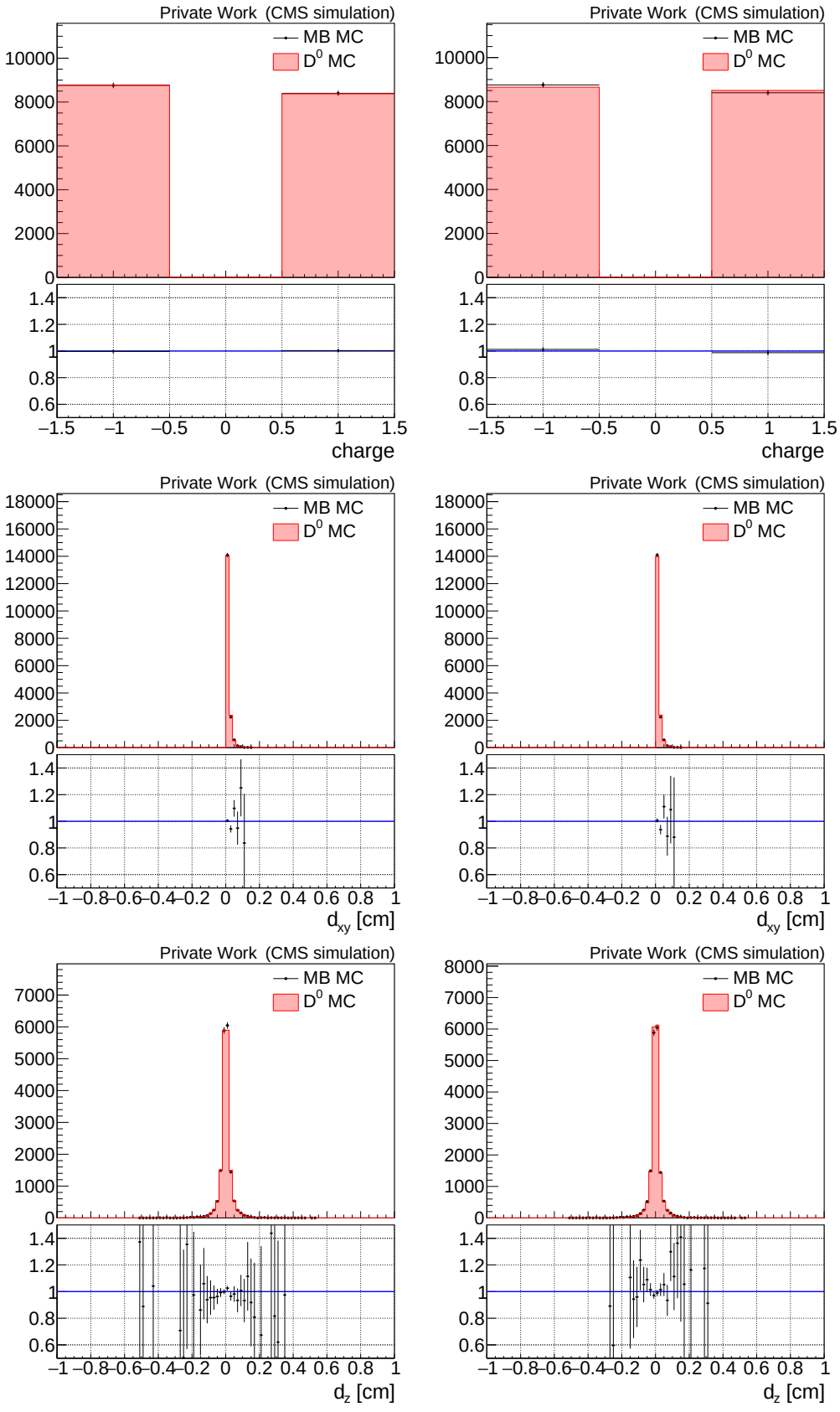


Figure E.19: K from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

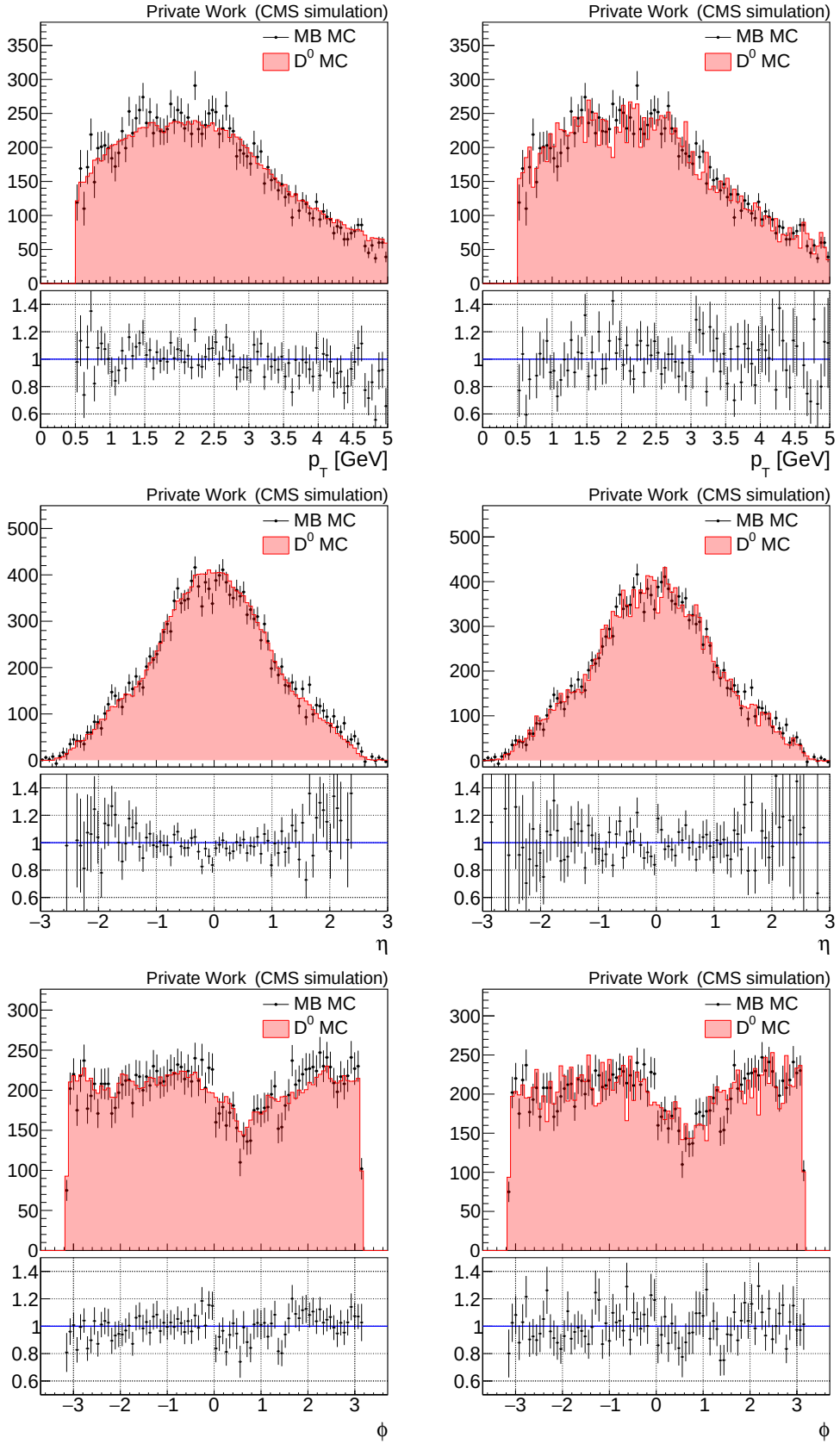


Figure E.20: π from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

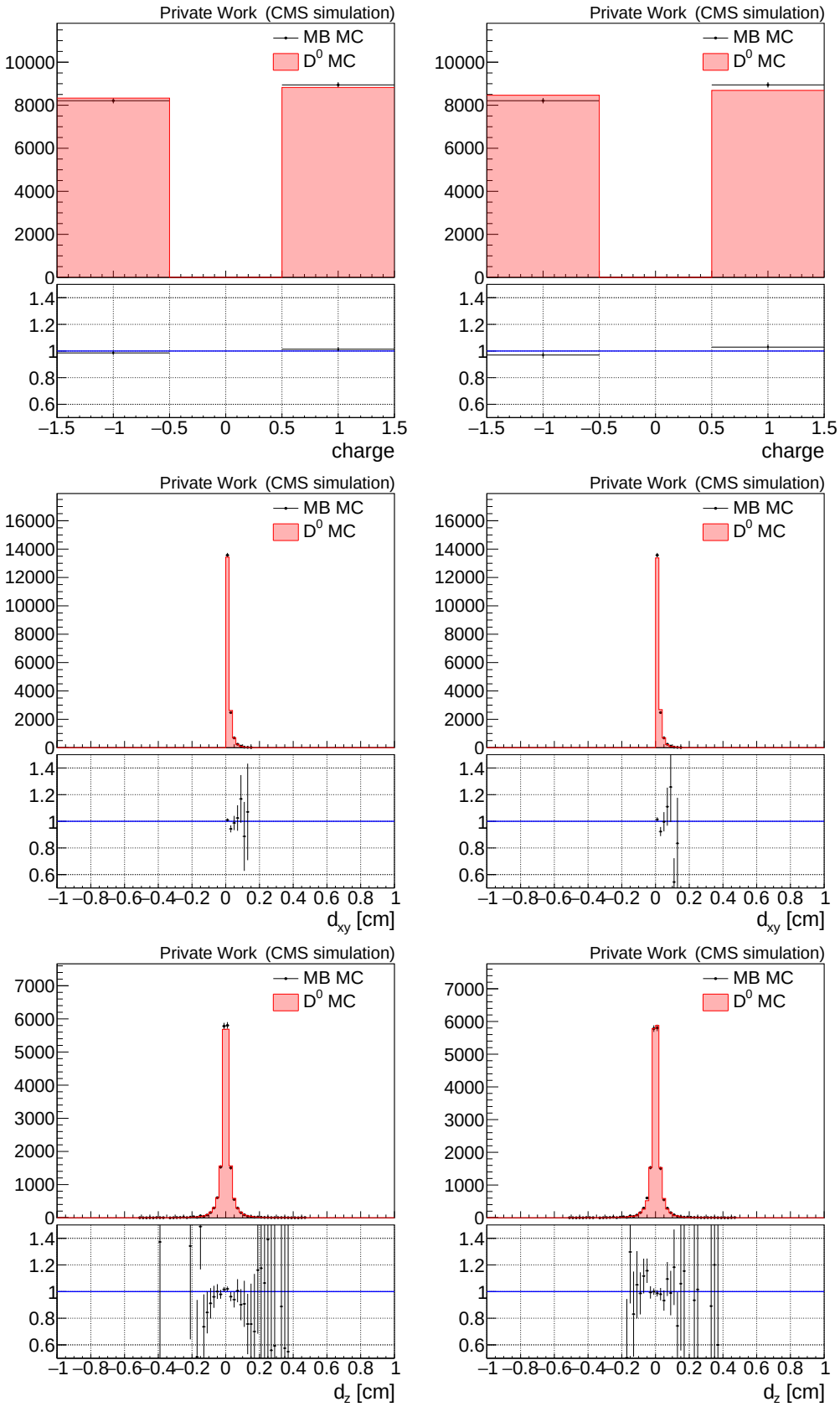


Figure E.21: π from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

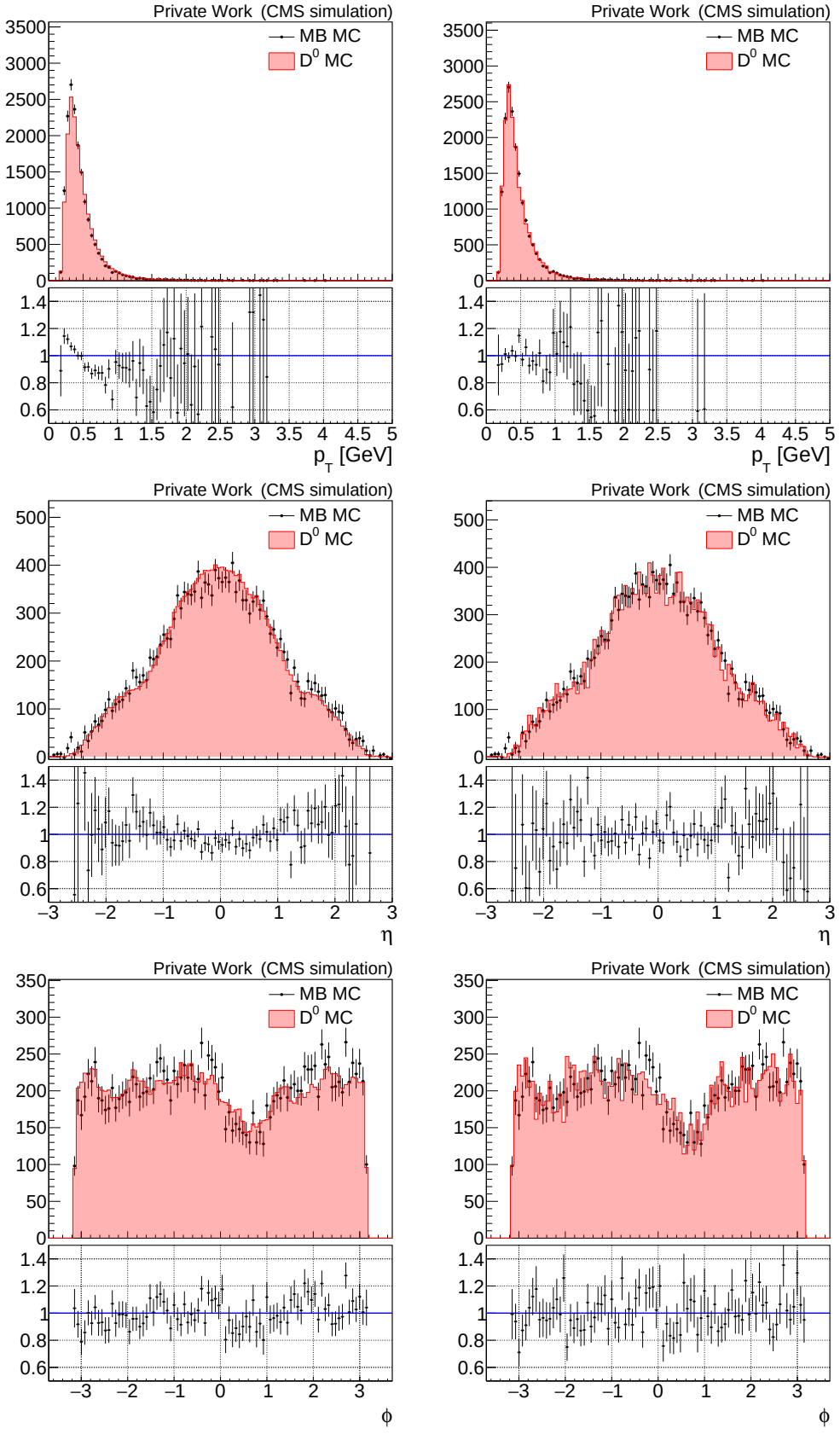


Figure E.22: π_s from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

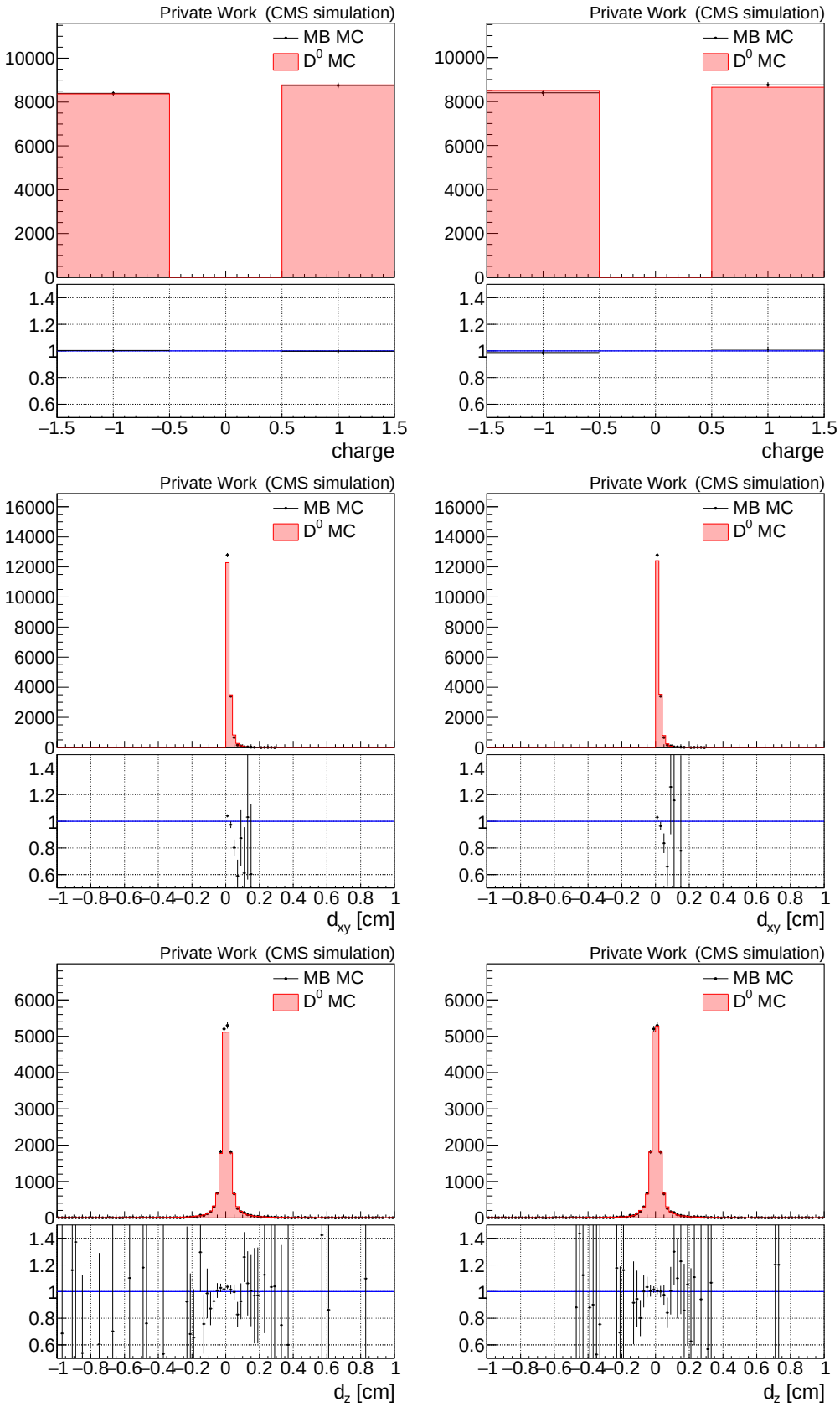


Figure E.23: π_s from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

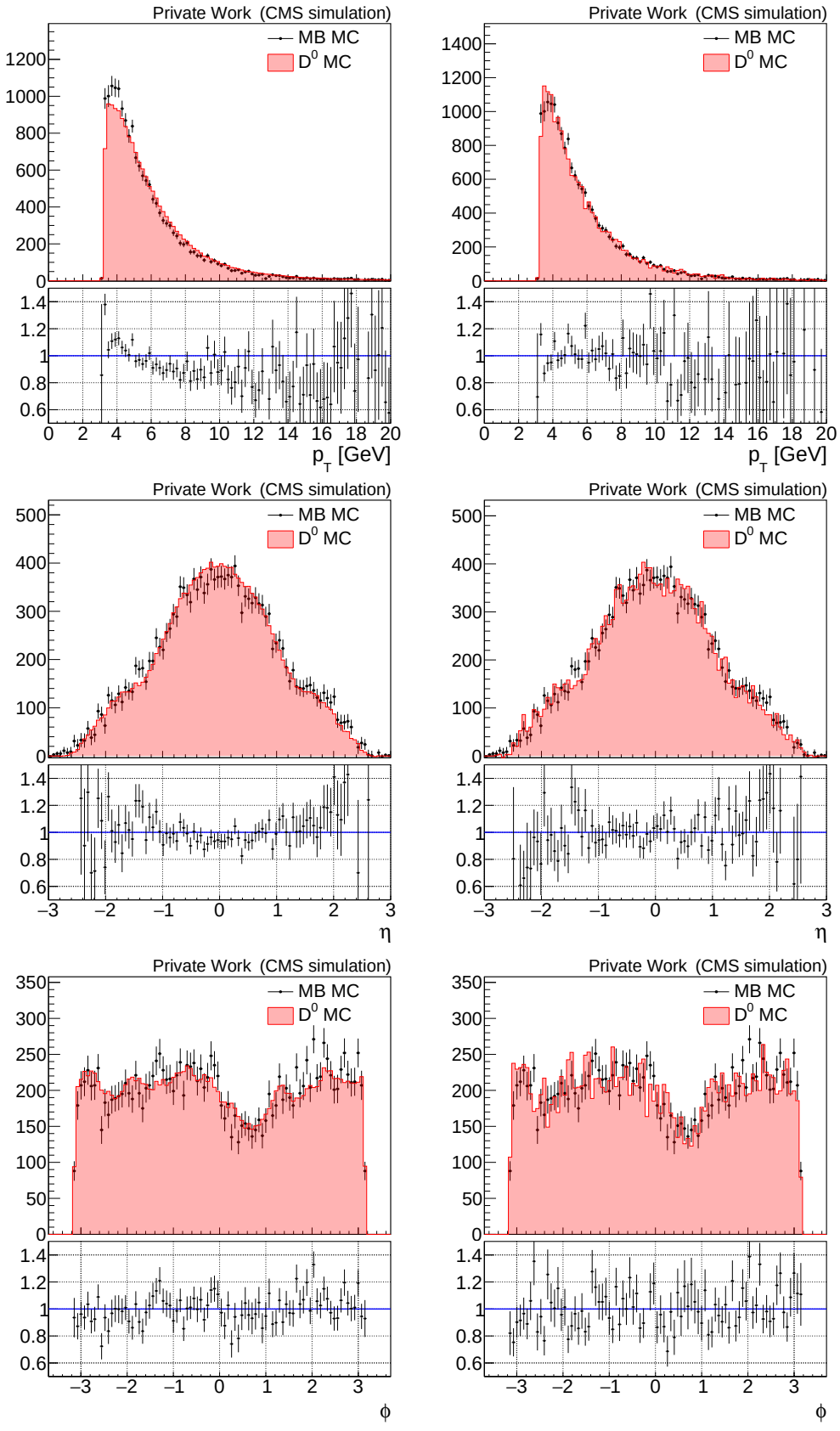


Figure E.24: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

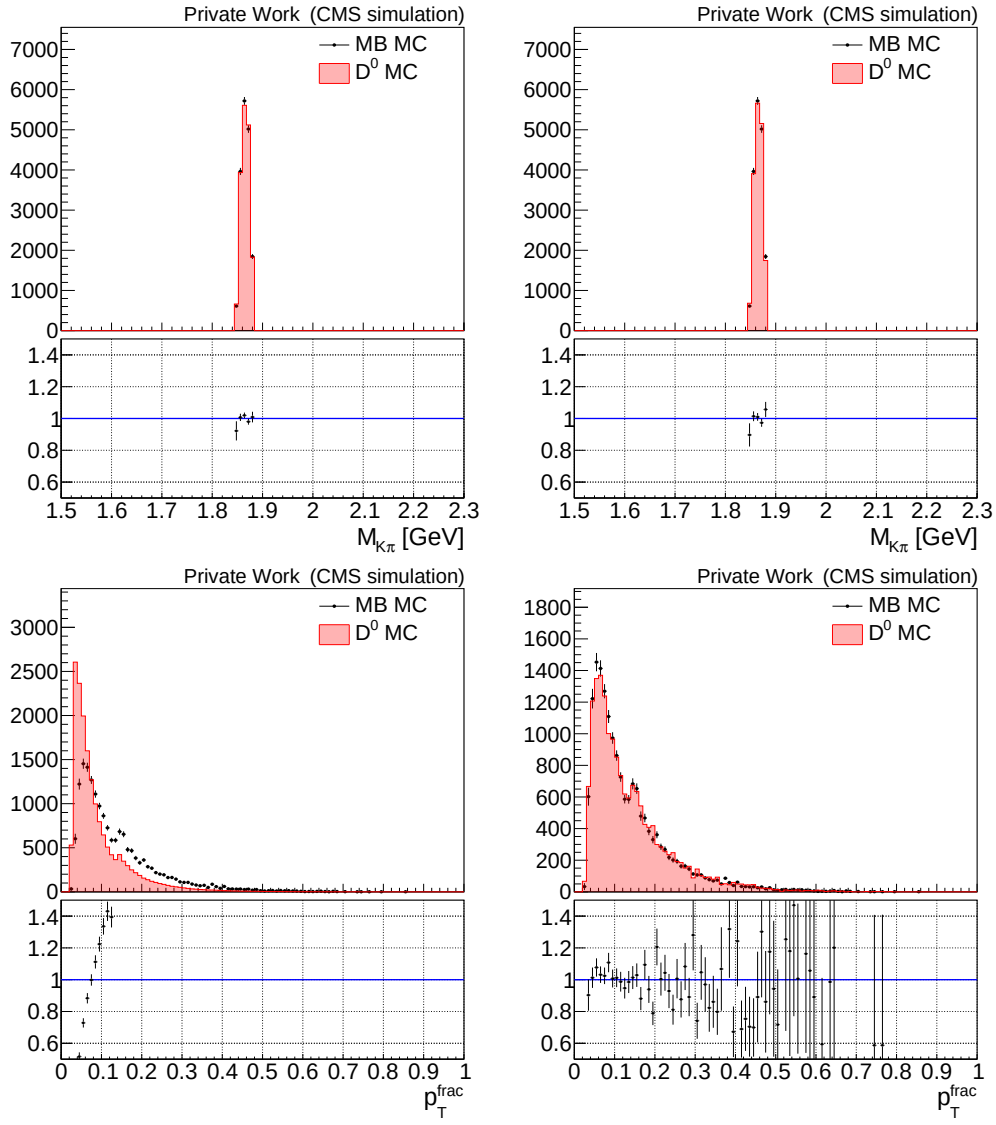


Figure E.25: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

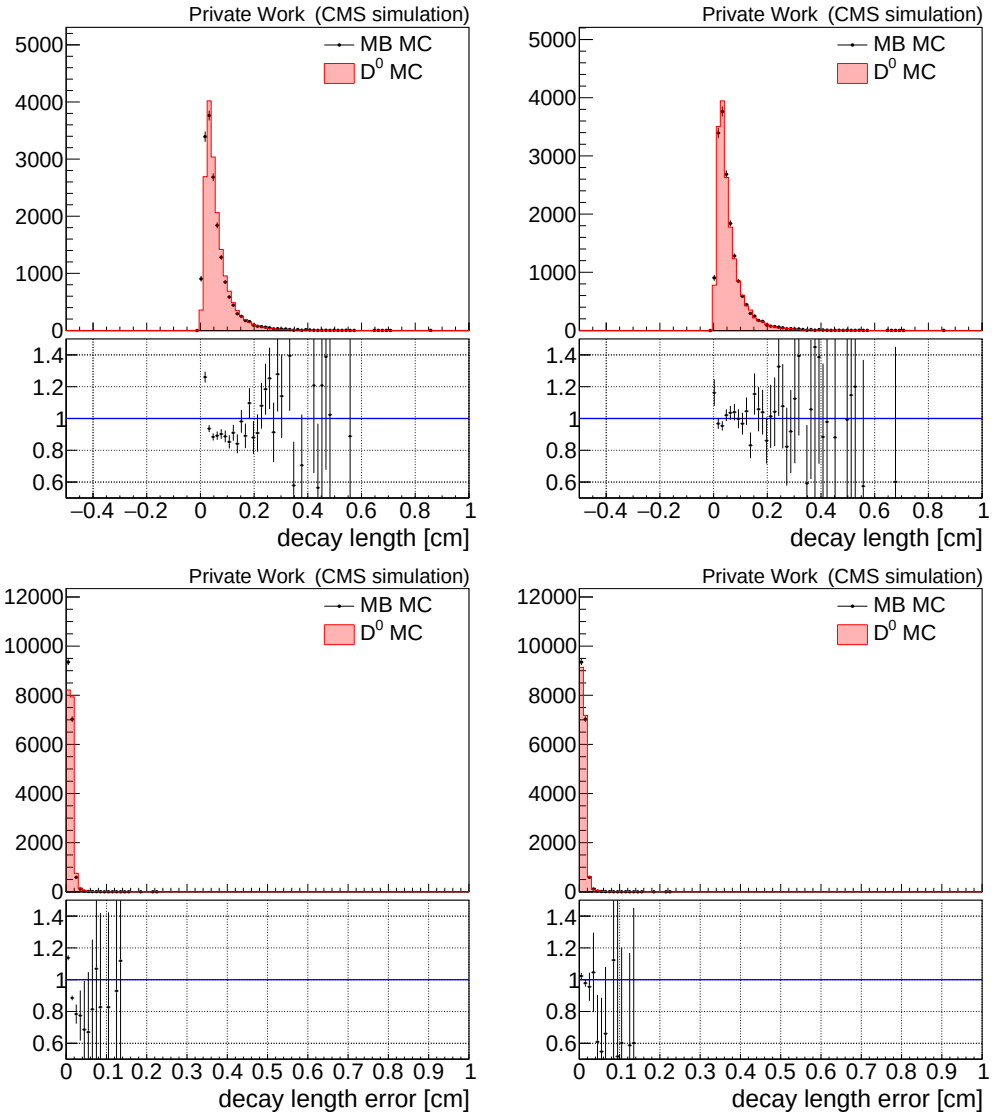


Figure E.26: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

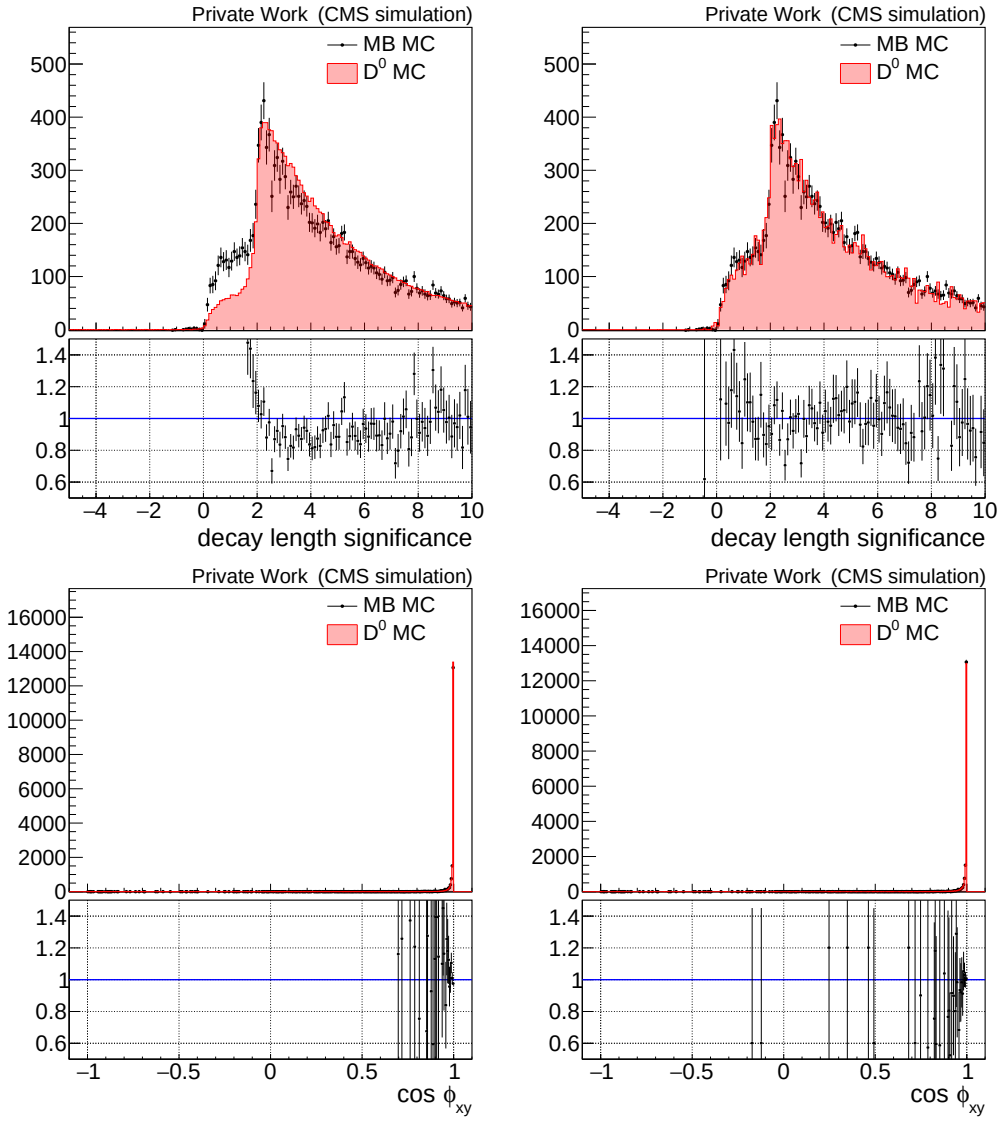


Figure E.27: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

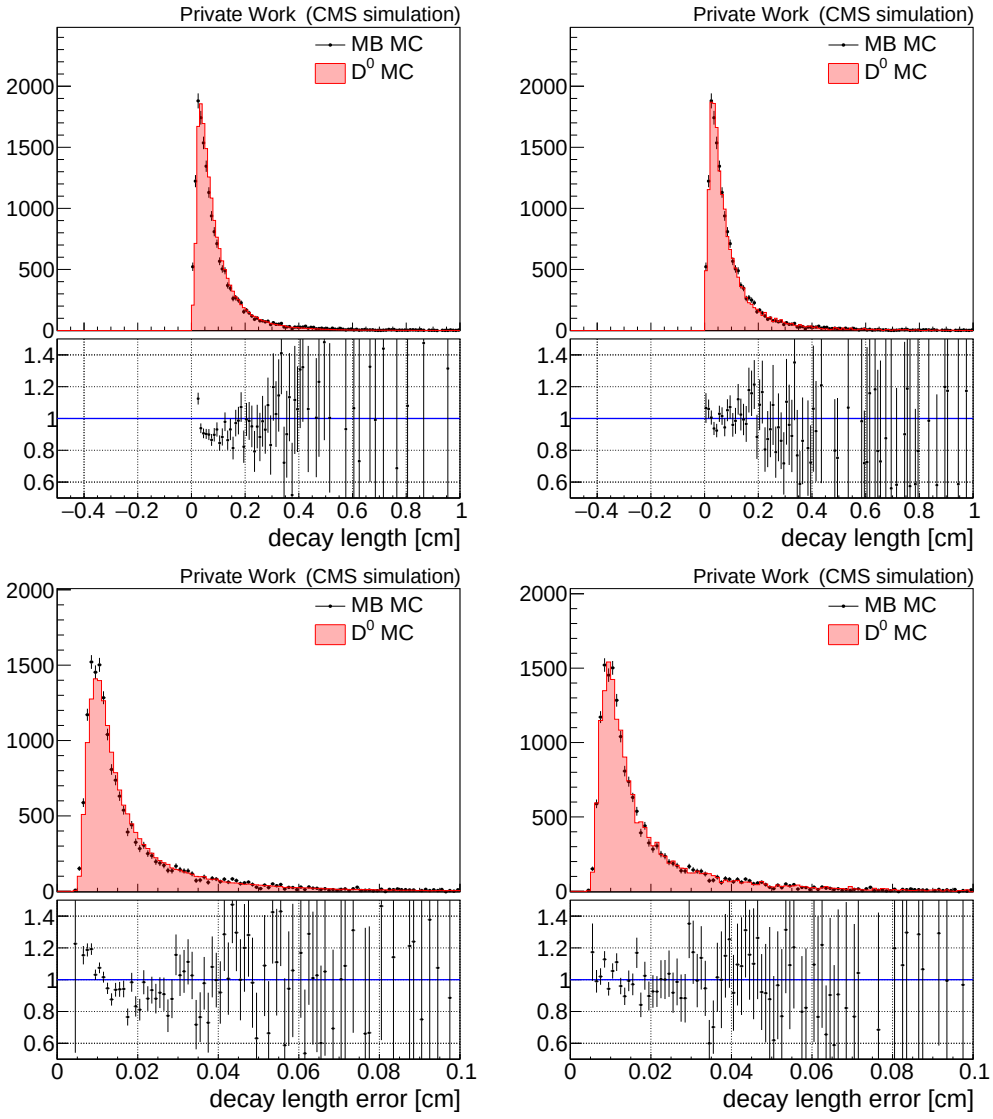


Figure E.28: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

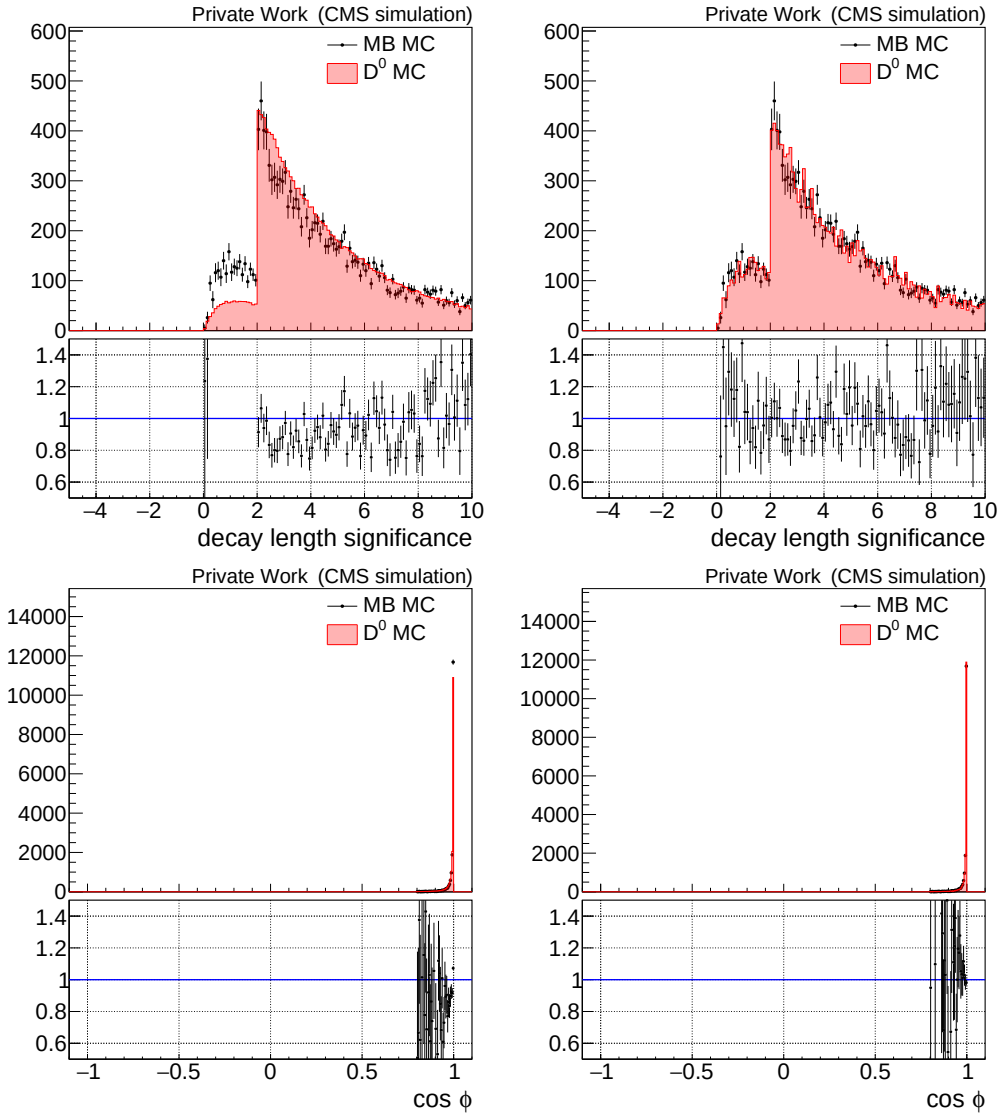


Figure E.29: D^0 from D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

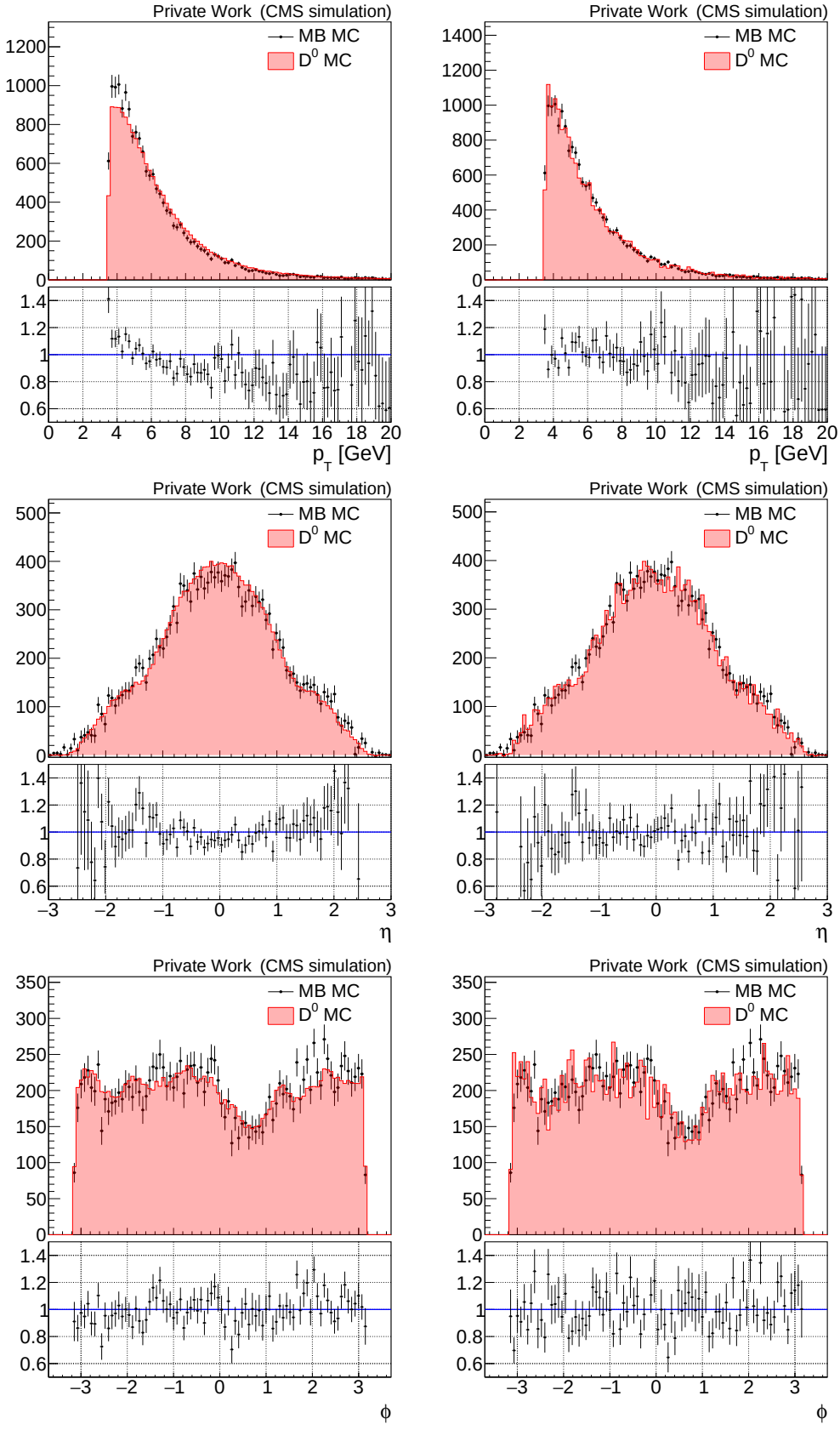


Figure E.30: D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

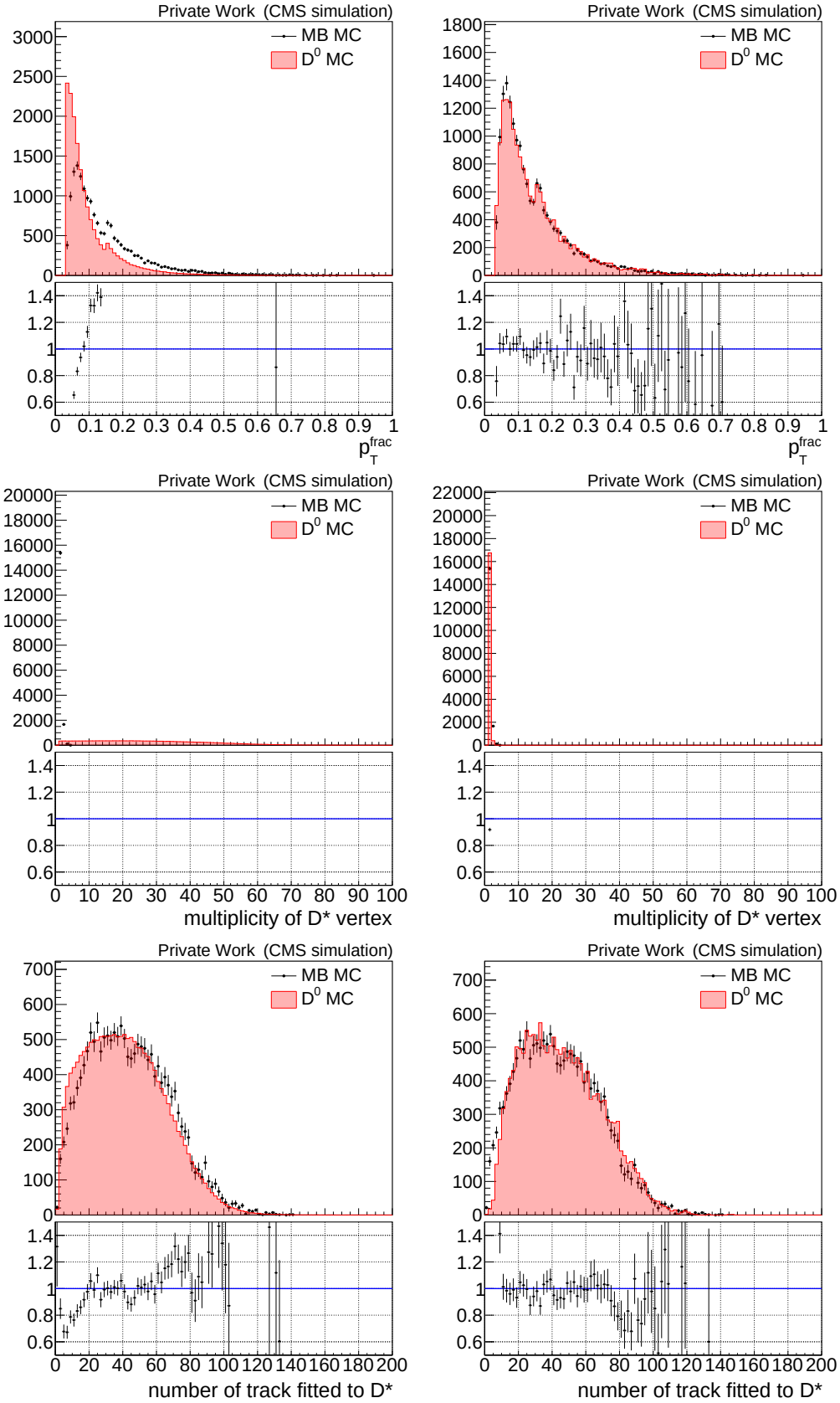


Figure E.31: D^* in $p_T > 3.5$ GeV before (left) and after (right) reweighting.

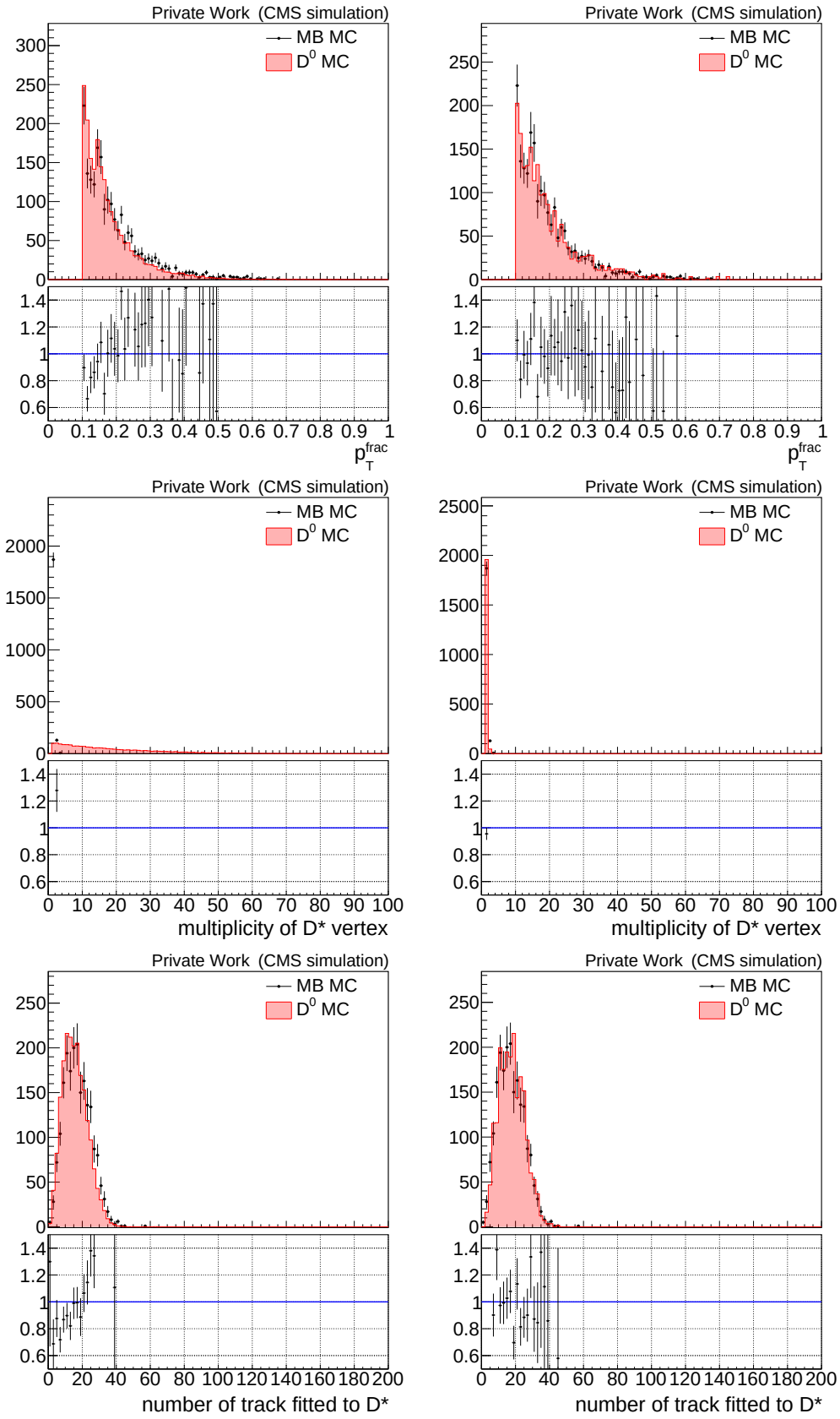


Figure E.32: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

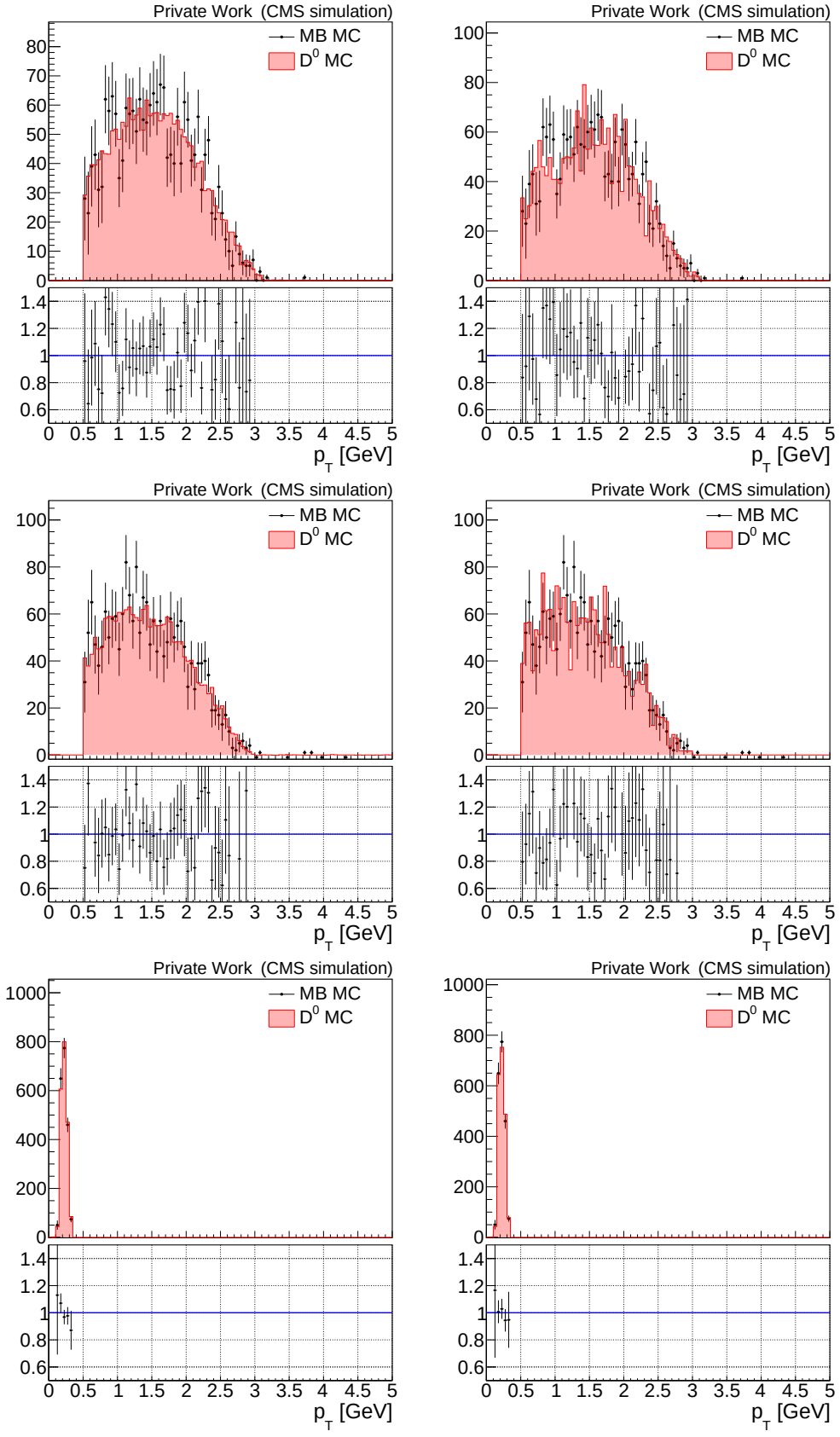


Figure E.33: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

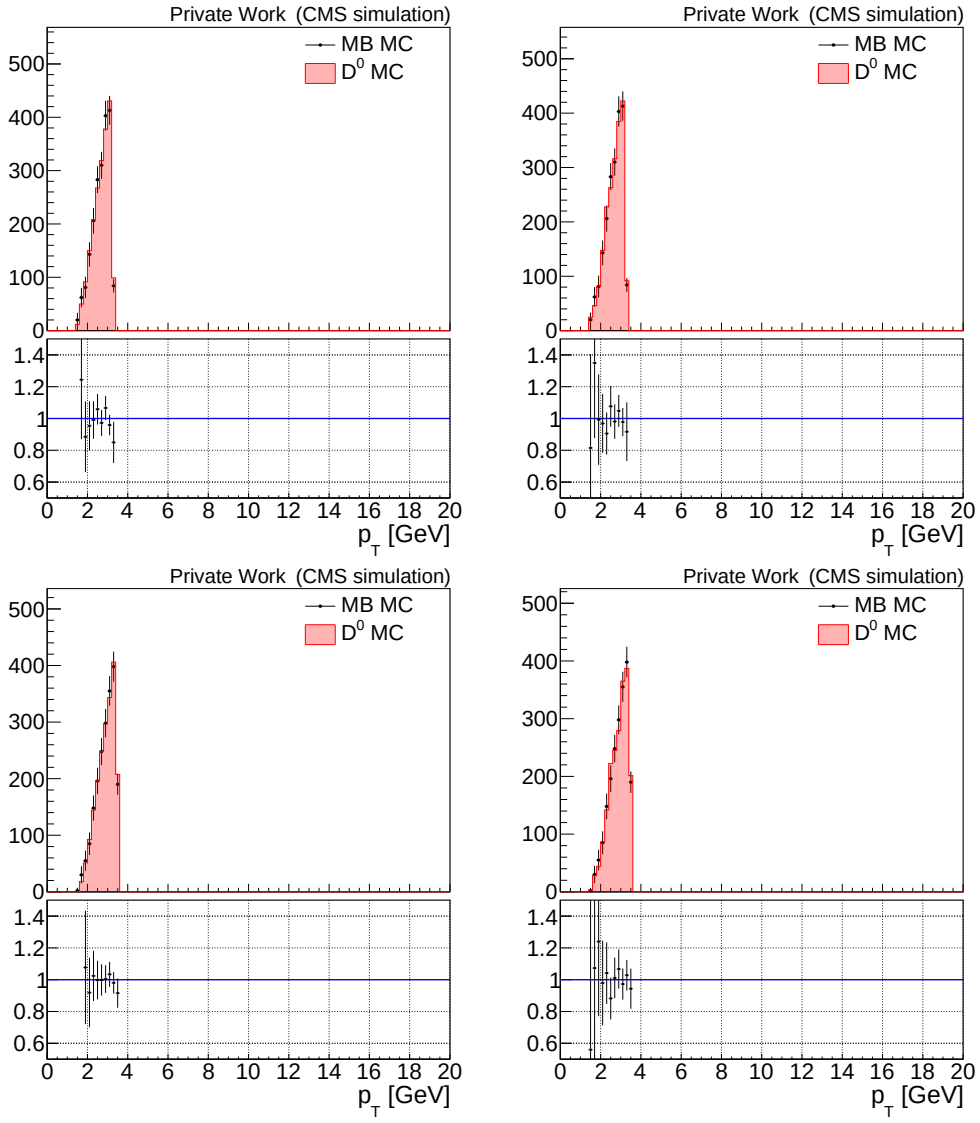


Figure E.34: $p_T < 3.5$ GeV before (left) and after (right) reweighting.

F 13 TeV MC Result for π_s^+ Tracking Efficiency

In this appendix, D^{*+} and D^0 signal are shown on the 13 TeV D^0 and MinimumBias MC sample. Ratios of the number of the signals are also shown for each sample. The results shown here were used to the π_s^+ efficiency measurement in Appendix C.

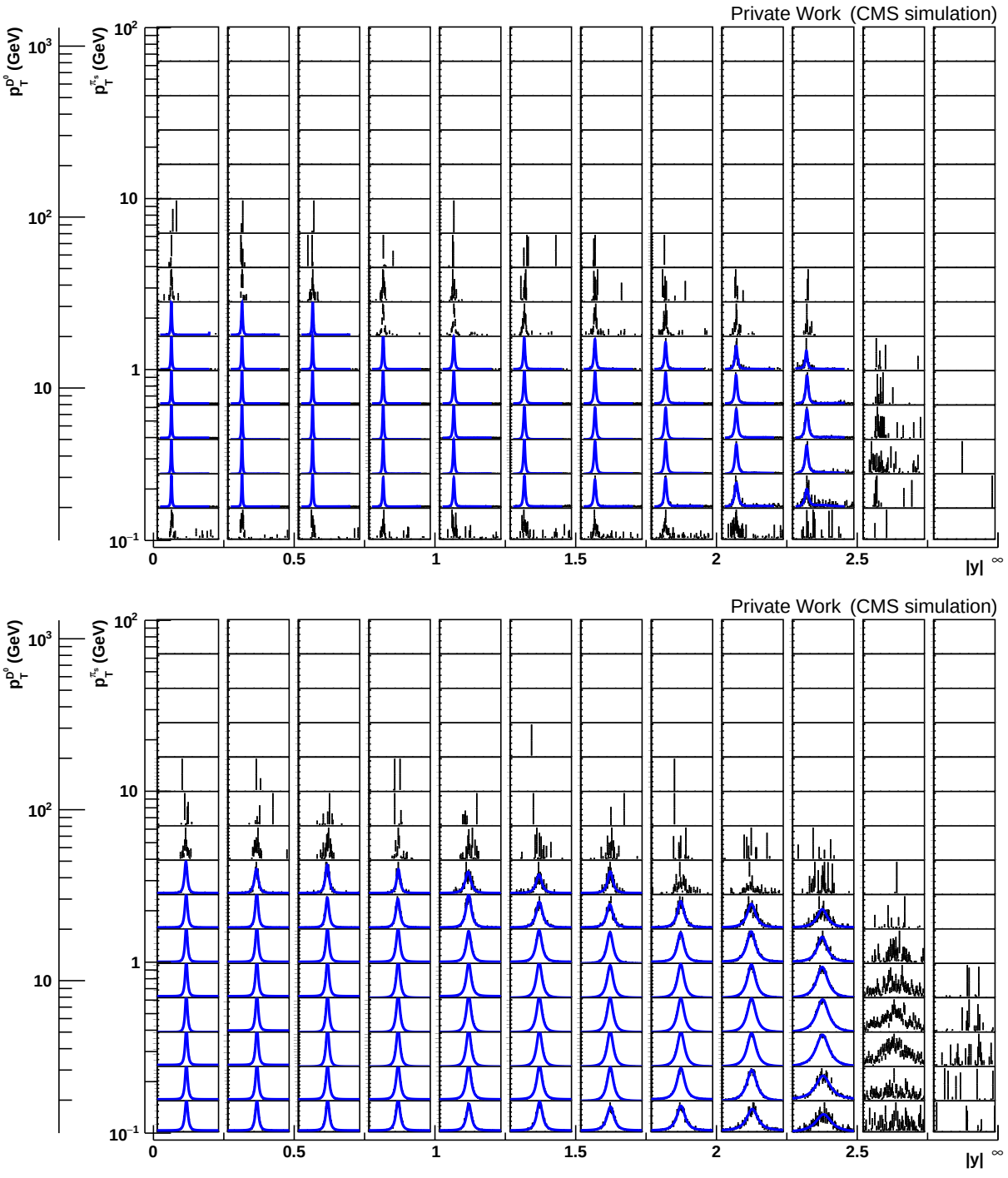


Figure F.1: D^{*+} (top) and D^0 (bottom) reconstruction on the 13 TeV D^0 MC sample for π_s^+ efficiency measurement.

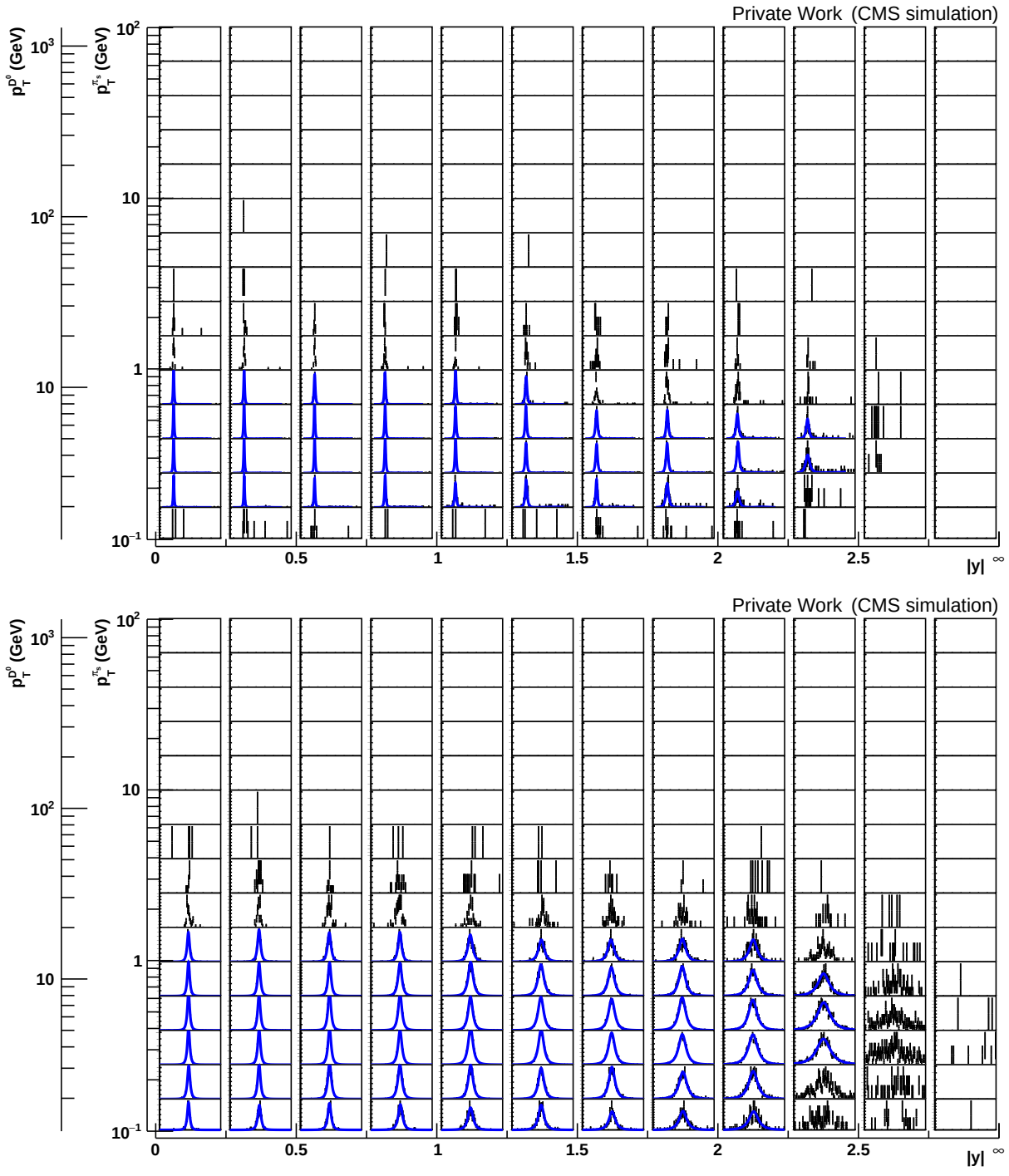


Figure F.2: D^{*+} (top) and D^0 (bottom) reconstruction on the 13 TeV MinimumBais MC sample for π_s^+ efficiency measurement.

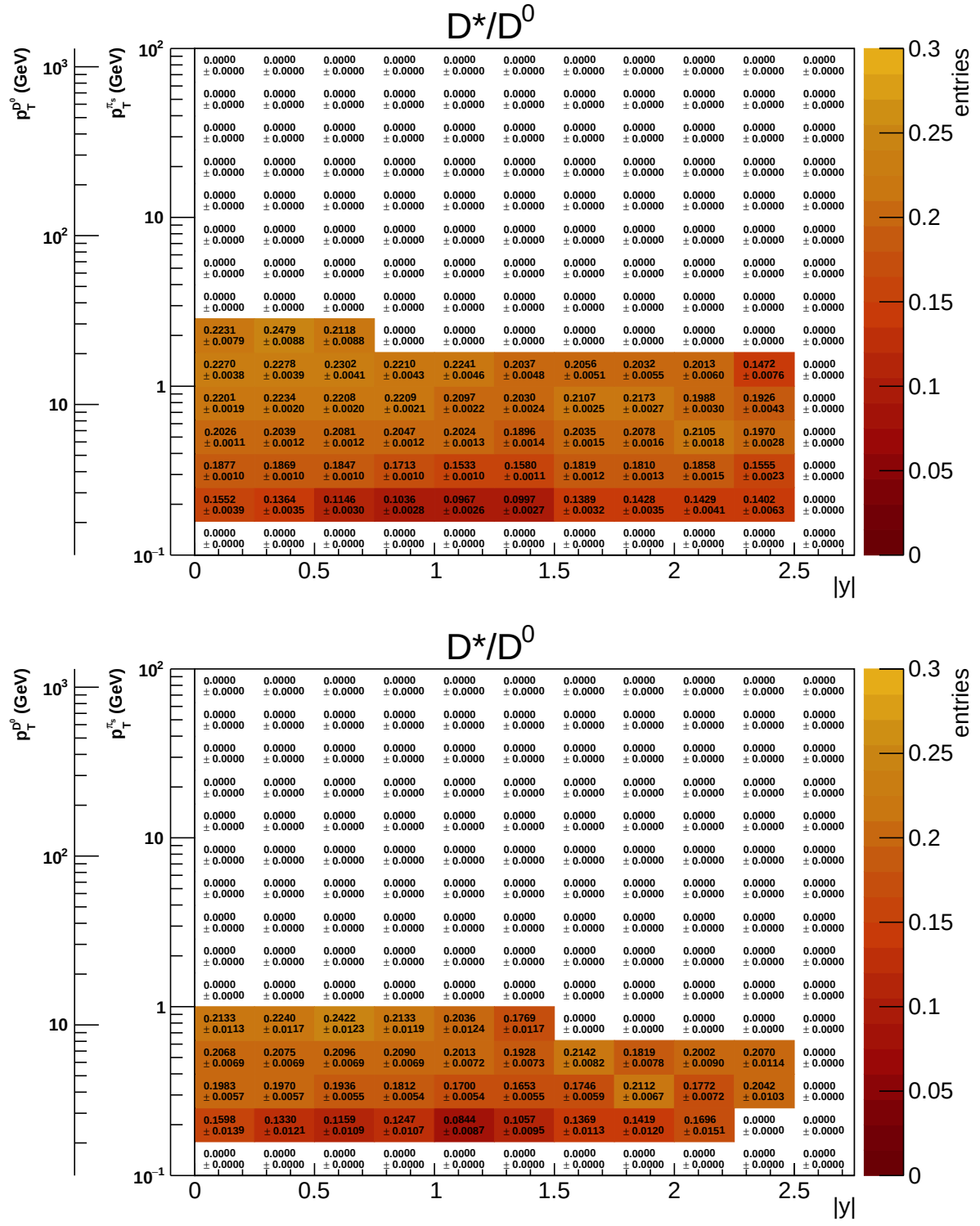


Figure F.3: Ratios of the number of the D^{*+} and D^0 events on the 13 TeV D^0 (top) and MinimumBias (bottom) MC sample.

G | χ^2 Tables for Data-Driven FONLL

In this Appendix, the χ^2s calculated by Eq.(6.22) for ddFONLL are collected. Figure G.1-G.36 show the 2-dimensional χ^2 tables for the μ_r and α_K variations. The least χ^2s of each table are summarized in Figure G.37-G.43.

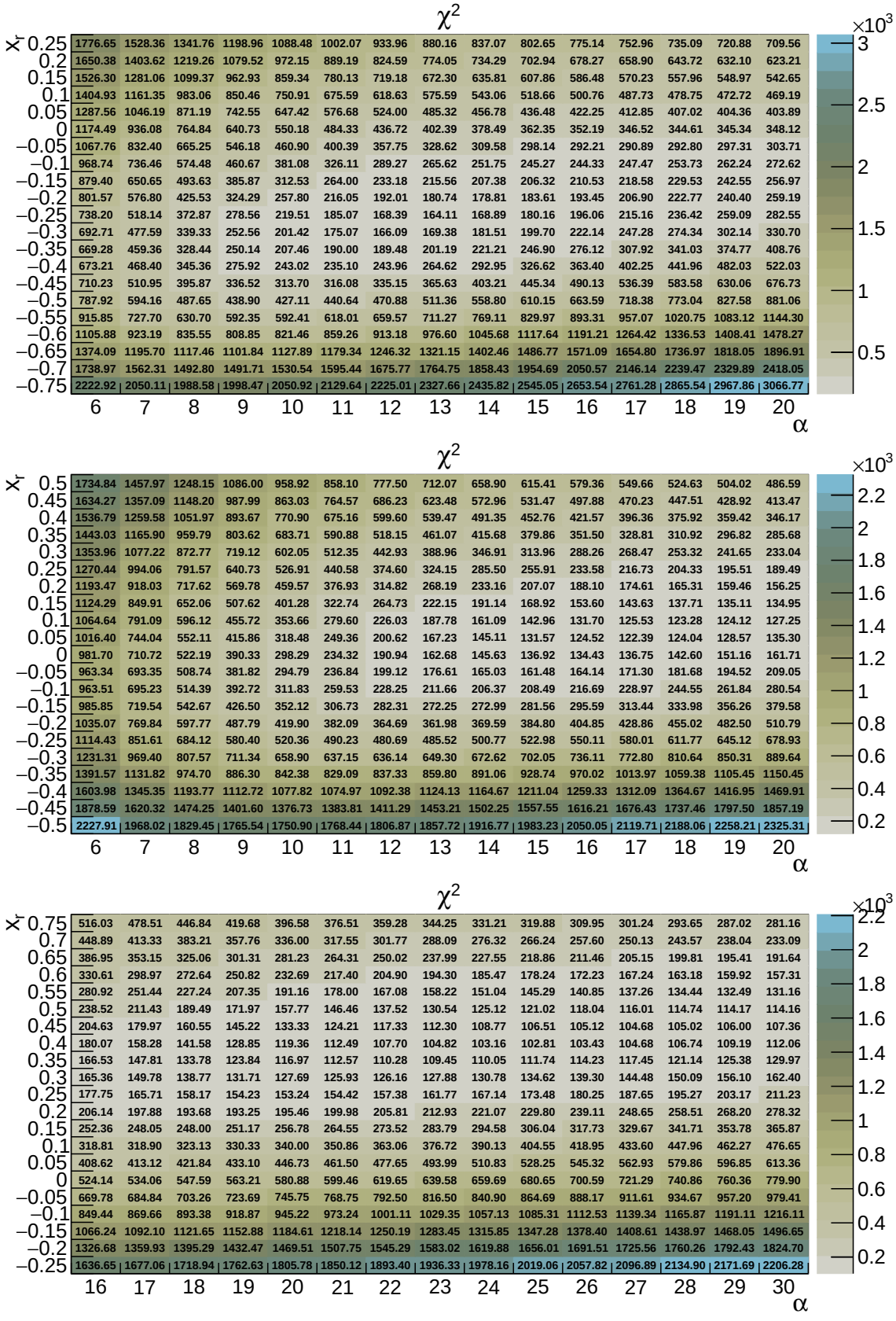


Figure G.1: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.3$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

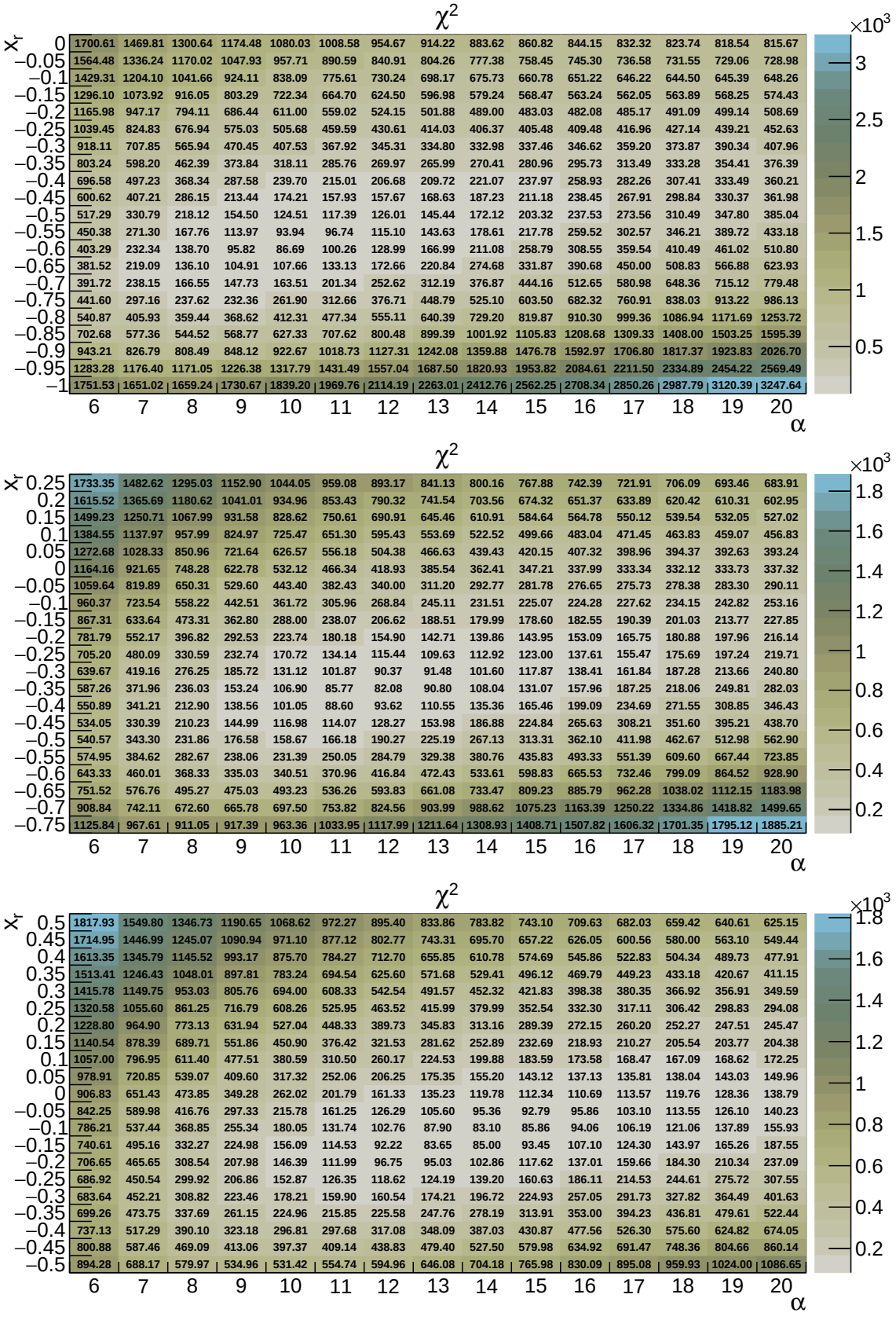


Figure G.2: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.5$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

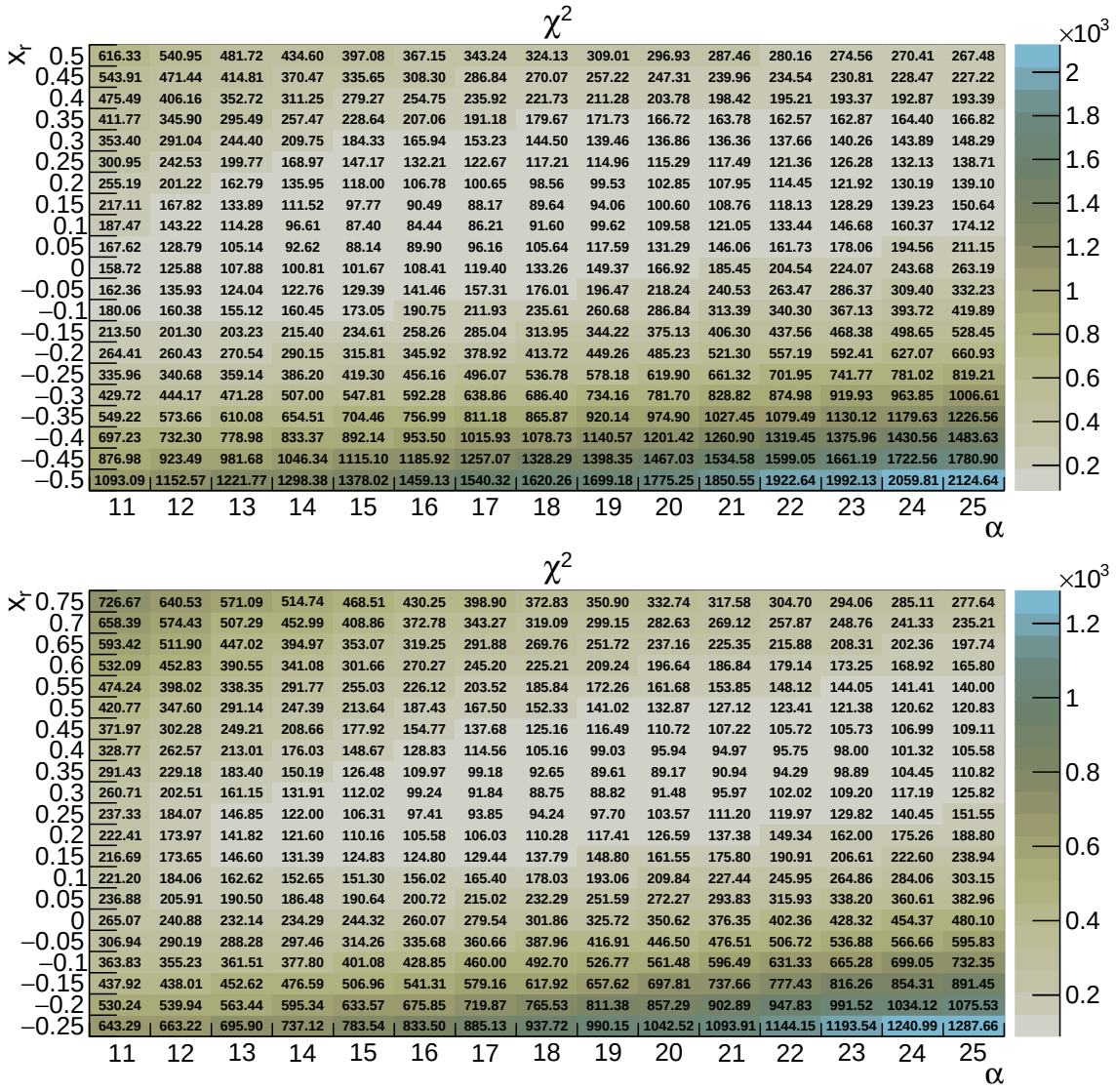


Figure G.3: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.5$ GeV and x_f =0.75 (top), 1 (bottom).

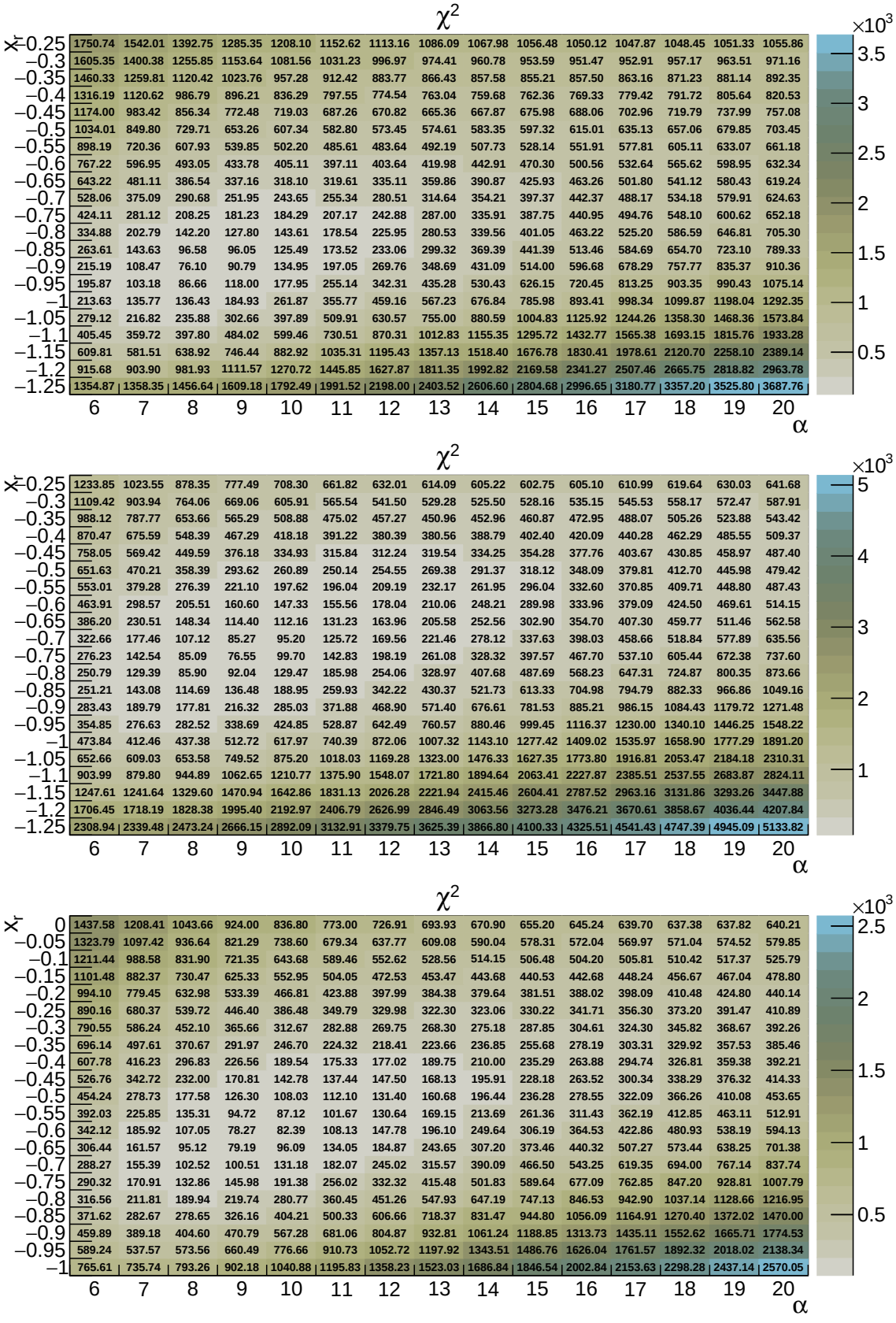


Figure G.4: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.7$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

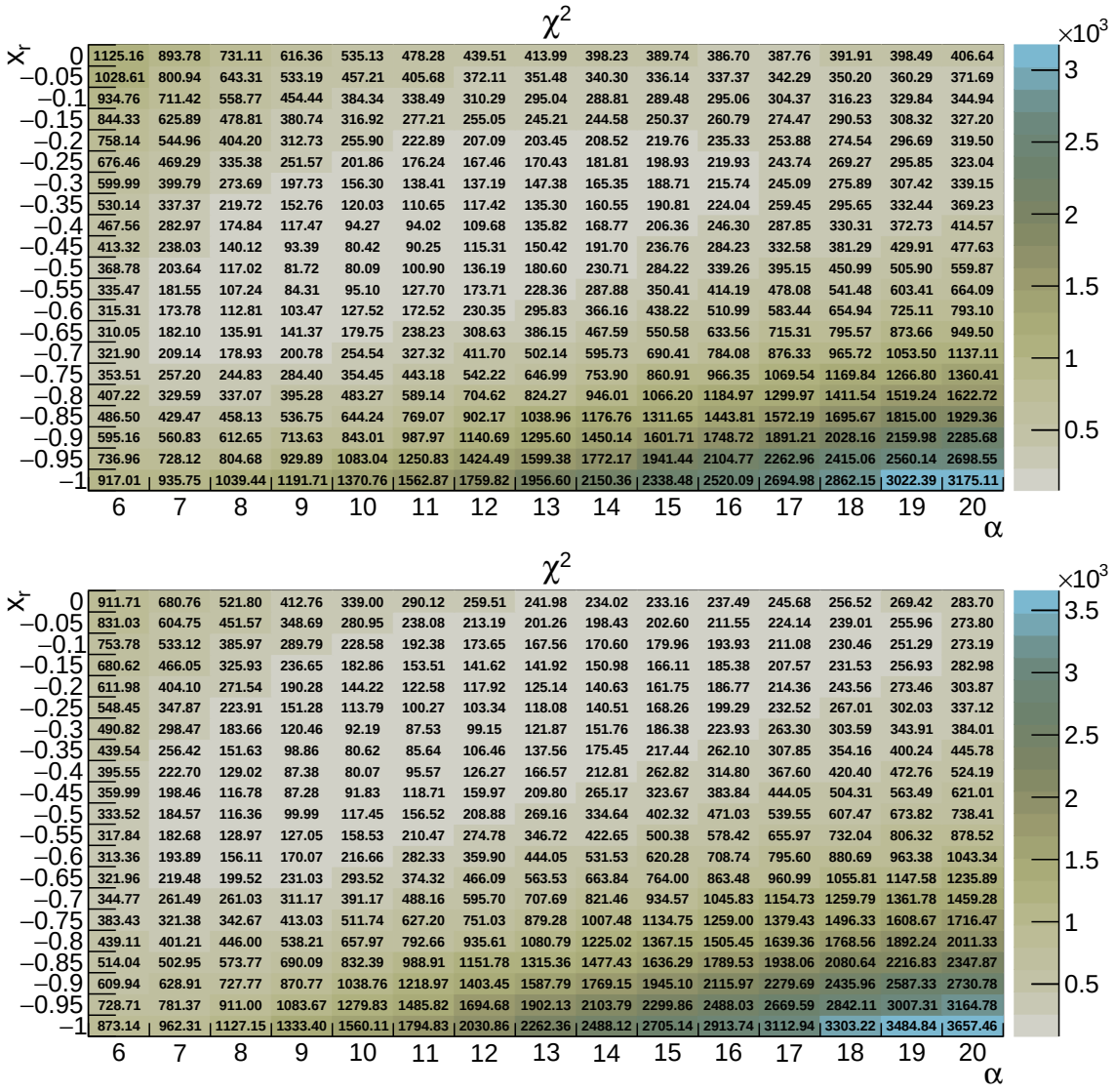


Figure G.5: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.7$ GeV and x_f =0.75 (top), 1 (bottom).

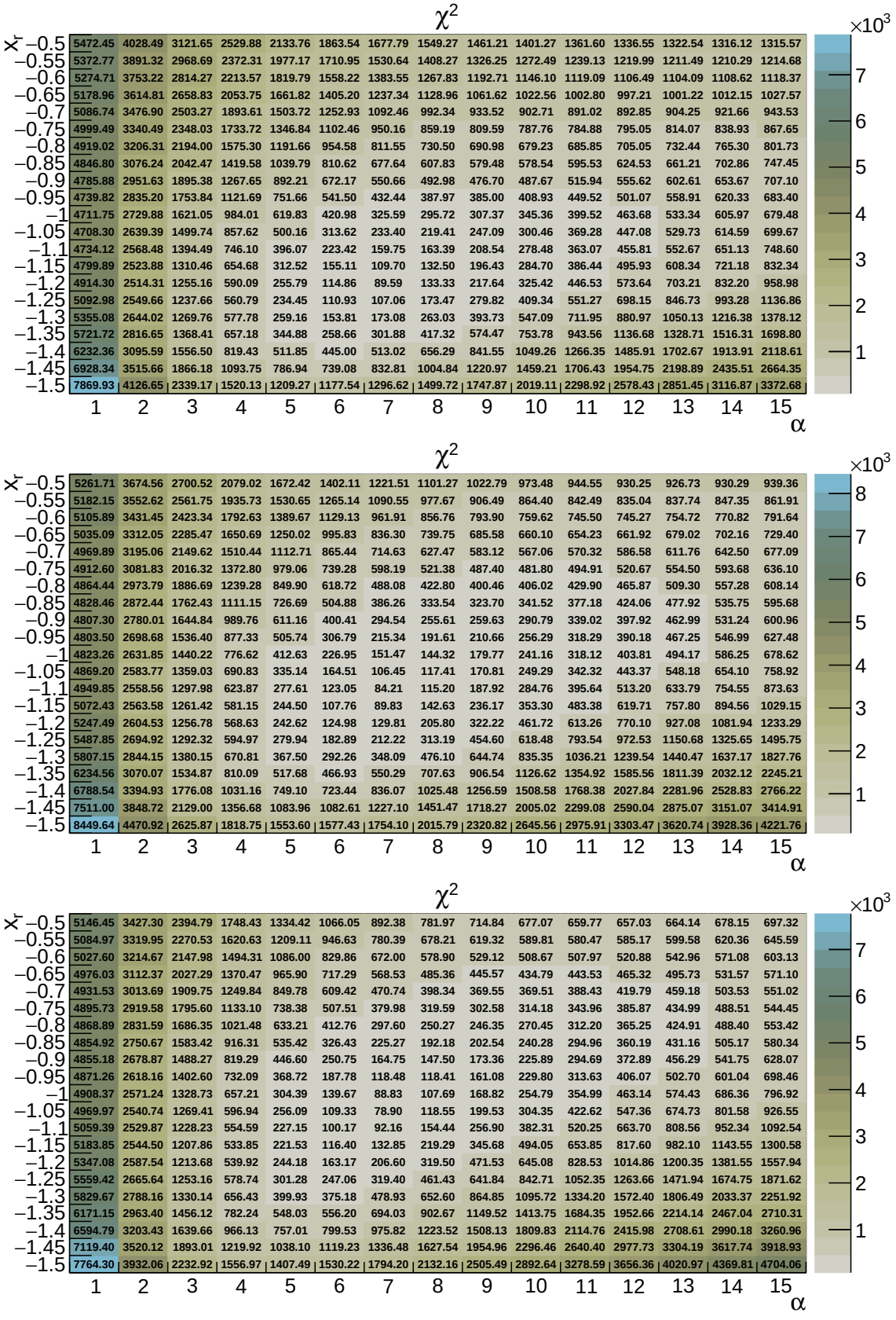


Figure G.6: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.9$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

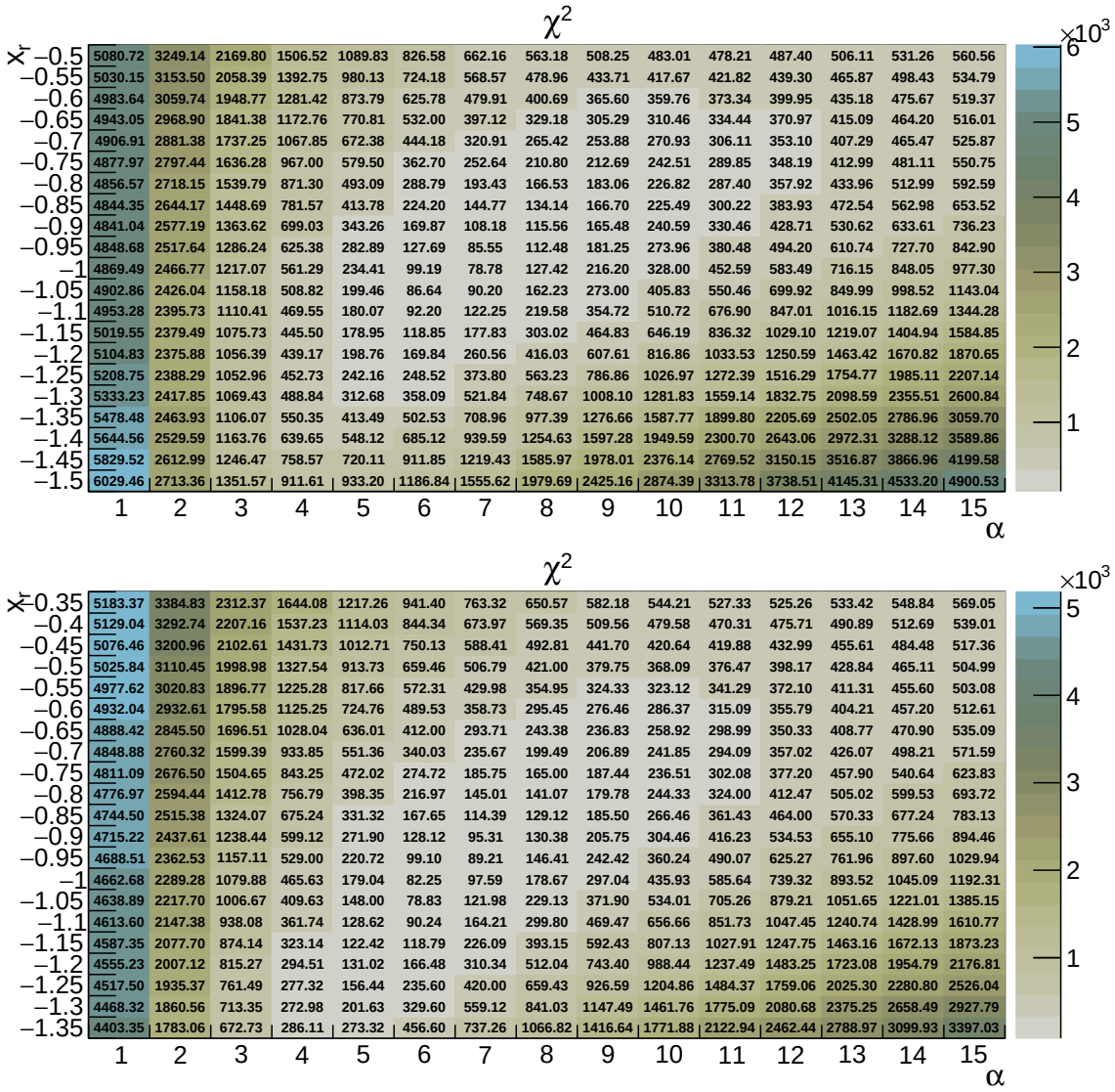


Figure G.7: χ^2 tables at $\sqrt{s} = 5$ TeV with the universality assumption. $m_c = 1.9$ GeV and x_f =0.75 (top), 1 (bottom).

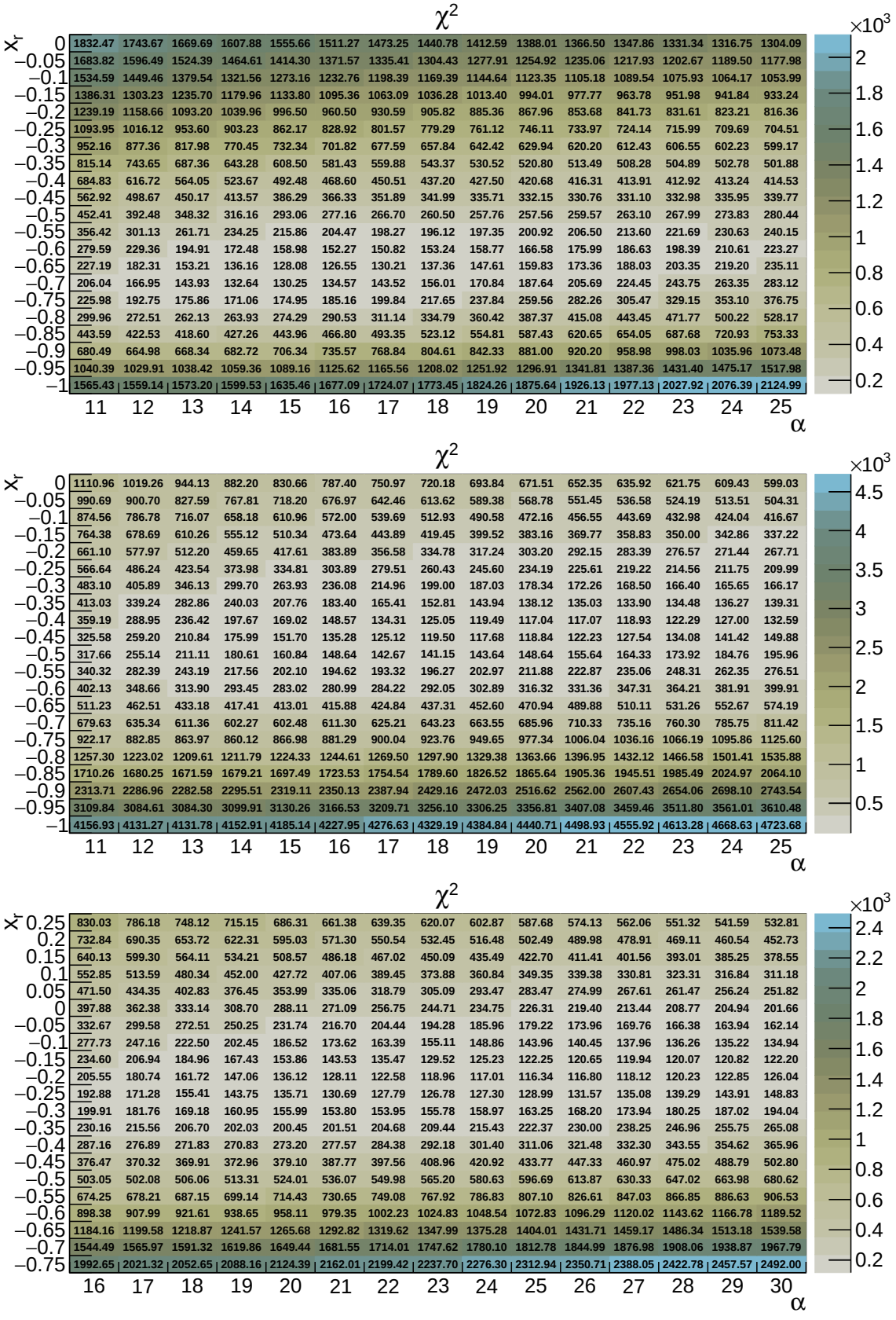


Figure G.8: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.3$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

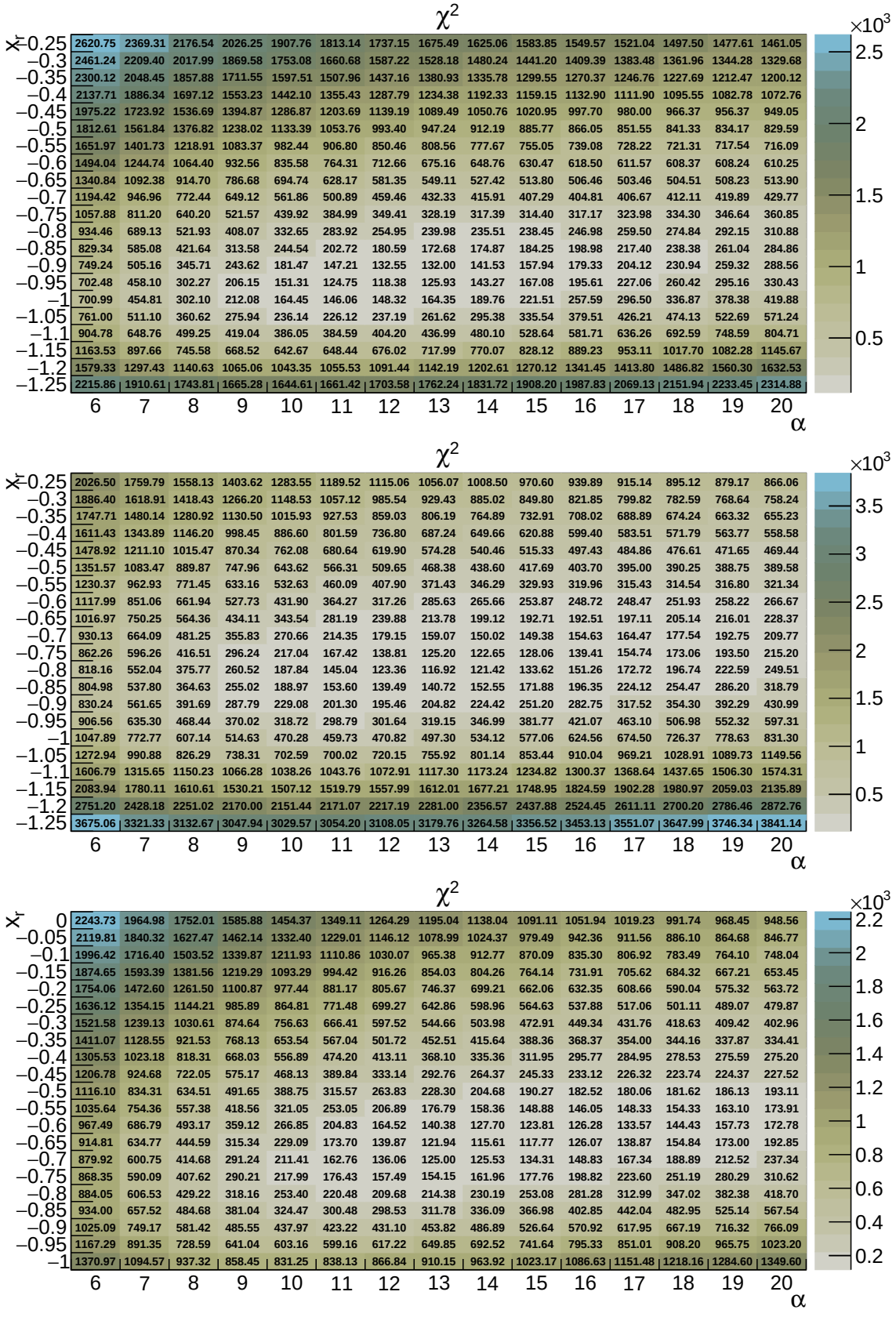


Figure G.9: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

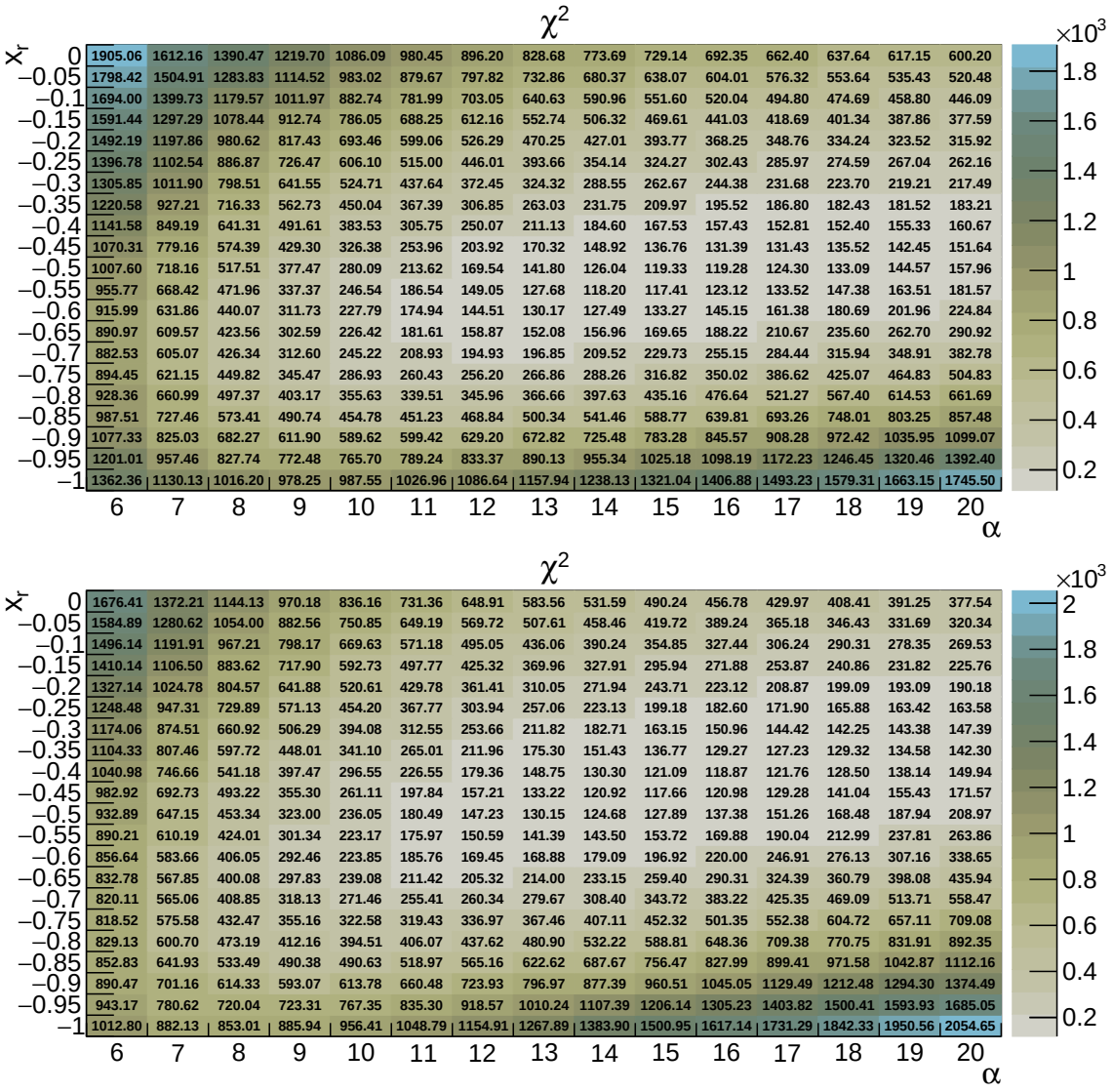


Figure G.10: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0.75 (top) and 1 (bottom).

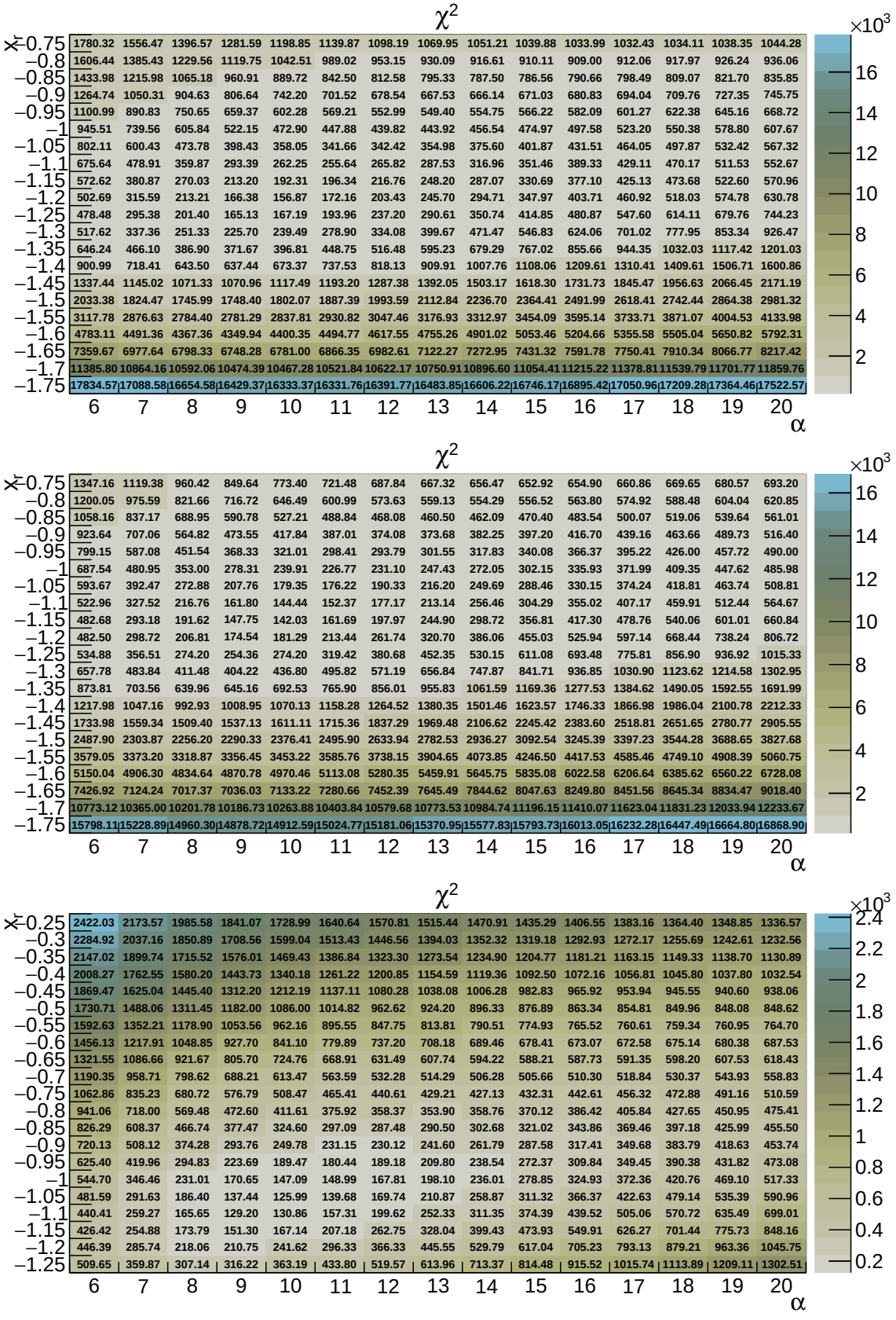


Figure G.11: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

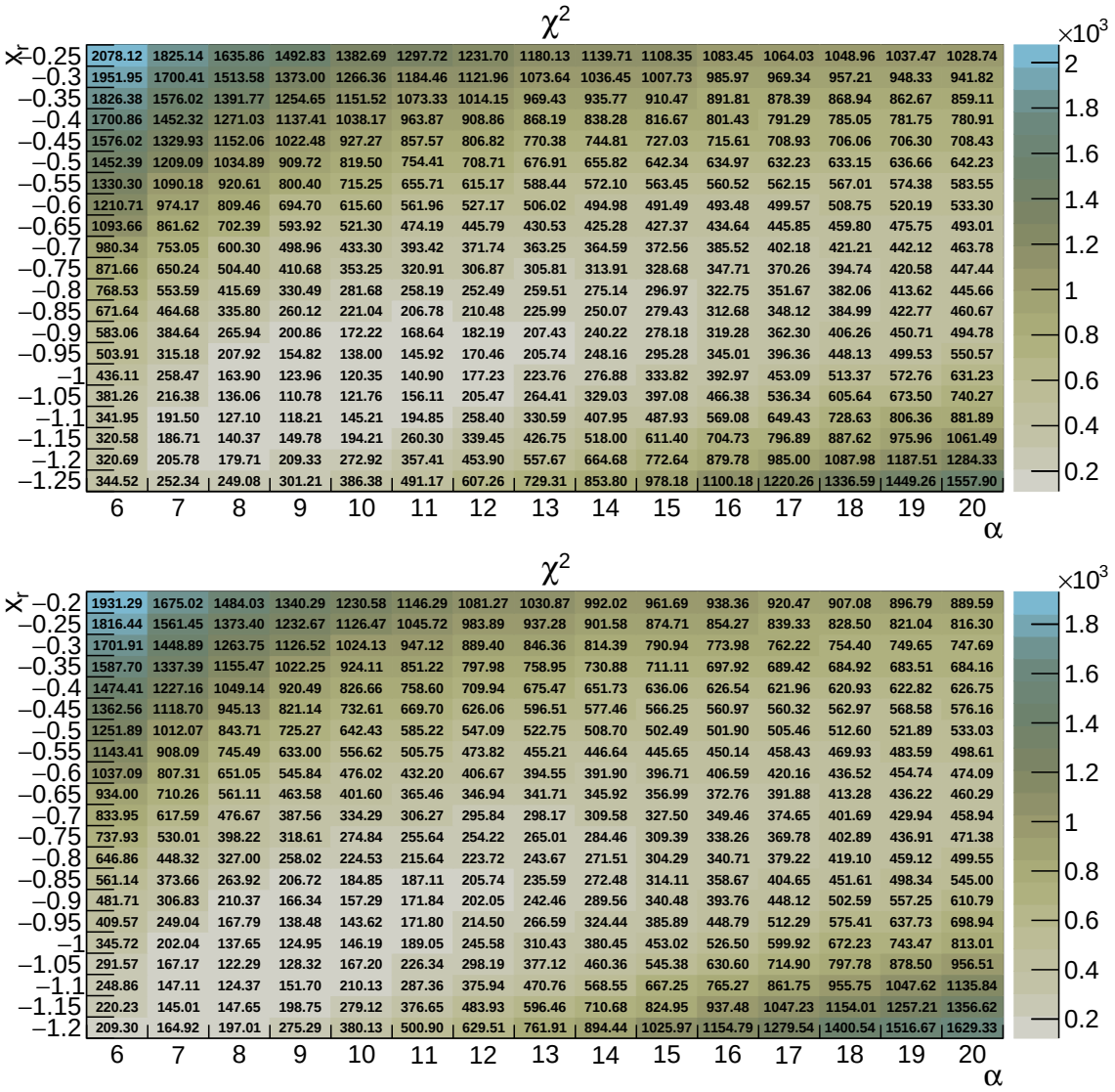


Figure G.12: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0.75 (top) and 1 (bottom).

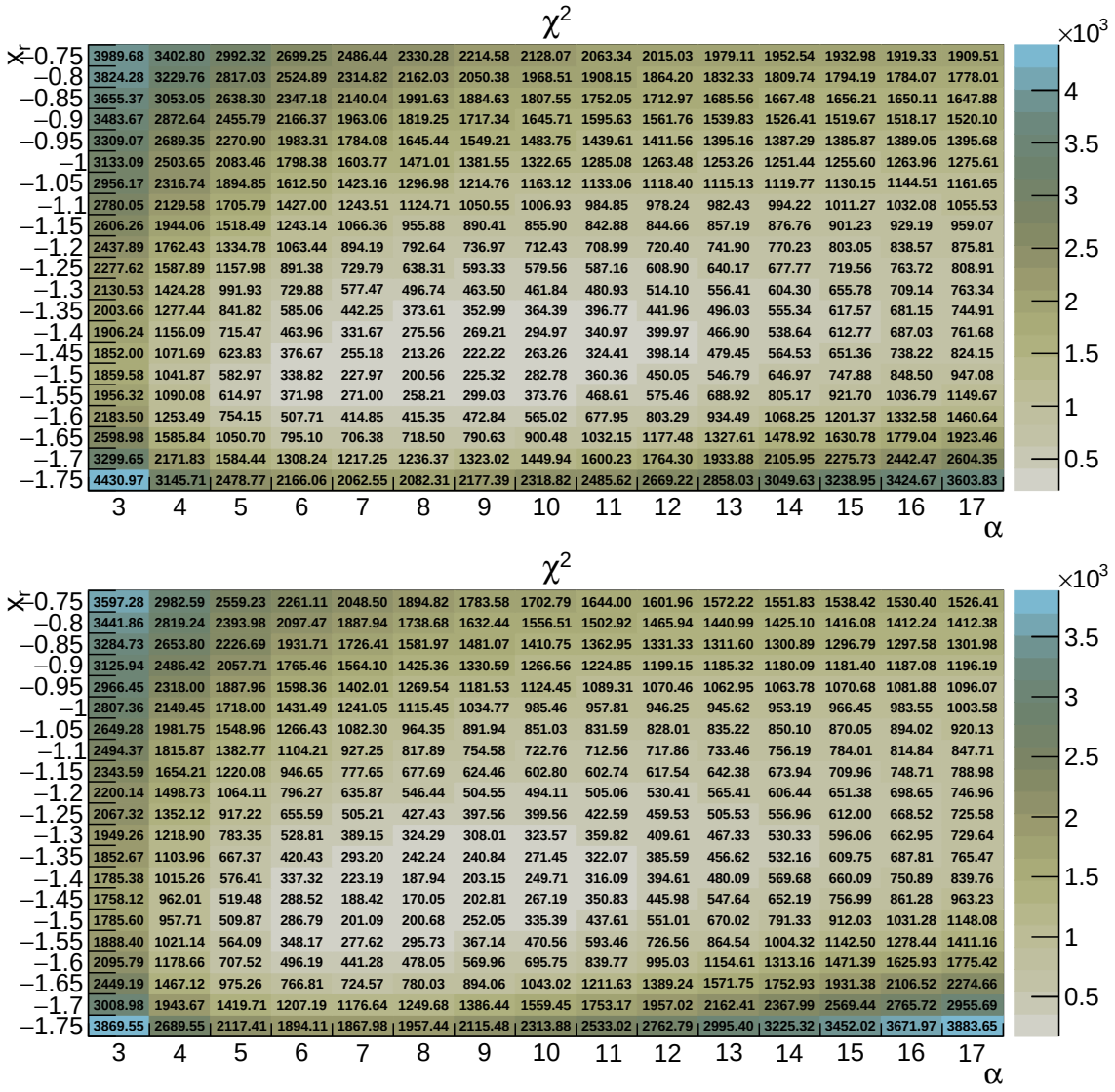


Figure G.13: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0 (top) and 0.25 (bottom).

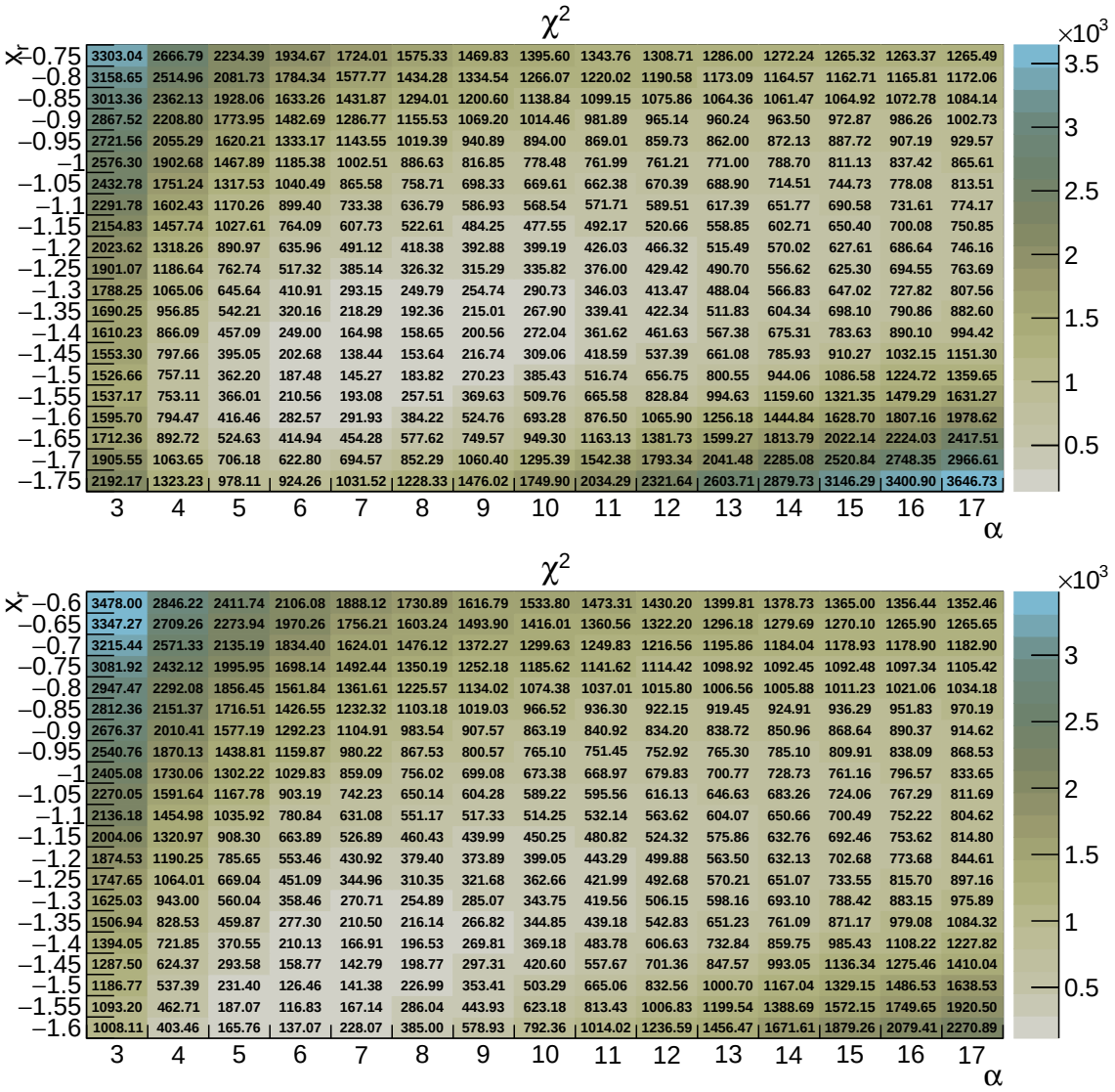


Figure G.14: χ^2 tables at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0.5 (top) and 0.75 (bottom).

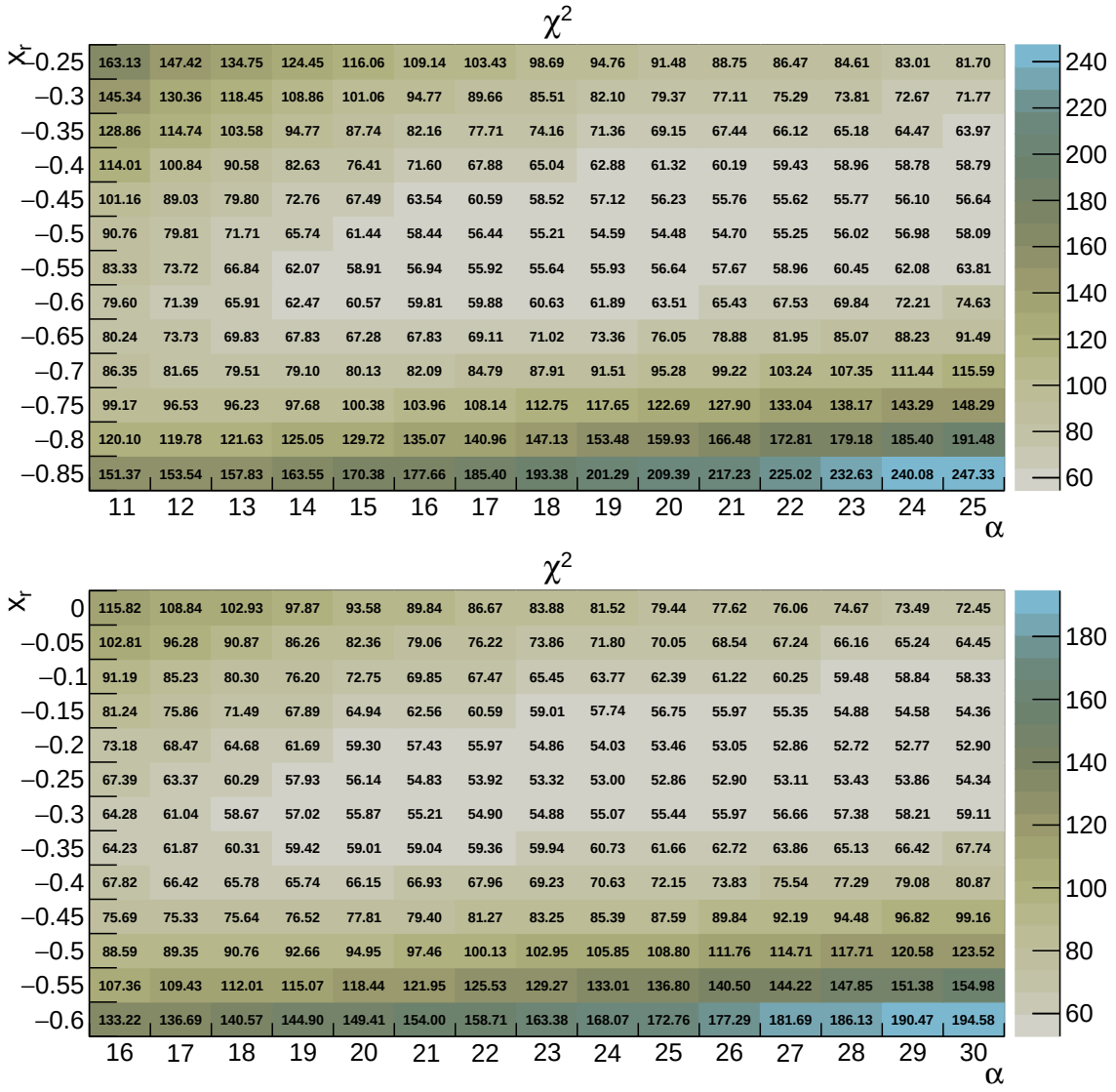


Figure G.15: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.3$ GeV and x_f =0 (top) and 0.25 (bottom).

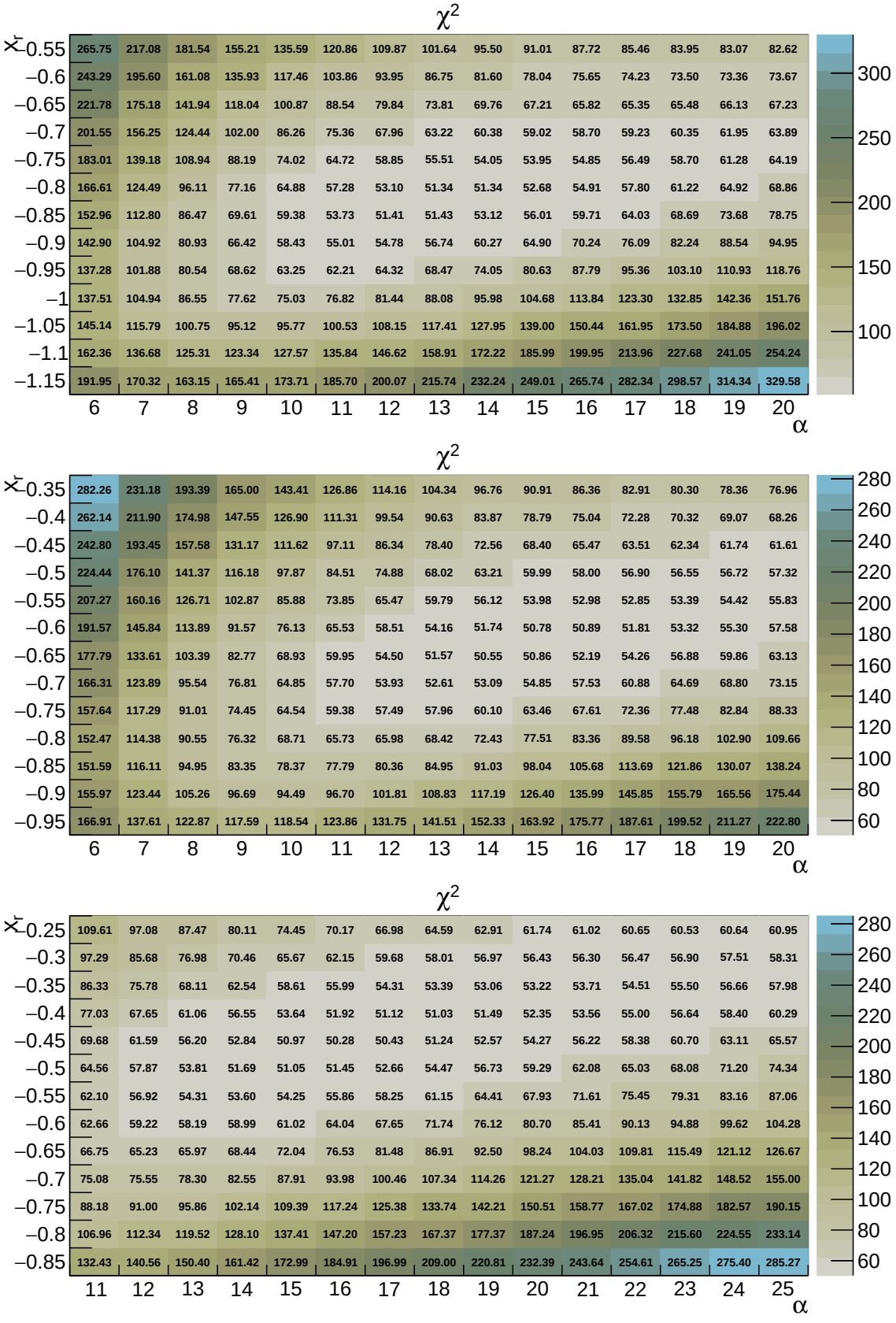


Figure G.16: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

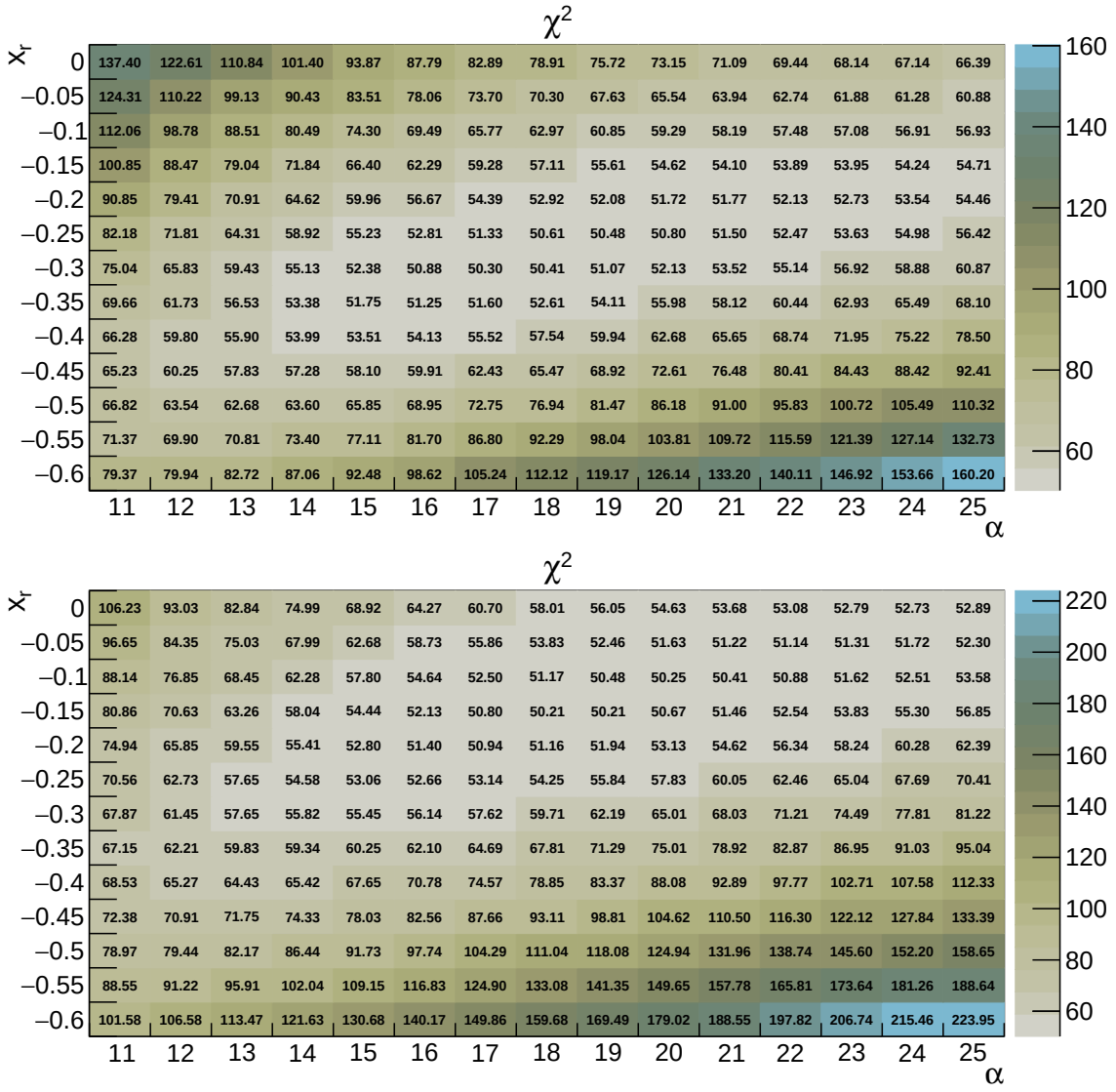


Figure G.17: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0.75 (top) and 1 (bottom).

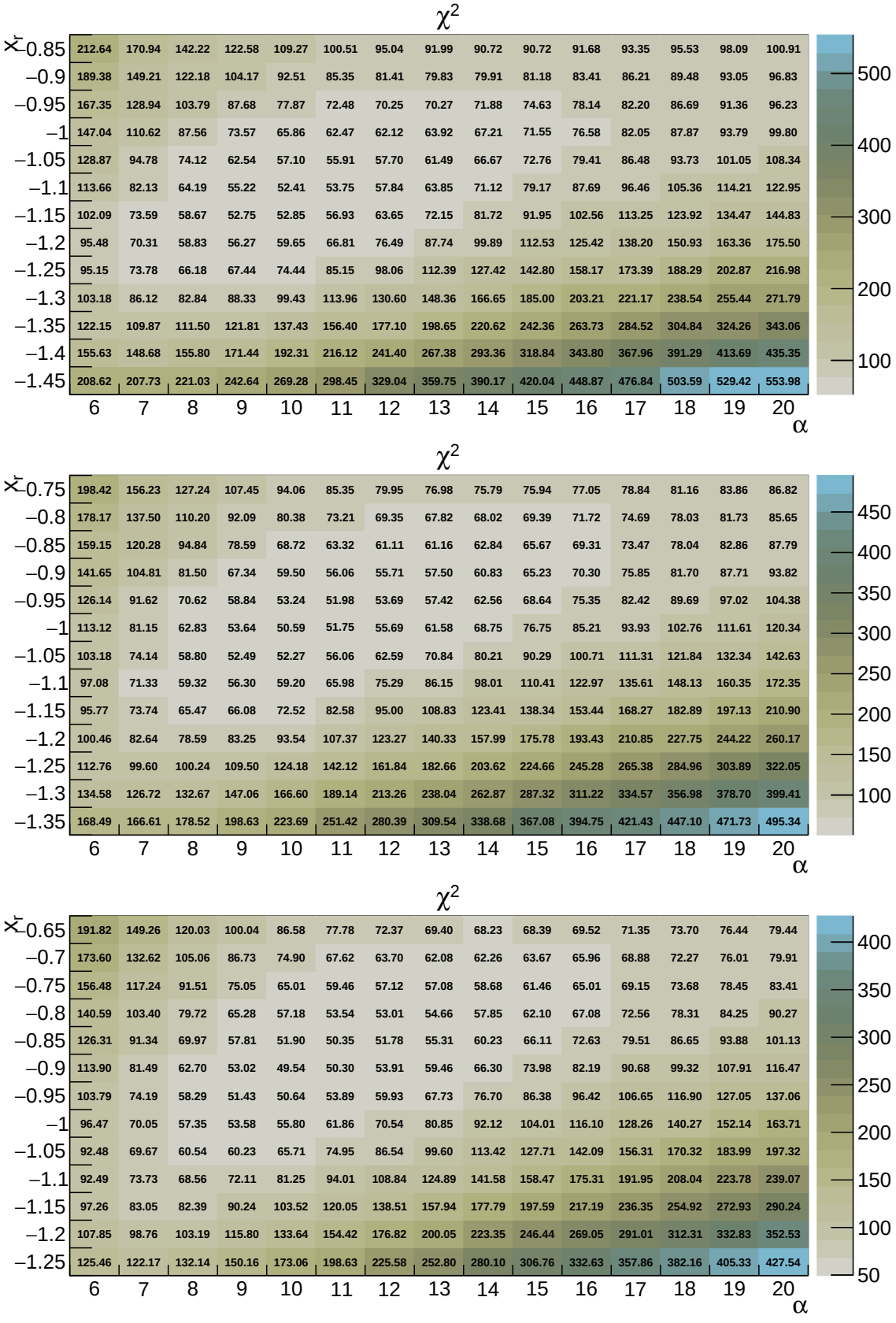


Figure G.18: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

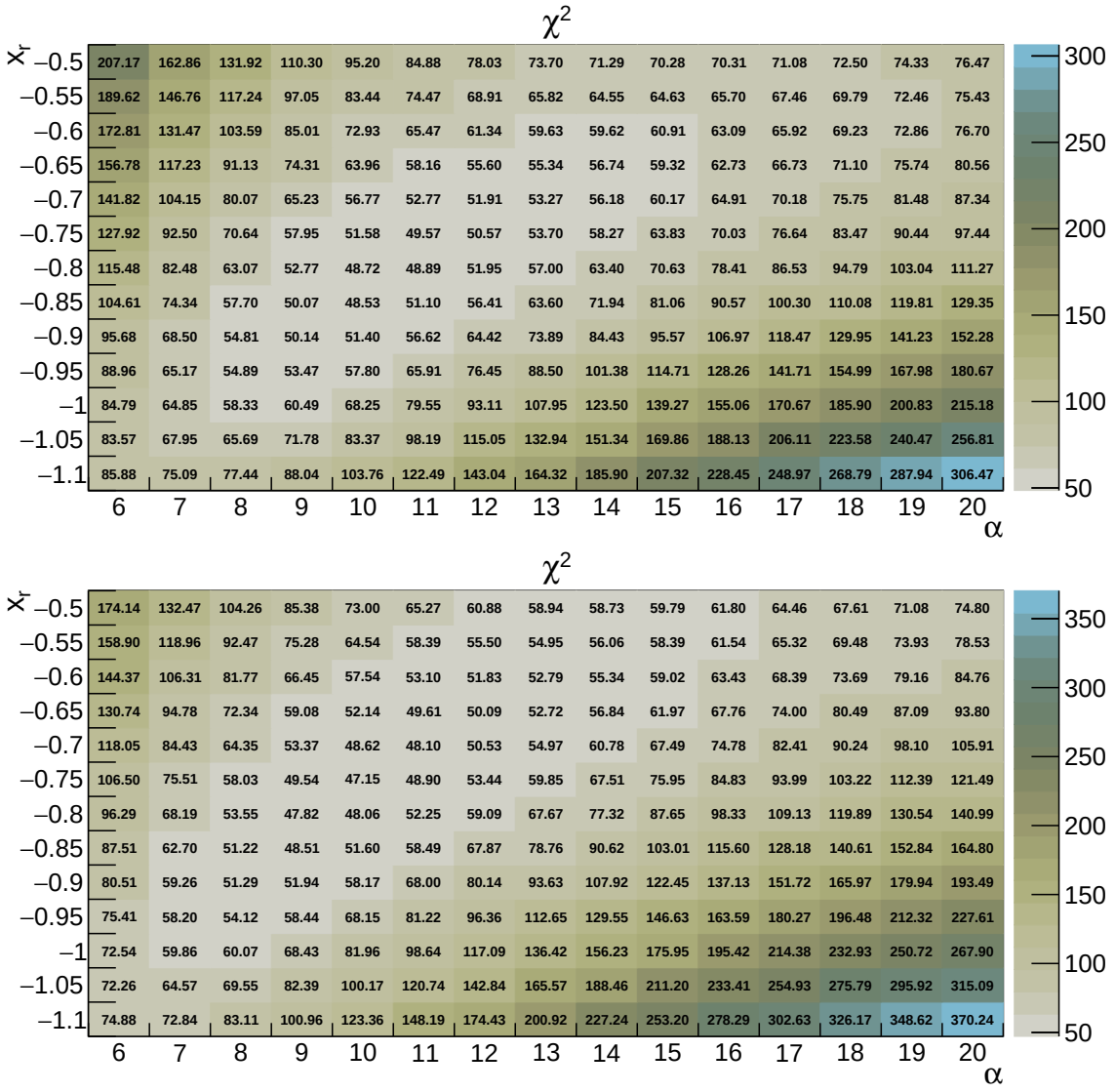


Figure G.19: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0.75 (top) and 1 (bottom).

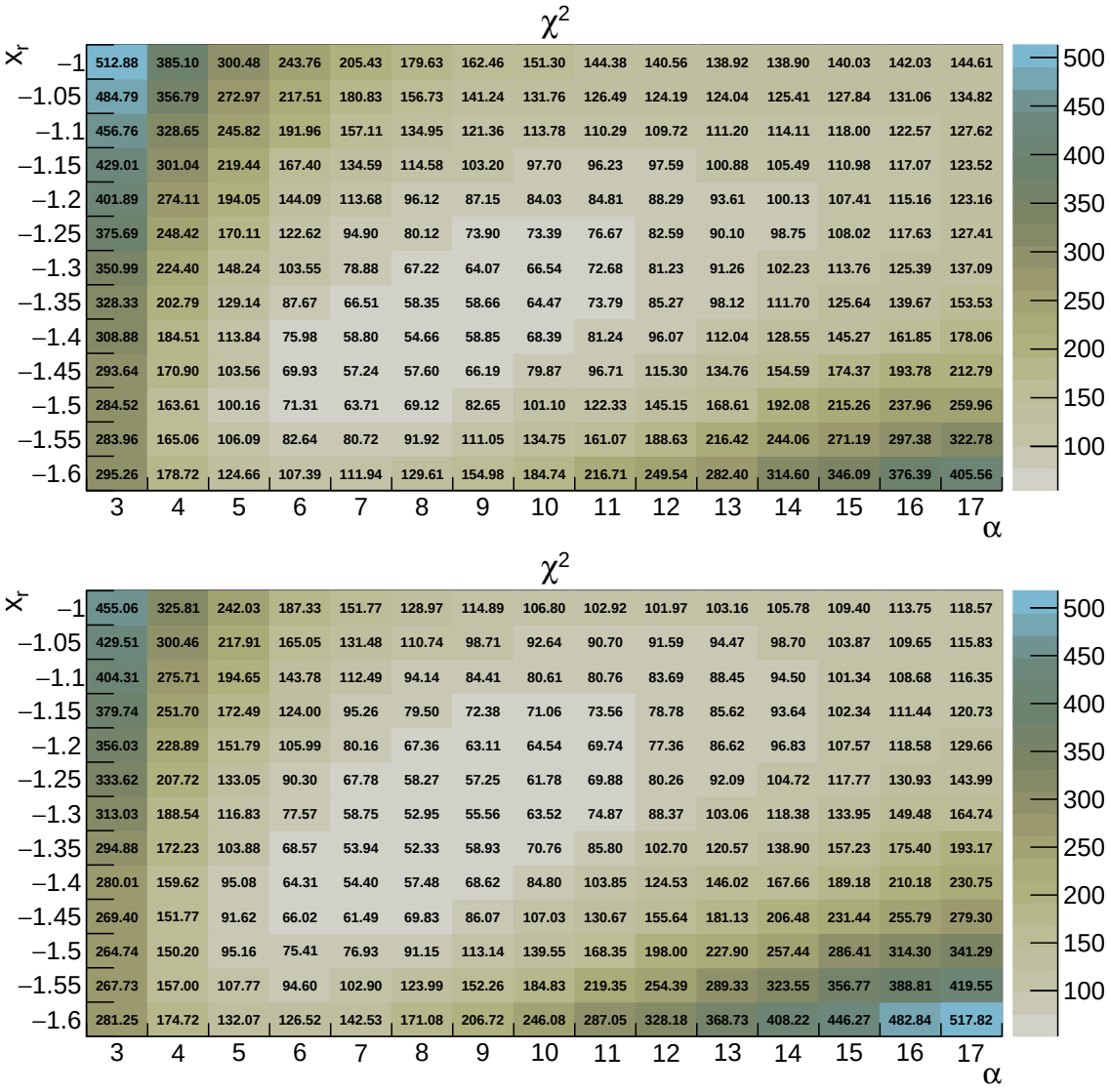


Figure G.20: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0 (top) and 0.25 (bottom).

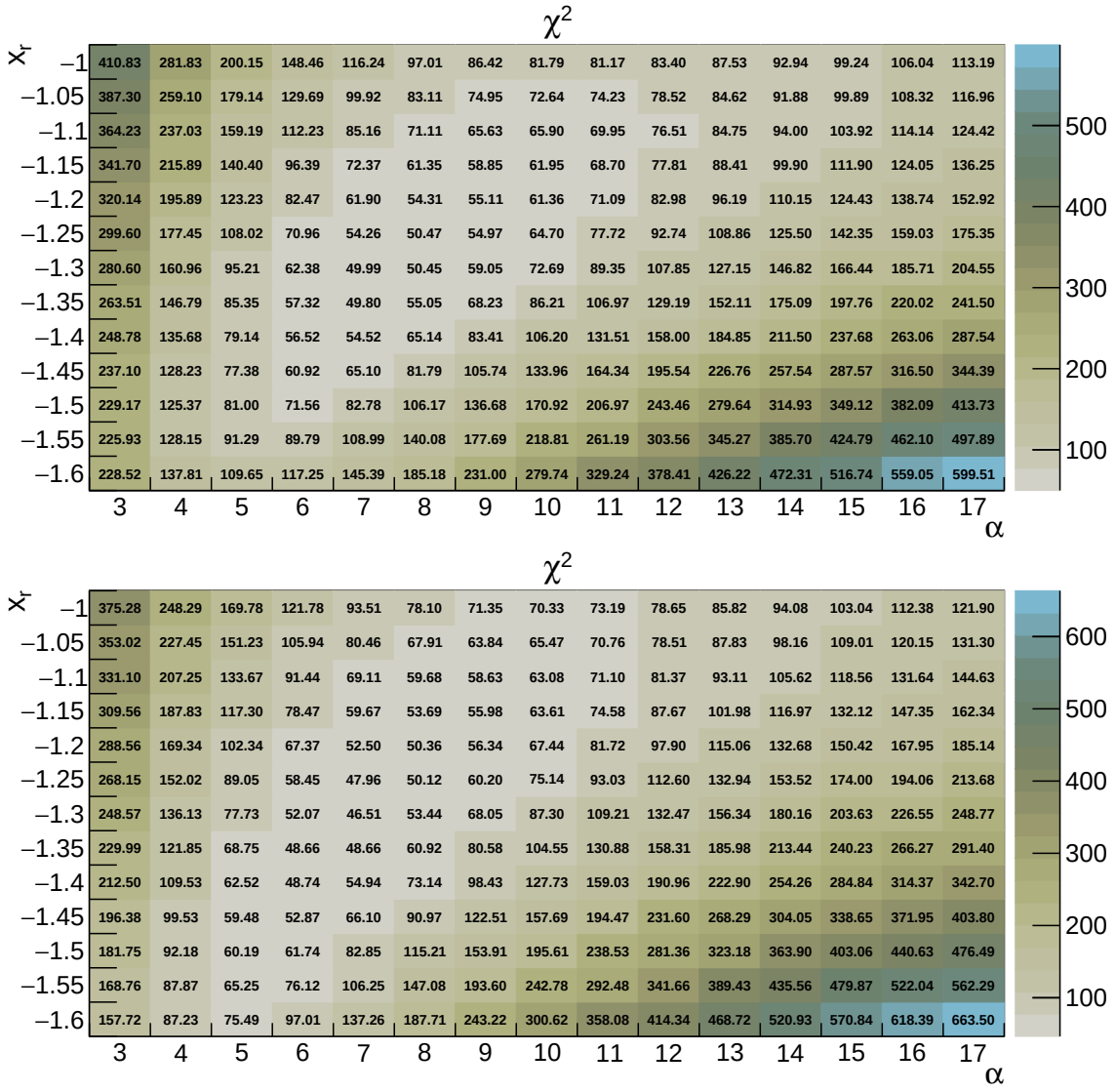


Figure G.21: χ^2 tables at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0.5 (top) and 0.75 (bottom).

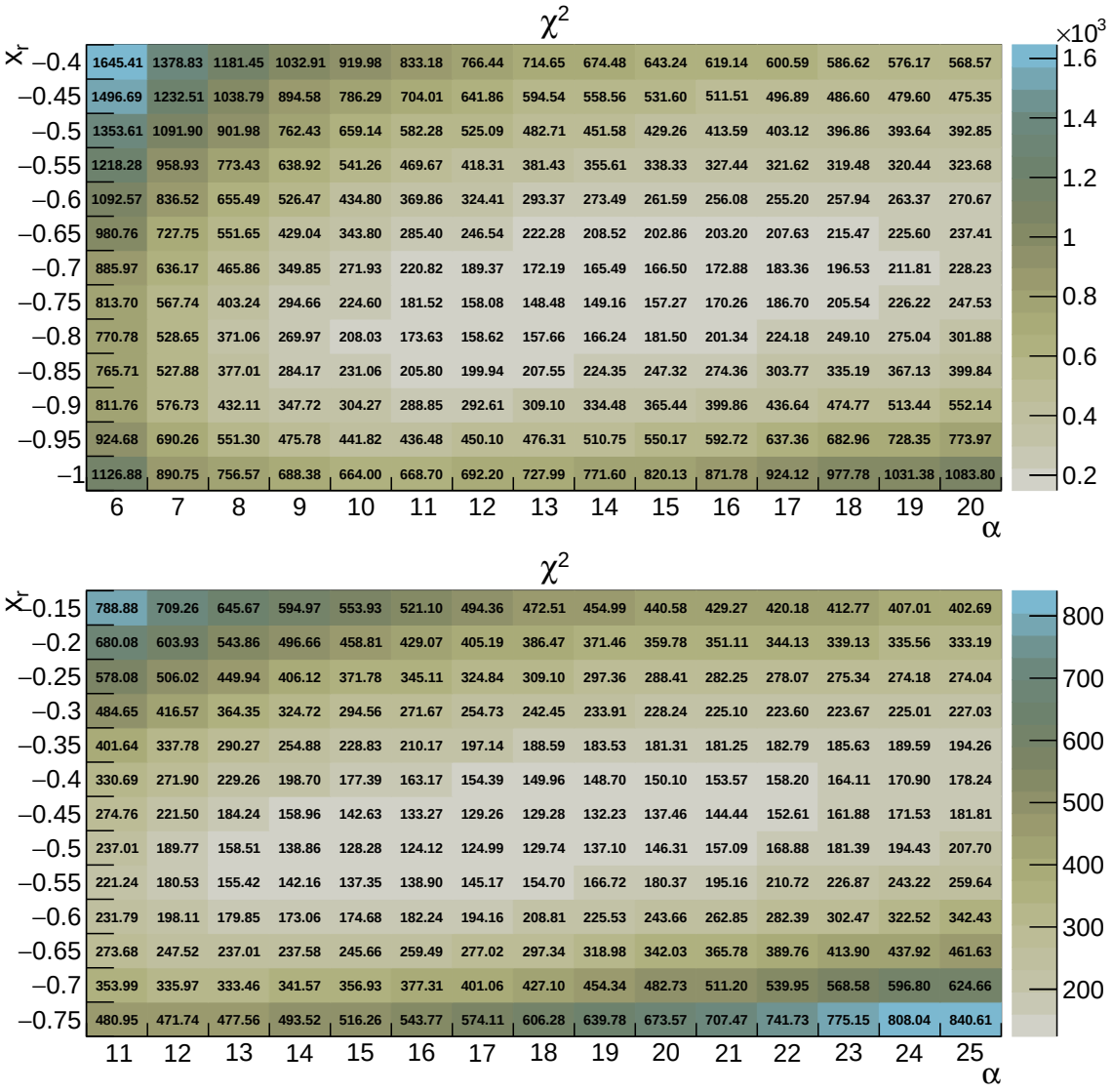


Figure G.22: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.3$ GeV and x_f =0 (top) and 0.25 (bottom).

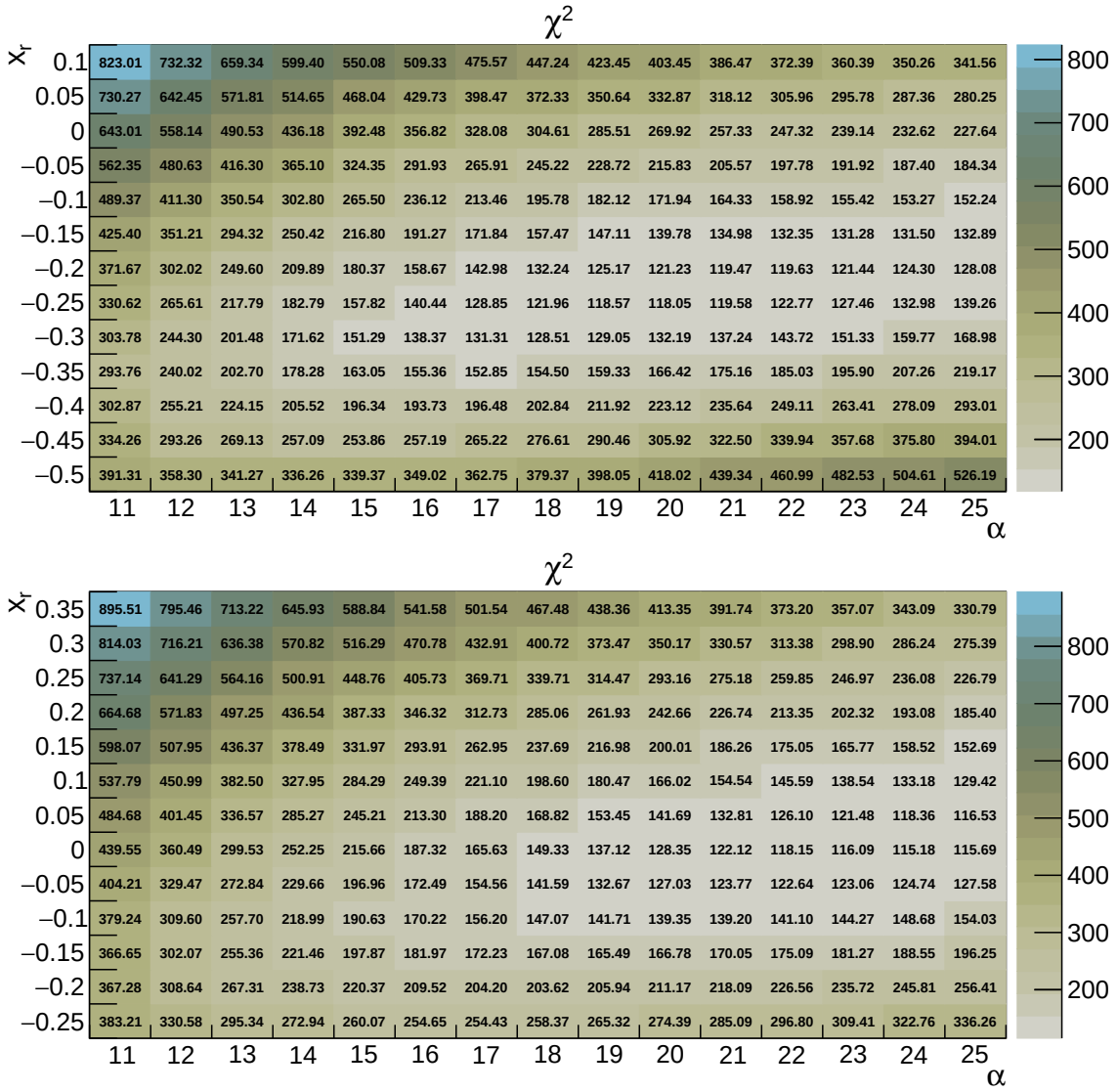


Figure G.23: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.3$ GeV and x_f =0.5 (top), 0.75 (bottom).

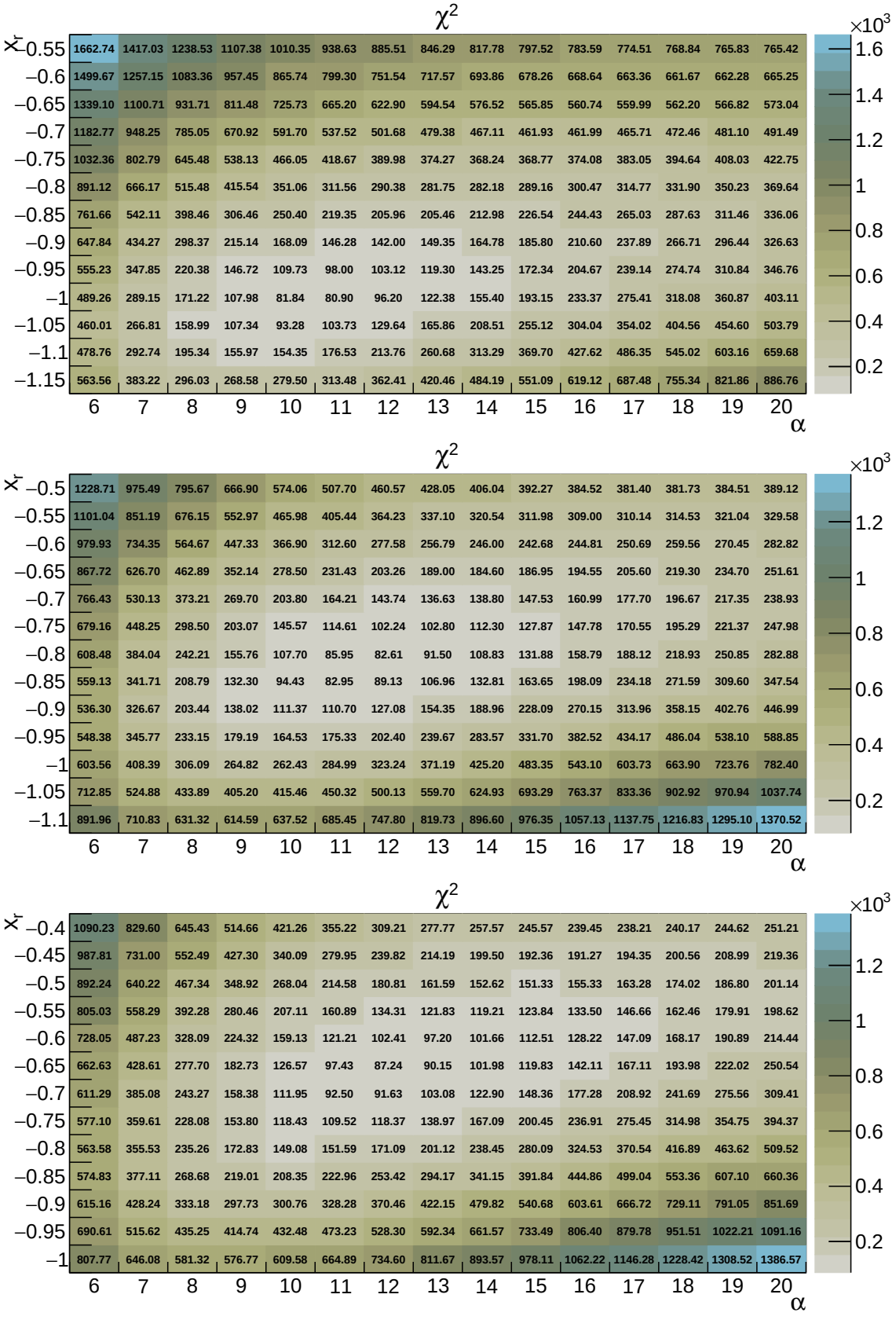


Figure G.24: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

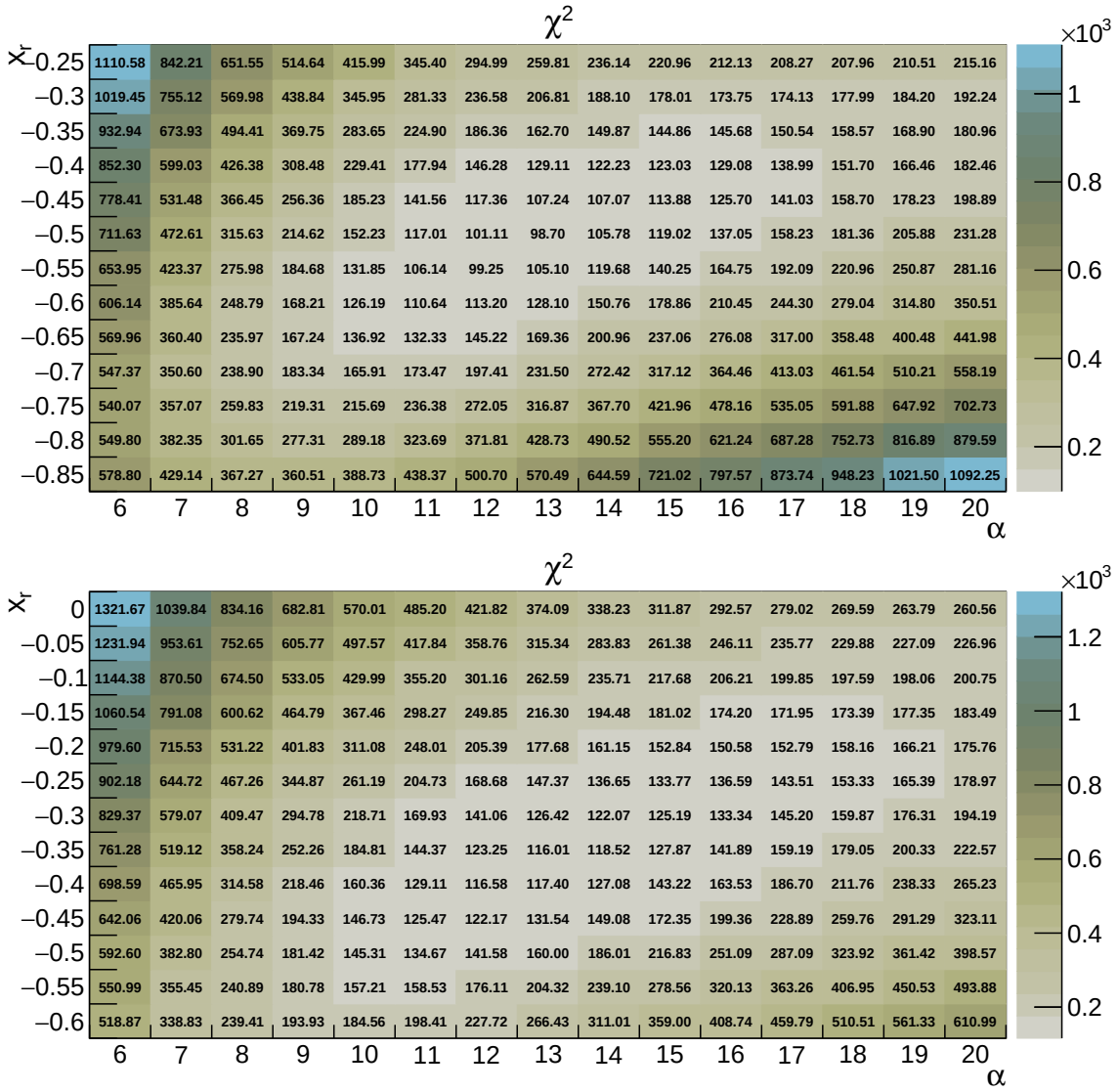


Figure G.25: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0.75 (top) and 1 (bottom).

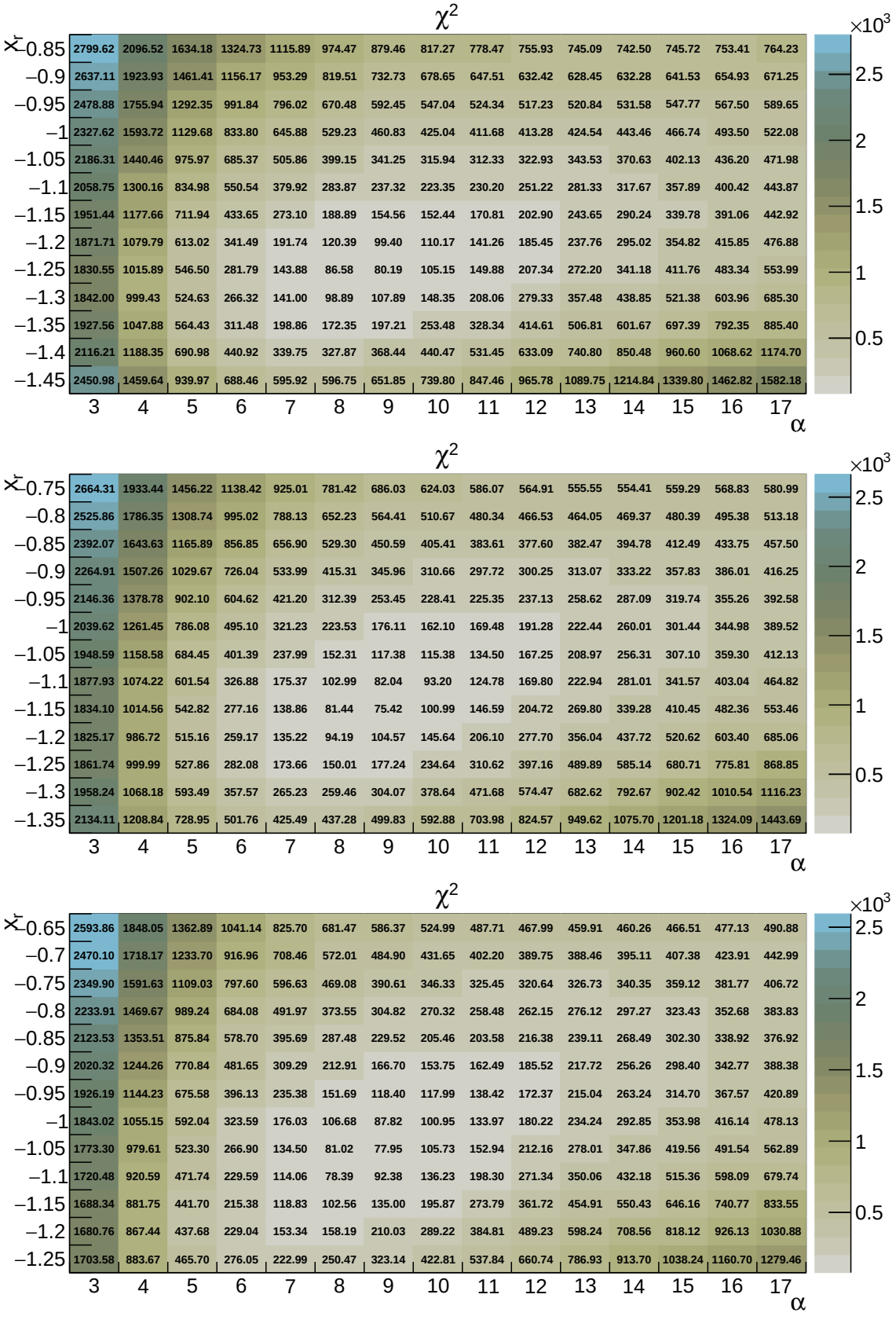


Figure G.26: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

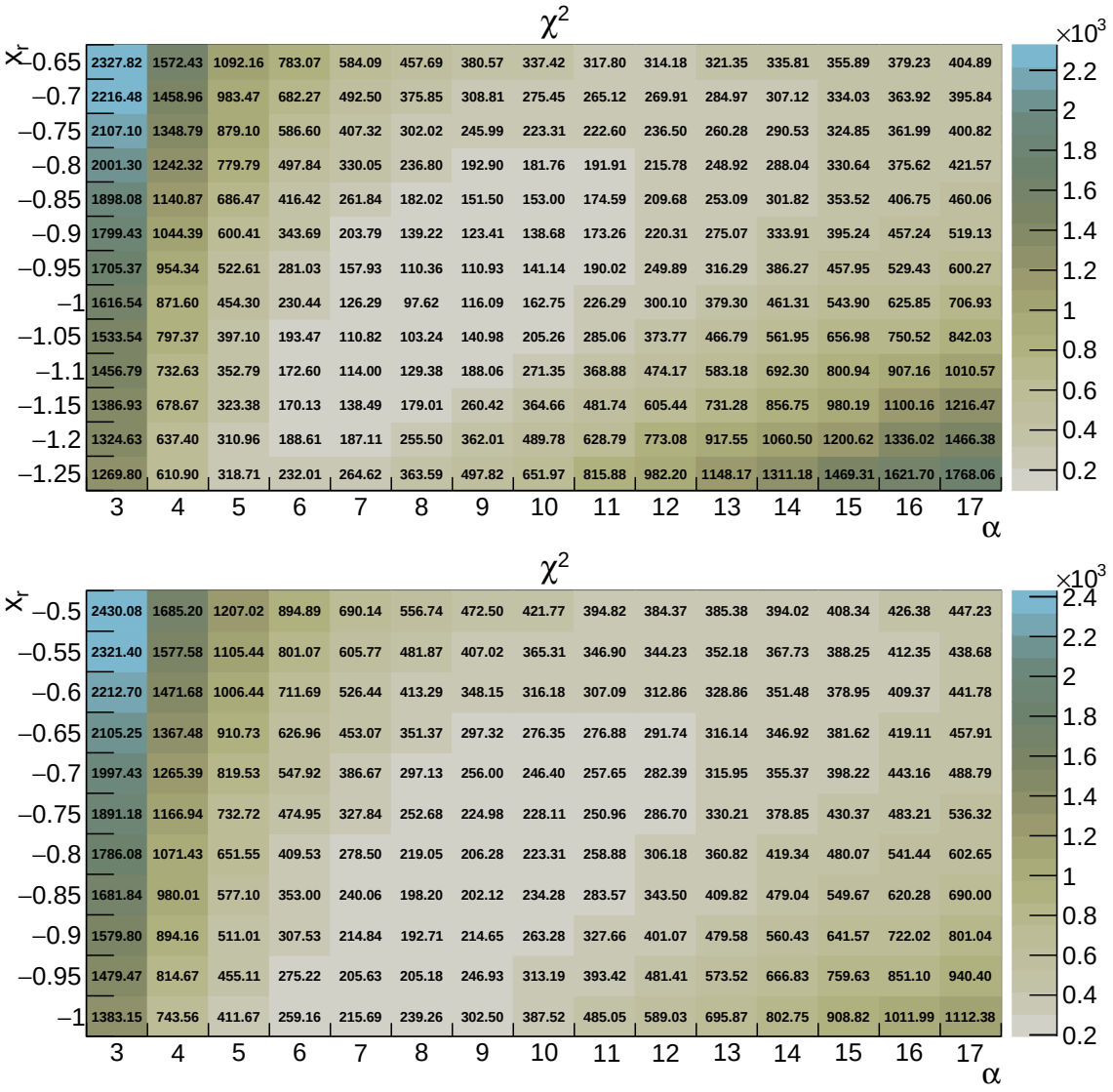


Figure G.27: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0.75 (top) and 1 (bottom).

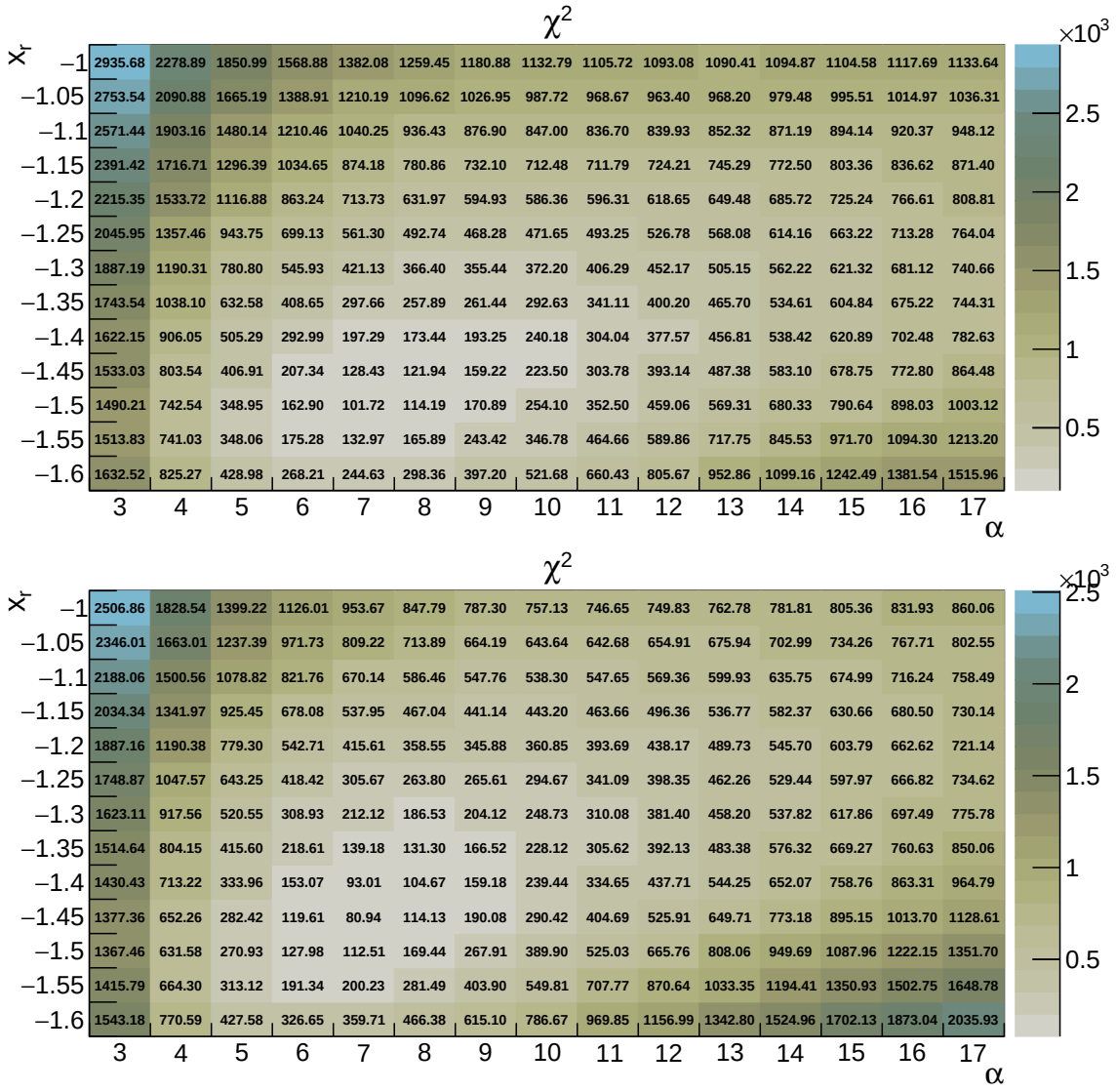


Figure G.28: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0 (top) and 0.25 (bottom).

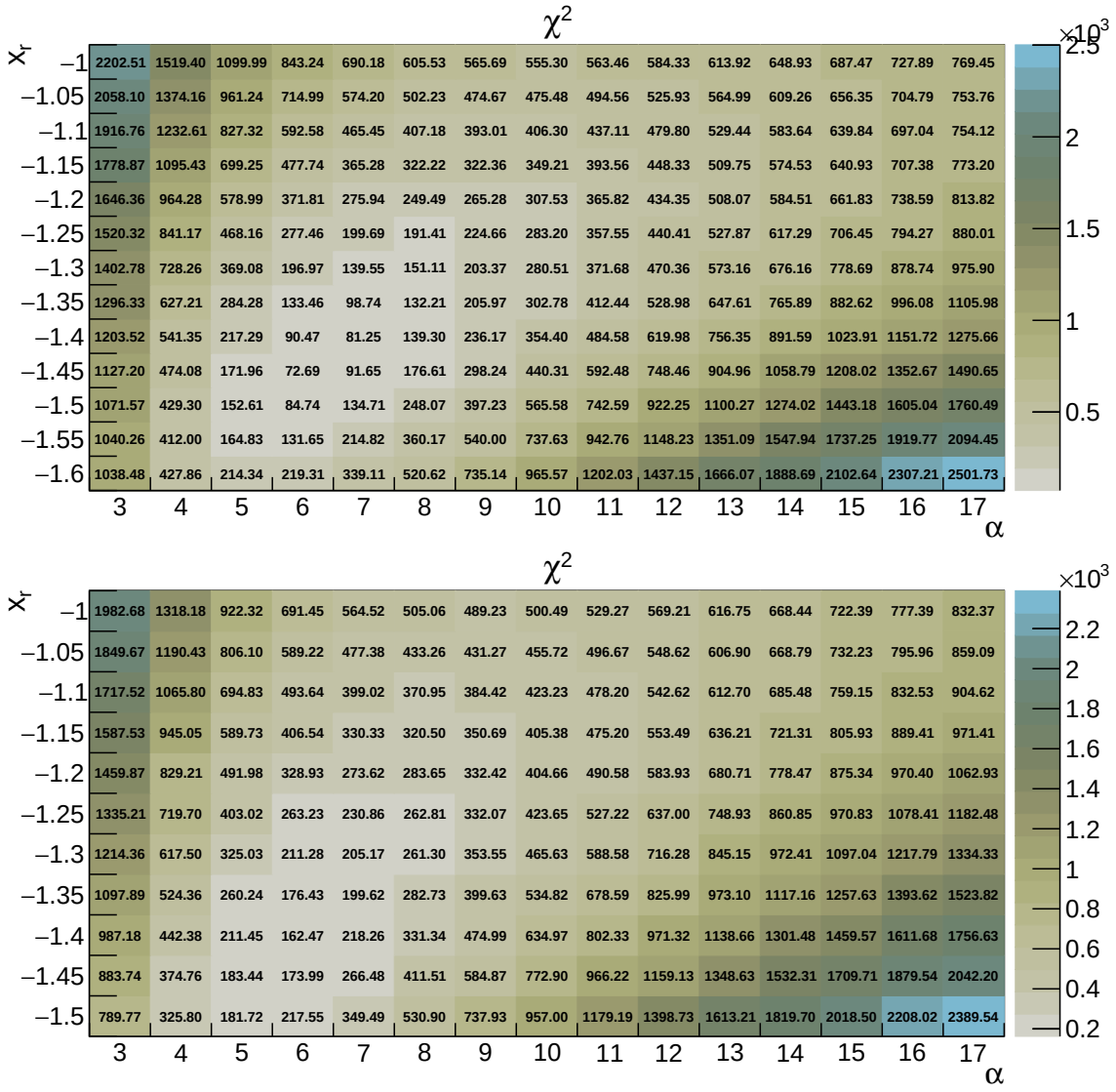


Figure G.29: χ^2 tables at $\sqrt{s} = 13$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0.5 (top) and 0.75 (bottom).

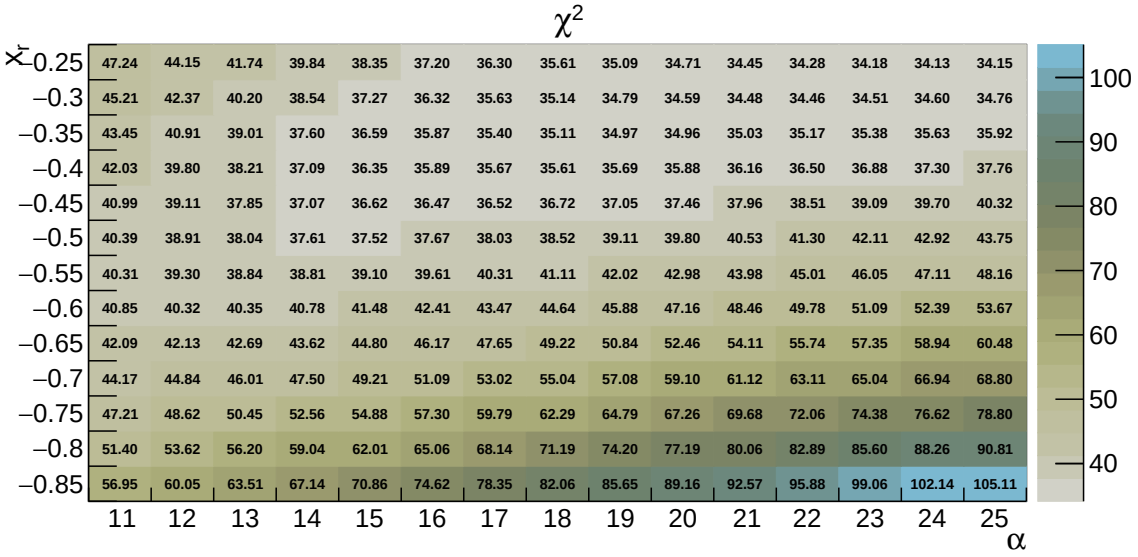


Figure G.30: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.3$ GeV and x_f =0.

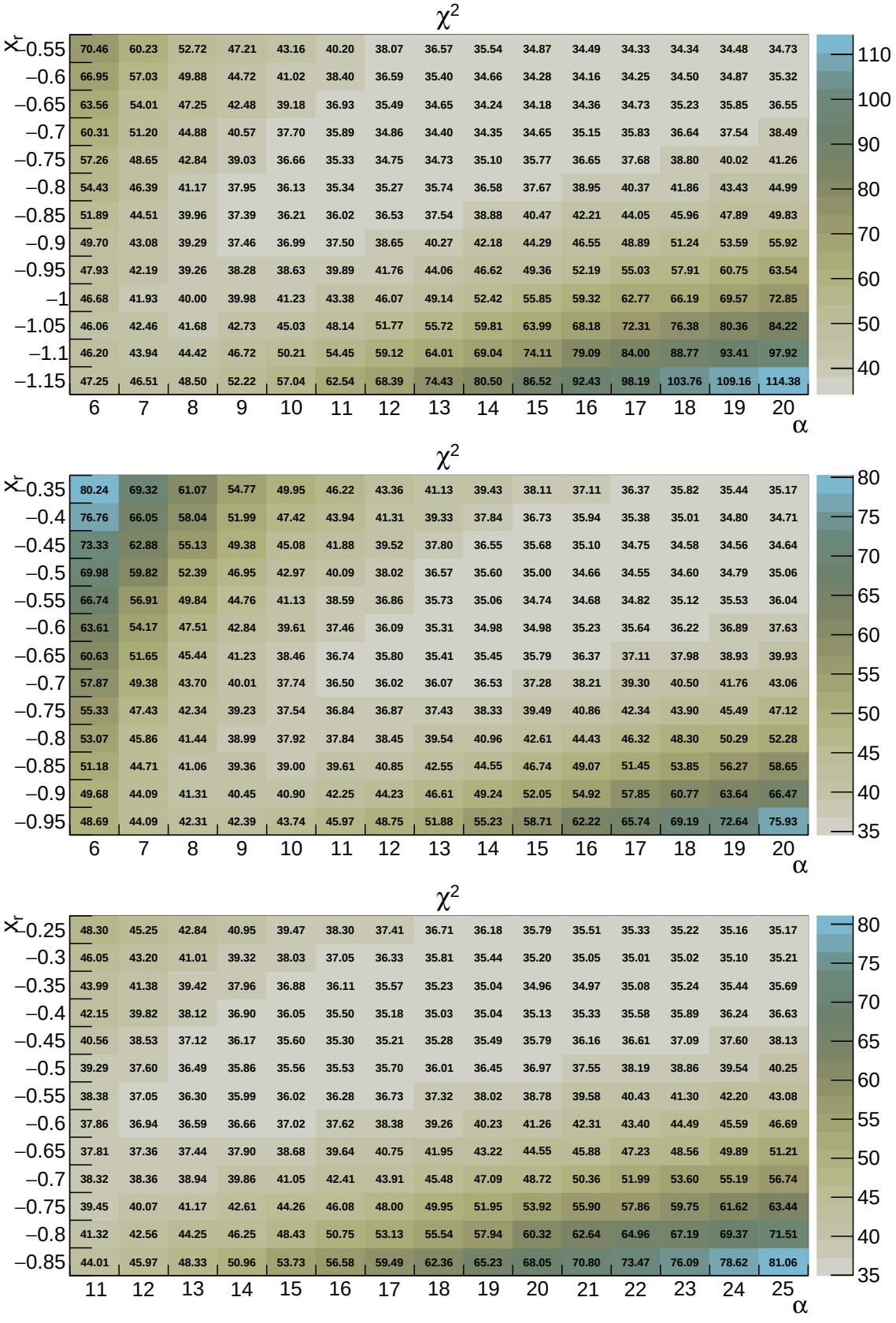


Figure G.31: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

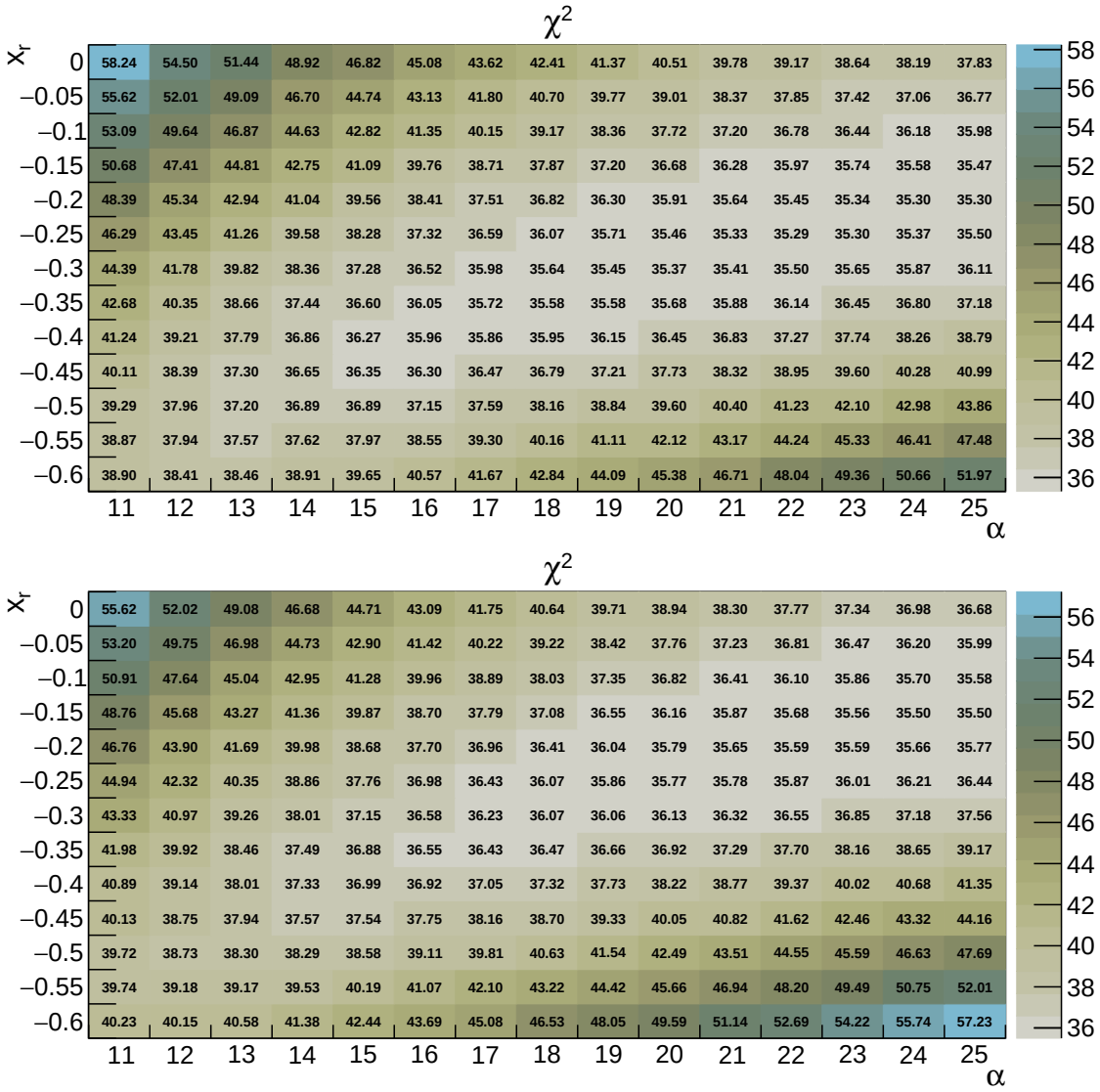


Figure G.32: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.5$ GeV and x_f =0.75 (top) and 1 (bottom).

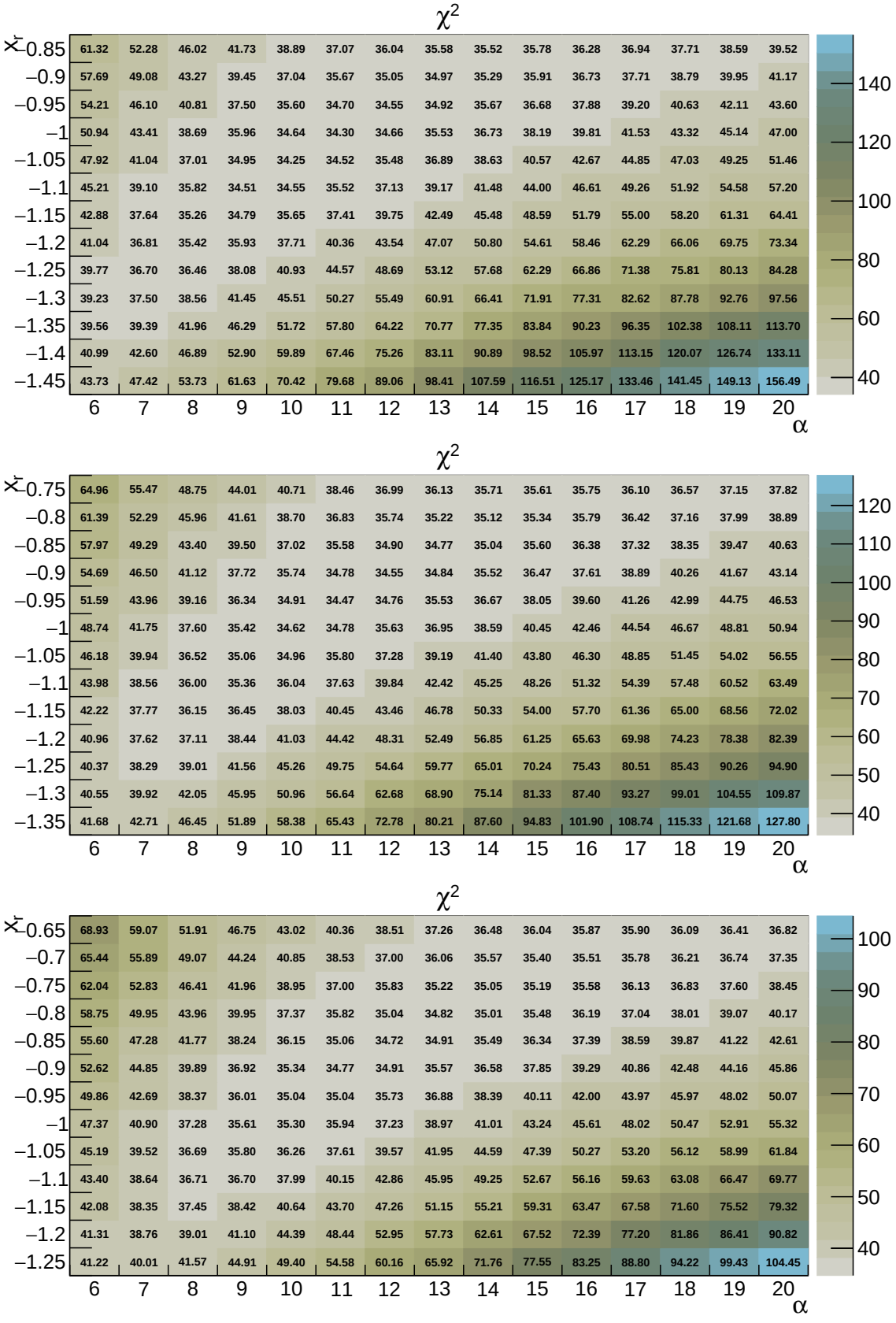


Figure G.33: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0 (top), 0.25 (middle) and 0.5 (bottom).

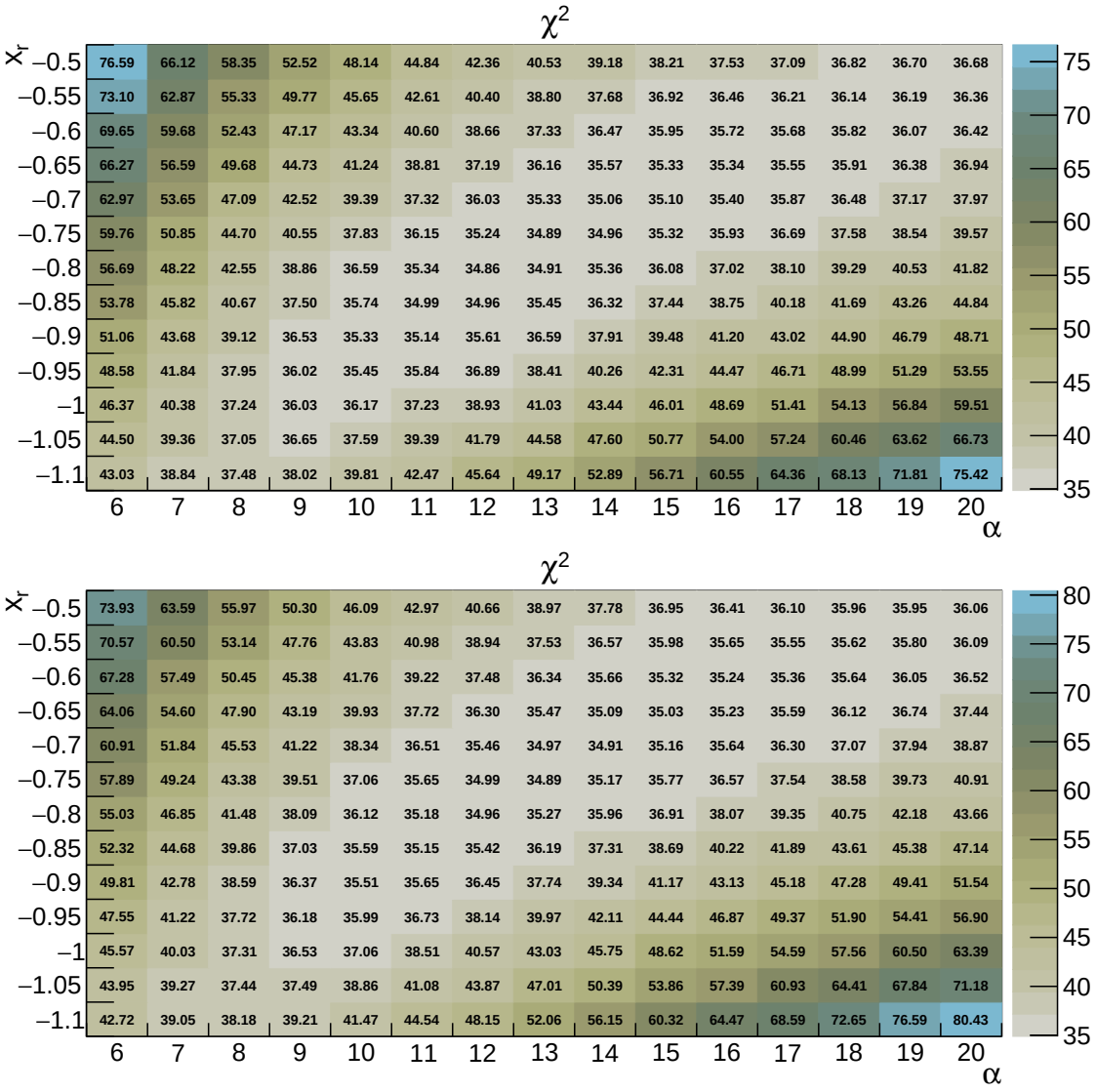


Figure G.34: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.7$ GeV and x_f =0.75 (top) and 1 (bottom).

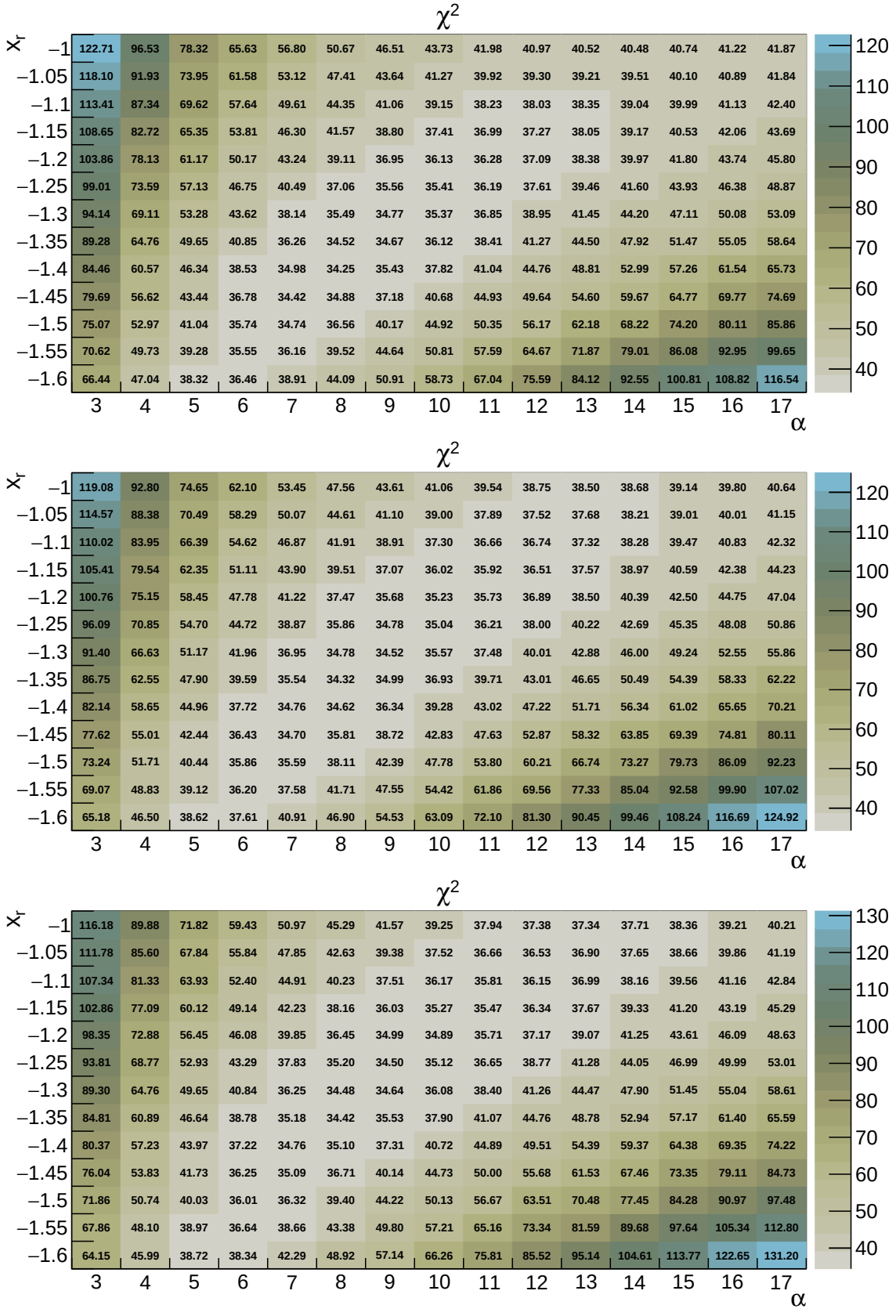


Figure G.35: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0 (top) and 0.25 (middle) and 0.5 (bottom).

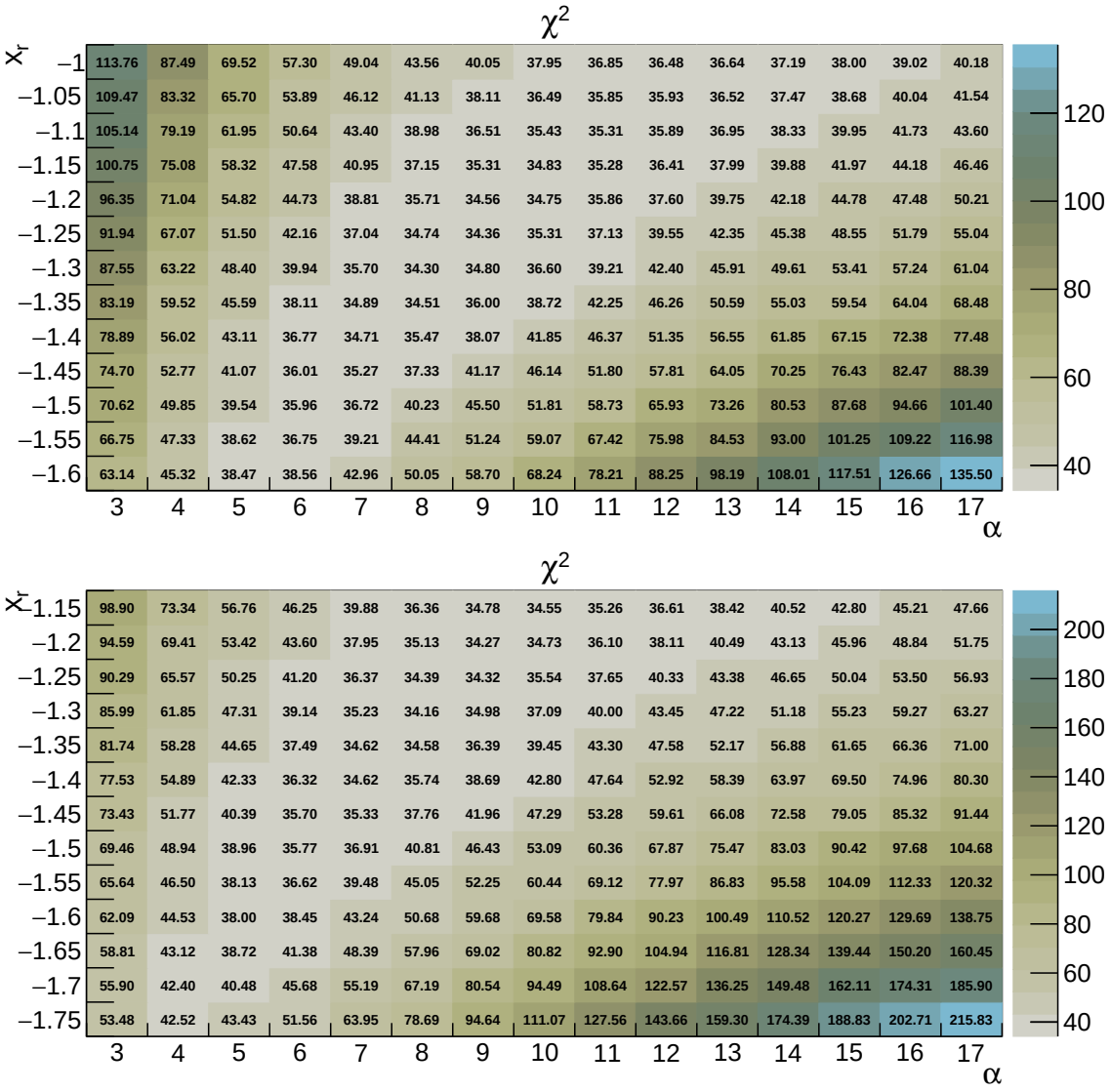


Figure G.36: χ^2 tables at $\sqrt{s} = 0.9$ TeV with the non-universal charm fragmentation. $m_c = 1.9$ GeV and x_f =0.75 (top) and 1 (bottom).

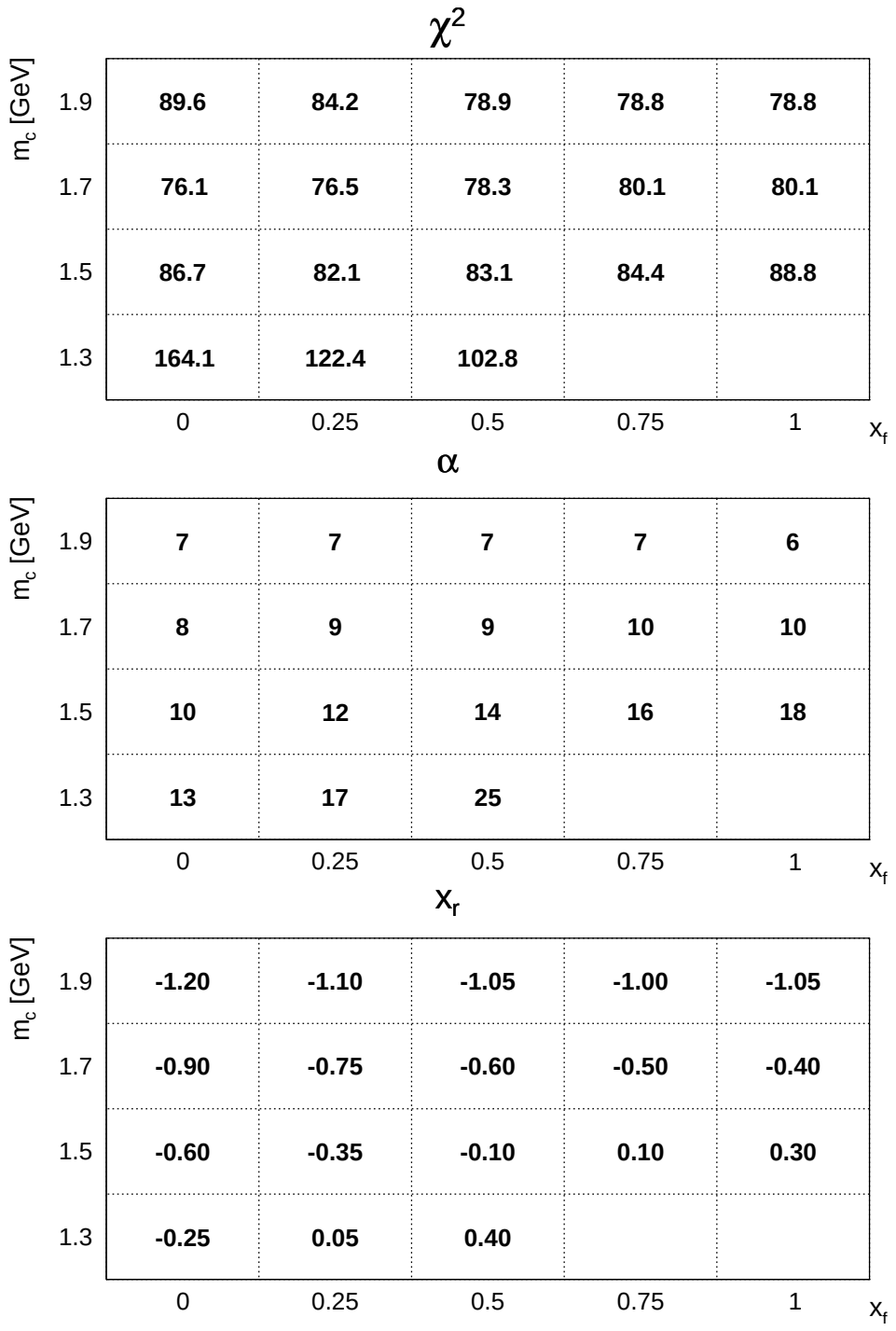


Figure G.37: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} =$ 5 TeV with the universality assumption.

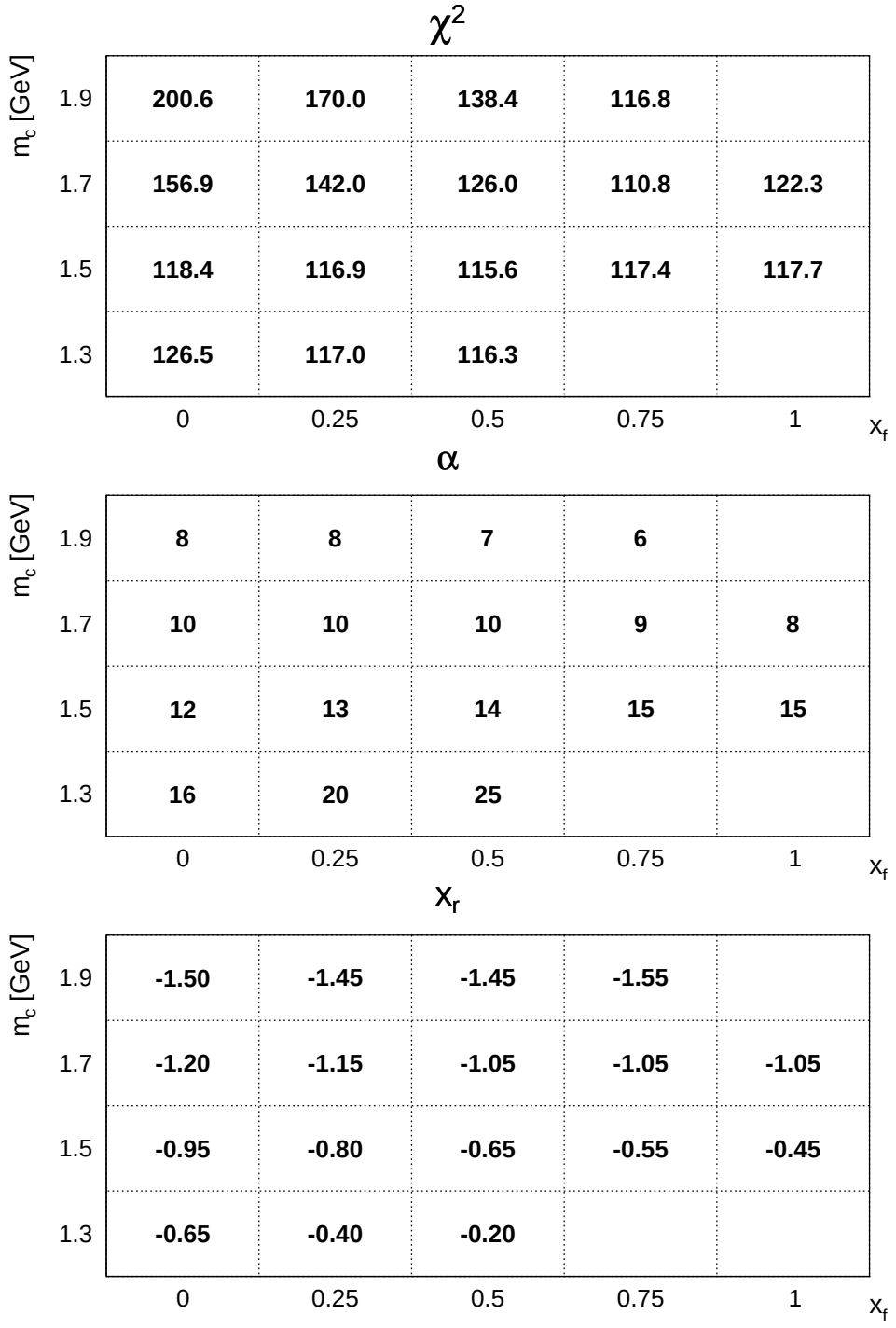


Figure G.38: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} =$ 5 TeV with the non-universal charm fragmentation. The global least χ^2 is determined to be 110.8 with the number of degrees freedom is 52. $\Delta\chi^2 = 10.0$ with S factor 1.46.

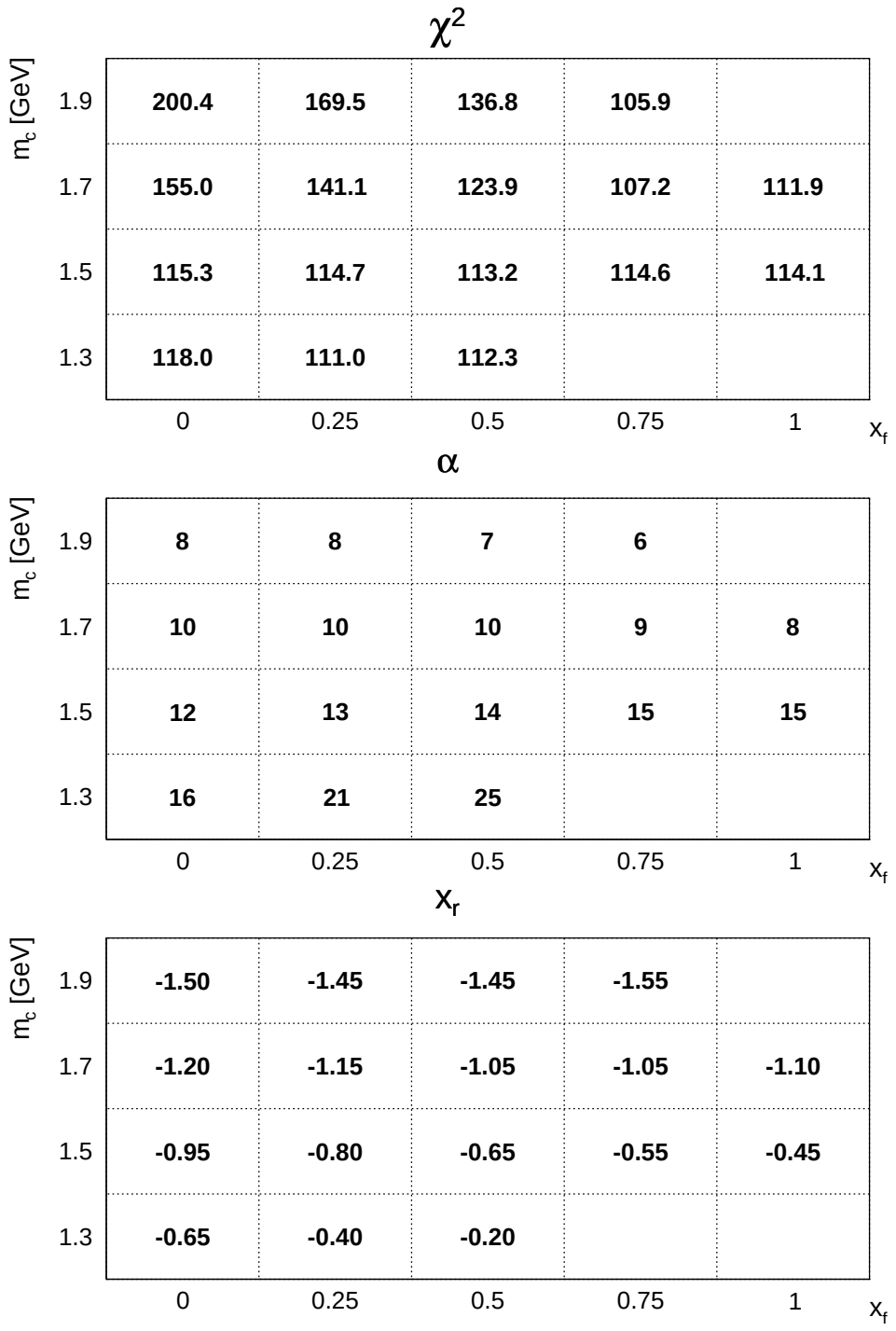


Figure G.39: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} = 5$ TeV with the non-universal charm fragmentation, but excluding the $p_T < 1$ GeV bins of ALICE. This table was produced just as a reference and not used to derive the total charm cross section at $\sqrt{s} = 5$ TeV.

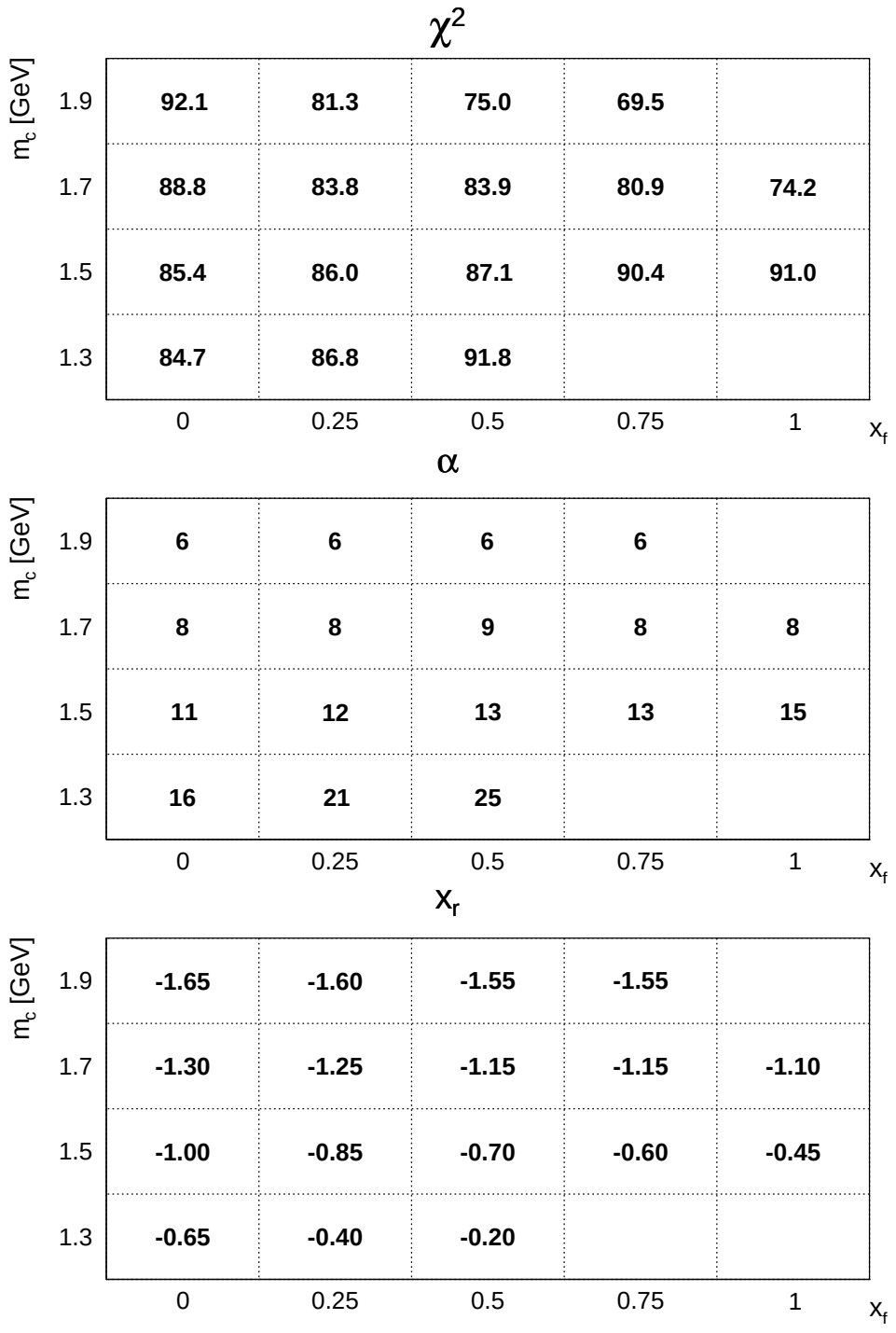


Figure G.40: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} =$ 5 TeV with the non-universal charm fragmentation, but excluding all the $p_T < 1$ GeV bins. This table was produced just as a reference and not used to derive the total charm cross section at $\sqrt{s} = 5$ TeV.

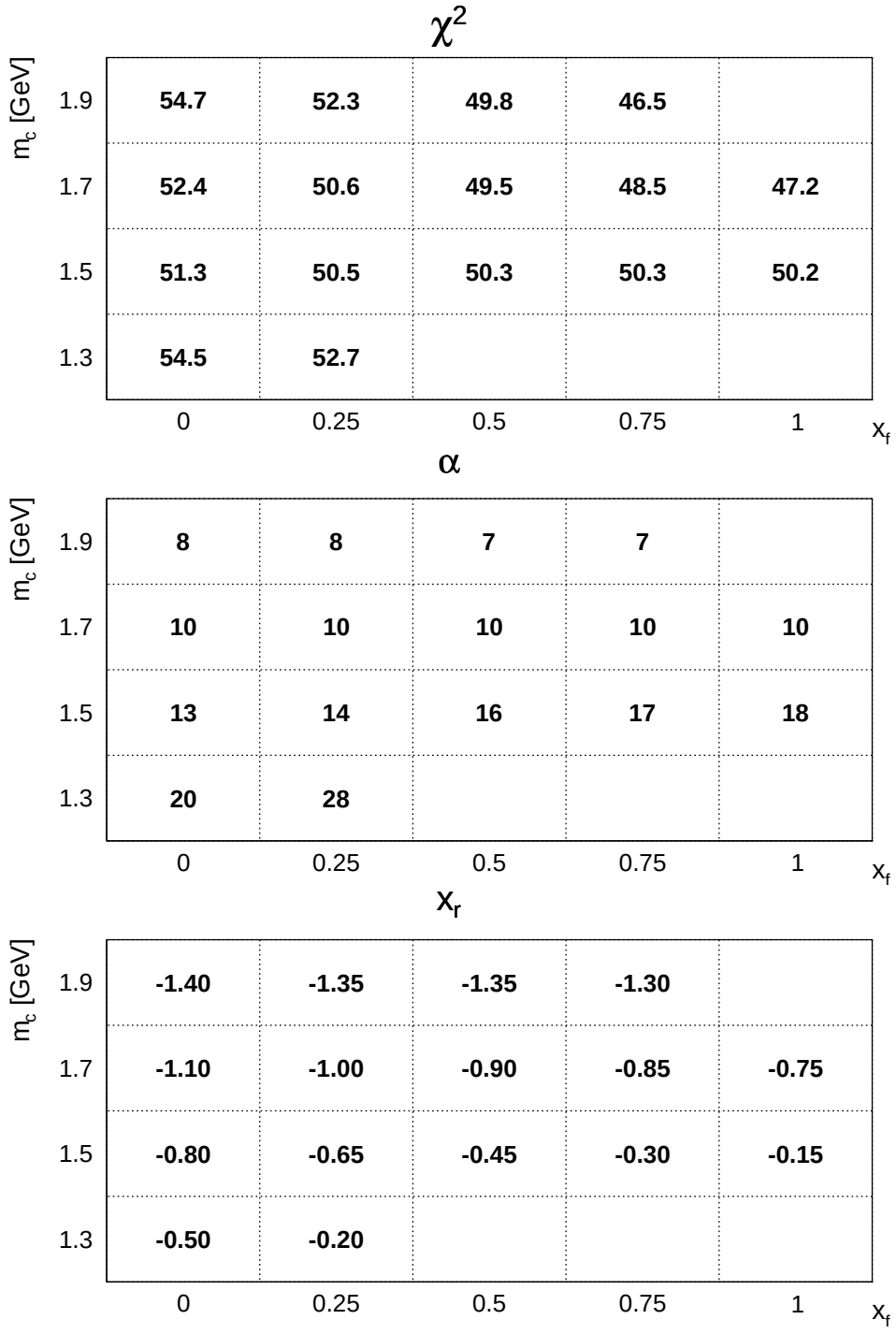


Figure G.41: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} = 7$ TeV with the non-universal charm fragmentation. The global least χ^2 is not determined within the chosen parameters of x_f and m_c. Since similar patterns of χ^2 distribution were observed in Figure G.39 and G.40 with missing (part of) $p_T < 1$ GeV data inputs, the global least χ^2 is chosen to be 48.5 by taking the same μ_f and m_c as the ones determined in Figure G.38. With the number of degrees freedom is 72, $\Delta\chi^2 = 4.7$ (without applying a S factor).

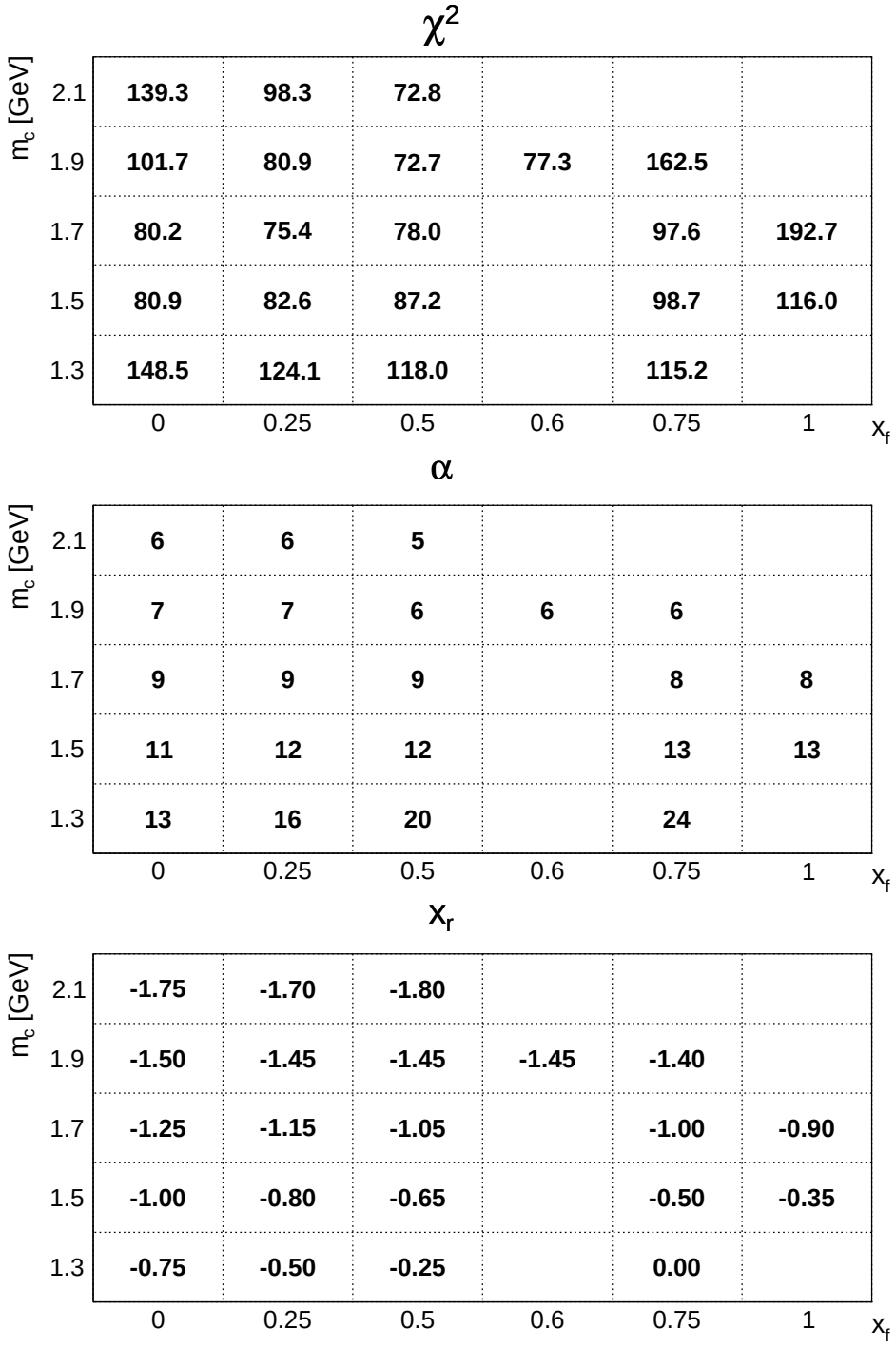


Figure G.42: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} =$ 13 TeV with the non-universal charm fragmentation. The global least χ^2 is determined to be 72.7 with the number of degrees freedom is 54. $\Delta\chi^2 = 6.3$ with S factor 1.16.

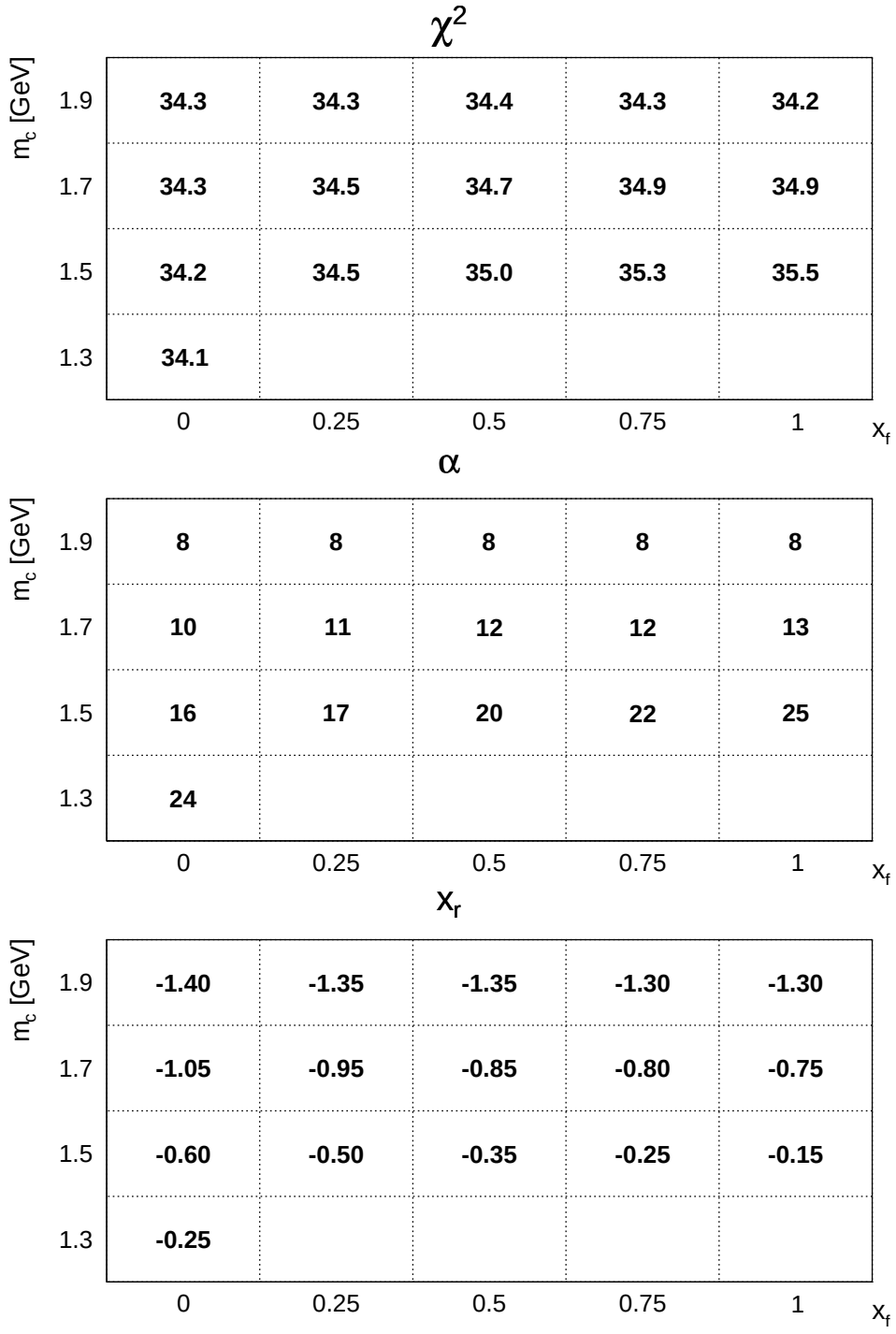


Figure G.43: The least χ^2s (top), α_Ks (middle) and x_rs (bottom) with fixed x_f and m_c at $\sqrt{s} =$ 0.9 TeV with the non-universal charm fragmentation. Within the chosen parameters of x_f and m_c, almost no variations are observed in the least χ^2s. The global least χ^2 was chosen to be 34.9 with the same μ_f and m_c as the ones determined in Figure G.38, while all the least χ^2s were taken to determine χ^2 uncertainty for the total charm cross section at $\sqrt{s} = 0.9$ TeV. The number of degrees freedom is 42.

H Cross Sections as a Function of p_T and $|y|$

In this Appendix, the D^0 or D^{*+} cross sections at $\sqrt{s}$ = 5, 7, 13 and 0.9 TeV are collected as a function of p_T and $|y|$.

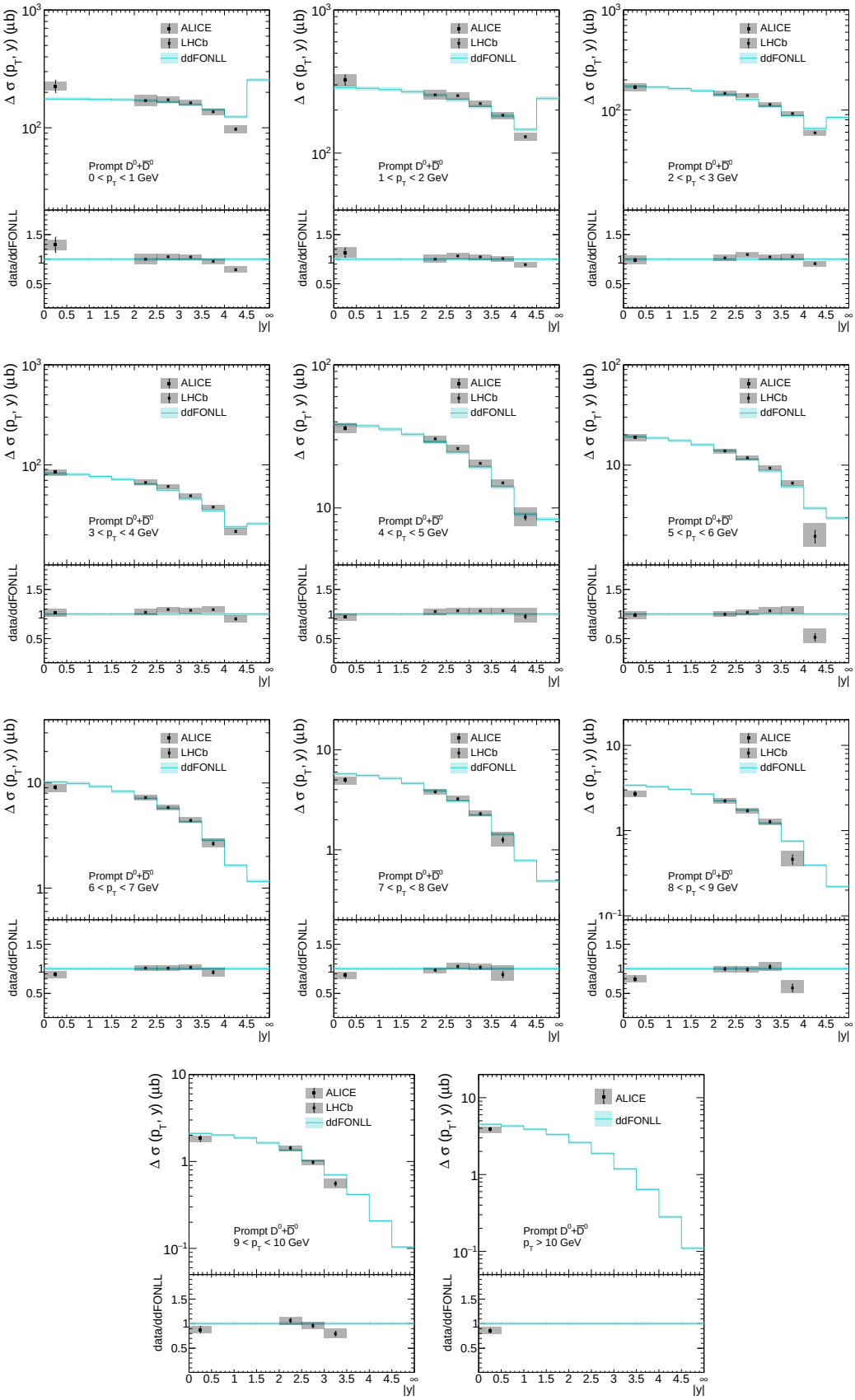


Figure H.1: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The ddFONLL was derived with the universality assumption and is shown with the f^{uni} uncertainty.

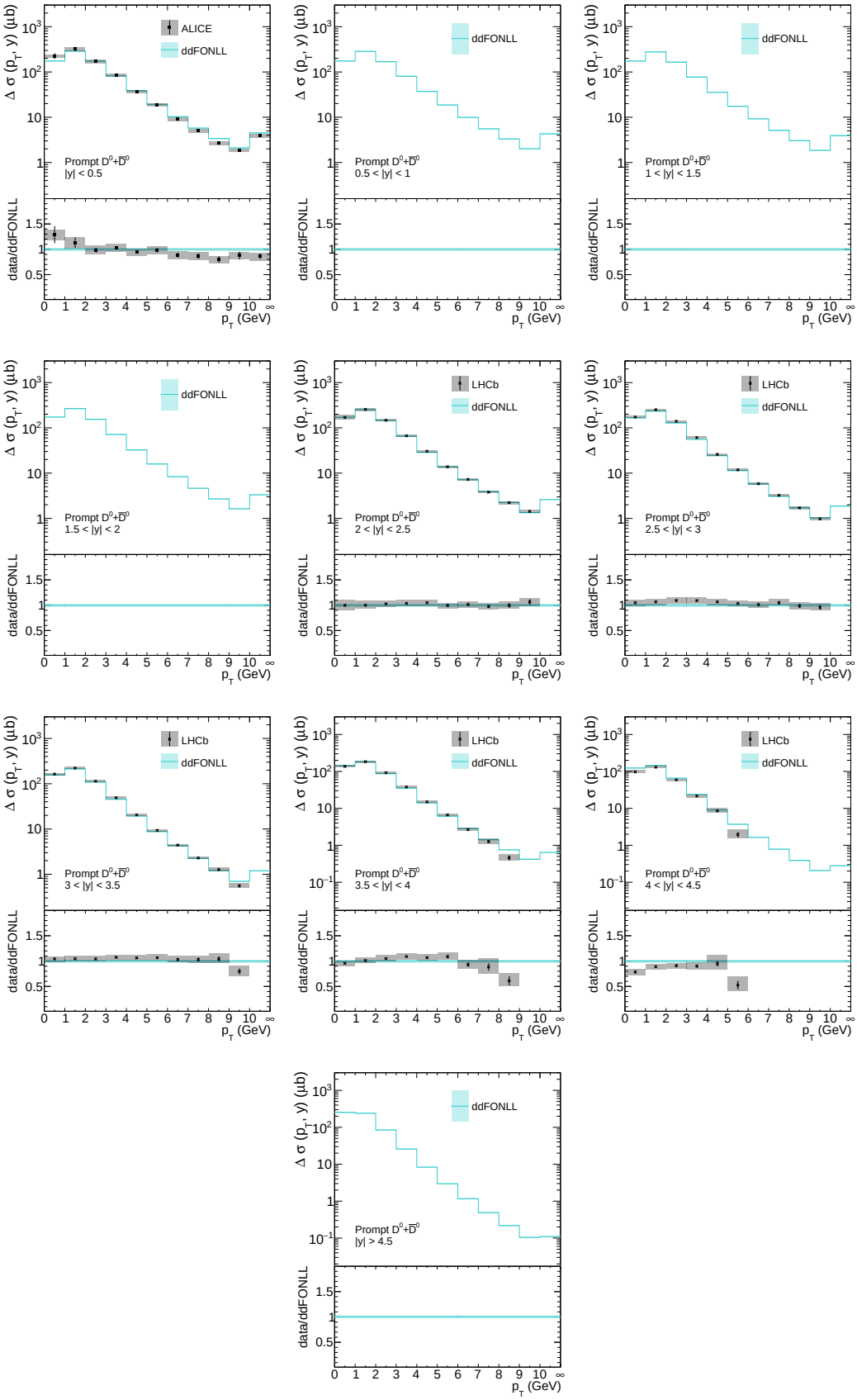


Figure H.2: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL was derived with the universality assumption and is shown with f^{uni} uncertainty.

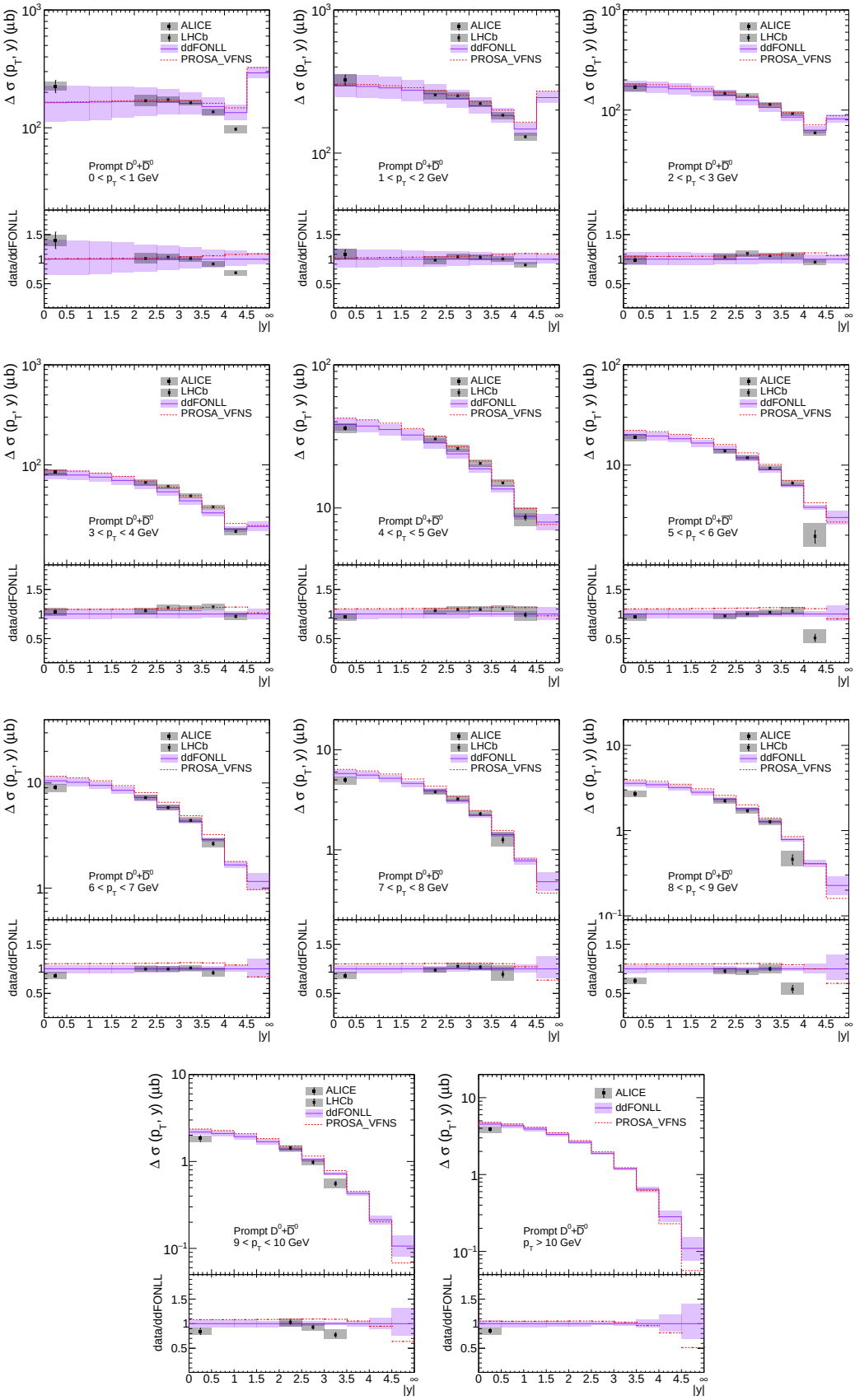


Figure H.3: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with the PDF uncertainty.

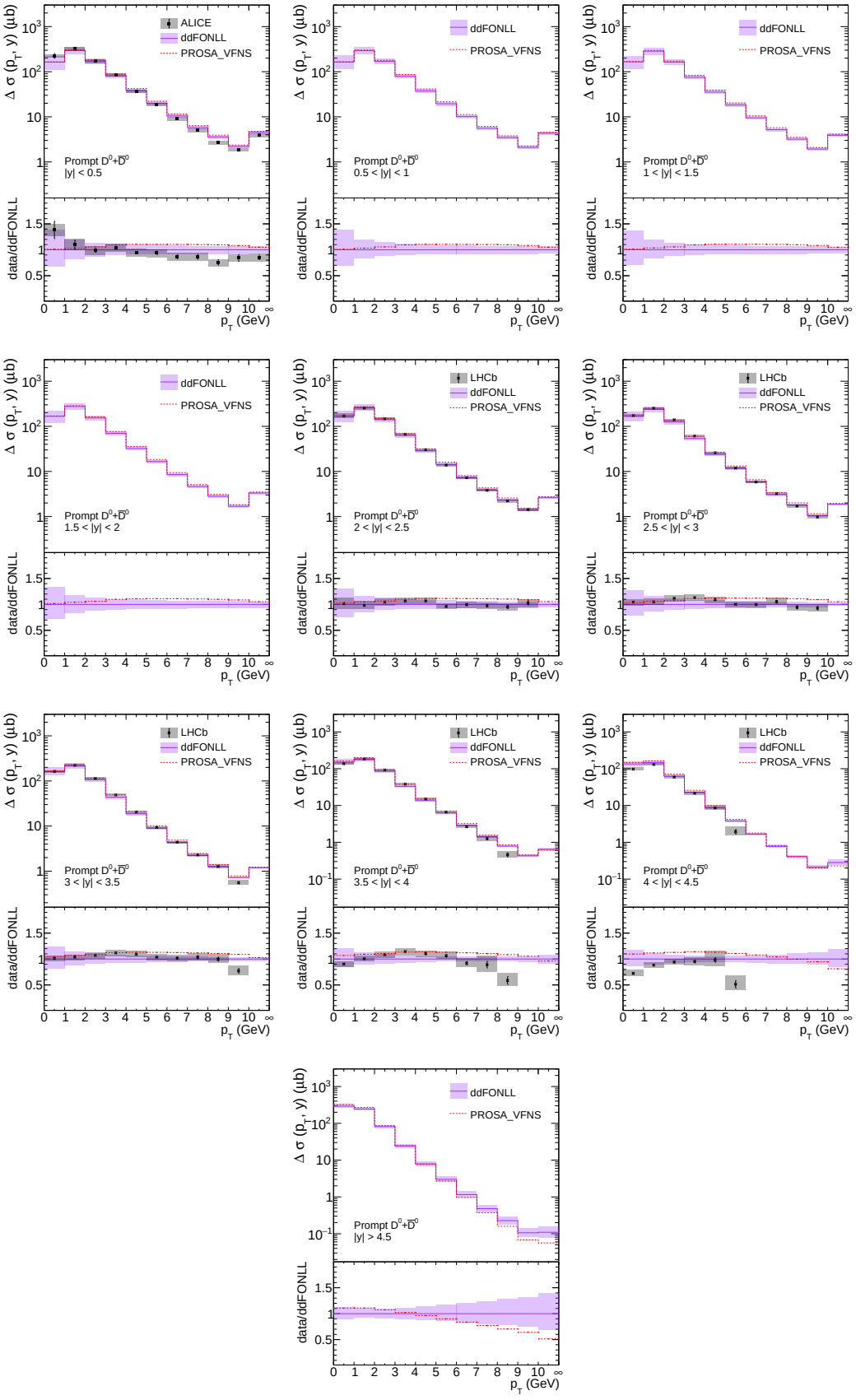


Figure H.4: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with the PDF uncertainty.

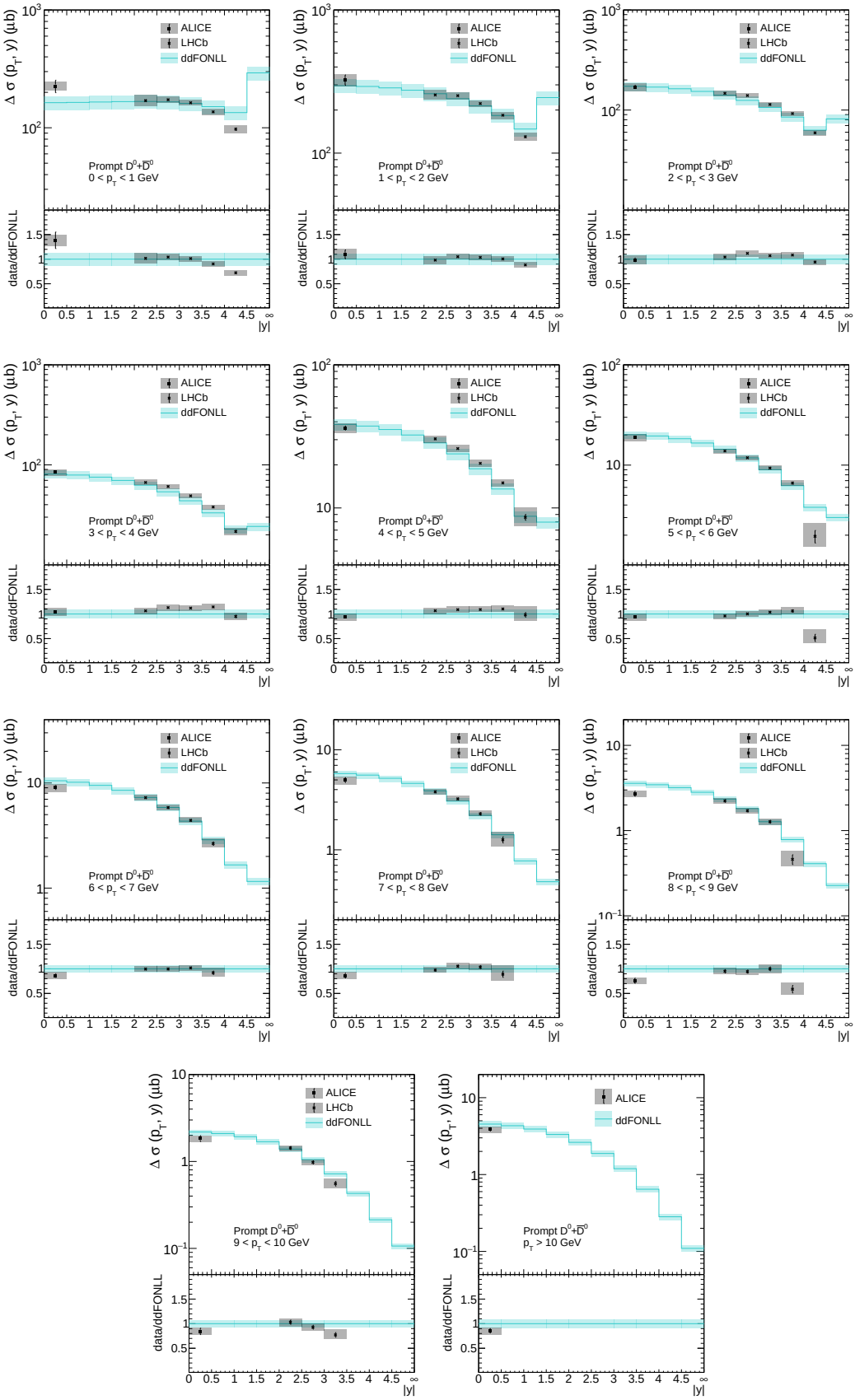


Figure H.5: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

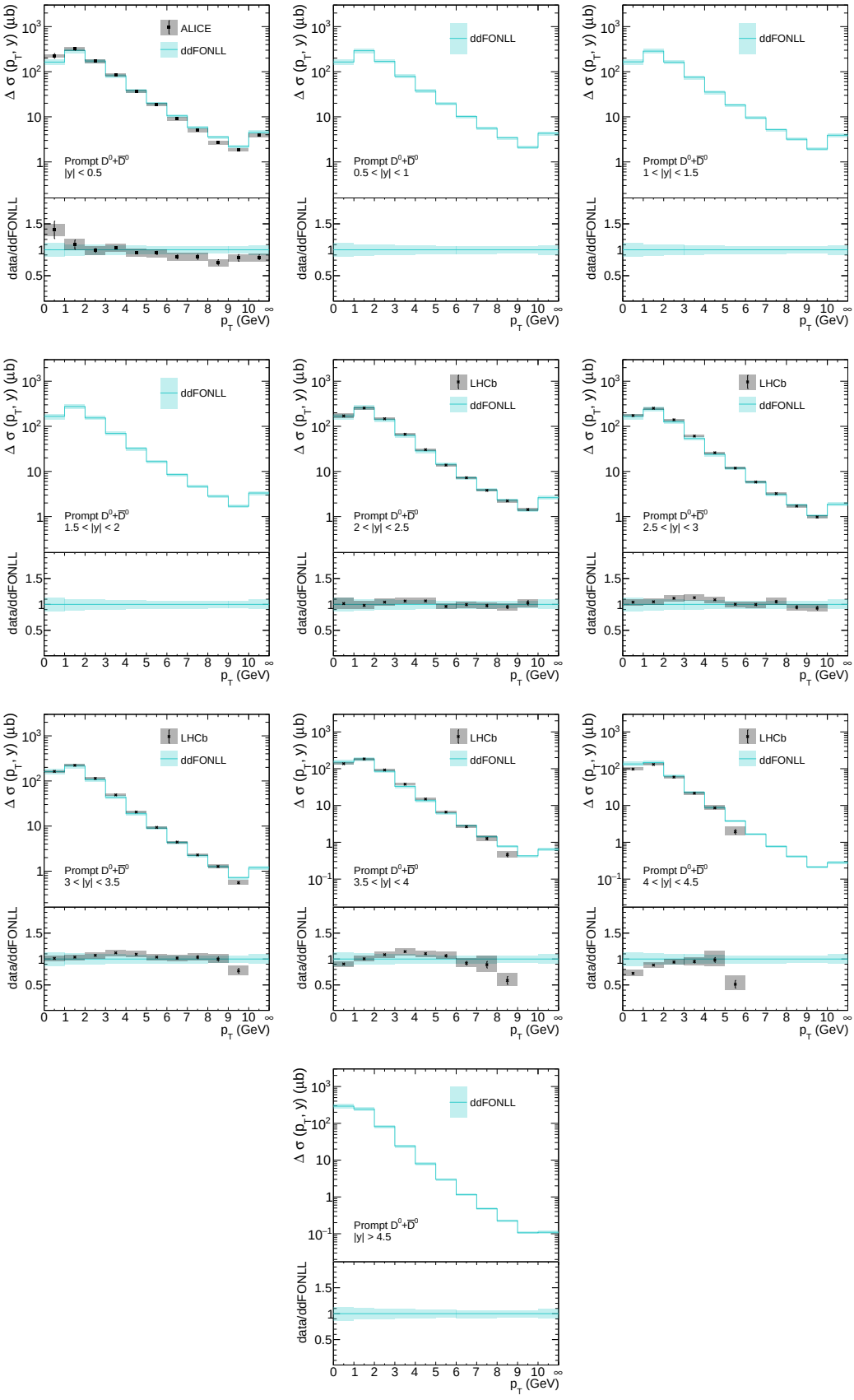


Figure H.6: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

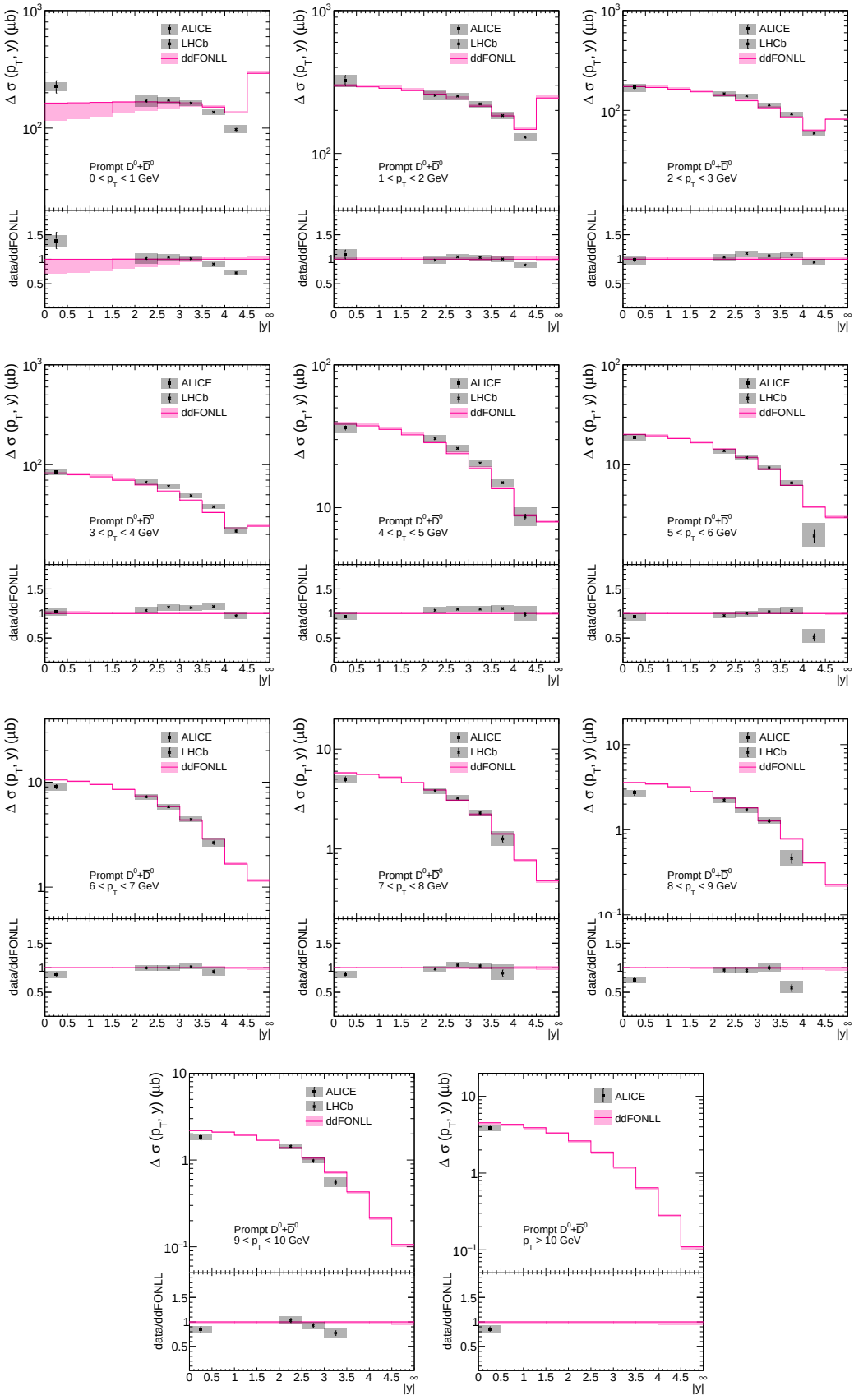


Figure H.7: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

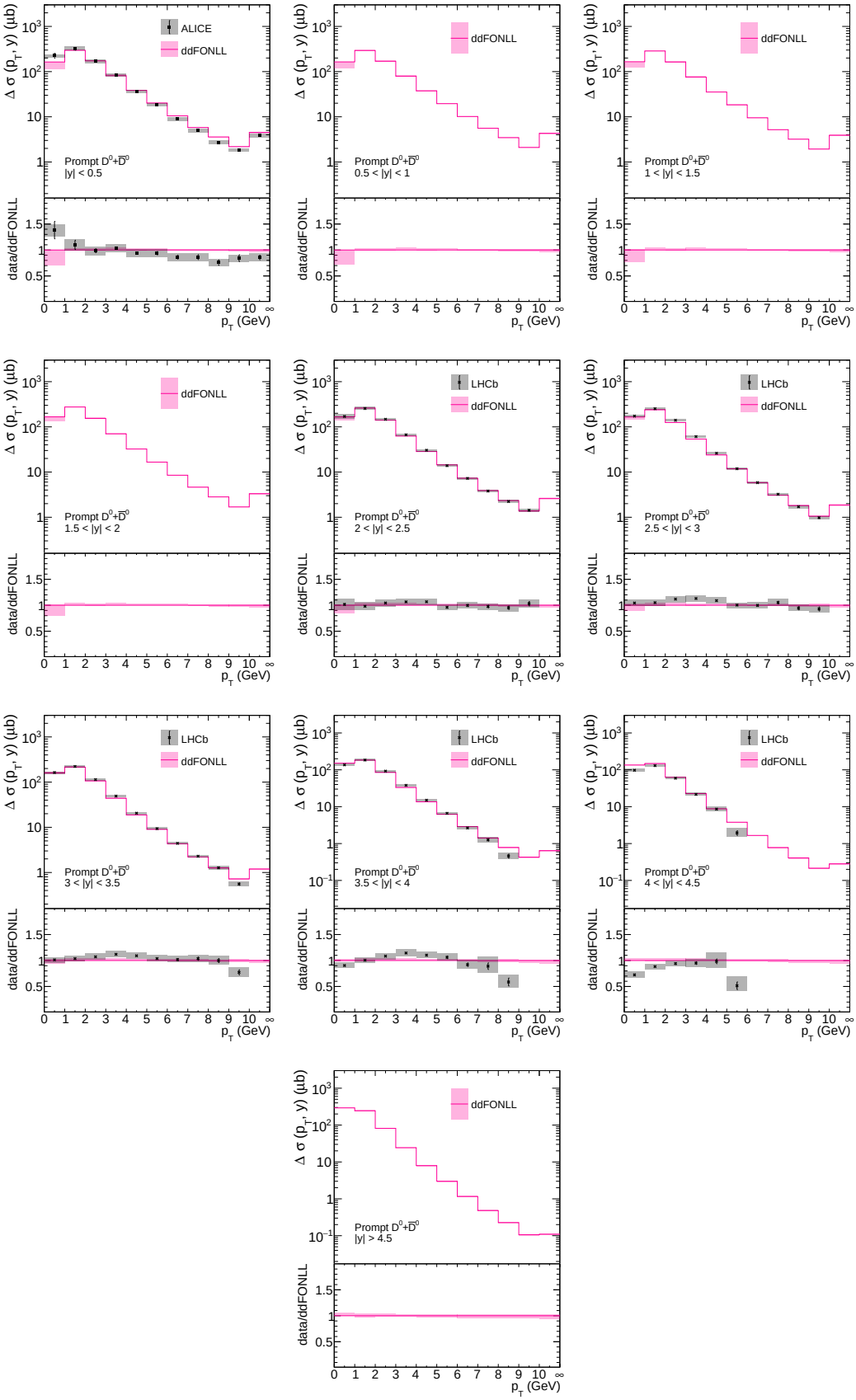


Figure H.8: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

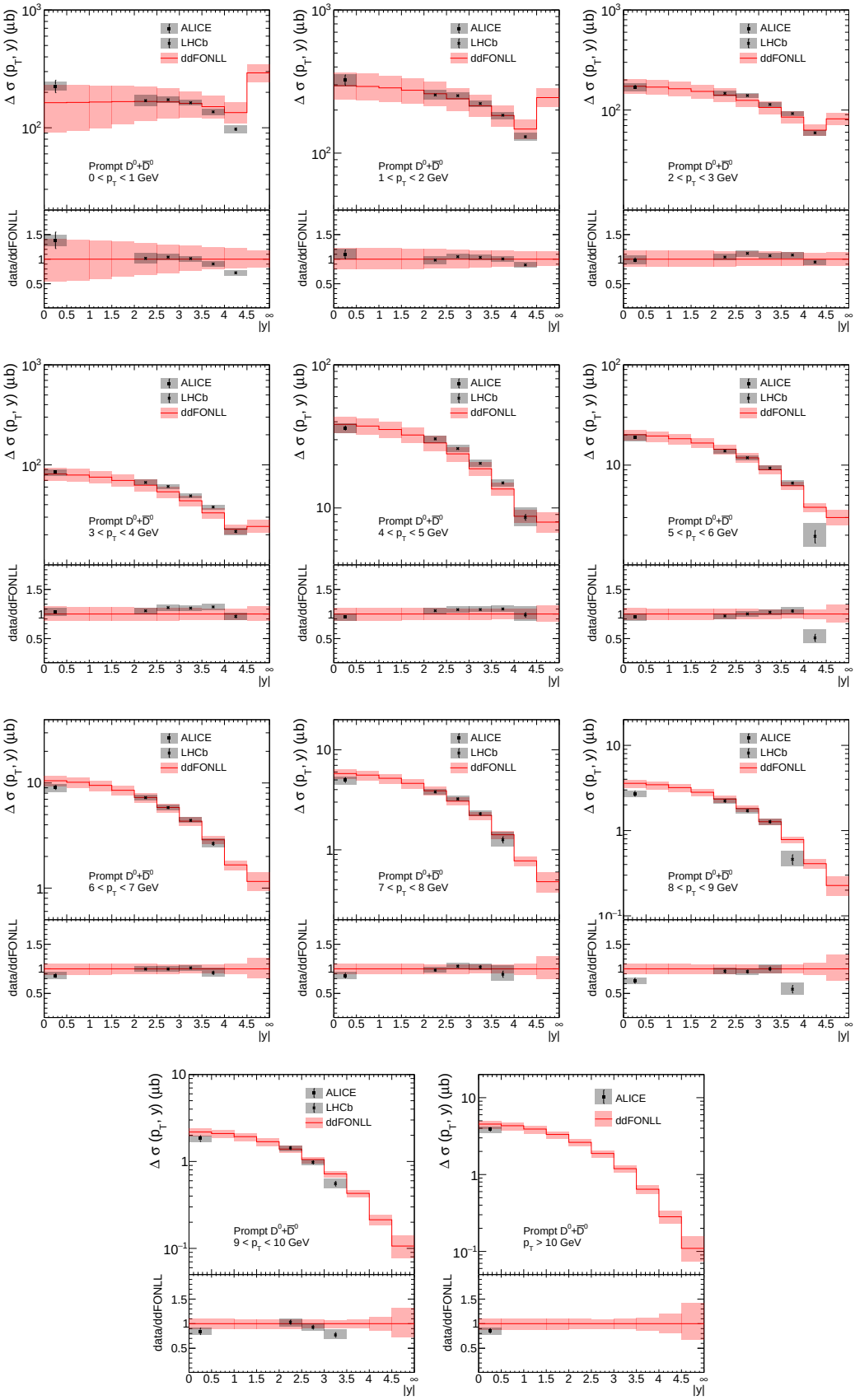


Figure H.9: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

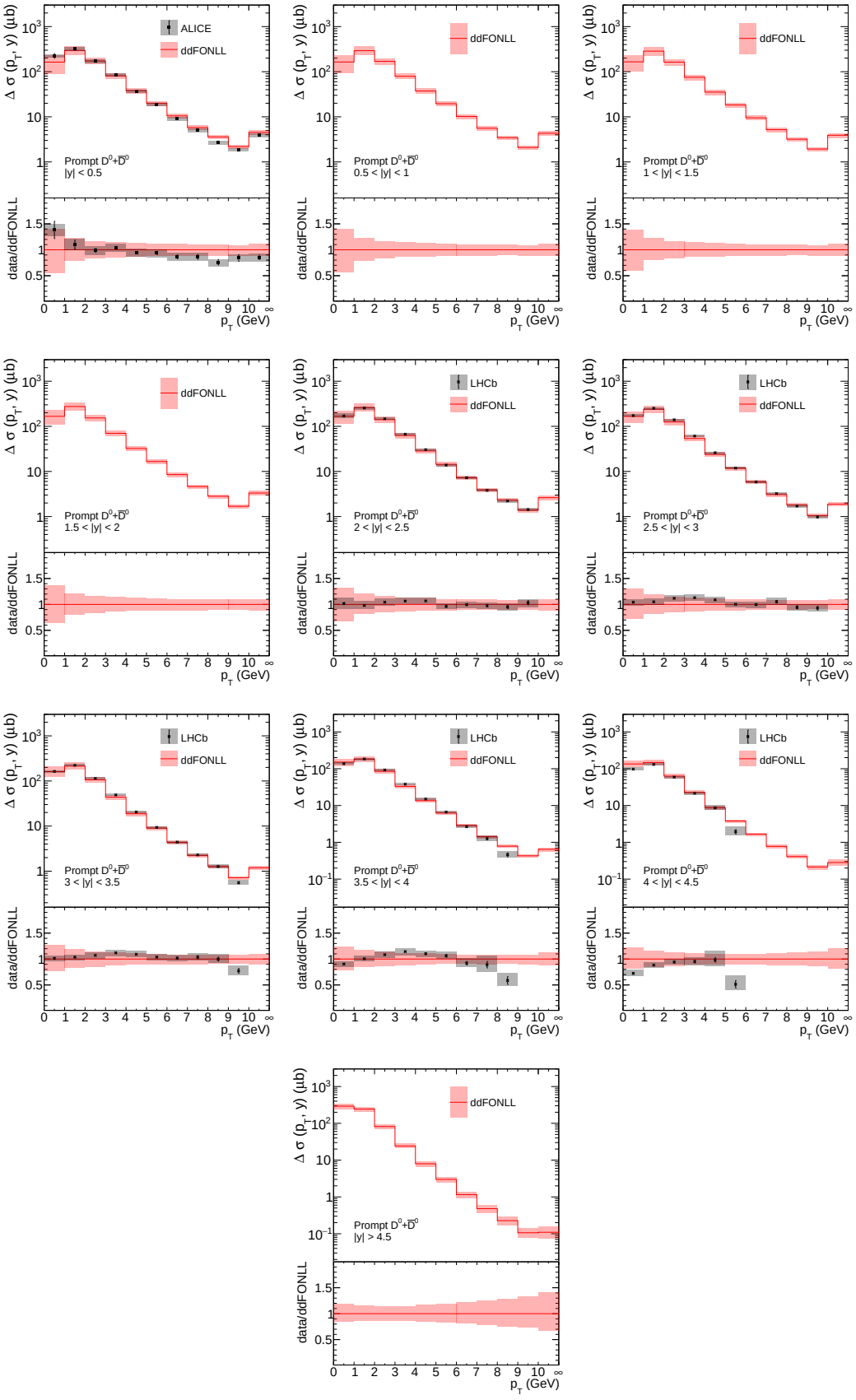


Figure H.10: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 5$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

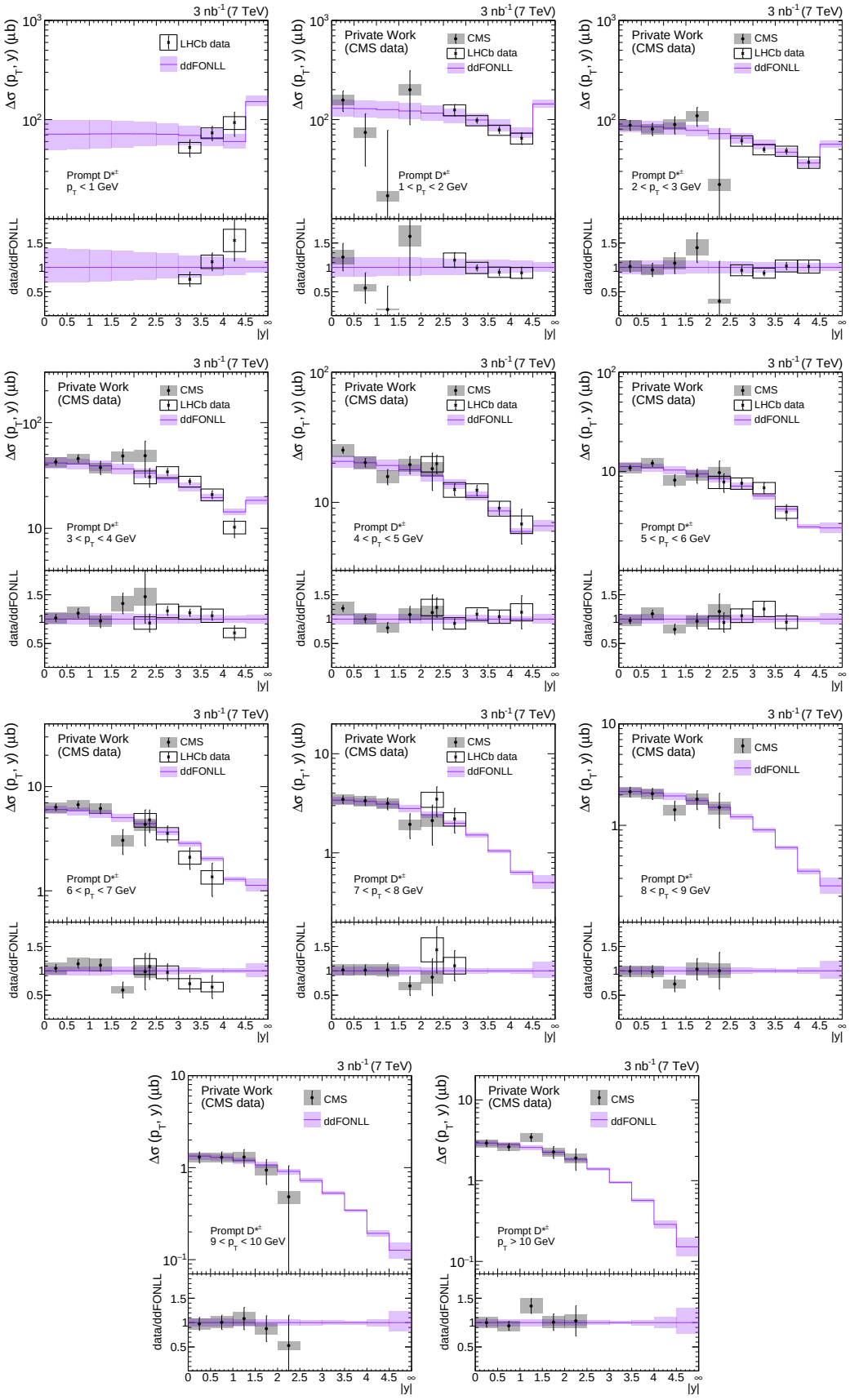


Figure H.11: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

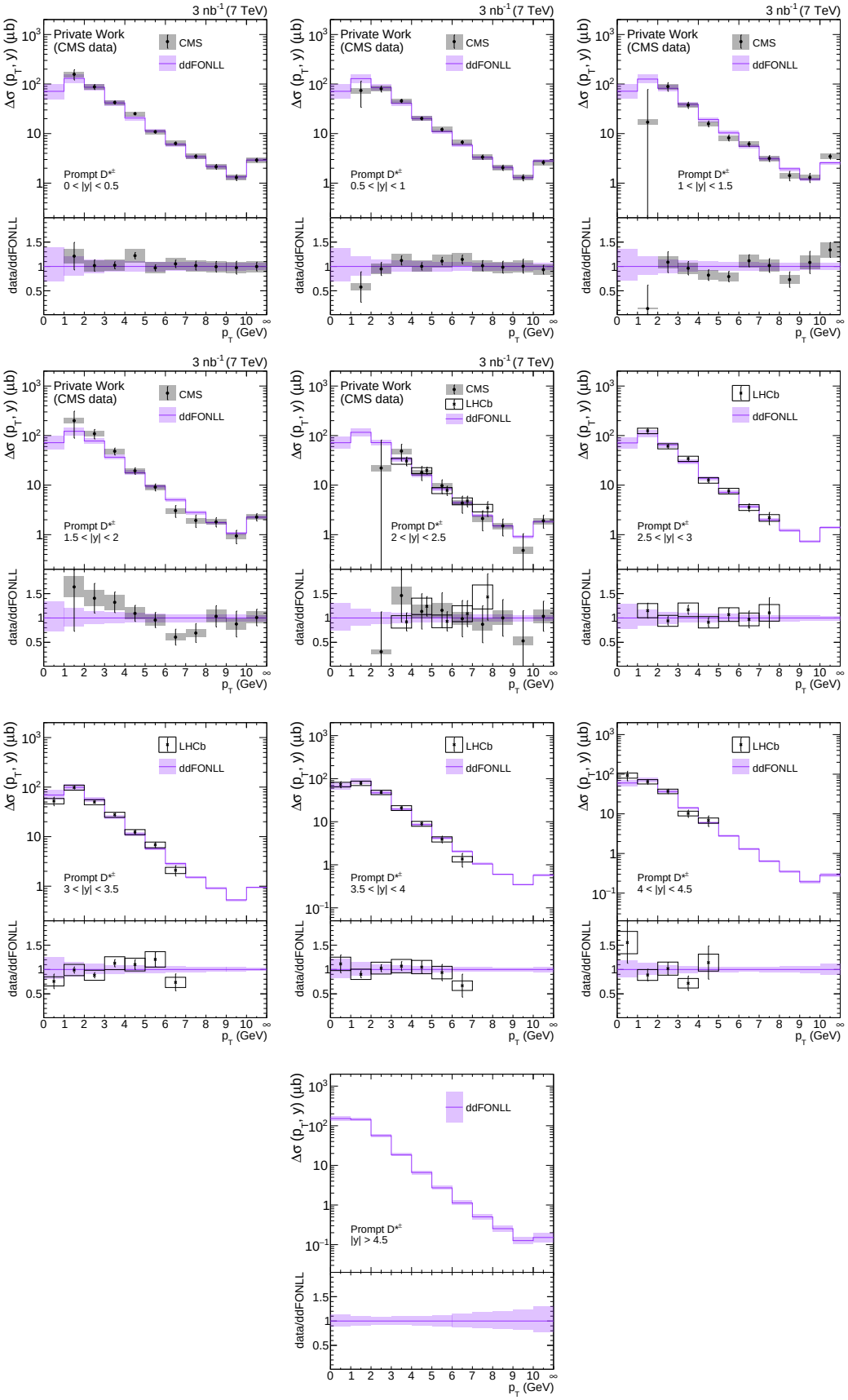


Figure H.12: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

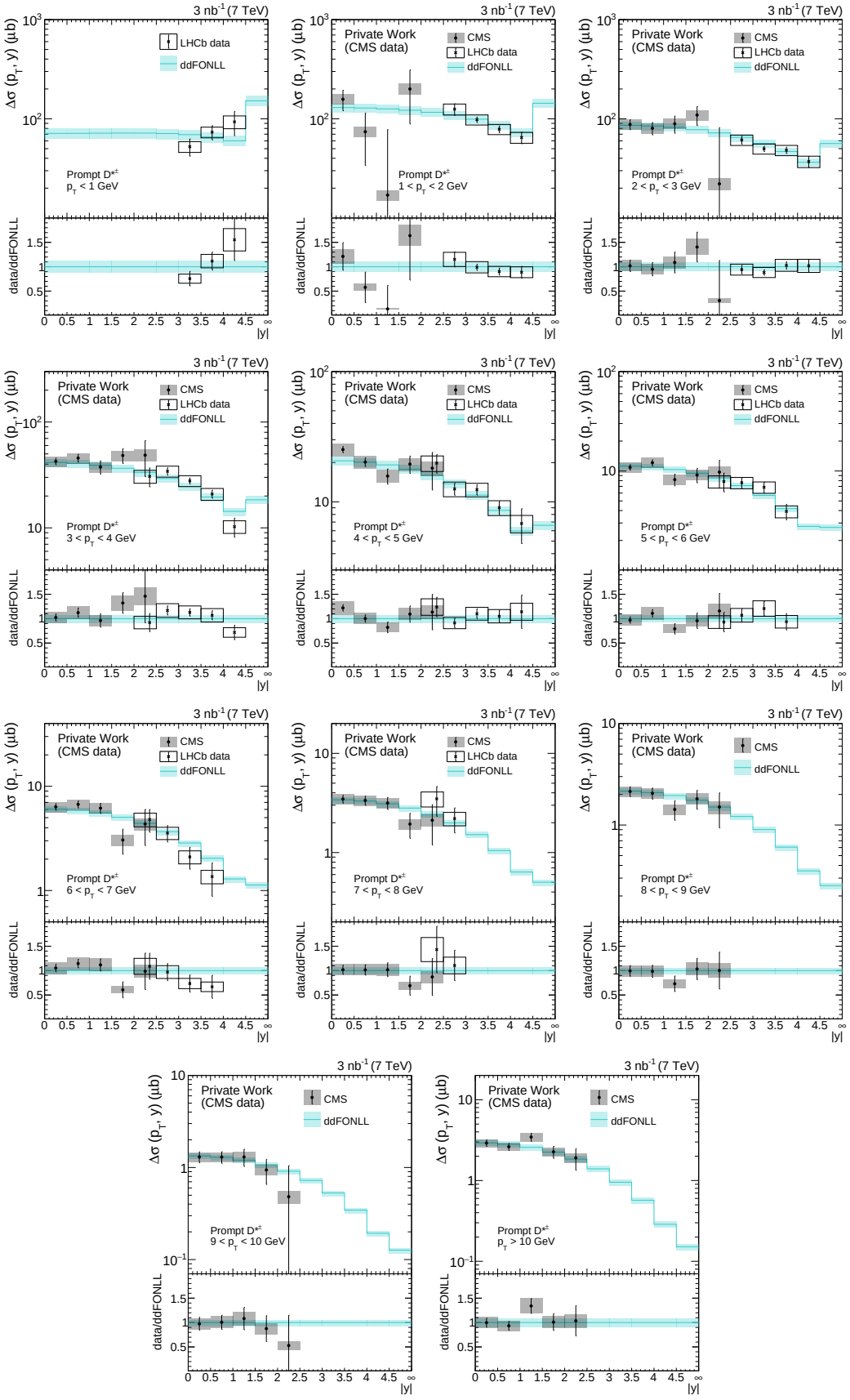


Figure H.13: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

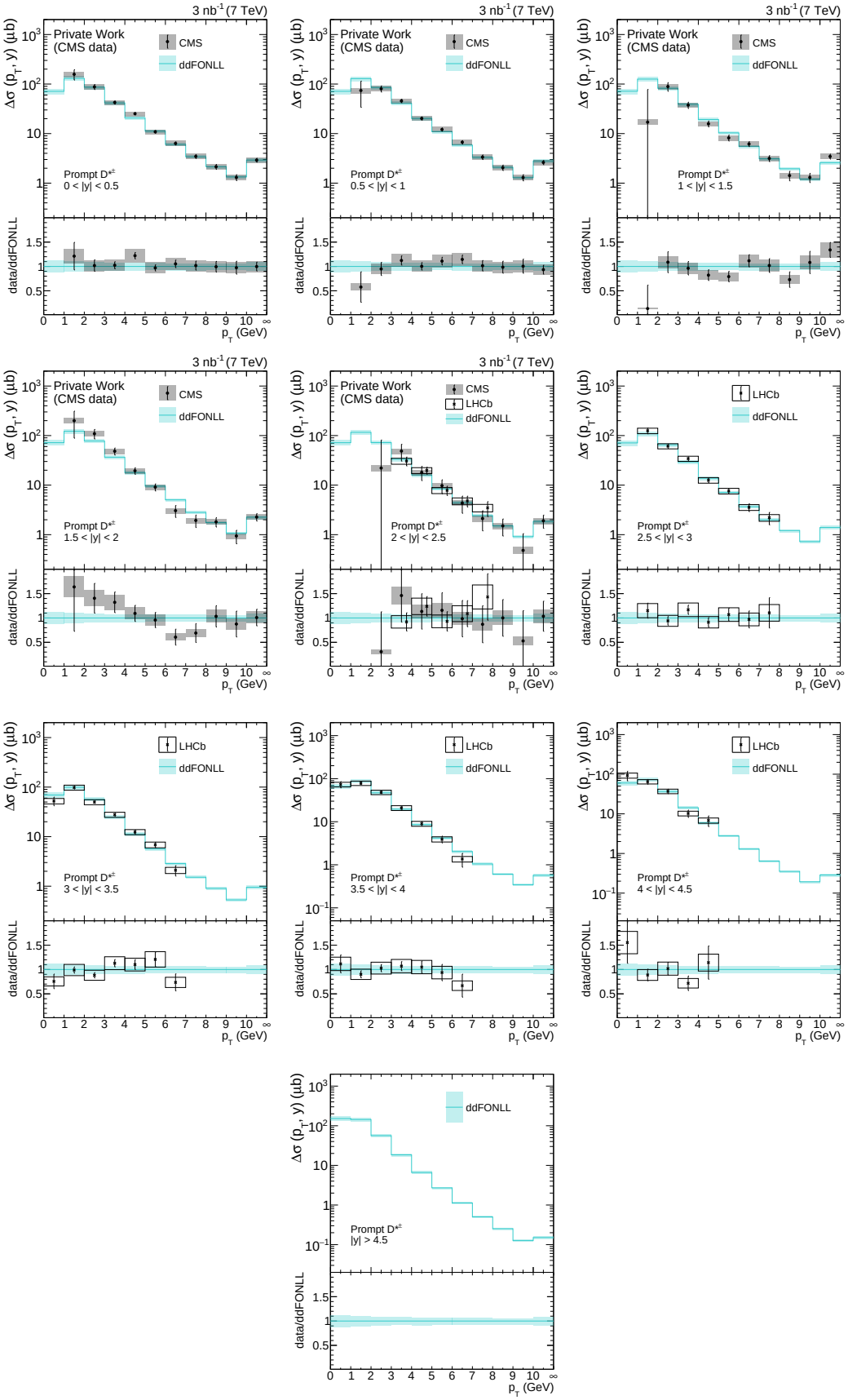


Figure H.14: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

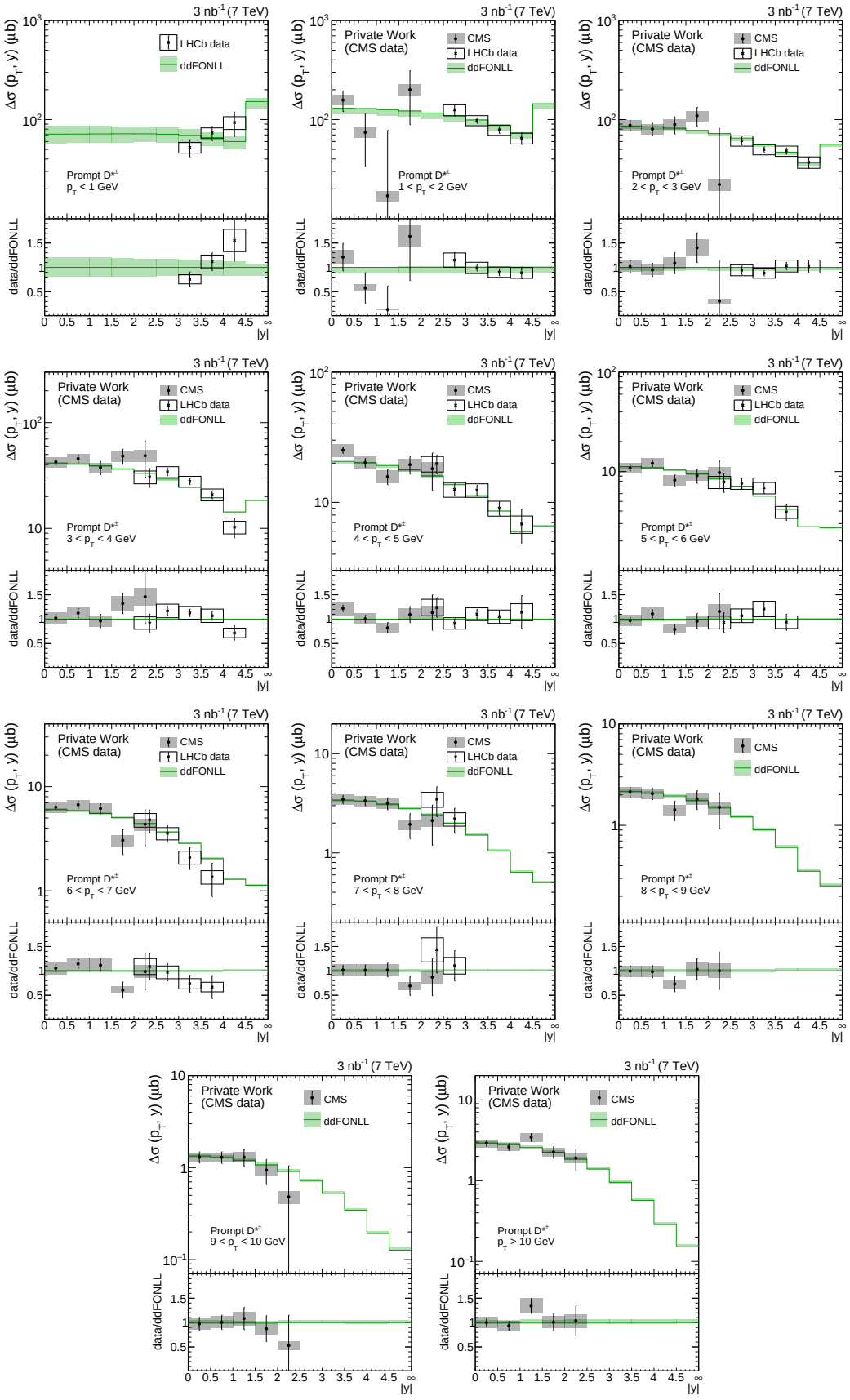


Figure H.15: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

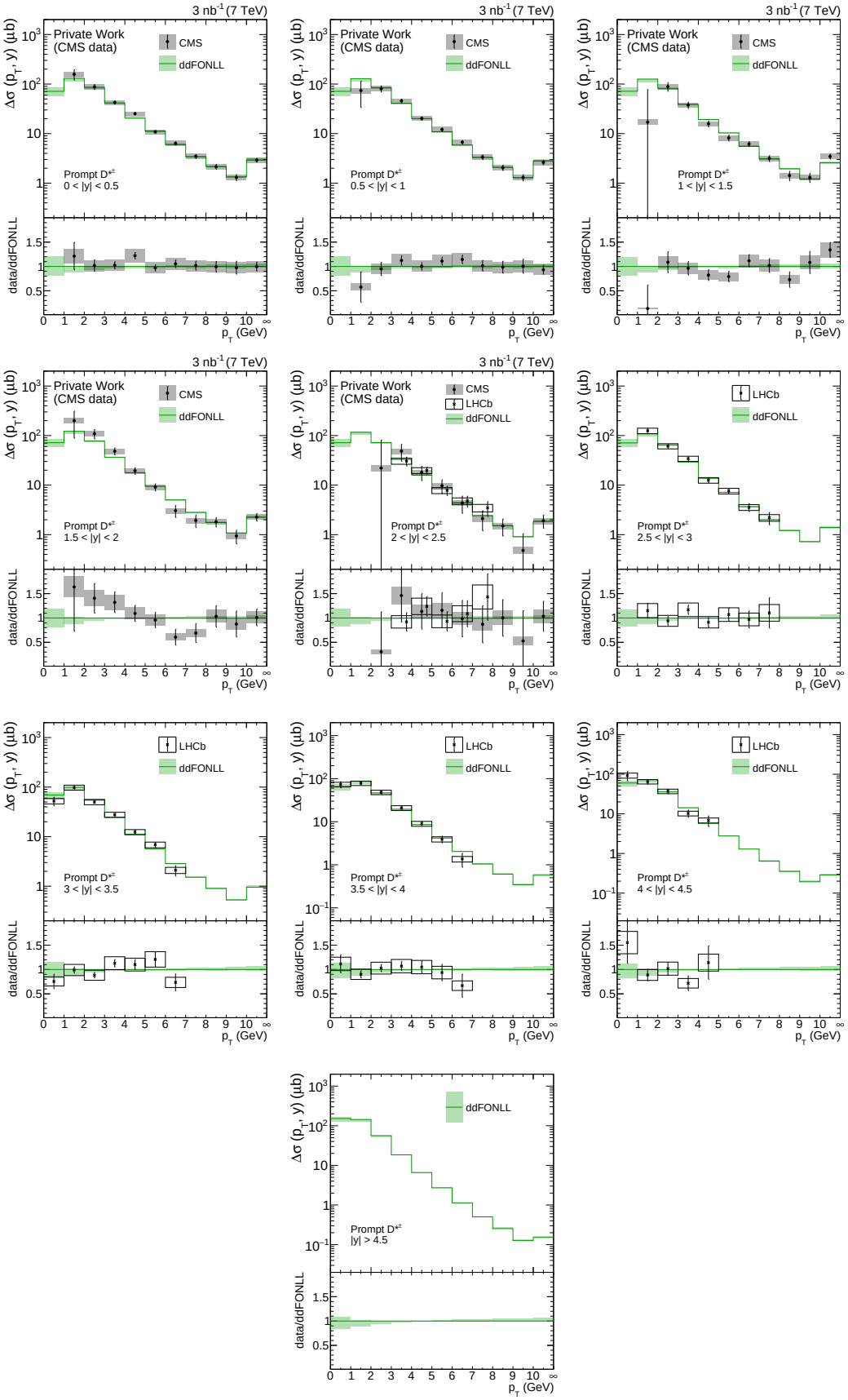


Figure H.16: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

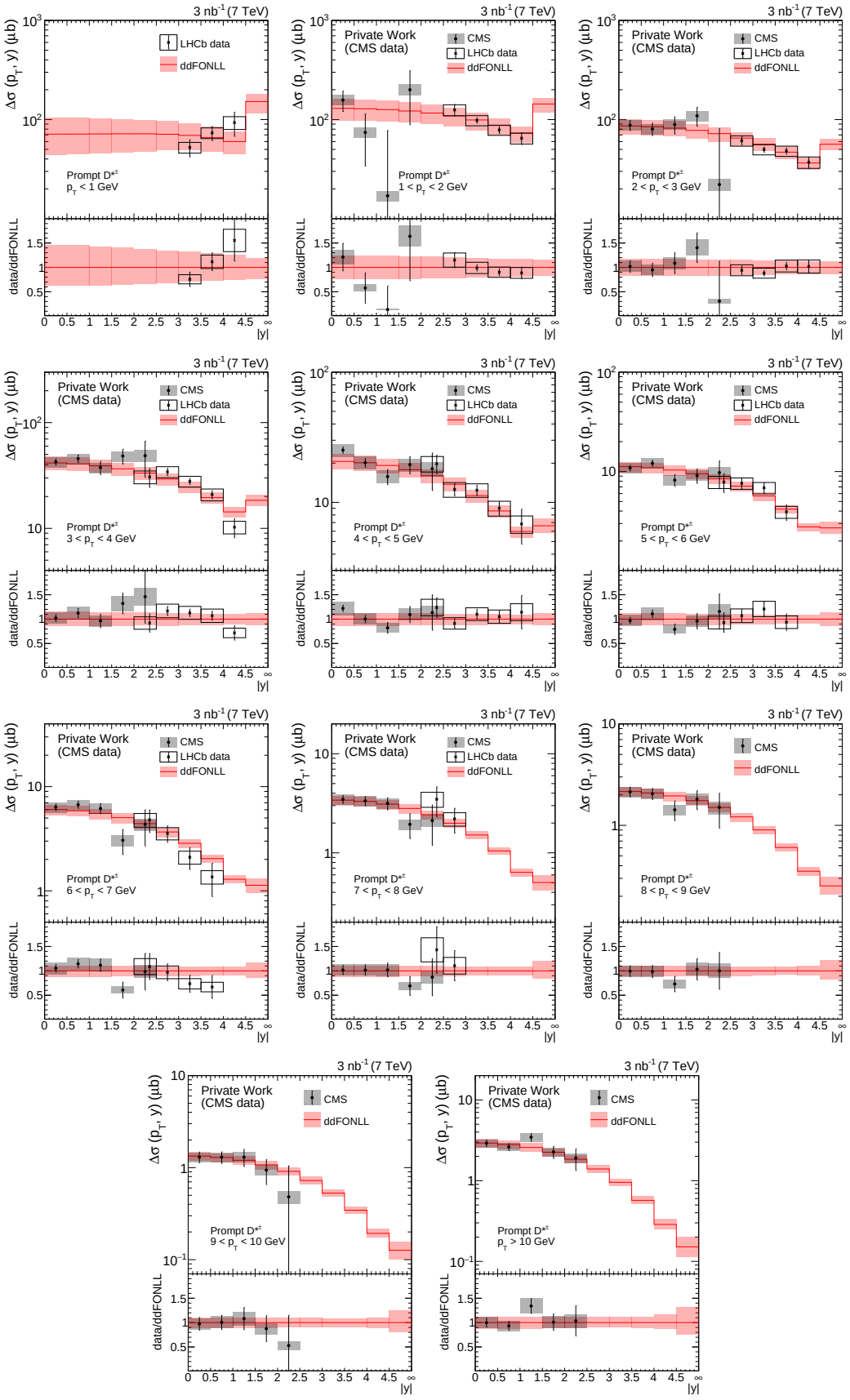


Figure H.17: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

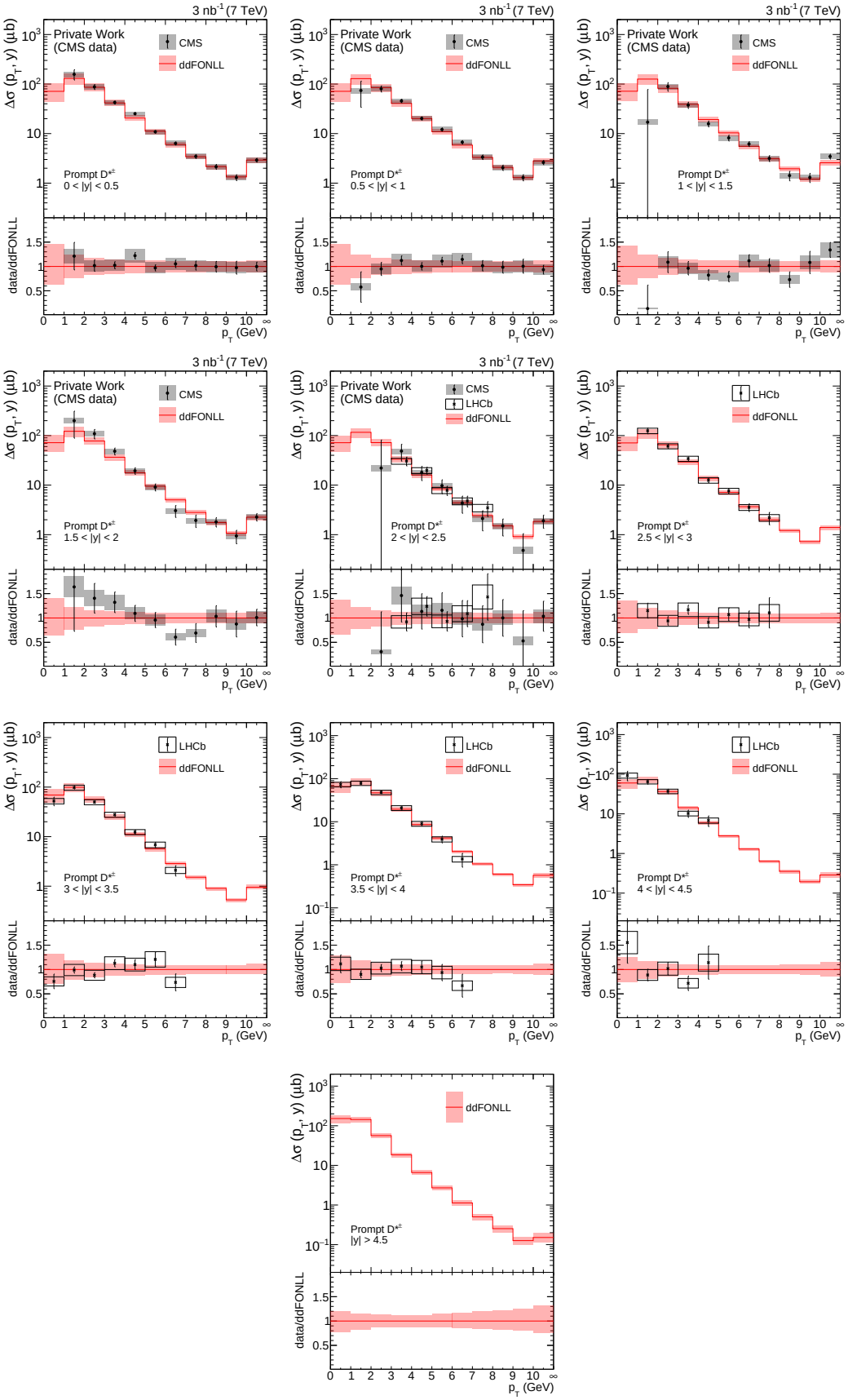


Figure H.18: $D^{*\pm}$ cross sections at $\sqrt{s} = 7$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

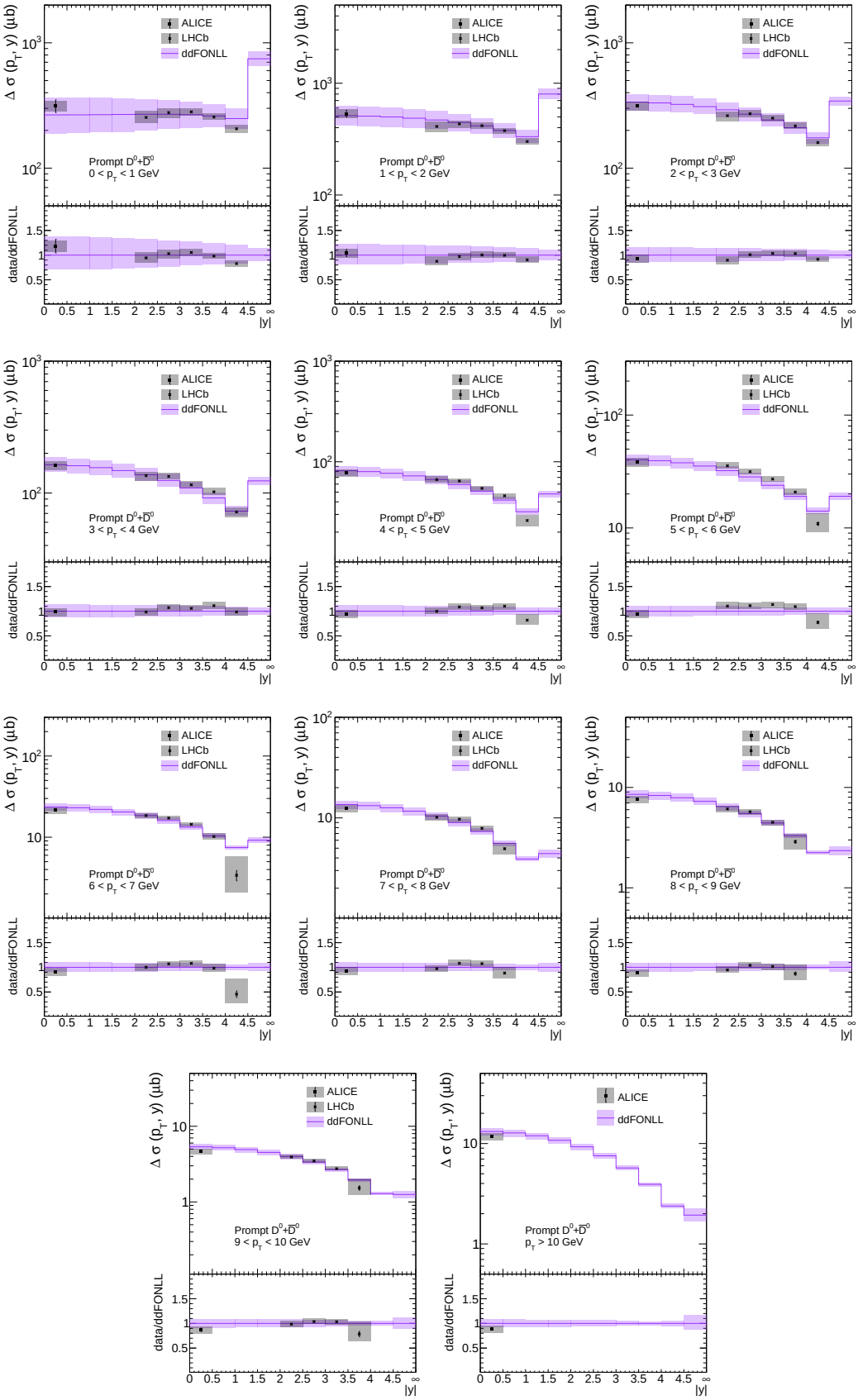


Figure H.19: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

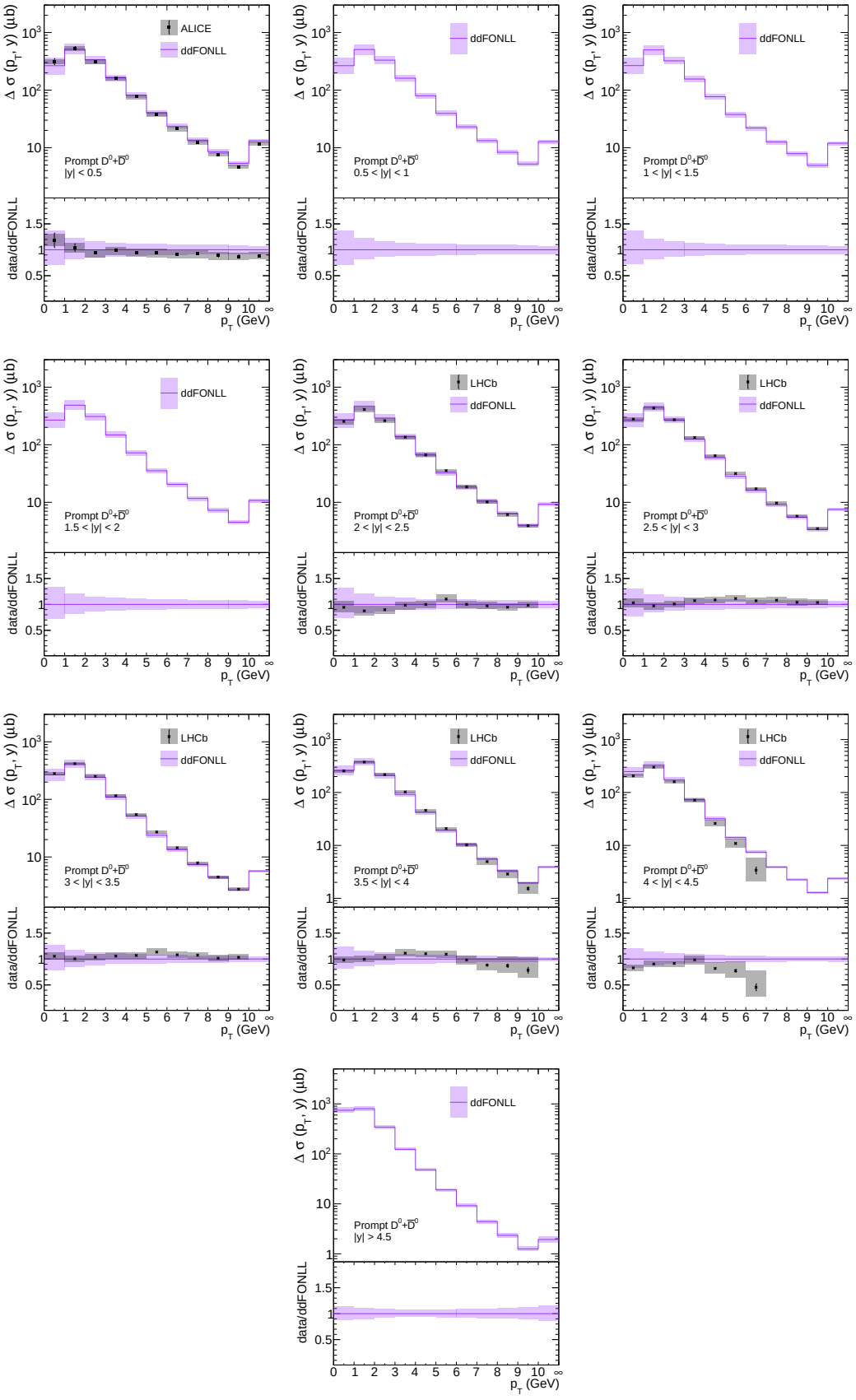


Figure H.20: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

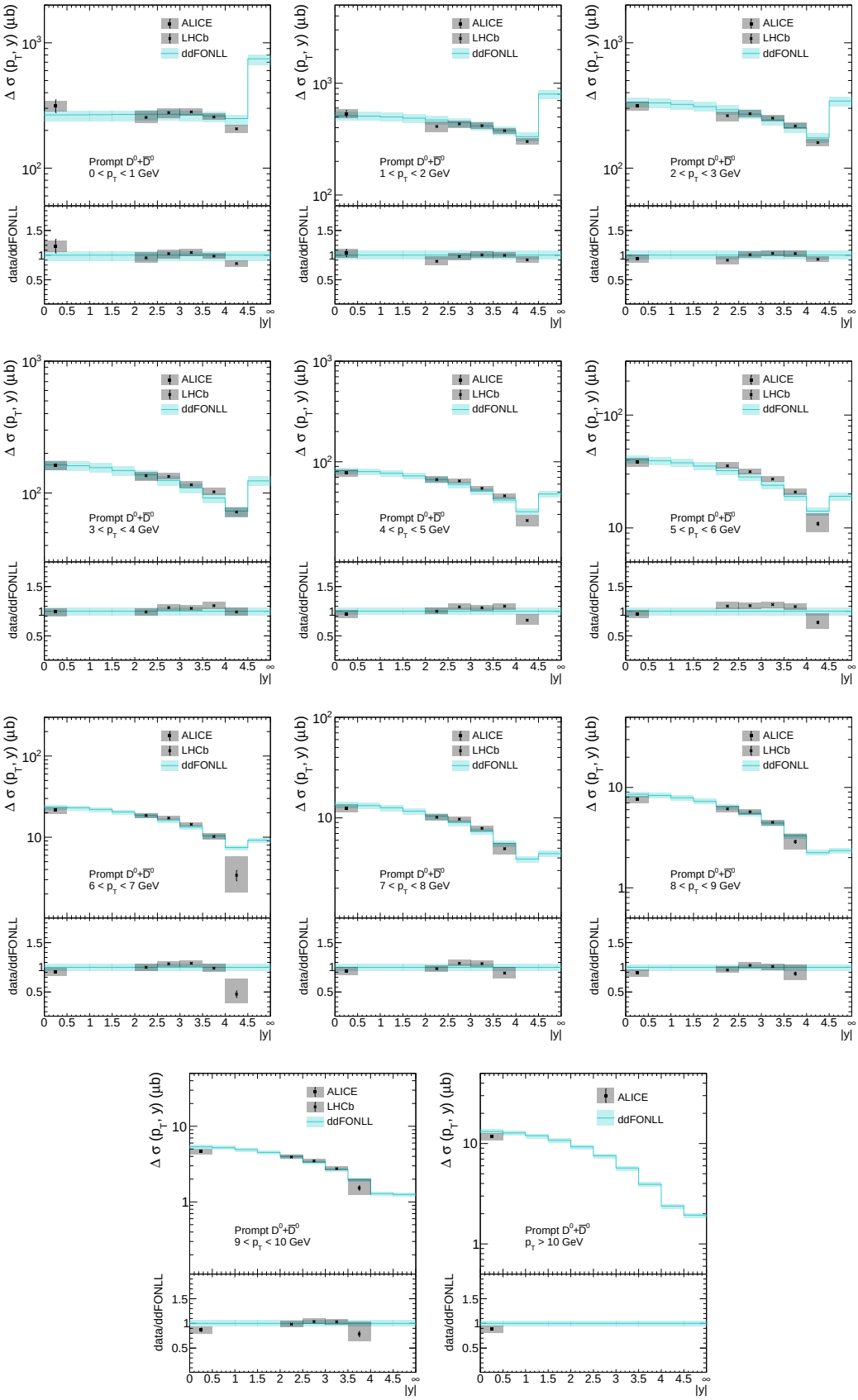


Figure H.21: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

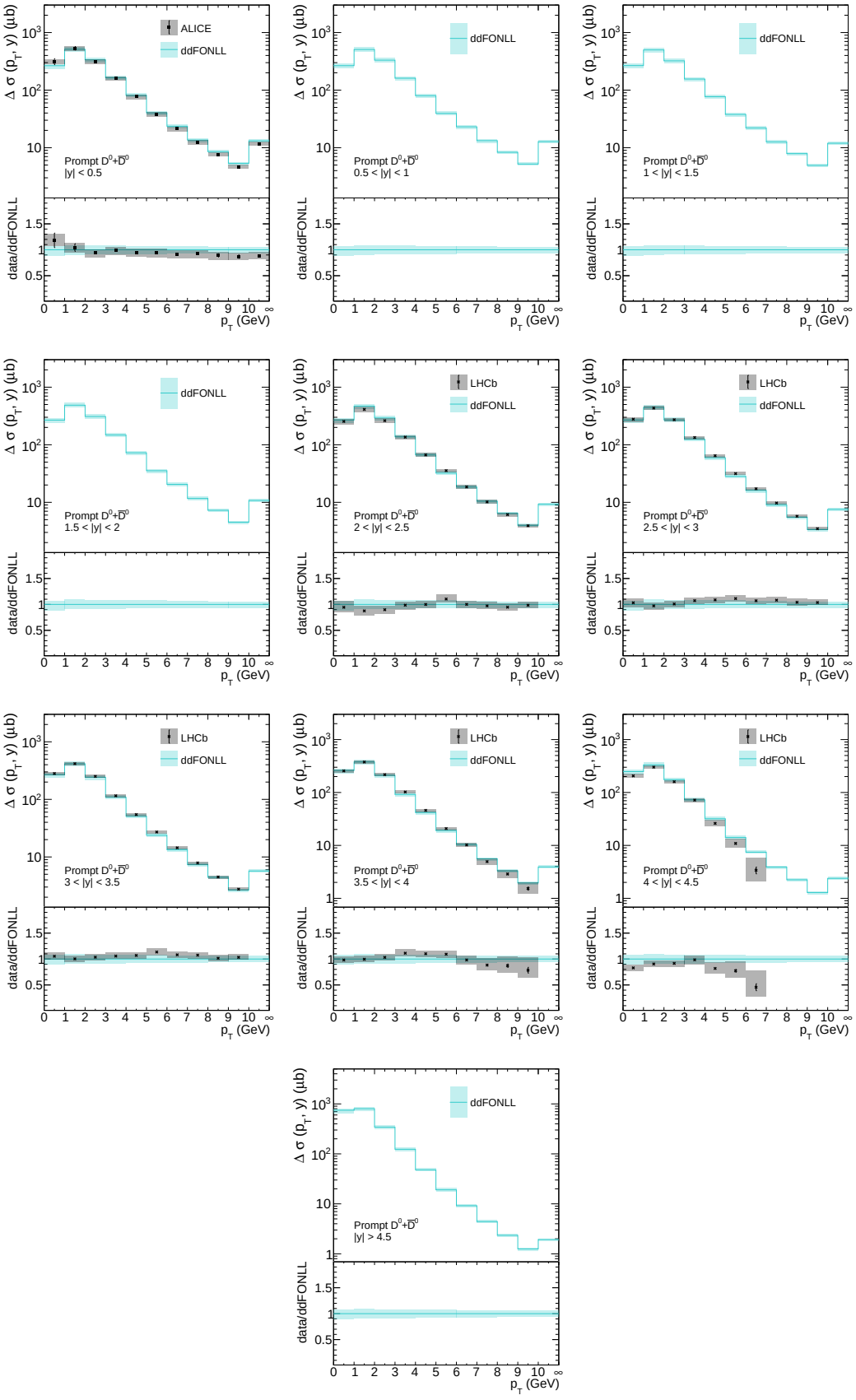


Figure H.22: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

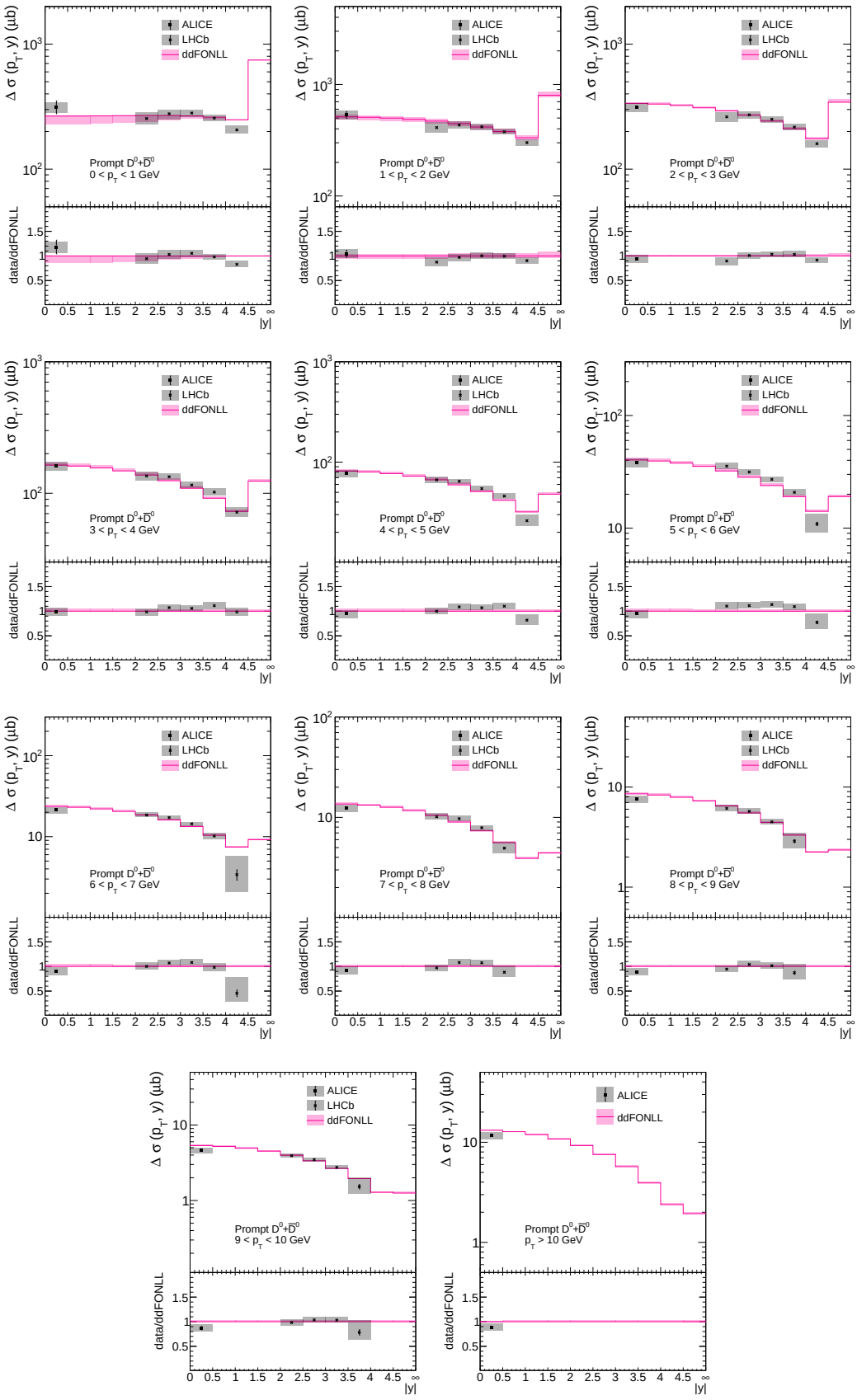


Figure H.23: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

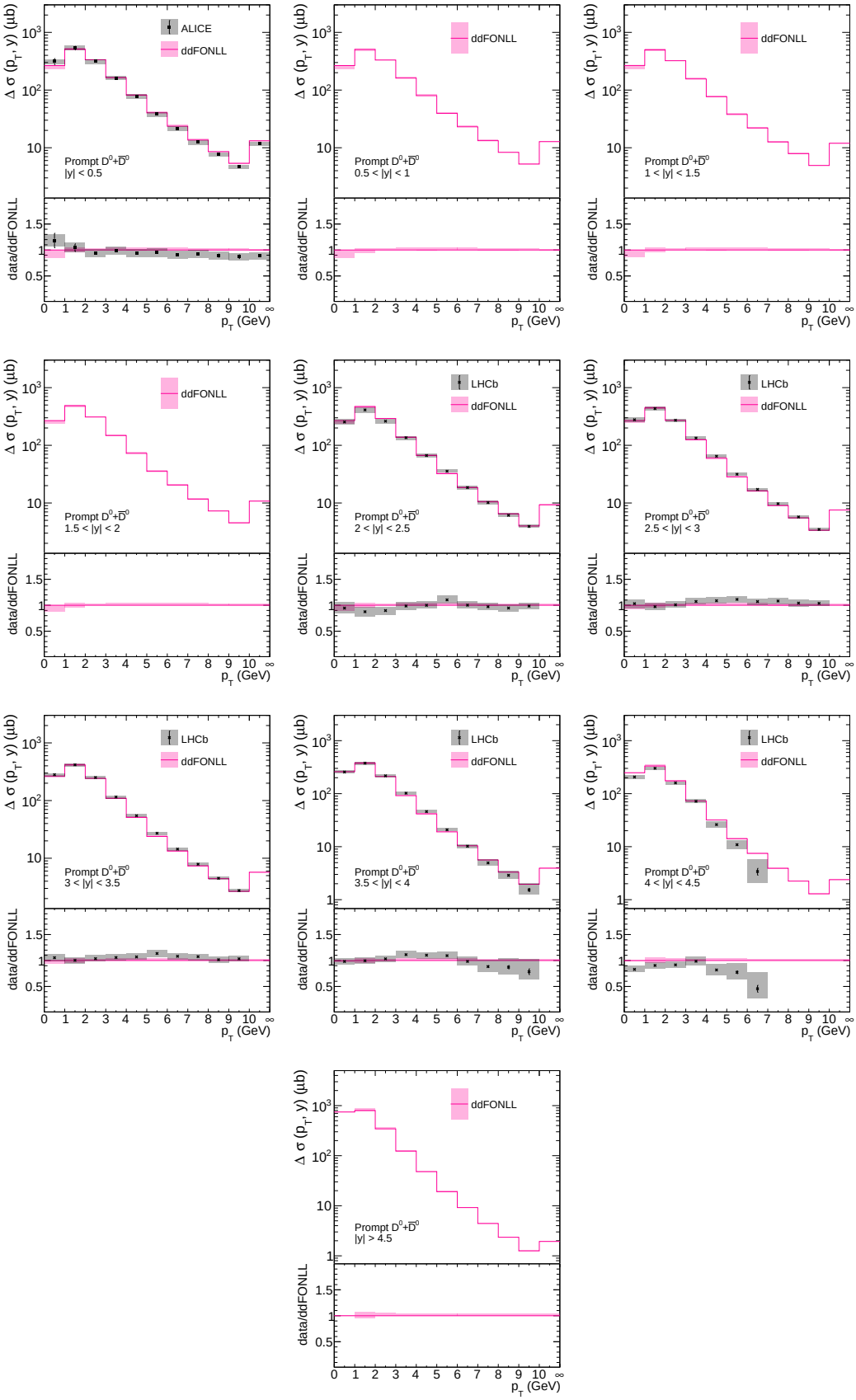


Figure H.24: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

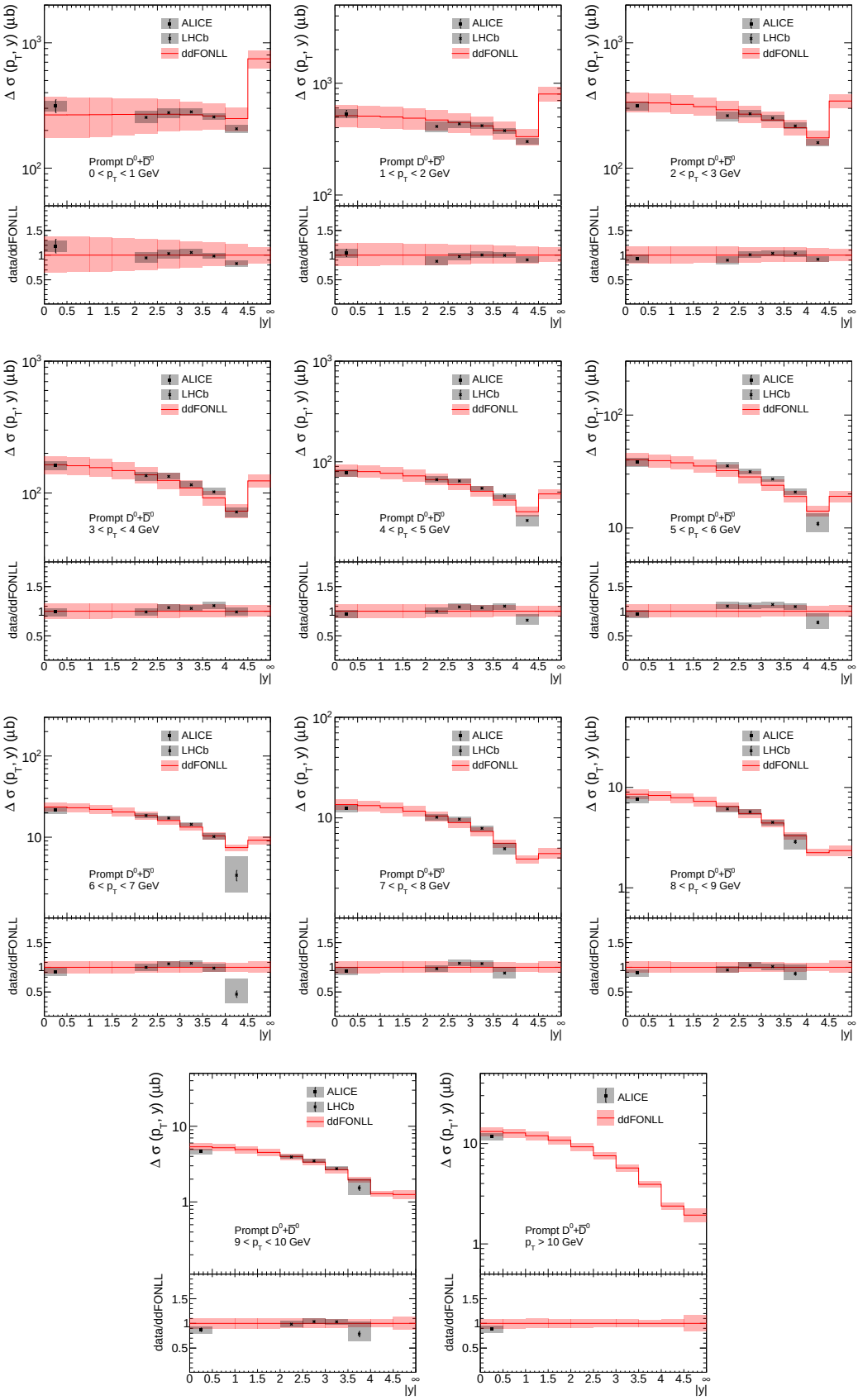


Figure H.25: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

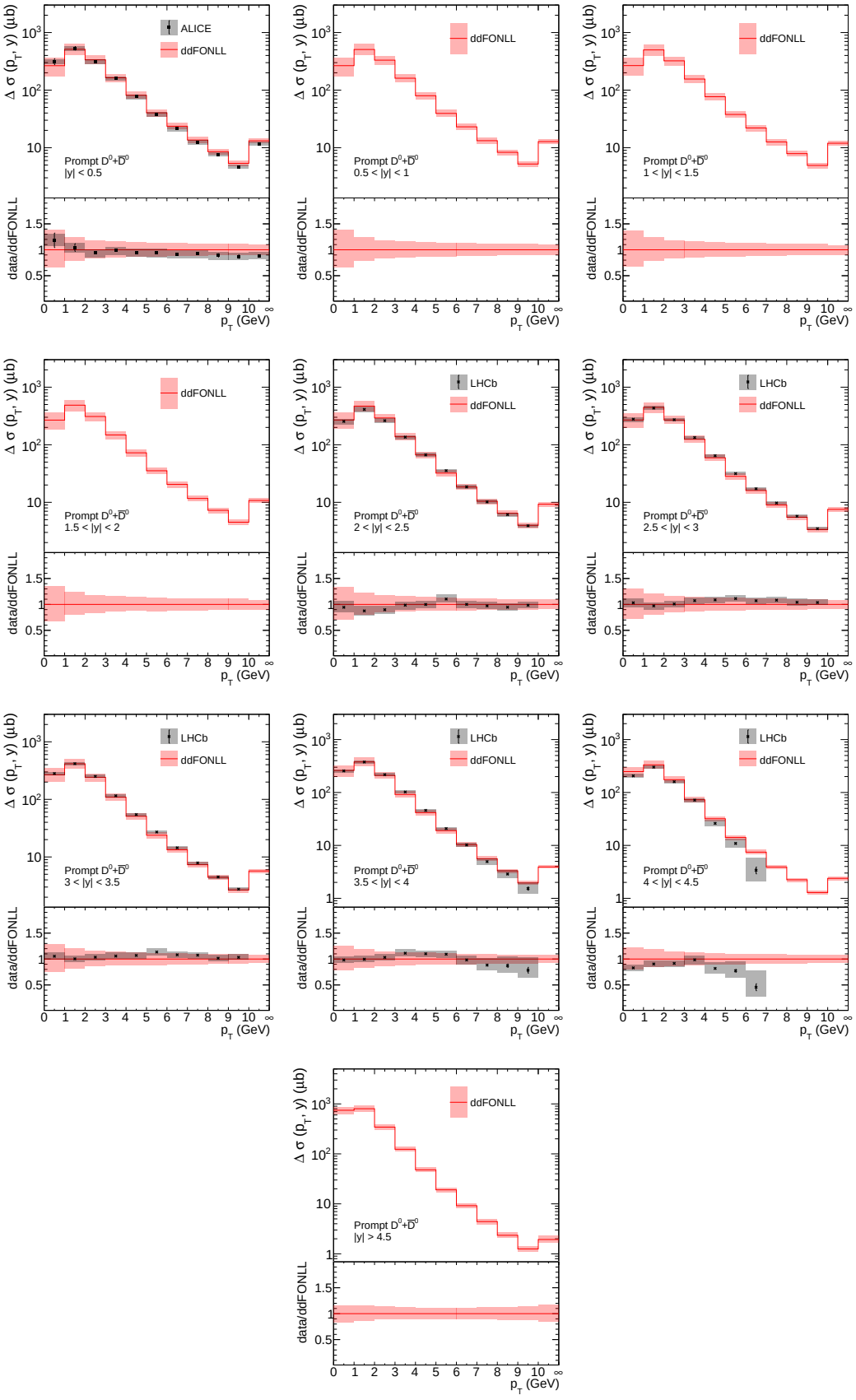


Figure H.26: $D^0 + \overline{D}^0$ cross sections at $\sqrt{s} = 13$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

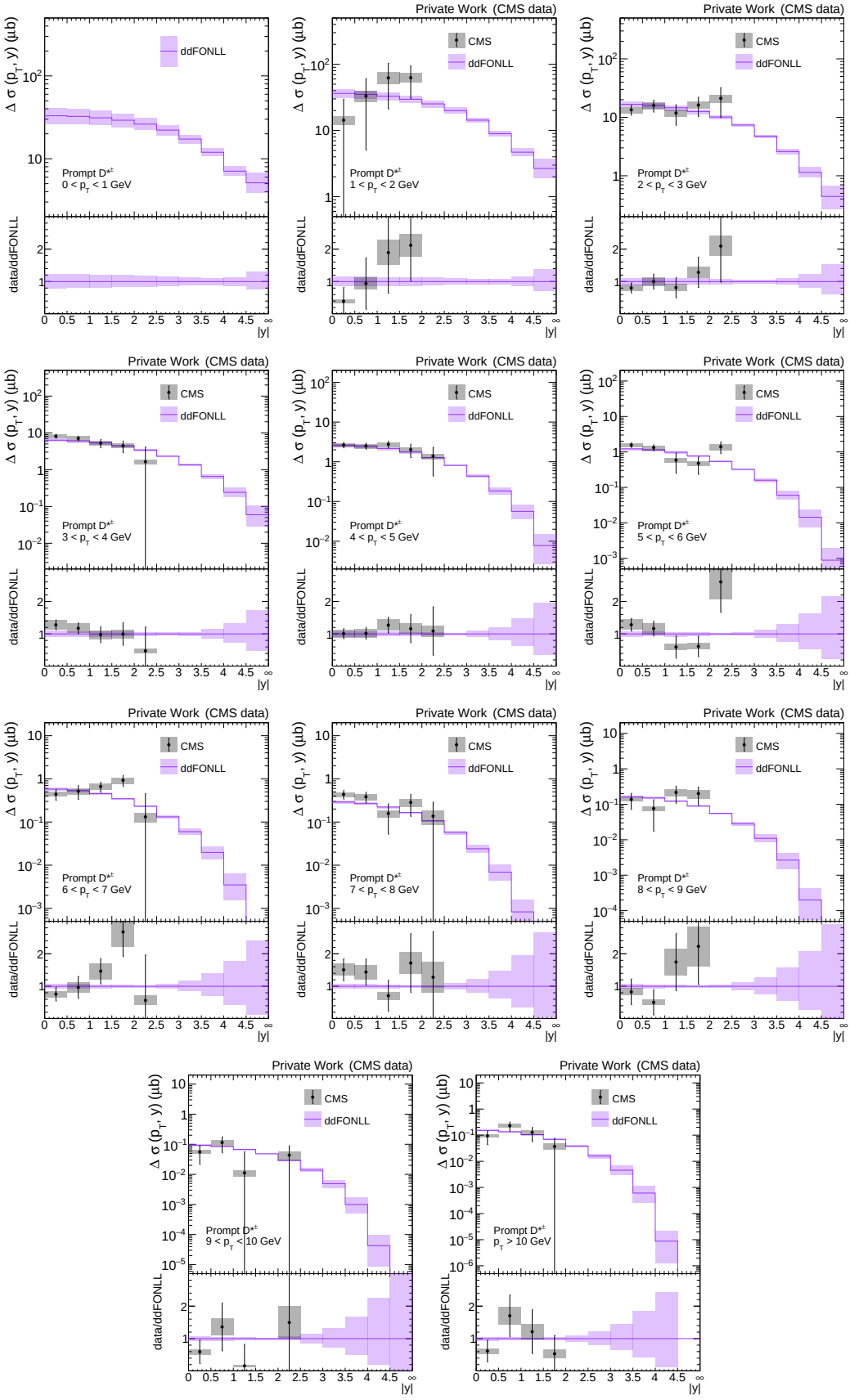


Figure H.27: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

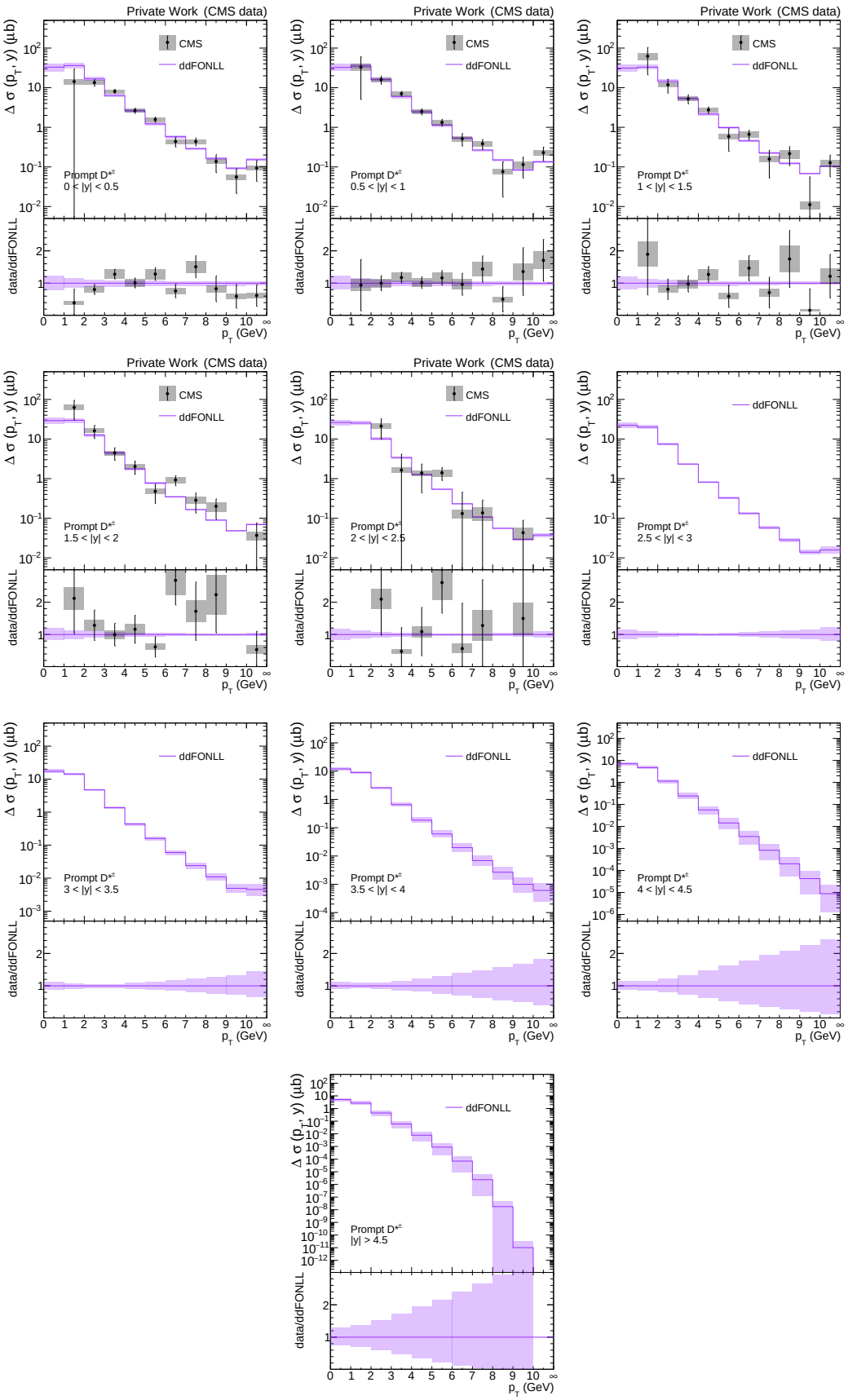


Figure H.28: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with PDF uncertainty.

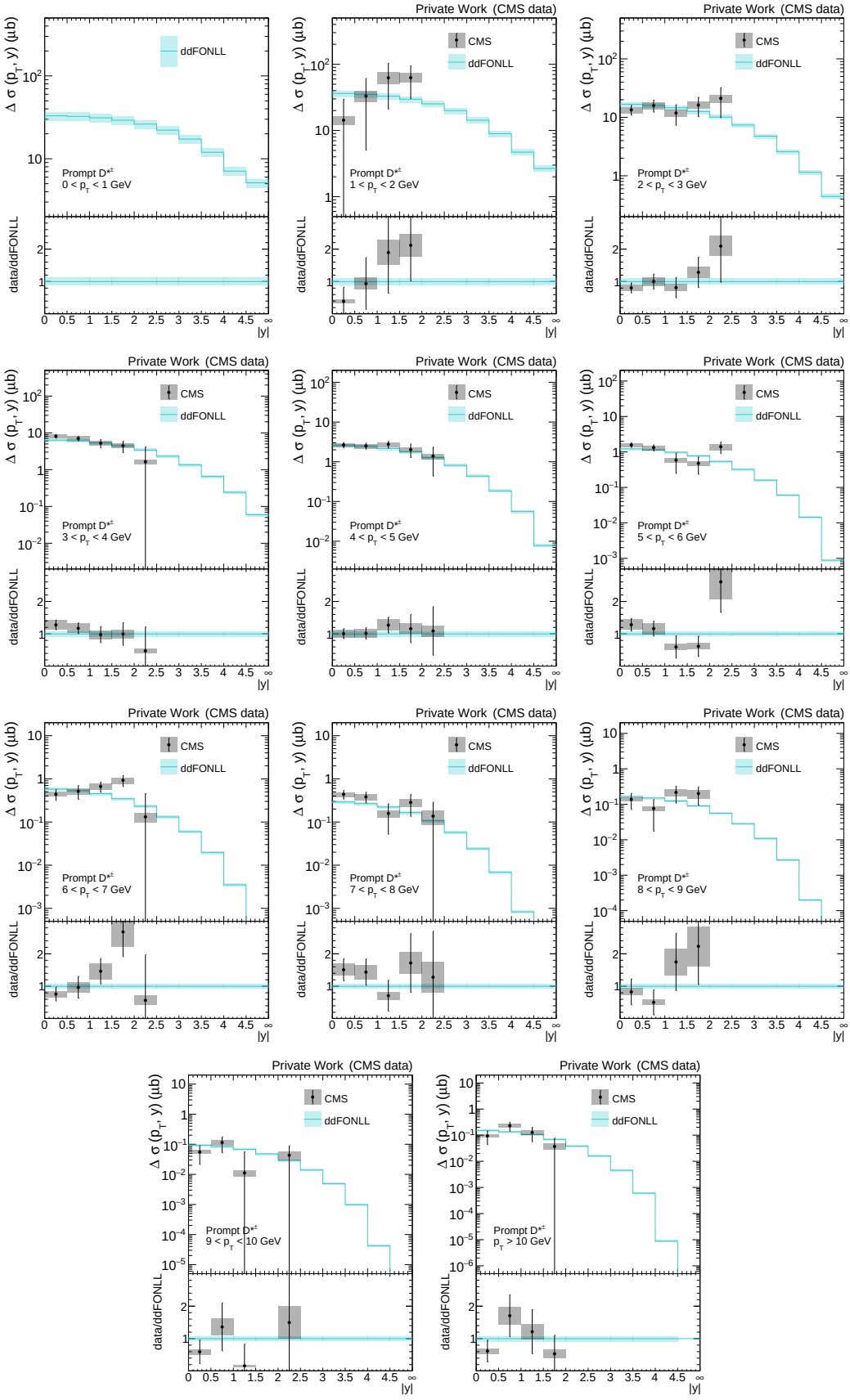


Figure H.29: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

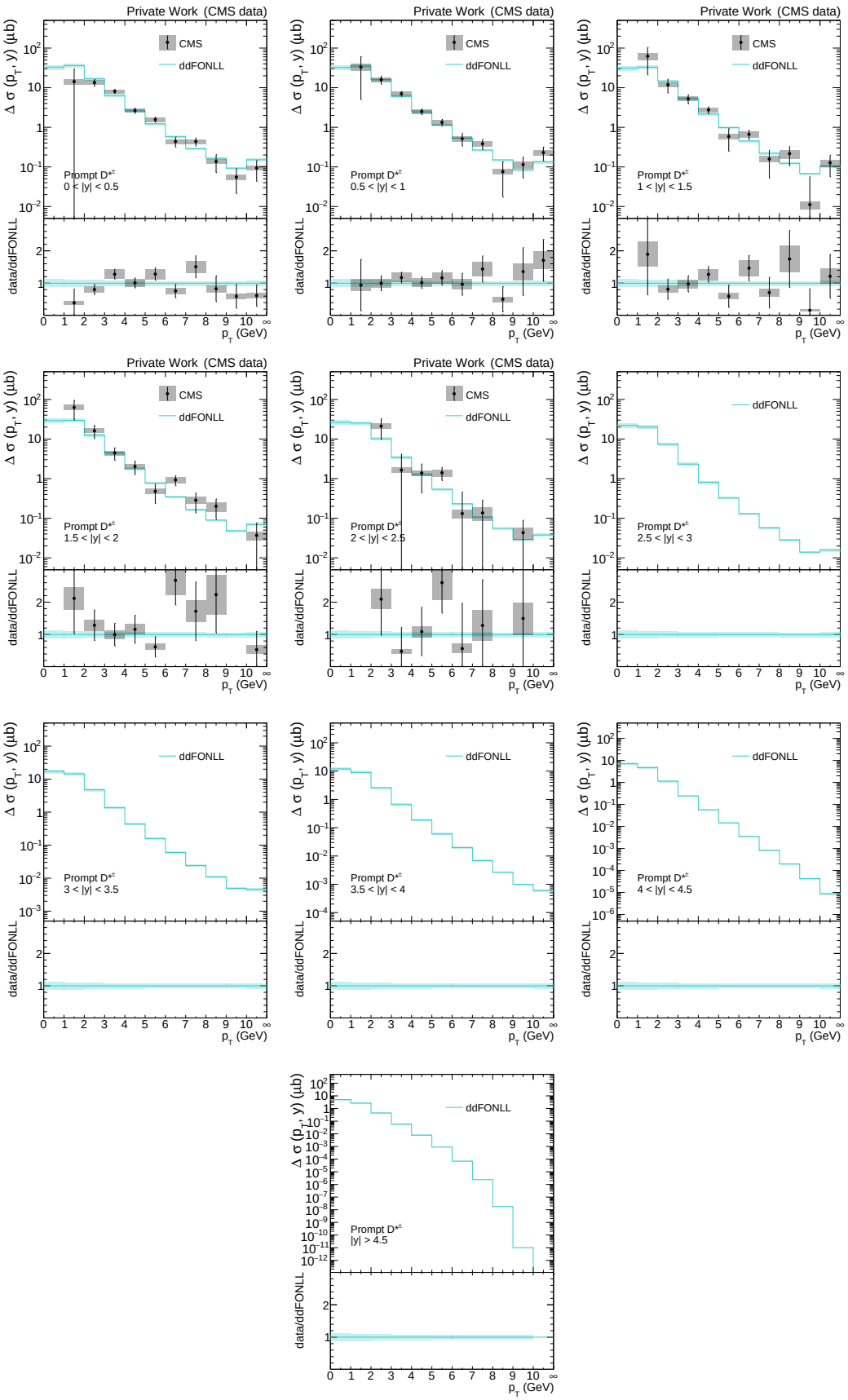


Figure H.30: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with $\tilde{f}$ uncertainty.

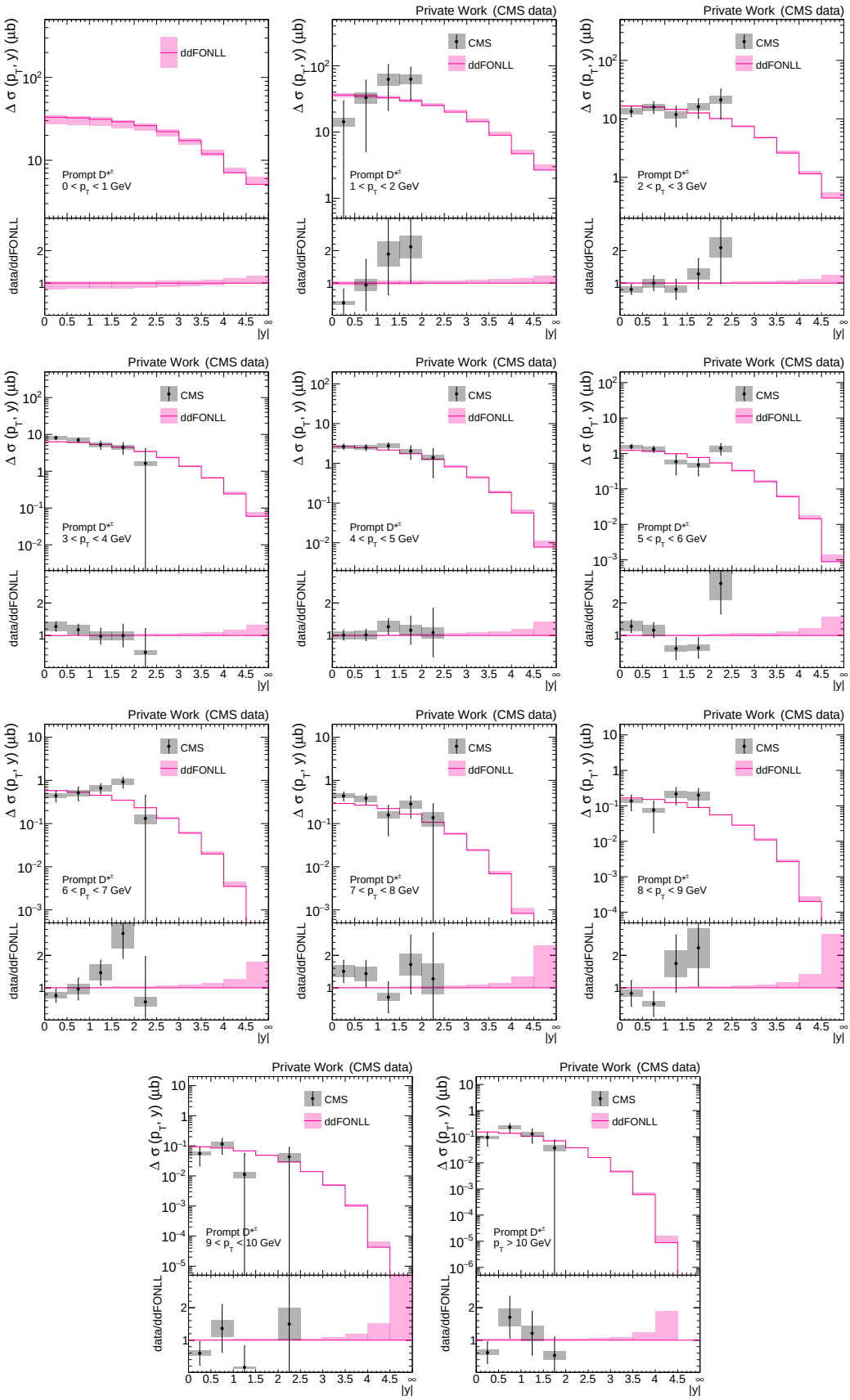


Figure H.31: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

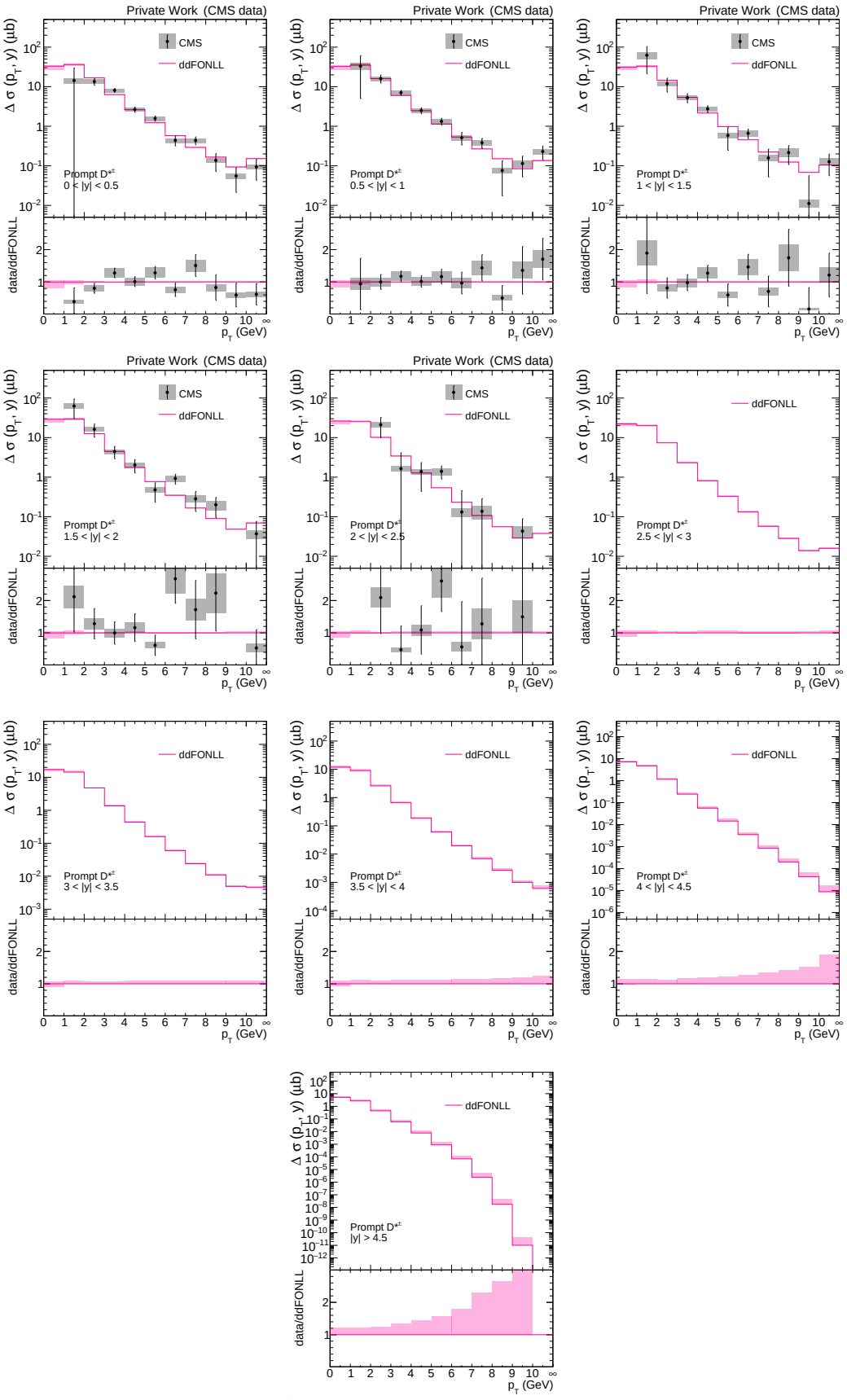


Figure H.32: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with x_f, x_r, α_K and m_c uncertainty.

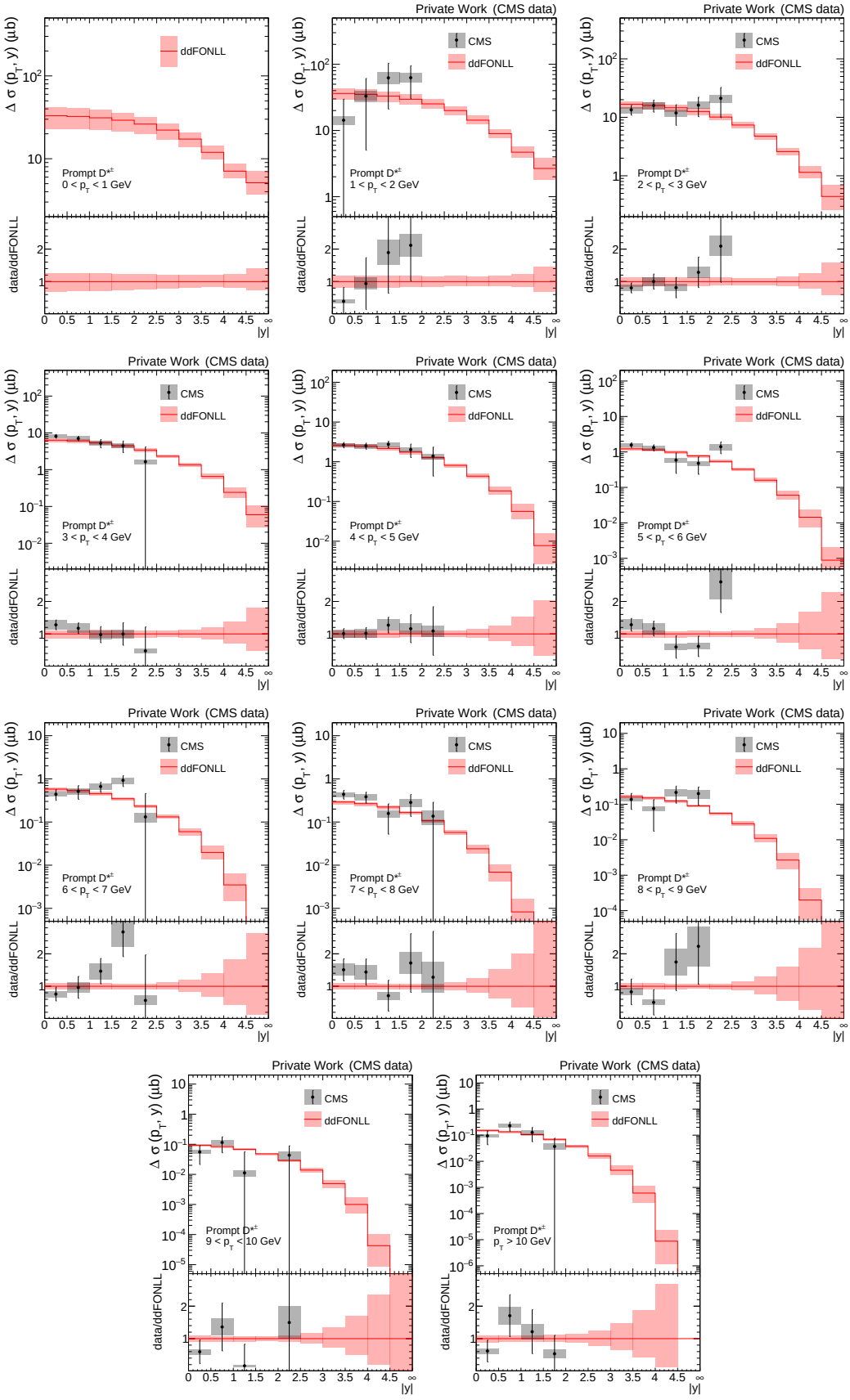


Figure H.33: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of $|y|$. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

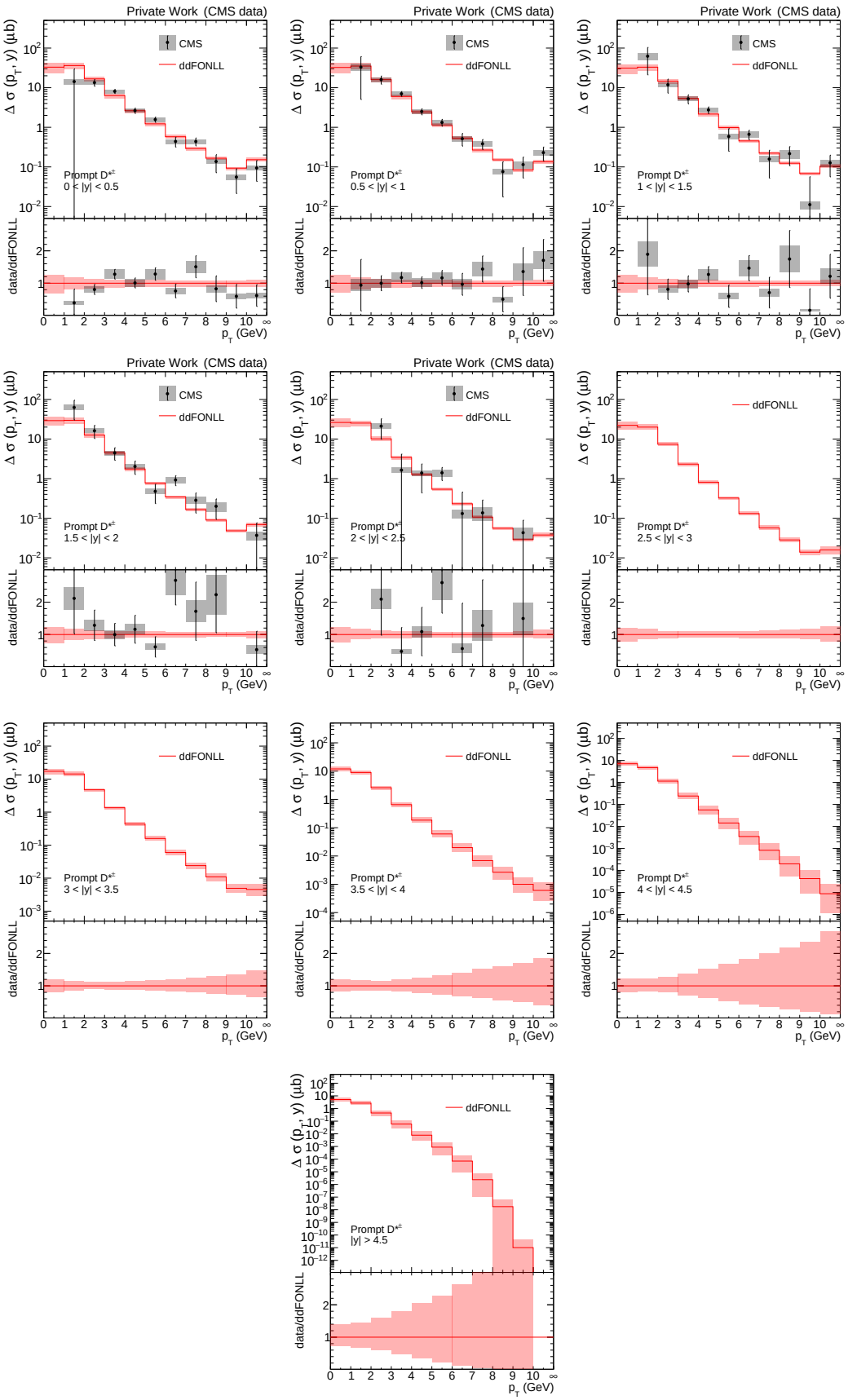


Figure H.34: $D^{*\pm}$ cross sections at $\sqrt{s} = 0.9$ TeV as a function of p_T. The ddFONLL was derived with the non-universal charm fragmentation and is shown with total uncertainty.

Propagation of Uncertainties for $\tilde{f}$

In this thesis, for a function f of n variables $x_1, x_2, \cdots, x_n$, uncertainties of the variables were propagated to uncertainty of f by following a general form:

$$\Sigma^f = \mathbf{J}\Sigma^{\mathbf{x}}\mathbf{J}^{\mathrm{T}} \tag{I.1}$$

where Σ^f and $\Sigma^{\mathbf{x}}$ are the variance–covariance matrix of f and the variables, respectively, and $\mathbf{J}$ is the Jacobian matrix. Here the variance-covariance matrix is composed of $\sigma_{x_i}\sigma_{x_j}\rho_{x_ix_j}$ where $\sigma_{x_i}(\sigma_{x_j})$ is the standard deviation of $x_i(x_j)$, and $\rho_{x_ix_j}$ is the correlation between x_i and x_j. With σ_{x_i} given to be the uncertainty of x_i, the square root of Σ^f gives the uncertainty of f.

For example, for a function

$$f(x_1, x_2) = x_1 + x_2, \tag{I.2}$$

the variance-covariance matrix of the variables is

$$\Sigma^{\mathbf{x}} = \begin{pmatrix} \sigma_{x_1}\sigma_{x_1}\rho_{x_1x_1} & \sigma_{x_1}\sigma_{x_2}\rho_{x_1x_2} \\ \sigma_{x_2}\sigma_{x_1}\rho_{x_2x_1} & \sigma_{x_2}\sigma_{x_2}\rho_{x_2x_2} \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix} \tag{I.3}$$

assuming x_1 and x_2 are fully uncorrelated ($\rho_{x_1x_2} = 0$). The Jacobian matrix of this function is

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix}. \tag{I.4}$$

Therefore, the variance-covariance matrix of f becomes

$$\Sigma^f = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \sigma_{x_1}^2 + \sigma_{x_2}^2. \tag{I.5}$$

Therefore, the uncertainty of f is given by $\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}$.

I.1 $f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}$ uncertainty

To calculate the uncertainty of the fragmentation fractions ($f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}$ defined by Eq.(6.19)) of the unmeasured charm-hadrons in e^+e^- collisions, a function $f \equiv f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}$ is defined by the variables, $f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}$ and $f^{uni}_{\Lambda_c^+}$ as

$$f(f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}, f^{uni}_{\Lambda_c^+}) = 1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+} - f^{uni}_{\Lambda_c^+} \tag{I.6}$$

Then the variance-covariance matrix of the variables is

$$\Sigma^{\mathbf{x}} = \begin{pmatrix} \sigma^2_{f^{uni}_{D^0}} & 0 & 0 & 0 \\ 0 & \sigma^2_{f^{uni}_{D^+}} & 0 & 0 \\ 0 & 0 & \sigma^2_{f^{uni}_{D_s^+}} & 0 \\ 0 & 0 & 0 & \sigma^2_{f^{uni}_{\Lambda_c^+}} \end{pmatrix}, \tag{I.7}$$

and the Jacobian matrix is

$$\mathbf{J} = \left(\frac{\partial f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{\partial f^{uni}_{D^0}} \quad \frac{\partial f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{\partial f^{uni}_{D^+}} \quad \frac{\partial f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{\partial f^{uni}_{D_s^+}} \quad \frac{\partial f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{\partial f^{uni}_{\Lambda_c^+}} \right) = \begin{pmatrix} -1 & -1 & -1 & -1 \end{pmatrix} \tag{I.8}$$

assuming the uncertainties of the variables are fully uncorrelated. Then the variance-covariance matrix of f becomes

$$\Sigma^f = \sigma^2_{f^{uni}_{D^0}} + \sigma^2_{f^{uni}_{D^+}} + \sigma^2_{f^{uni}_{D_s^+}} + \sigma^2_{f^{uni}_{\Lambda_c^+}} = 0.016^2 + 0.009^2 + 0.005^2 + 0.004^2 = 0.000378. \tag{I.9}$$

A resulting uncertainty of $f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}$ is $\sqrt{0.000378} = 0.019$.

I.2 $\frac{f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{f^{uni}_{\Lambda_c^+}}$ uncertainty

The uncertainites in Figure 6.3 were derived assuming all the measurement uncertainties are fully uncorrelated. Specifically, the uncertainties of the ratios $f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}$ to $f^{uni}_{\Lambda_c^+}$ of e^+e^- data were calculated by defining a function $f \equiv \frac{f^{uni}_{\Xi_c^0+\Xi_c^++\Omega_c^0}}{f^{uni}_{\Lambda_c^+}}$ with the variables, $f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}$ and $f^{uni}_{\Lambda_c^+}$ as

$$f(f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}, f^{uni}_{\Lambda_c^+}) = \frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+} - f^{uni}_{\Lambda_c^+}}{f^{uni}_{\Lambda_c^+}}. \tag{I.10}$$

The partial derivatives for the Jacobian matrix are

$$\frac{\partial f}{\partial f^{uni}_{D^0}} = \frac{\partial f}{\partial f^{uni}_{D^+}} = \frac{\partial f}{\partial f^{uni}_{D_s^+}} = -\frac{1}{f^{uni}_{\Lambda_c^+}} \tag{I.11}$$

and

$$\frac{\partial f}{\partial f^{uni}_{\Lambda_c^+}} = -\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni,2}_{\Lambda_c^+}}. \tag{I.12}$$

Then, with $\Sigma^{\mathbf{x}}$ defined by Eq.(I.7), the variance-covariance matrix of f is

$$\Sigma^f = \left(\frac{f^{uni}_{BY}}{f^{uni}_{\Lambda_c^+}} \right)^2 \left\{ \left(\frac{\sigma_{f^{uni}_{BY}}}{f^{uni}_{BY}} \right)^2 + \left(\frac{\sigma_{f^{uni}_{\Lambda_c^+}}}{f^{uni}_{\Lambda_c^+}} \right)^2 \right\} \tag{I.13}$$

where

$$f^{uni}_{BY} = 1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}, \quad \sigma_{f^{uni}_{BY}} = \sqrt{\sigma^2_{f^{uni}_{D^0}} + \sigma^2_{f^{uni}_{D^+}} + \sigma^2_{f^{uni}_{D_s^+}}}. \tag{I.14}$$

I.3 $\tilde{f}_{D^0}$ uncertainty

The $\tilde{f}_{D^0}$ of Eq.(6.17) can be rewritten as

$$\tilde{f}_{D^0} = \left\{ \frac{f^{uni}_{\mathrm{MS}}}{f^{uni}_{D^0}} + \frac{f^{uni}_{\mathrm{BY}}}{f^{uni}_{\Lambda_c^+}} R \right\}^{-1}. \tag{I.15}$$

Therefore a function $f \equiv \tilde{f}_{D^0}$ is defined with the variables $R, f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}$ and $f^{uni}_{\Lambda_c^+}$ as

$$f(R, f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}, f^{uni}_{\Lambda_c^+}) = \left\{ 1 + \frac{f^{uni}_{D^+} + f^{uni}_{D_s^+}}{f^{uni}_{D^0}} + \left(\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni}_{\Lambda_c^+}} \right) R \right\}^{-1}. \tag{I.16}$$

With assuming that R, $f_{D^0}^{uni}$, $f_{D^+}^{uni}$, $f_{D_s^+}^{uni}$ and $f_{\Lambda_c^+}^{uni}$ are fully uncorrelated, the variance-covariance matrix is

$$\Sigma^{\mathbf{x}} = \begin{pmatrix} \sigma_R^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{f_{D^0}^{uni}} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{f_{D^+}^{uni}} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{f_{D_s^+}^{uni}} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{f_{\Lambda_c^+}^{uni}} \end{pmatrix}. \tag{I.17}$$

The partial derivatives of the Jacobian matrix are

$$\frac{\partial f}{\partial R} = -\tilde{f}_{D^0}^2 \left(\frac{1 - f_{D^0}^{uni} - f_{D^+}^{uni} - f_{D_s^+}^{uni}}{f_{\Lambda_c^+}^{uni}} \right), \tag{I.18}$$

$$\frac{\partial f}{\partial f_{D^0}^{uni}} = +\tilde{f}_{D^0}^2 \left(\frac{f_{D^+}^{uni} + f_{D_s^+}^{uni}}{f_{D^0}^{uni,2}} + \frac{R}{f_{\Lambda_c^+}^{uni}} \right), \tag{I.19}$$

$$\frac{\partial f}{\partial f_{D^+}^{uni}} = -\tilde{f}_{D^0}^2 \left(\frac{1}{f_{D^0}^{uni}} - \frac{R}{f_{\Lambda_c^+}^{uni}} \right), \tag{I.20}$$

$$\frac{\partial f}{\partial f_{D_s^+}^{uni}} = -\tilde{f}_{D^0}^2 \left(\frac{1}{f_{D^0}^{uni}} - \frac{R}{f_{\Lambda_c^+}^{uni}} \right), \tag{I.21}$$

and

$$\frac{\partial f}{\partial f_{\Lambda_c^+}^{uni}} = +\tilde{f}_{D^0}^2 \left(\frac{1 - f_{D^0}^{uni} - f_{D^+}^{uni} - f_{D_s^+}^{uni}}{f_{\Lambda_c^+}^{uni,2}} \right) R. \tag{I.22}$$

Similarly as shown above, the uncertainties of $\tilde{f}_{D^0}$ were calculated based on these components but as a function of p_T this time (R is a function of p_T) and results are shown in Figure 6.5.

I.4 $\tilde{f}_{\Lambda_c^+}$ uncertainty

The $\tilde{f}_{\Lambda_c^+}$ of Eq.(6.18) can be rewritten as

$$\tilde{f}_{\Lambda_c^+} = \left\{ \frac{f_{\mathrm{MS}}^{uni}}{f_{D^0}^{uni}} \frac{1}{R} + \frac{f_{\mathrm{BY}}^{uni}}{f_{\Lambda_c^+}^{uni}} \right\}^{-1} \tag{I.23}$$

and a function $f \equiv \tilde{f}_{\Lambda_c^+}$ is defined with the same variables as the ones of the $\tilde{f}_{D^0}$:

$$f(R, f_{D^0}^{uni}, f_{D^+}^{uni}, f_{D_s^+}^{uni}, f_{\Lambda_c^+}^{uni}) = \left\{ \left(1 + \frac{f_{D^+}^{uni} + f_{D_s^+}^{uni}}{f_{D^0}^{uni}} \right) \frac{1}{R} + \frac{1 - f_{D^0}^{uni} - f_{D^+}^{uni} - f_{D_s^+}^{uni}}{f_{\Lambda_c^+}^{uni}} \right\}^{-1}. \tag{I.24}$$

Thus, assuming the uncertainties of the variables to be again fully uncorrelated, the same $\Sigma^{\mathbf{x}}$ of Eq.(I.17) is applied also for the $\tilde{f}_{\Lambda_c^+}$. The partial derivatives of the Jacobian matrix for the $\tilde{f}_{\Lambda_c^+}$ are

$$\frac{\partial f}{\partial R} = +\tilde{f}_{\Lambda_c^+}^2 \left(1 + \frac{f_{D^+}^{uni} + f_{D_s^+}^{uni}}{f_{\Lambda_c^+}^{uni}} \right) \frac{1}{R^2}, \tag{I.25}$$

$$\frac{\partial f}{\partial f_{D^0}^{uni}} = +\tilde{f}_{\Lambda_c^+}^2 \left\{ \left(\frac{f_{D^+}^{uni} + f_{D_s^+}^{uni}}{f_{D^0}^{uni,2}} \right) \frac{1}{R} + \frac{1}{f_{\Lambda_c^+}^{uni}} \right\}, \tag{I.26}$$

$$\frac{\partial f}{\partial f^{uni}_{D^+}} = -\tilde{f}^2_{\Lambda_c^+}\left(\frac{1}{f^{uni}_{D^0}}\frac{1}{R} - \frac{1}{f^{uni}_{\Lambda_c^+}}\right), \tag{I.27}$$

$$\frac{\partial f}{\partial f^{uni}_{D_s^+}} = -\tilde{f}^2_{\Lambda_c^+}\left(\frac{1}{f^{uni}_{D^0}}\frac{1}{R} - \frac{1}{f^{uni}_{\Lambda_c^+}}\right), \tag{I.28}$$

and

$$\frac{\partial f}{\partial f^{uni}_{\Lambda_c^+}} = +\tilde{f}^2_{\Lambda_c^+}\left(\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni,2}_{\Lambda_c^+}}\right). \tag{I.29}$$

The resulting uncertainties are shown again in Figure 6.5.

I.5 $\tilde{f}_{D^{*+}}$ uncertainty

The $\tilde{f}_{D^{*+}}$ of Eq.(6.20) can be rewritten as

$$\tilde{f}_{D^{*+}} = \left\{ \frac{f^{uni}_{\overline{\text{MS}}}}{f^{uni}_{D^{*+}}} + \frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}}}\frac{f^{uni}_{\text{BY}}}{f^{uni}_{\Lambda_c^+}}R \right\}^{-1}. \tag{I.30}$$

This has one more variable ($f^{uni}_{D^{*+}}$) compared to the ones of the previous two cases, and a function $f \equiv \tilde{f}_{D^{*+}}$ is defined to be

$$f(R, f^{uni}_{D^0}, f^{uni}_{D^+}, f^{uni}_{D_s^+}, f^{uni}_{\Lambda_c^+}, f^{uni}_{D^{*+}}) = \left\{ \frac{f^{uni}_{D^0} + f^{uni}_{D^+} + f^{uni}_{D_s^+}}{f^{uni}_{D^{*+}}} + \frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}}}\left(\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni}_{\Lambda_c^+}}\right)R \right\}^{-1}. \tag{I.31}$$

The variance-covariance matrix of the variables is then

$$\Sigma^{\mathbf{x}} = \begin{pmatrix} \sigma^2_R & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{f^{uni}_{D^0}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{f^{uni}_{D^+}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{f^{uni}_{D_s^+}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{f^{uni}_{\Lambda_c^+}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{f^{uni}_{D^{*+}}} \end{pmatrix} \tag{I.32}$$

assuming the uncertainties of the variables are fully uncorrelated. The partial derivatives of the Jacobian matrix are

$$\frac{\partial f}{\partial R} = -\tilde{f}^2_{D^{*+}}\left\{ \frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}}}\left(\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni}_{\Lambda_c^+}}\right)\right\}, \tag{I.33}$$

$$\frac{\partial f}{\partial f^{uni}_{D^0}} = -\tilde{f}^2_{D^{*+}}\left\{ \left(\frac{1 - 2f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni}_{D^{*+}} f^{uni}_{\Lambda_c^+}}\right)R + \frac{1}{f^{uni}_{D^{*+}}}\right\}, \tag{I.34}$$

$$\frac{\partial f}{\partial f^{uni}_{D^+}} = +\tilde{f}^2_{D^{*+}}\left\{ \left(\frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}} f^{uni}_{\Lambda_c^+}}\right)R - \frac{1}{f^{uni}_{D^{*+}}}\right\}, \tag{I.35}$$

$$\frac{\partial f}{\partial f^{uni}_{D_s^+}} = +\tilde{f}^2_{D^{*+}}\left\{ \left(\frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}} f^{uni}_{\Lambda_c^+}}\right)R - \frac{1}{f^{uni}_{D^{*+}}}\right\}, \tag{I.36}$$

$$\frac{\partial f}{\partial f^{uni}_{\Lambda_c^+}} = +\tilde{f}^2_{D^{*+}}\left\{ \frac{f^{uni}_{D^0}}{f^{uni}_{D^{*+}}}\left(\frac{1 - f^{uni}_{D^0} - f^{uni}_{D^+} - f^{uni}_{D_s^+}}{f^{uni,2}_{\Lambda_c^+}}\right)R\right\}, \tag{I.37}$$

and

$$\frac{\partial f}{\partial f_{D^{*+}}^{uni}} = +\tilde{f}_{D^{*+}}^{2}\left\{\frac{f_{D^0}^{uni}}{f_{D^{*+}}^{uni,2}}\left(\frac{1-f_{D^0}^{uni}-f_{D^+}^{uni}-f_{D_s^+}^{uni}}{f_{\Lambda_c^+}^{uni}}\right)R+\frac{f_{D^0}^{uni}+f_{D^+}^{uni}+f_{D_s^+}^{uni}}{f_{D^{*+}}^{uni,2}}\right\}. \tag{I.38}$$

Again the resulting uncertainties as a function of p_T are shown in Figure 6.5.

J | Reference α

In this thesis, the BCFY functions [46] (having single parameter r for all three D^{*+}, D^0 and D^+) were taken as the reference of the non-perturbative fragmentation spectrum for D^{*+} and D^0 in e^+e^- collisions. However, extrapolation of the D^{*+} and D^0 measurements was introduced by using one of the general functions, Kartvelishvili (having single parameter α_K), instead of the BCFY functions. The best α_K relative r can be found in [46] for D^{*+}. For $r = 0.0988$ (the closest one to $r = 0.1$ used as default for the FONLL predictions in this thesis), $\alpha_K = 9.58$ refering to Table 4 in [46]. Meanwhile, the best α_K is not shown in [46] for D^0. Therefore, I estimated the best α_K relative to r for D^0, by calculating χ^2. The χ^2 was calculated by taking difference of the FONLL predictions between using the Kartvelishvili function and using the BCFY function, while uncertainties were defined to be theoretical uncertainties with the BCFY applied. I checked that this χ^2 calculation gives very close results with the ones provided in [46] for D^{*+}: $\alpha_K = 9.47$ (refer to the left of Figure J.1). Eventually, I used $\alpha_K = 9.5$ and 6.1 for D^{*+} and D^0, respectively, in Figure 6.7 (refer to Figure J.1).

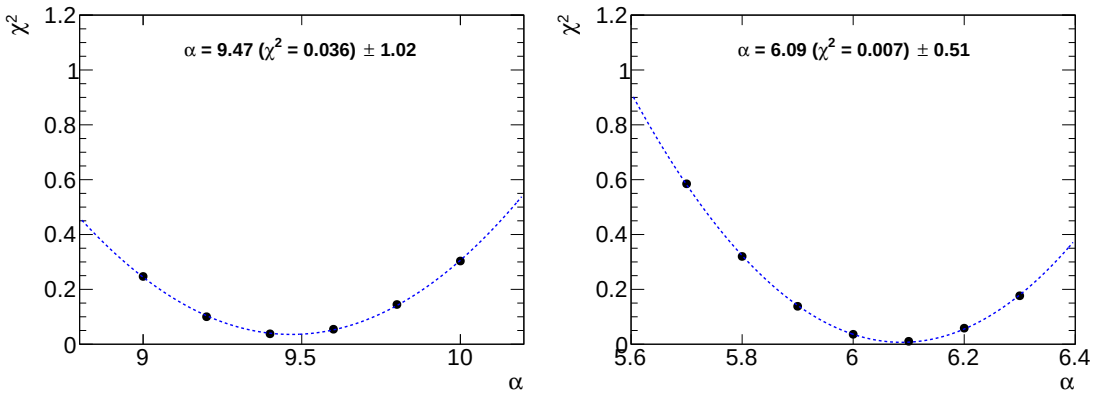


Figure J.1: Estimated α_K (labeled by "α") compared to BCFY D^{*+} (left) and D^0 (right) function with $r = 0.0988$.

K | Renormalized QCD Theory

This appendix is taken and adapted from [50].

To eliminate ultraviolet divergences in the QCD perturbation theory, renormalization should be introduced. Since the renormalization scale μ_r is an arbitrary parameter, physical observable R is independent of the scale and thus the following equation should be satisfied:

$$\mu_r^2 \frac{d}{d\mu_r^2} R(Q^2/\mu_r^2, \alpha_s) \equiv \left(\mu_r^2 \frac{\partial}{\partial \mu_r^2} + \mu_r^2 \frac{\partial \alpha_s}{\partial \mu_r^2} \frac{\partial}{\partial \alpha_s} \right) R = 0. \tag{K.1}$$

Here R is a function of dimensionless parameter, Q^2/μ_r^2, and expanded in α_s:

$$R = R_1 \alpha_s + R_2 \alpha_s^2 + \cdots . \tag{K.2}$$

If new definitions

$$t = \ln\left(\frac{Q^2}{\mu_r^2}\right), \quad \text{and} \quad \mu_r^2 \frac{\partial \alpha_s}{\partial \mu_r^2} \equiv \beta(\alpha_s) \tag{K.3}$$

are introduced, Eq.(K.1) can be rewritten as

$$\left(-\frac{\partial}{\partial t} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s} \right) R(e^t, \alpha_s) = 0. \tag{K.4}$$

Using Eq.(K.3) and the running coupling $\alpha_s(Q^2)$, the following relation is derived:

$$\ln\left(\frac{Q^2}{\mu_r^2}\right) = \int_{\mu_r^2}^{Q^2} \frac{d\mu_r'^2}{\mu_r'^2} = \int_{\alpha_s(\mu_r^2)}^{\alpha_s(Q^2)} \frac{d\alpha_s(\mu_r')}{\beta(\alpha_s(\mu_r'))}. \tag{K.5}$$

The coefficients for the β function are determined from loop corrections to the strong coupling constant and the results up to two-loop are written in [39] as

$$\beta(\alpha_s) = -b_0 \alpha_s^2 - b_1 \alpha_s^3 + \mathcal{O}(\alpha_s^4) \tag{K.6}$$

where

$$b_0 = \frac{33 - 2n_f}{12\pi}, \qquad b_1 = \frac{153 - 19n_f}{24\pi^2}. \tag{K.7}$$

Then the running of the strong coupling constant can be determined using the experimentally well-known value, $\alpha_s(\mu_r^2)$ as an initial value for the differential equation and up to bare correction with b_0 coefficient, solution is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_r^2)}{1 + b_0 \alpha_s(\mu_r^2) \ln\left(\frac{Q^2}{\mu_r^2}\right)}. \tag{K.8}$$

Using b_1 at NLO, the higher order of α_s is also possible.

Returning again to the question about the physical observable R, it should be independent of the scale variation. $R(1, \alpha_s(Q^2))$ is a solution for Eq.(K.4), i.e. if Q is introduced as μ_r then the theory should show no difference between different choices of μ_r. However if the theory

is truncated by only few orders, the scale dependence is remained in the prediction. If $\alpha_s(Q)$ defined by Eq.(K.8) is expanded at $Q \sim \mu_r$, $R(1, \alpha_s(Q^2))$ can be rewritten as

$$\begin{aligned} R(1,\alpha_s(Q^2)) &= R_1\alpha_s(\mu_r^2)\left[1-\alpha_s(\mu_r^2)b_0\ln\left(\frac{Q^2}{\mu_r^2}\right)+\alpha_s^2(\mu_r^2)b_0^2\ln^2\left(\frac{Q^2}{\mu_r^2}\right)-\cdots\right] \\ &\quad + R_2\alpha_s^2(\mu_r^2)\left[1-2\alpha_s(\mu_r^2)b_0\ln\left(\frac{Q^2}{\mu_r^2}\right)+\ldots\right]^2+\cdots \\ &= \alpha_s(\mu_r^2)R_1+\alpha_s^2(\mu_r^2)\left[R_2-b_0R_1\ln\left(\frac{Q^2}{\mu_r^2}\right)\right] \\ &\quad + \alpha_s^3(\mu_r^2)\left[R_3-2b_0R_2\ln\left(\frac{Q^2}{\mu_r^2}\right)+b_0^2R_1\ln^2\left(\frac{Q^2}{\mu_r^2}\right)\right]+\mathcal{O}(\alpha_s^4). \end{aligned} \tag{K.9}$$

For the inclusive total cross section shown in Section 2.4.1, there are two scales, μ_r which is defined for the renormalization of mass singularities, and μ_f at which the theory is factorized into hard-scattering process and collinear parton distributions. Then there are two equations to be introduced. One is the RG-equation for the running coupling, of which the known-solution up to two-loop, i.e. up to the coefficient b_1, is

$$\begin{aligned} \alpha_s(Q^2) = \alpha_s(\mu_r^2)\Bigg[1 &- \alpha_s(\mu_r^2)\left\{b_0\ln\left(\frac{Q^2}{\mu_r^2}\right)\right\} \\ &+ \alpha_s^2(\mu_r^2)\left\{b_0^2\ln^2\left(\frac{Q^2}{\mu_r^2}\right)-b_1\ln\left(\frac{Q^2}{\mu_r^2}\right)\right\} \\ &+ \alpha_s^3(\mu_r^2)\left\{3b_0b_1\ln^2\left(\frac{Q^2}{\mu_r^2}\right)\right\}+\mathcal{O}(\alpha_s^4)\Bigg] \end{aligned} \tag{K.10}$$

where this is again the expansion of Eq.(K.8) at $Q \sim \mu_r$. The other equation is the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) evolution equation for PDFs, of which solution is known as [137, p.681]

$$f_i(x,Q) = f_i(x,\mu) - \frac{\alpha_s(\mu)}{2\pi}\int_x^1 \frac{d\xi}{\xi} f_i(\xi,\mu)P_{qq}\left(\frac{x}{\xi}\right)\ln\left(\frac{Q^2}{\mu^2}\right) \tag{K.11}$$

where the splitting functions are expanded as

$$P_{ij} = P_{ij}^{(0)} + \alpha_s P_{ij}^{(1)} + \alpha_s^2 P_{ij}^{(2)} + \mathcal{O}(\alpha_s^3). \tag{K.12}$$

Then considering the scale-independence of σ with the RG-equation for α_s and the DGLAP equations for PDFs at $\mu_r = \mu_f = \mu$, the scaling functions are expanded as

$$\begin{aligned} f_{ij} = f_{ij}^{(0)} &+ 4\pi\alpha_s\left\{f_{ij}^{(10)}+f_{ij}^{(11)}\ln\left(\frac{\mu^2}{m^2}\right)\right\} \\ &+ (4\pi\alpha_s)^2\left\{f_{ij}^{(20)}+f_{ij}^{(21)}\ln\left(\frac{\mu^2}{m^2}\right)+f_{ij}^{(22)}\ln^2\left(\frac{\mu^2}{m^2}\right)\right\}+\mathcal{O}(\alpha_s^3) \end{aligned} \tag{K.13}$$

up to NNLO with $Q = m$. Here the coefficient function for the logarithm is, e.g., [61]

$$\begin{aligned} f_{ij}^{(11)} = \frac{1}{8\pi^2}\Bigg[&4\pi b f_{ij}^{(0)}(\rho) \\ &- \int_\rho^1 dz_1 \sum_k f_{kj}^{(0)}\left(\frac{\rho}{z_1}\right)P_{ki}^{(0)}(z_1) - \int_\rho^1 dz_2 \sum_k f_{ik}^{(0)}\left(\frac{\rho}{z_2}\right)P_{kj}^{(0)}(z_2)\Bigg]. \end{aligned} \tag{K.14}$$

Bibliography

[1] S. Catani *et al.*, *Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO*, *JHEP*, **07**, (2019), 100, [arXiv:1906.06535].

[2] S. Catani *et al.*, *Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD*, *JHEP*, **03**, (2021), 029, [arXiv:2010.11906].

[3] M. L. Mangano, P. Nason and G. Ridolfi, *Heavy-quark correlations in hadron collisions at next-to-leading order*, *Nucl. Phys. B*, **373**, (1992), 295-345, [10.1016/0550-3213(92)90435-E].

[4] M. Cacciari *et al.*, *Theoretical predictions for charm and bottom production at the LHC*, *JHEP*, **1210**, (2012), 137, [arXiv:1205.6344].

[5] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer and M. Wiedermann, *HATHOR - HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput.Phys.Commun.*, **182**, (2011), 1034-1046, [arXiv:1007.1327].

[6] **ALICE** Collaboration, B. Abelev *et al.*, *Measurement of charm production at central rapidity in proton-proton collisions at* $\sqrt{s} = 2.76$ *TeV*, *JHEP*, **07**, (2012), 191, [arXiv:1205.4007].

[7] **ALICE** Collaboration, S. Acharya *et al.*, *Measurement of* D^0, D^+, D^{*+} *and* D_s^+ *production in pp collisions at* $\sqrt{s} = 5.02$ *TeV with ALICE*, *Eur.Phys.J.*, **C79**, (2019), 388, [arXiv:1901.07979].

[8] **ALICE** Collaboration, S. Acharya *et al.*, Λ_c^+ *production in pp and in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys.Rev.C*, **104**, (2021), 054905, [arXiv:2011.06079].

[9] **ALICE** Collaboration, S. Acharya *et al.*, *Measurement of beauty and charm production in pp collisions at* $\sqrt{s} = 5.02$ *TeV via non-prompt and prompt D mesons*, *JHEP*, **2105**, (2021), 220, [arXiv:2102.13601].

[10] **ALICE** Collaboration, *First measurement of* Λ_c^+ *production down to* $p_T = 0$ *in pp and p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, CERN-EP-2022-261, (2022), [arXiv:2211.14032].

[11] **ALICE** Collaboration, B. Abelev *et al.*, *Measurement of charm production at central rapidity in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP*, **01**, (2012), 128, [arXiv:1111.1553].

[12] **ALICE** Collaboration, J. Adam *et al.*, *D-meson production in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV and in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Phys.Rev.C*, **94**, (2016), 5, 054908, [arXiv:1605.07569].

[13] **ALICE** Collaboration, S. Acharya *et al.*, *Measurement of D-meson production at mid-rapidity in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Eur.Phys.J.*, **C77**, (2017), 550, [arXiv:1702.00766].

[14] **ALICE** Collaboration, S. Acharya *et al.*, *Measurement of Prompt* D^0, Λ_c^+, *and* $\Sigma_c^{0,++}$*(2455) Production in Proton-Proton Collisions at* $\sqrt{s} = 13$ *TeV*, *Phys.Rev.Lett.*, **128**, (2022), 1, 012001, [arXiv:2106.08278].

[15] **ALICE** Collaboration, S. Acharya *et al.*, *Charm production and fragmentation fractions at midrapidity in pp collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **12** (2023) 086, [arXiv:2308.04877].

[16] **LHCb** Collaboration, R. Aaij *et al.*, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 5$ *TeV*, *JHEP*, **1706**, (2017), 147, [arXiv:1610.02230].

[17] **LHCb** Collaboration, R. Aaij *et al.*, *Prompt charm production in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Nucl.Phys.B*, **871**, (2013), 1-20, [arXiv:1302.2864].

[18] **LHCb** Collaboration, R. Aaij *et al.*, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP*, **05**, (2017), 074, [arXiv:1510.01707].

[19] **ATLAS** Collaboration, G. Aad *et al.*, *Measurement of* $D^{*\pm}$, $D^{\pm}$ *and* $D_s^{\pm}$ *meson production cross sections in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, *Nucl.Phys.B*, **907**, (2016), 717-763, [arXiv:1512.02913].

[20] **CMS** Collaboration, A. M Sirunyan *et al.*, *Nuclear modification factor of* D^0 *mesons in PbPb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys.Lett.B*, **782**, (2018), 474-496, [arXiv:1708.04962].

[21] **CMS** Collaboration, A. M Sirunyan *et al.*, *Production of* Λ_c^+ *baryons in proton-proton and lead-lead collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys.Lett.B*, **803**, (2020), 135328, [arXiv:1906.03322].

[22] **CMS** Collaboration, A. Tumasyan *et al.*, *Study of charm hadronization with prompt* Λ_c^+ *baryons in proton-proton and lead-lead collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *JHEP*, **01**, (2024), 128, [arXiv:2307.11186].

[23] **CMS** Collaboration, *Measurement of prompt open-charm production cross sections in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP*, **11**, (2021), 225, [arXiv:2107.01476].

[24] M.V. Garzelli, L. Kemmler, S. Moch and O. Zenaiev, *Heavy-flavor hadro-production with heavy-quark masses renormalized in the* $\overline{MS}$*, MSR and on-shell schemes*, *JHEP*, **2104**, (2021), 043, [arXiv:2009.07763].

[25] **ALICE** Collaboration, S. Acharya *et al.*, *Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC*, *Phys.Rev.D* **105** (2022) L011103, [arXiv:2105.06335].

[26] **ALICE** Collaboration, *Exploring the non-universality of charm hadronisation through the measurement of the fraction of jet longitudinal momentum carried by* Λ_c^+ *baryons in pp collisions*, CERN-EP-2023-005, [arXiv:2301.13798].

[27] **ALICE** Collaboration, *Measurement of the non-prompt D-meson fraction as a function of multiplicity in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, CERN-EP-2023-018, [arXiv:2302.07783].

[28] Y. Yang and A. Geiser, *A novel phenomenological approach to total charm cross section measurements at the LHC*, *PoS*, EPS-HEP2023, (2024), 367, [arXiv:2311.07523].

[29] C. Bierlich *et al.*, *Open charm production cross section from combined LHC experiments in pp collisions at* $\sqrt{s} = 5.02$ *TeV*, (2023), [arXiv:2311.11426].

[30] T. Sjöstrand *et al.*, *An introduction to PYTHIA 8.2*, *Comput.Phys.Commun.*, **191**, (2015), 159-177, [arXiv:1410.3012].

[31] C. Bierlich *et al.*, *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys.Codeb.*, **2022**, (2022), 8, [arXiv:2203.11601].

[32] **CMS** Collaboration, *Measurement of double differential and total charm cross sections at 7 TeV*, CMS-PAS-BPH-22-007, (2024), `https://cds.cern.ch/record/2905307`.

[33] L. Boyle, `https://upload.wikimedia.org/wikipedia/commons/2/2f/Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png`.

[34] M. Kuhlmann, *Quantum Field Theory*, *The Stanford Encyclopedia of Philosophy (Summer 2023 Edition)*, (2023), `https://plato.stanford.edu/archives/sum2023/entries/quantum-field-theory/`.

[35] *Why Feynman Diagrams Are So Important*, *Quanta Magazine*, (2016), `https://www.quantamagazine.org/why-feynman-diagrams-are-so-important-20160705/`.

[36] R. P. Feynman, *The Theory of Positrons*, *Phys. Rev.*, **76**, (1949), 749 [10.1103/PhysRev.76.749].

[37] P. Langacker, *The standard model and beyond*, *CRC Press, 2nd edition*, (2017), [10.1201/b22175].

[38] M. Thomson, *Modern Particle Physics*, *Cambridge University Press*, (2013), [10.1017/CBO9781139525367].

[39] R. K. Ellis, W. J. Stirling, B. R. Webber, *QCD and collider physics*, *Cambridge University Press*, **8**, (2011), [10.1017/CBO9780511628788].

[40] **Particle Data Group**, P. A. Zyla *et al.*, *Review of Particle Physics*, *Prog.Theor.Exp.Phys.*, **2020**, (2020), 083C01, [10.1093/ptep/ptaa104].

[41] Y. Kato and T. Iijima, *Open charm hadron spectroscopy at B-factories*, *Prog.Part.Nucl.Phys.*, **105**, (2019), 61-81, [arXiv:1810.03748].

[42] **Particle Data Group**, R. L. Workman *et al.*, *Review of Particle Physics*, *Prog.Theor.Exp.Phys.*, **2022**, (2022), 083C01, [10.1093/ptep/ptac097].

[43] M. Cacciari, M. Greco and P. Nason, *The* p_T *spectrum in heavy-flavour hadroproduction*, *JHEP*, **9805**, (1998), 007, [arXiv:hep-ph/9803400].

[44] C. Peterson *et al.*, *Scaling violations in inclusive* e^+e^- *annihilation spectra*, *Phys.Rev.D*, **27**, (1983), 105, [10.1103/PhysRevD.27.105].

[45] V. G. Kartvelishvili *et al.*, *On the Fragmentation Functions of Heavy Quarks Into Hadrons*, *Phys.Lett.B*, **78**, (1978), 615-617, [10.1016/0370-2693(78)90653-6].

[46] M. Cacciari, P. Nason and C. Oleari, *A study of heavy flavoured meson fragmentation functions in* e^+e^- *annihilation*, *JHEP*, **0604**, (2006), 006, [arXiv:hep-ph/0510032].

[47] M. Lisovyi, A. Verbytskyi and O. Zenaiev, *Combined analysis of charm-quark fragmentation-fraction measurements*, *Eur.Phys.J.*, **C76**, (2016), 397, [arXiv:1509.01061].

[48] **Belle** Collaboration, R. Seuster *et al.*, *Charm hadrons from fragmentation and B decays in e^+e^- annihilation at $\sqrt{s}$ = 10.6 GeV*, *Phys.Rev.D*, **73**, (2006), 032002, [arXiv:hep-ex/0506068].

[49] K. G. Chetyrkin, R. V. Harlander and J. H. Kuhn, *Quartic mass corrections to R_{had} at $\mathcal{O}(\alpha_s^3)$*, *Nucl.Phys.B*, **634**, (2002), 413-414, [arXiv:hep-ph/0005139].

[50] Y. Yang, *Heavy-Quark Hadro-Production at Next-to-Leading-Order in QCD and Beyond*, *Master's thesis*, University of Hamburg, (2020).

[51] M. Czakon and A. Mitov, *Inclusive Heavy Flavor Hadroproduction in NLO QCD: the Exact Analytic Result*, *Nucl. Phys. B*, **824**, (2010), 111-135, [arXiv:0811.4119].

[52] V. Ahrens *et al.*, *RG-improved single-particle inclusive cross sections and forward-backward asymmetry in $t\bar{t}$ production at hadron colliders*, *JHEP*, **09**, (2011), 070, [arXiv:1103.0550].

[53] A. Grozin, *Lectures on Soft-Collinear Effective Theory*, *Proceedings, HQ*, (2016), 121-149, [arXiv:1611.08828].

[54] T. Becher, *Les Houches Lectures on Soft-Collinear Effective Theory, Les Houches summer school: EFT in Particle Physics and Cosmology*, *Proceedings, Les Houches summer school*, (2020), [arXiv:1803.04310].

[55] S. Moch, P. Uwer and A. Vogt, *On top-pair hadro-production at next-to-next-to-leading order*, *Phys. Lett. B*, **714**, (2012), 48-54, [arXiv:1203.6282].

[56] J. C. Collins and R. K. Ellis, *Heavy-quark production in very high energy hadron collisions*, *Nucl. Phys. B*, **360**, (1991), 3, [10.1016/0550-3213(91)90288-9].

[57] S. Catani, M. Ciafaloni and F. Hautmann, *High energy factorization and small-x heavy flavour production*, *Nucl. Phys. B*, **366**, (1991), 135, [10.1016/0550-3213(91)90055-3].

[58] S. Catani and F. Hautmann, *High-Energy Factorization and Small-x Deep Inelastic Scattering Beyond Leading Order*, *Nucl. Phys. B*, **427**, (1994), 475-524, [arXiv:hep-ph/9405388].

[59] G. Camici and M. Ciafaloni, *k-Factorization and small-x anomalous dimensions*, *Nucl. Phys. B*, **496**, (1997), 305-336, [arXiv:hep-ph/9701303].

[60] R. D. Ball and R. K. Ellis, *Heavy quark production at high energy*, *JHEP*, **05**, (2001), 053, [arXiv:hep-ph/0101199].

[61] P. Nason, S. Dawson and R. K. Ellis, *The total cross section for the production of heavy quarks in hadronic collisions, Nucl. Phys. B Nucl. Phys. B*, **303**, (1988), 607-633, [10.1016/0550-3213(88)90422-1].

[62] P. Bärnreuther, M. Czakon and A. Mitov, *Precent level precision physics at the Tevatron: first genuine NNLO QCD corrections to $q\bar{q} \to t\bar{t}+X$*, *Phys.Rev.Lett.*, **109**, (2012), 132001, [arXiv:1204.5201].

[63] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP*, **12**, (2012), 054, [arXiv:1207.0236].

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

[65] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP*, **01**, (2013), 080, [arXiv:1210.6832].

[66] U. Langenfeld, S. Moch and P. Uwer, *Measuring the running top-quark mass, Phys. Rev. D Phys.Rev.D*, **80**, (2009), 054009, [arXiv:0906.5273].

[67] P. M. Nadolsky *et al.*, *Implications of CTEQ global analysis for collider observables*, *Phys.Rev.D*, **78**, (2008), 013004, [arXiv:0802.0007].

[68] **PROSA** Collaboration, O. Zenaiev *et al.*, *Improved constraints on parton distributions using LHCb, ALICE and HERA heavy-flavour measurements and implications for the predictions for prompt atmospheric-neutrino fluxes*, *JHEP*, **2004**, (2020), 118, [arXiv:1911.13164].

[69] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys.Rept.*, **97**, (1983), 31-145, [10.1016/0370-1573(83)90080-7].

[70] T. Sjostrand, *Jet Fragmentation of Nearby Partons*, *Nucl.Phys.B*, **248**, (1984), 469-502, [10.1016/0550-3213(84)90607-2].

[71] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, *Nucl.Phys.B*, **238**, (1984), 492-528, [10.1016/0550-3213(84)90333-X].

[72] M. Bahr *et al.*, *Herwig++ Physics and Manual*, *Eur.Phys.J.C*, **58**, (2008), 639-707, [arXiv:0803.0883].

[73] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, *JHEP*, **02**, (2009), 007, [arXiv:0811.4622].

[74] R. Bailey, C. Benvenuti1, S. Myers, D. Treille, *The LEP collider*, *C. R. Physique*, **3**, (2002), 1107-1120, [10.1016/S1631-0705(02)01402-0].

[75] A. Straessner, *The LEP Experiments*, *Electroweak Physics at LEP and LHC. Springer Tracts in Modern Physics, Springer*, **235**, (2010), [10.1007/978-3-642-05169-2_2].

[76] D. Rice *et al.*, *Parameters for low energy operation of CESR*, *Proceedings of the 2001 Particle Accelerator Conference*, **1**, (2001), 374-376, [10.1109/PAC.2001.987518].

[77] T. Mundzeck, *The three lives of DORIS: from charm quarks to cell biology*, *CERN Courier.*, **27**, (2012), `https://cerncourier.com/a/the-three-lives-of-doris-from-charm-quarks-to-cell-biology/`.

[78] G. A. Loew, *The SLAC linear collider and a few ideas on future linear colliders*, *Proceedings of the 1984 Linear Accelerator Conference*, (1984), 282-287, [SLAC-PUB-3327].

[79] S.-I. Kurokawa, *Control system design for KEKB accelerators*, *Proceedings Particle Accelerator Conference*, **4**, (1995), 2205-2207, [10.1109/PAC.1995.505499].

[80] M. Klein and R. Yoshida, *Collider Physics at HERA*, *Prog.Part.Nucl.Phys.* **61** (2008) 343-393, [arXiv:0805.3334].

[81] S. Myers, *The Large Hadron Collider 2008-2013*, *Int.J.Mod.Phys.A* **28** (2013) 1330035, [10.1142/S0217751X13300354].

[82] S. Braibant, G. Giacomelli and M. Spurio, *Particles and Fundamental Interactions Undergraduate lecture notes in physics*, (2012), 498, [10.1007/978-94-007-2464-8].

[83] **CMS** Collaboration, *Illustration of the performance of the CMS tracker and reconstruction on early Run 3 data, on the example of D^* meson reconstruction*, CMS-DP-2022-024, (2022), `http://cds.cern.ch/record/2815408?ln=en`.

[84] **ALEPH** Collaboration, R. Barate *et al.*, *Study of charm production in Z decays*, *Eur.Phys.J.C*, **16**, (2000), 597-611, [arXiv:hep-ex/9909032].

[85] **CLEO** Collaboration, M. Artuso *et al.*, *Charm meson spectra in e^+e^- annihilation at 10.5 GeV c.m.e.*, *Phys.Rev.D*, **70**, (2004), 112001, [arXiv:hep-ex/0402040].

[86] **H1** Collaboration, A. Aktas *et al.*, *Inclusive production of D^+, D^0, D_s^+ and D^{*+} mesons in deep inelastic scattering at HERA*, *Eur.Phys.J.C* **38** (2005) 447-459, [arXiv:hep-ex/0408149].

[87] **ALICE** Collaboration, S. Acharya *et al.*, *Measurement of the production cross section of prompt Ξ_c^0 baryons at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV*, *JHEP*, **2110**, (2021), 159, [arXiv:2105.05616].

[88] **ZEUS** Collaboration, H. Abramowicz *et al.*, *Measurement of charm fragmentation fractions in photoproduction at HERA*, *JHEP*, **1309**, (2013), 58, [arXiv:1306.4862].

[89] Heavy Flavor Averaging Group (**HFLAV**), Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, *Eur.Phys.J.*, **C81**, (2021), 226, [arXiv:1909.12524].

[90] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of f_s/f_u variation with proton-proton collision energy and B-meson kinematics*, *Phys.Rev.Lett.*, **124**, (2020), 122002, [arXiv:1910.09934].

[91] **CMS** Collaboration, *Measurement of the dependence of the hadron production fraction ratio f_s/f_u on B meson kinematic variables in proton-proton collisions at $\sqrt{s} = 13$ TeV*, CERN-EP-2022-248, [arXiv:2212.02309].

[92] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of b-hadron fractions in 13 TeV pp collisions*, *Phys.Rev.D.*, **100**, (2019), 031102, [arXiv:1902.06794].

[93] **CMS** Collaboration, G. L. Bayatian *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, [CERN-LHCC-2006-001].

[94] **CMS** Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, *JINST*, **3**, (2008), S08004, [10.1088/1748-0221/3/08/S08004].

[95] S. G. López for the **CMS** Collaboration, *CMS Detector Performance during LHC Run 1 and projections for Run 2*, *Nucl.Part.Phys.Proc.* **273-275** (2016) 1048-1054, [10.1016/j.nuclphysbps.2015.09.164].

[96] M. Ressegotti for the **CMS** Collaboration, *Overview of the CMS Detector Performance at LHC Run 2*, *Universe* **5** (2019) 18, [10.3390/universe5010018].

[97] **CMS** Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, CMS-PRF-21-001, CERN-EP-2023-136, [arXiv:2309.05466].

[98] T. Sakuma and T. McCauley, *Detector and Event Visualization with SketchUp at the CMS Experiment*, *J. Phys. Conf. Ser.*, **513**, (2014), 022032, [arXiv:1311.4942].

[99] I. Neutelings, `https://tikz.net/axis3d_cms`.

[100] D. Campi and D. Stickland, *CMS: a super solenoid is ready for business*, *CERN Courier.*, **47**, (2007), https://cerncourier.com/a/cms-a-super-solenoid-is-ready-for-business-2/.

[101] B. Cure, *CMS magnet warm up*, https://cms.cern/news/cms-magnet-warm.

[102] S. Chatrchyan, *Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays*, *JINST*, **5**, (2010), T03021, [arXiv:0910.5530].

[103] **CMS** Collaboration, V. Karimkäi *et al.*, *The CMS tracker system project : Technical Design Report*, CERN-LHCC-98-006, CMS-TDR-5, http://cds.cern.ch/record/368412.

[104] **CMS** Collaboration, *Silicon Pixels*, https://cms.cern/detector/identifying-tracks/silicon-pixels.

[105] **CMS** Collaboration, D. A. M. Dominguez *et al.*, *CMS Technical Design Report for the Pixel Detector Upgrade*, CERN-LHCC-2012-016, CMS-TDR-011, [10.2172/1151650].

[106] **CMS** Collaboration, W. Adam *et al.*, *The CMS Phase-1 pixel detector upgrade*, *JINST*, **16**, (2021), P02027, [10.1088/1748-0221/16/02/P02027].

[107] **CMS** Collaboration, P. Bloch *et al.*, *Changes to CMS ECAL electronics : addendum to the Technical Design Report*, CERN-LHCC-2002-027, CMS-TDR-4-add-1, http://cds.cern.ch/record/581342.

[108] K. Bunkowski and V. Sharma, *Introduction to Level-1 Trigger System*, Trigger Shifter Tutorial, 2022, https://indico.cern.ch/event/1127015/contributions/4730347/.

[109] R. Bainbridge, *Recording and reconstructing 10 billion unbiased b hadron decays in CMS*, *EPJ Web Conf.*, **245**, (2020), 01025, [10.1051/epjconf/202024501025].

[110] T. Speer, K. Prokofiev, R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Vertex Fitting in the CMS Tracker*, CMS-NOTE-2006-032, (2006), http://cds.cern.ch/record/927395.

[111] F. Sikler, *Low p_T Hadronic Physics with CMS*, *Int.J.Mod.Phys.E*, **16**, (2007), 1819-1825, [10.1142/S0218301307007052].

[112] 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, [arXiv:1405.0301].

[113] M. Dobbs and J. B. Hansen, *The HepMC C++ Monte Carlo event record for High Energy Physics*, *Comput.Phys.Commun.*, **134**, (2001), 41-46, http://cds.cern.ch/record/684090.

[114] K. Pedro, *Current and Future Performance of the CMS Simulation*, *EPJ Web Conf.*, **214**, (2019), 02036, [10.1051/epjconf/201921402036].

[115] **GEANT4** Collaboration, S. Agostinelli *et al.*, *GEANT4–a simulation toolkit*, *Nucl.Instrum.Meth.A*, **506**, (2003), 250-303, [10.1016/S0168-9002(03)01368-8].

[116] J. Allison *et al.*, *Recent developments in Geant4*, *Nucl.Instrum.Meth.A*, **835**, (2016), 186-225, [10.1016/j.nima.2016.06.125].

[117] C. D. Jones, M. Paterno, J. Kowalkowski, L. Sexton-Kennedy, W. Tanenbaum, *The new CMS event data model and framework*, *Proceedings of International Conference on Computing in High Energy and Nuclear Physics, CHEP*, **06**, (2006), https://indico.cern.ch/event/408139/contributions/979800/.

[118] CMS OpenData, *CMS Monte Carlo production overview*, `https://opendata.cern.ch/docs/cms-mc-production-overview`.

[119] N. Z. Jomhari, *Measurement of Total and Differential Charm Cross Sections at 7 TeV with the CMS Detector*, *PhD U. Hamburg*, (2022), DESY-THESIS-2022-006, [10.3204/PUBDB-2022-01666].

[120] J. Metwally, *Double differential charm and beauty production cross section measurements at 5 TeV in CMS at the LHC*, *PhD U. Hamburg*, (2022), DESY-THESIS-2022-019, [10.3204/PUBDB-2022-04972].

[121] L. Della Penna, A. Geiser, V. Mariani and Y. Yang, *Measurement of double differential and total charm cross section at 900 GeV*, CMS AN-22-157, and personal communication with L. Della Penna.

[122] A. Geiser, N. Z. Jomhari, V. Mariani, J. Metwally and Y. Yang, *Measurement of double differential and total charm cross sections at 7 TeV*, CMS AN-19-241.

[123] CERN, *CERN Open Data Portal*, `https://opendata.cern.ch/docs/about`.

[124] A. Geiser, N. Z. Jomhari and J. Metwally, *Measurement of total and associated charm and beauty cross sections in the CMS experiment*, CMS AN-18-284.

[125] V. Sergeychik, S. Polikarpov and J. Martins, *CP violation measurement in charm meson decays*, CMS AN-20-054.

[126] CMS McM, `https://cms-pdmv-prod.web.cern.ch/mcm/public/restapi/requests/get/PPD-RunIISummer20UL18GEN-00001`.

[127] CMS McM, `https://cms-pdmv-prod.web.cern.ch/mcm/public/restapi/requests/get/BPH-RunIISummer20UL18GEN-00263`.

[128] **CMS** Collaboration, A. M Sirunyan *et al.*, *Studies of Beauty Suppression via Nonprompt D^0 Mesons in Pb-Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV*, *Phys.Rev.Lett.*, **123**, (2019), 022001, [arXiv:1810.11102].

[129] **CMS** Collaboration, *Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update*, CMS-PAS-EWK-11-001, (2011), `https://cds.cern.ch/record/1376102/`.

[130] **CMS** Collaboration, S. Chatrchyan *et al.*, *Study of the Inclusive Production of Charged Pions, Kaons, and Protons in pp Collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV*, *Eur.Phys.J.C*, **72**, (2012), 2164 [arXiv:1207.4724].

[131] A. Buckley *et al.*, *LHAPDF6: parton density access in the LHC precision era*, *Eur.Phys.J.C*, **75**, (2015), 132, [arXiv:1412.7420].

[132] P. M. Nadolsky *et al.*, `https://lhapdfsets.web.cern.ch/current/cteq66/cteq66.info`.

[133] A. Geiser, Y. Yang, O. Zenaiev and S. Moch, *Charm total cross sections and extraction of QCD parameters*, 31st International Workshop on Deep Inelastic Scattering, (2024), `https://lpsc-indico.in2p3.fr/event/3268/contributions/7343/`.

[134] S. Alekhin, J. Bluemlein, S. Moch, R. Placakyte, *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II*, *Phys.Rev.D*, **96**, (2017), 1, 014011, [arXiv:1701.05838].

[135] T. Cridge, L. A. Harland-Lang, A. D. Martin, R. S. Thorne, *An investigation of the α_s and heavy quark mass dependence in the MSHT20 global PDF analysis*, *Eur.Phys.J.C*, **81**, (2021), 8, 744, [arXiv:2106.10289].

[136] L. Gladilin, *Fragmentation fractions of c and b quarks into charmed hadrons at LEP*, *Eur.Phys.J.*, **C75**, (2015), 19, [arXiv:1404.3888].

[137] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, *Cambridge University Press*, (2014).