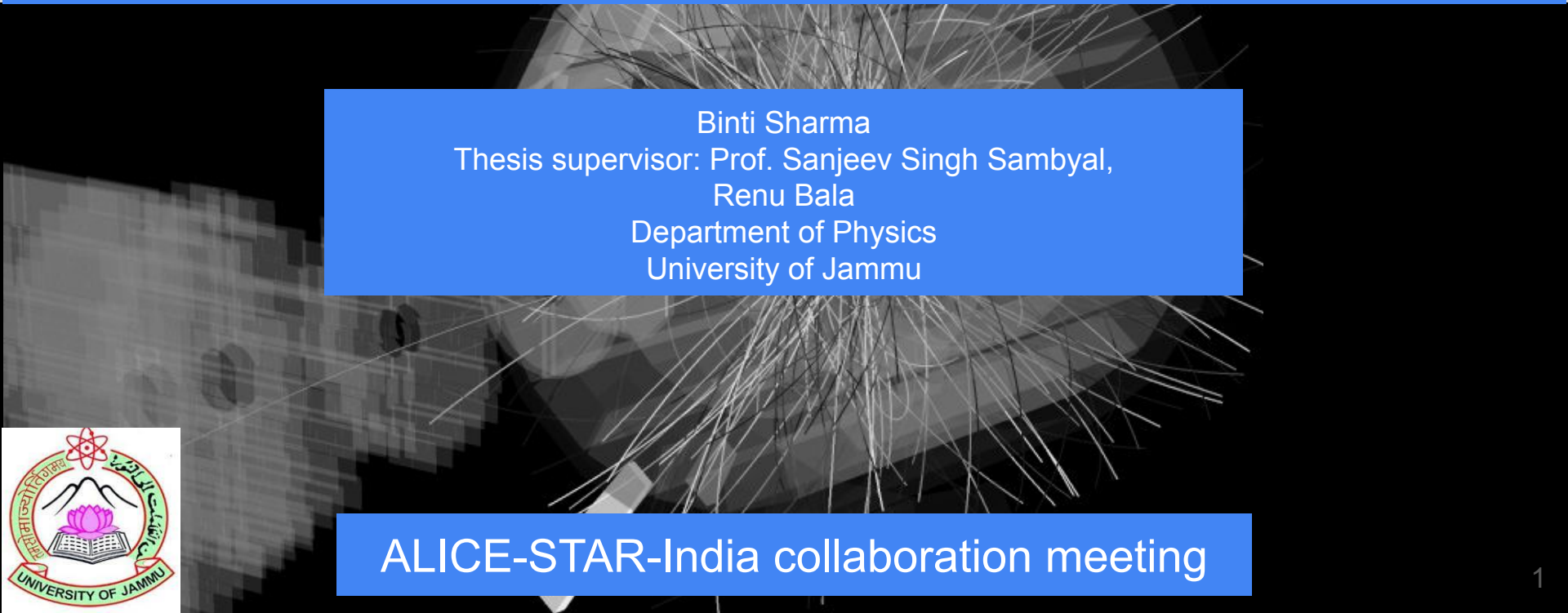
Non-prompt D^+ meson production in pp collisions at $\sqrt{s}$ = 13 TeV



Binti Sharma
Thesis supervisor: Prof. Sanjeev Singh Sambyal,
Renu Bala
Department of Physics
University of Jammu

ALICE-STAR-India collaboration meeting

Outline

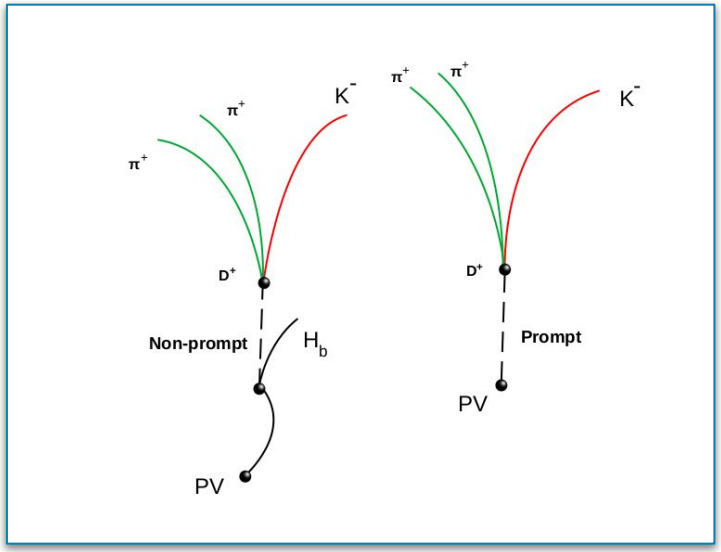
binti.sharma@cern.ch

- Physics motivation
- Data samples and analysis strategy
- Raw yields extraction
- Selection efficiency estimation
- Non-prompt fraction estimation
- Overview of systematics
- Production cross section
- Conclusions

Physics motivation

Non-prompt D^+ mesons

- Heavy Quarks(c, b) are produced in initial hard-scattering processes on a shorter time-scale than the QGP formation.
- They experience full evolution of the system, propagating and interacting with the medium constituents via elastic and inelastic scatterings.
- Study the beauty production in pp collisions
  - Non-prompt D^+ mesons come from B^0 and B^+
    - ➡ test *p*QCD theory
    - ➡ measure total b-bbar cross section in pp at 13 TeV
    - ➡ measure beauty-quark fragmentation-fraction to strange over non-strange



Data samples and analysis strategy

Samples:

- Data sample: pp collisions @ 13 TeV (2016, 2017, 2018); N_{ev} for norm = 1.836e+09
- MC sample: LHC20f4 for efficiencies, LHC20l1 for Machine Learing(ML) model training/testing

Analysis strategy:

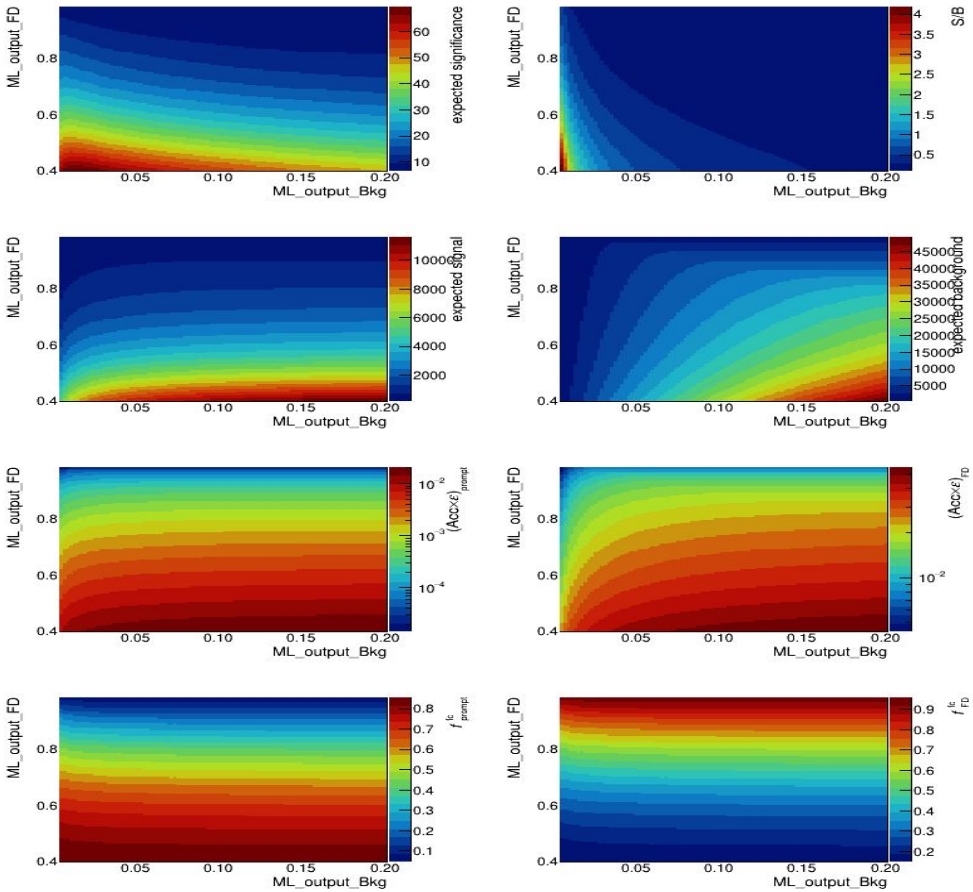
- Decay Channel $D^+ \rightarrow K^- \pi^+ \pi^+$ (with B.R. = (9.38±0.16)%)
- In particular, some preselections (which include single-track, topological, and PID selections), based on displaced decay-vertex topologies, were applied to select the D^+ candidates.
- Then the multi-class classification algorithm (hipe4ml) provided by XGBoost was used to separate the three contributions (prompt D^+, non-prompt D^+, and combinatorial background).
  - Raw yield extraction from fit to invariant-mass distribution
  - Selection efficiency from MC simulation
  - f_{FD} estimated via data driven method
    - ➡ production cross section

Topological and kinematic selections and Single Track Selections

Topological variables	pT intervals(GeV/*C*)	
	[1,5]	[5,50]
σ_{vertex} (μm)	<400	<600
Decay length (μm)	>300	>300
$\cos\theta_p$	>0.85	>0.75
$\cos\theta_p^{xy}$	>0.80	>0.70

- pseudorapidity interval $|\eta| < 0.8$
- $p_T > 0.3$ GeV/*c* in pp collisions
- $\chi^2/ndf < 2$
- at least 50, out of a maximum of 159, crossed rows in the TPC.
- ratio of crossed rows over findable clusters in the TPC larger than 0.8

Optimisation of ML selection for $D^+(2 < p_T < 3$ GeV/$c)$

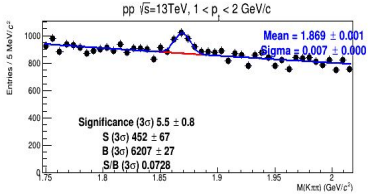


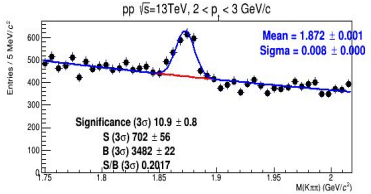
- Choice of ML-based selections is performed by estimating expected quantities for several threshold values on the ML output scores (NP, Bkg)
  - signal from FONLL
  - efficiencies from MC and non-prompt fraction from theory-driven method (f^c)
  - background from sidebands (only fraction of data)

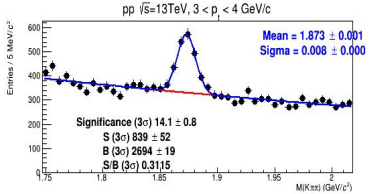
Central ML Selections

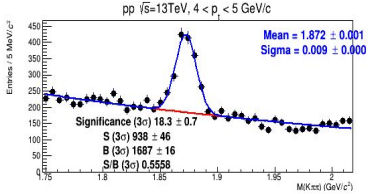
D^+ meson	p_T interval (GeV/c)									
	[1,2]	[2,3]	[3.4]	[4,5]	[5,6]	[6,8]	[8,10]	[10,12]	[12,16]	[16,24]
probability to be background <	0.03	0.025	0.025	0.040	0.040	0.030	0.040	0.050	0.050	0.050
probability to be non-prompt >	0.80	0.86	0.86	0.82	0.80	0.80	0.82	0.82	0.76	0.70

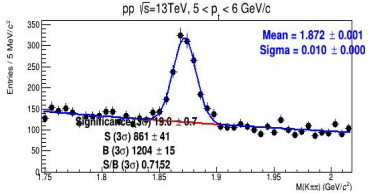
Raw yields for non-prompt D^+

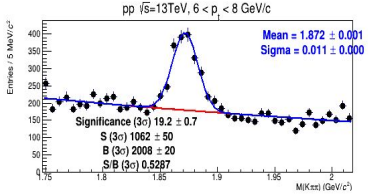


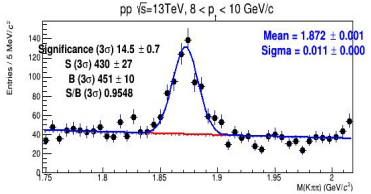


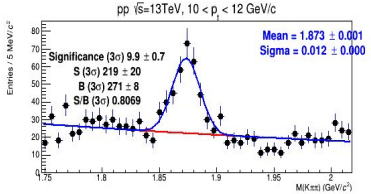


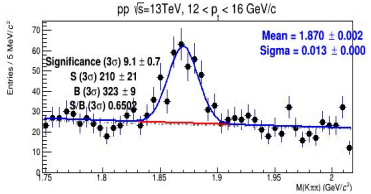


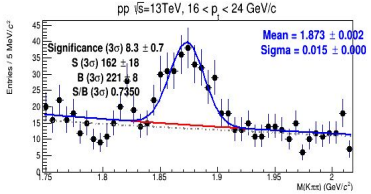






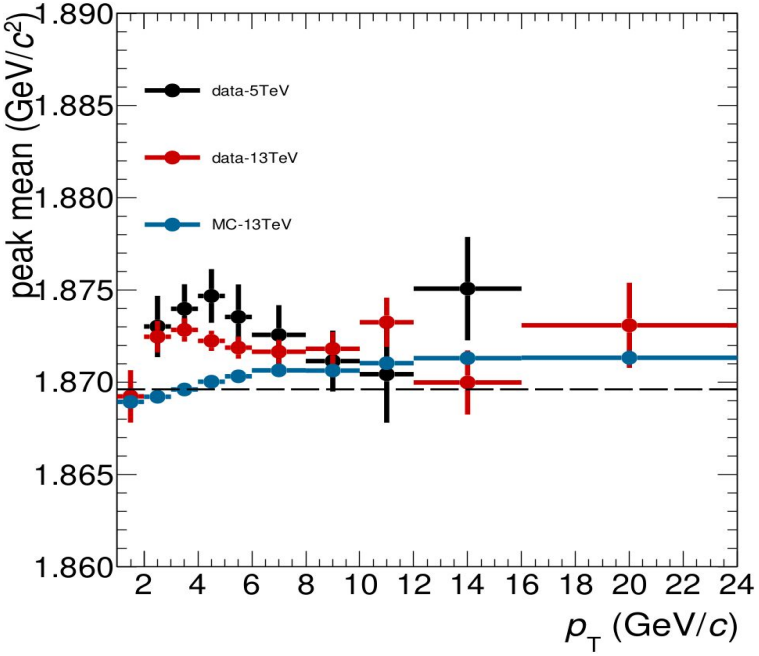


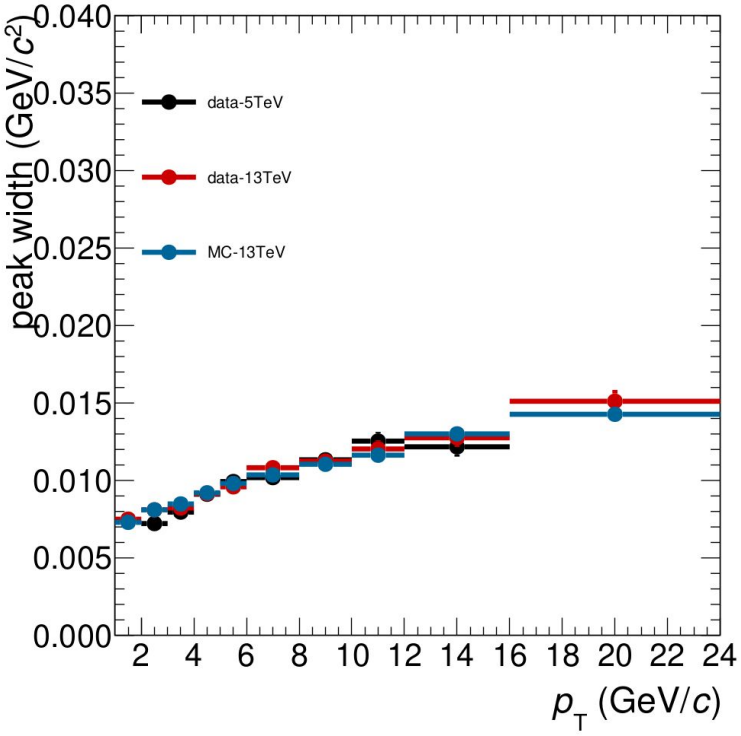




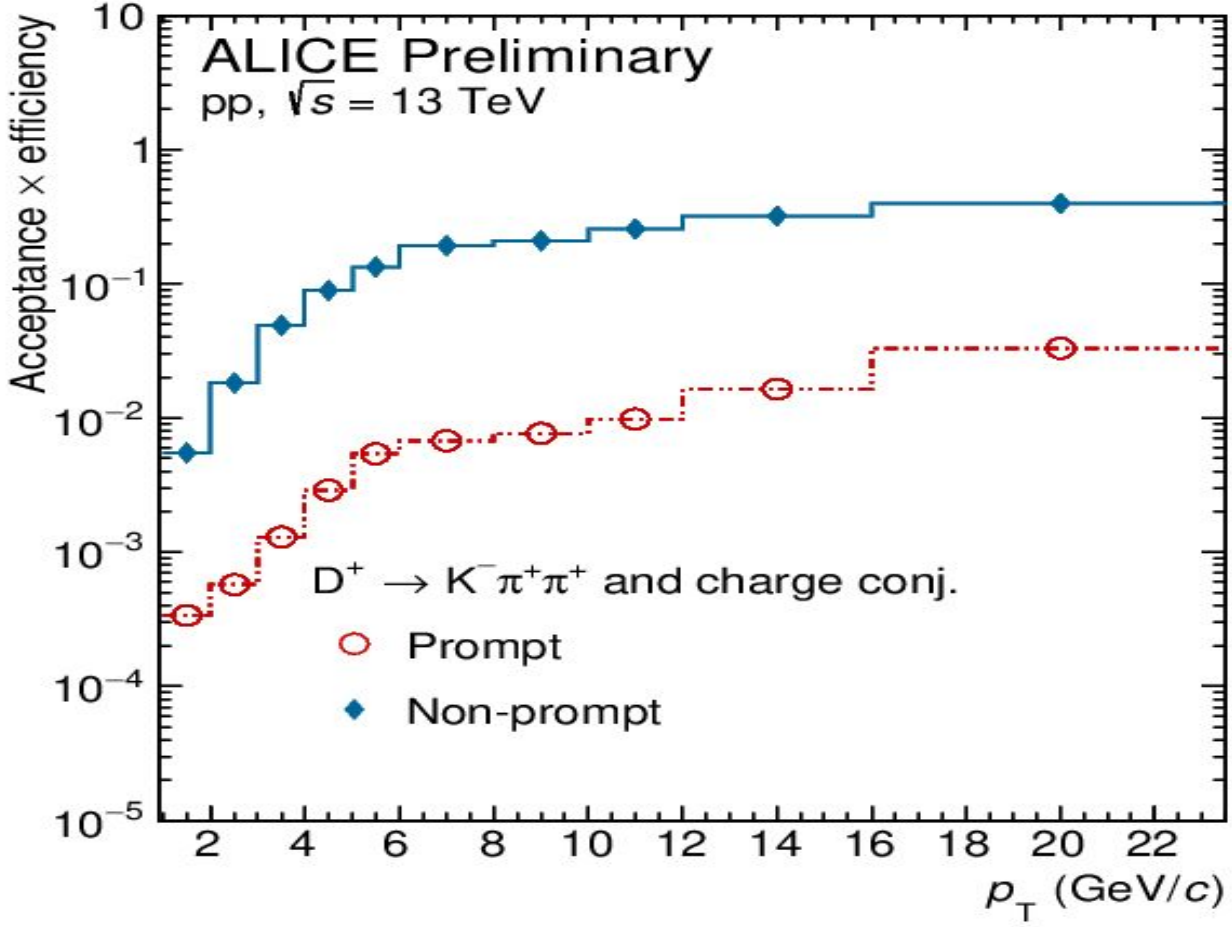
- Signif: [5.5, 10.9, 14.1, 18.3, 19.0, 19.2, 14.2, 9.9, 9.1, 8.3]
- Sigma fixed to prompt-enhanced results
- Good signal extraction up to 24 GeV/*c*
  - Gauss + Expo in full p_T range
  - [24-50] not accessible

Comparisons: D^+ fit parameters

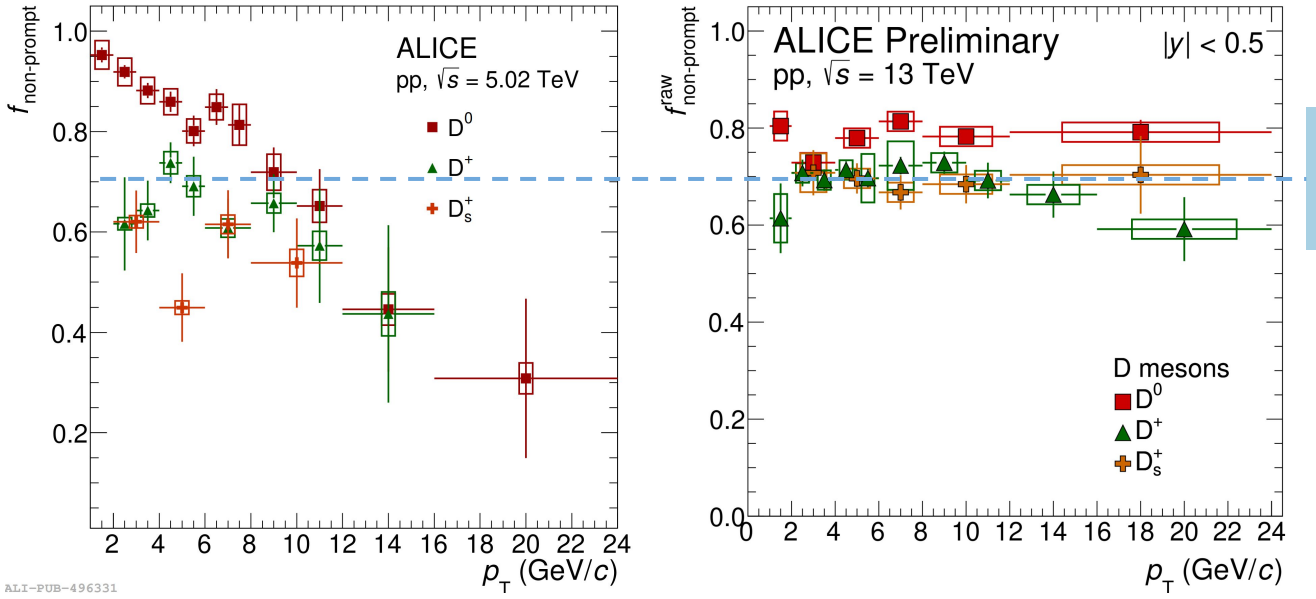




Efficiency times Acceptance



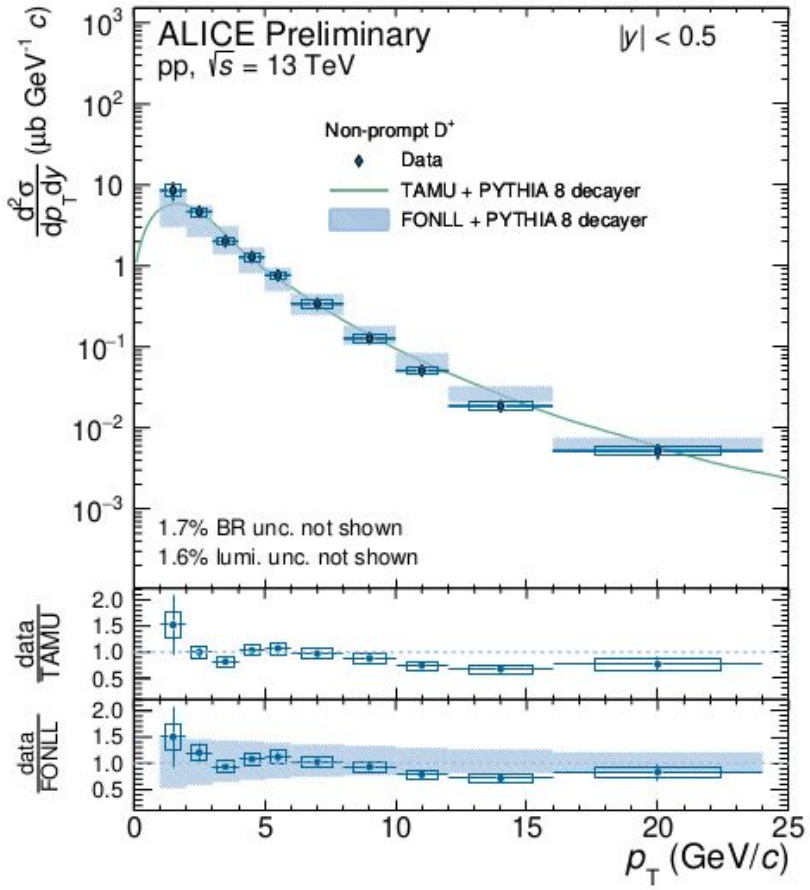
Non-prompt fraction



To drive the eye along 70% f_{FD}

- Same strategy adopted for non-prompt D meson measurements @ 5 TeV (arXiv:2102.13601)
  - data driven method based on selection criteria variation (more details in backup)
- Higher sample purity in pp collisions at 13 TeV (~70%) wrt 5 TeV

Cross section vs. FONLL

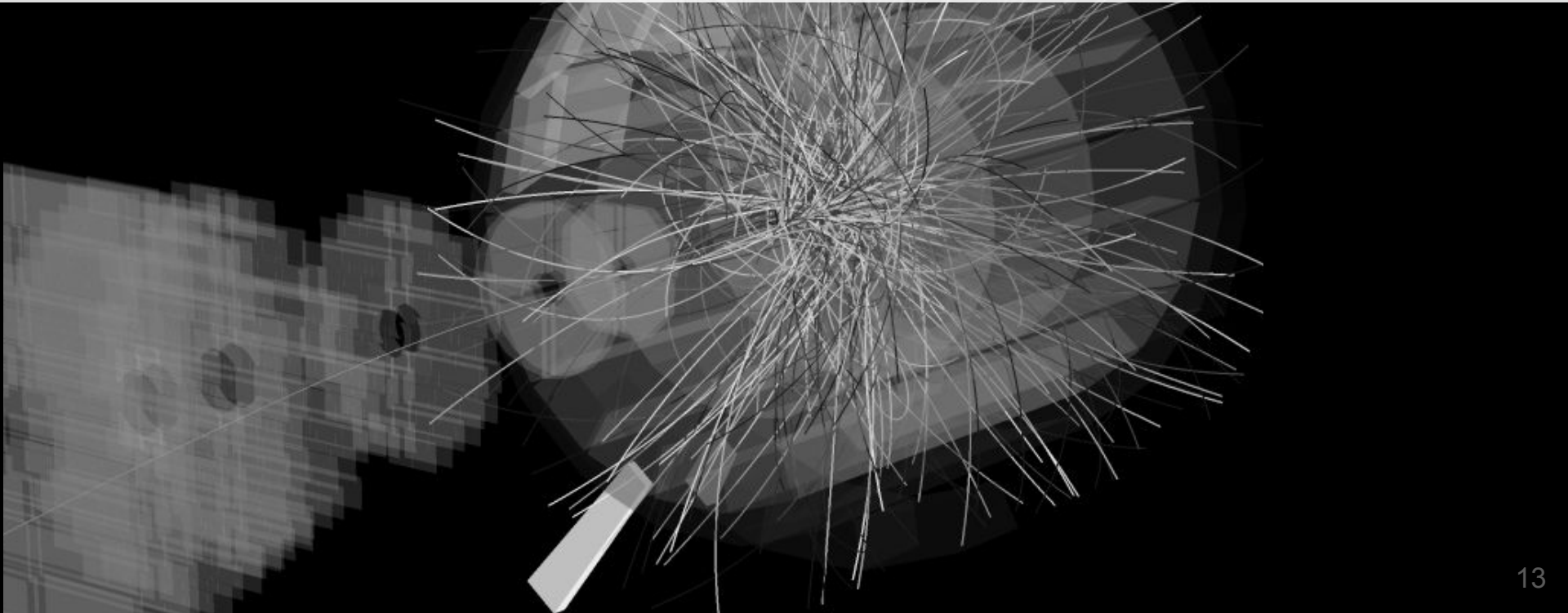


$$\frac{\mathrm{d}N(\mathrm{D}_{\text{non-prompt}})}{\mathrm{d}p_{\mathrm{T}}} = \frac{f_{\text{non-prompt}}(p_{\mathrm{T}}) \cdot N^{\mathrm{D}}_{\text{raw}}(p_{\mathrm{T}})}{2 \cdot \Delta p_{\mathrm{T}} \cdot c_{\Delta y}(p_{\mathrm{T}}) \cdot (\mathrm{Acc} \times \epsilon)_{\text{non-prompt}}(p_{\mathrm{T}}) \cdot \mathrm{BR} \cdot \mathcal{L}_{int}}$$

Comparison with:

- FONLL (B) + PYTHIA 8 (e^+e^- FF) as done at 5 TeV
  - good agreement
    - ➡ similar to what observed @ 5.02 TeV
- TAMU predictions from Min He and Ralf Rapp., Which adopts the p_T -differential beauty-quark cross section from FONLL along with the same fragmentation functions employed in FONLL and a statistical hadronisation approach for f (b $\rightarrow$ h_b).

Overview of systematic uncertainties



Summary of systematic uncertainties

D^+ meson	p_T interval (GeV/c)									
	[1,2]	[2,3]	[3,4]	[4,5]	[5,6]	[6,8]	[8,10]	[10,12]	[12,16]	[16,24]
Raw-yield extraction	5%	4%	4%	4%	4%	4%	4%	4%	5%	5%
ML selection efficiency	10%	6%	5%	4%	4%	4%	4%	4%	4%	4%
Non-prompt fraction	5%	3%	3%	3%	3%	3%	3%	3%	3%	3%
Tracking efficiency	6%	6%	6%	6%	7%	7%	7%	7%	8%	8%
PID efficiency	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.
MC p_T shape	7%	5%	3%	1%	negl.	negl.	negl.	negl.	2%	5%
Normalisation	1.6%									
Branching ratio	1.7%									

Taken from prompt D analyses

Summary

Measurement of non-prompt D^+ meson production in pp @ 13 TeV with ML multi-classification technique

- Non-prompt D^+ meson production cross section measured
- Systematic uncertainty estimation completed

TO DO

- D^+ meson systematic uncertainty on material budget to be updated for the paper
- Total b-bbar cross section in pp at 13 TeV
- Beauty-quark fragmentation fraction to strange over non-strange

AN: https://alice-notes.web.cern.ch/node/1347

Summary

Measurement of non-prompt D^+ meson production in pp @ 13 TeV with ML multi-classification technique

- Non-prompt D^+ meson production cross section measured
- Systematic uncertainty estimation completed

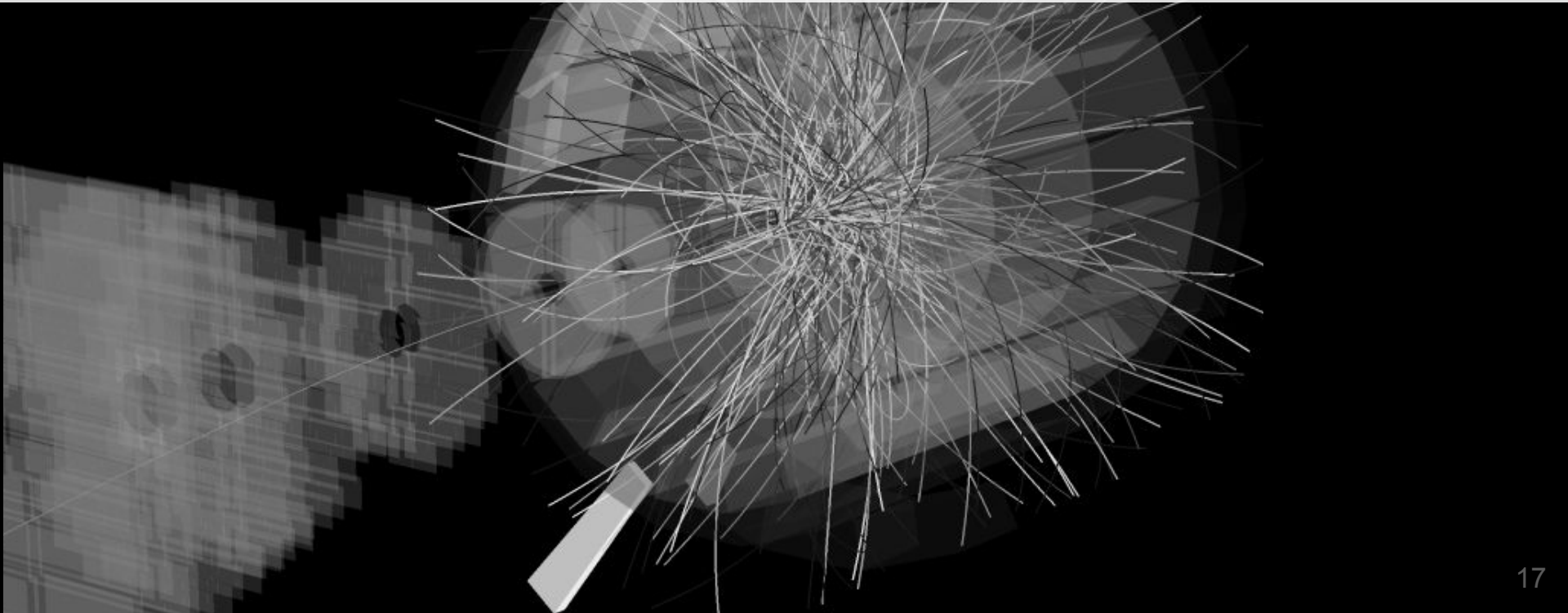
TO DO

- D^+ meson systematic uncertainty on material budget to be updated for the paper
- Total b-bbar cross section in pp at 13 TeV
- Beauty-quark fragmentation fraction to strange over non-strange

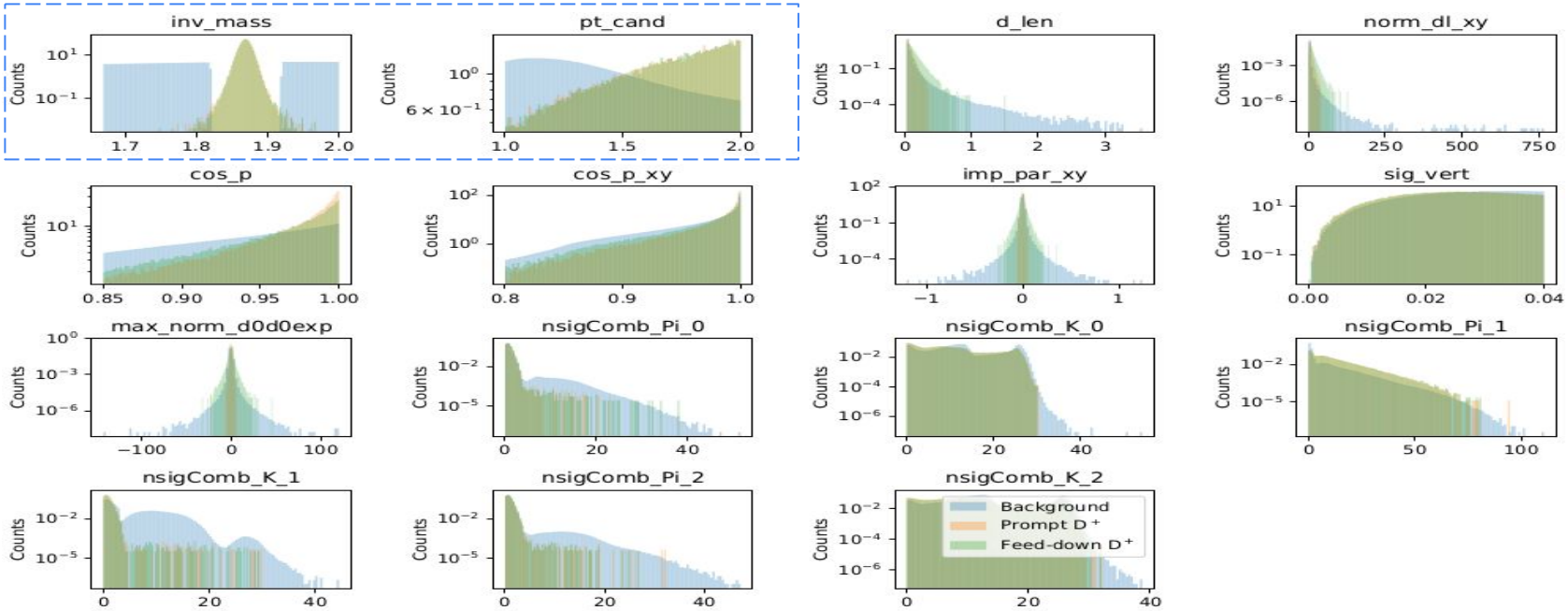
THANK YOU

AN: https://alice-notes.web.cern.ch/node/1347

Back up

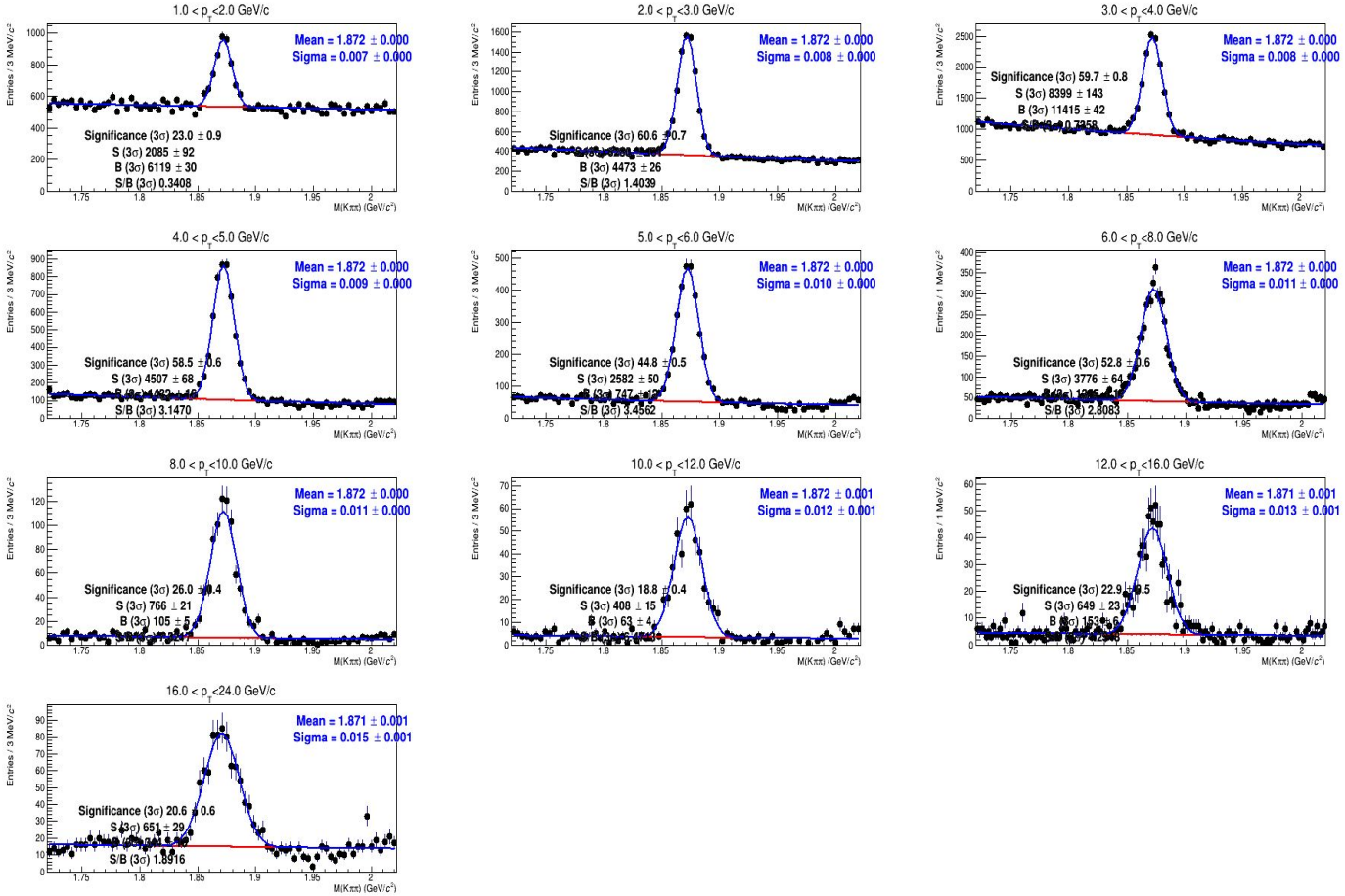


Training variable distributions



- All variables employed in the training (invariant mass and transverse momentum excluded)
- Training samples of prompt D^+ mesons and non-prompt D^+ mesons from MC
- Training samples of bkg with data from SB (D^+ meson)

Raw yields D^+ ("prompt enhanced" sample)



- Signal extraction from 1 to 24 GeV/c
- Significance between 18 and 59

Non-prompt fraction estimation via data-driven method

Same strategy adopted for non-prompt D^+ meson measurements @ 5 TeV (arXiv:2102.13601)

I). *n* set of ML-based selections with different prompt and non-prompt D^+ mesons contributions

$$\begin{cases} \epsilon^1_{prompt} \cdot N_{prompt} + \epsilon^1_{FD} \cdot N_{FD} = Y^1 \\ \vdots \\ \epsilon^n_{prompt} \cdot N_{prompt} + \epsilon^n_{FD} \cdot N_{FD} = Y^n \end{cases}$$

II). Each set is equivalent to an equation with 2 variables (N_p, N_{np})

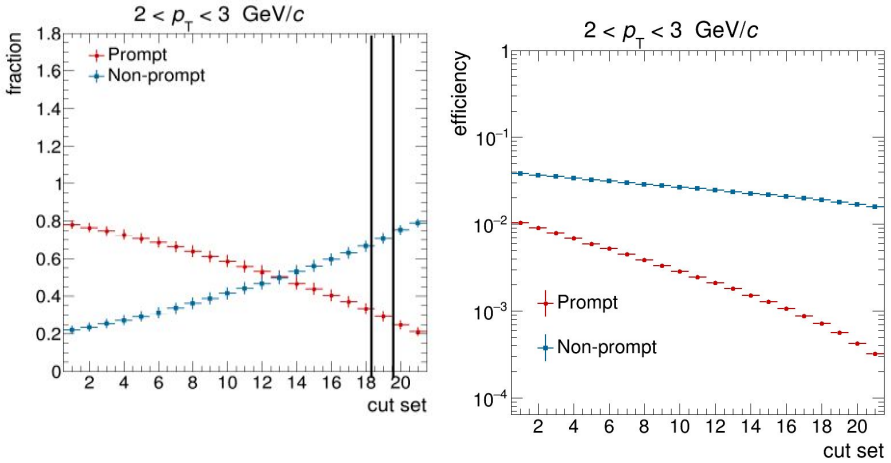
III). System of equations is overdetermined: approximated solution obtained by minimising a χ^2

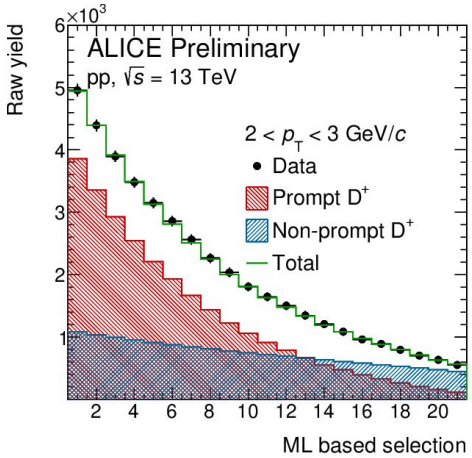
$$\begin{pmatrix} \epsilon^1_{prompt} & \epsilon^1_{FD} \\ \vdots & \vdots \\ \epsilon^n_{prompt} & \epsilon^n_{FD} \end{pmatrix} \times \begin{pmatrix} N_{\mathrm{p}} \\ N_{\mathrm{np}} \end{pmatrix} - \begin{pmatrix} Y1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \delta_2 \\ \vdots \\ \delta_n \end{pmatrix}$$

IV). From the approximated solution (N_p, N_{np}), the non-prompt fraction can be estimated

$$f^j_{FD} = \frac{\varepsilon^j_{FD} N_{FD}}{\varepsilon^j_{FD} N_{FD} + \varepsilon^j_{prompt} N_{prompt}}$$

Non-prompt fraction estimation (D^+, $2 < p_T < 3$ GeV/c)

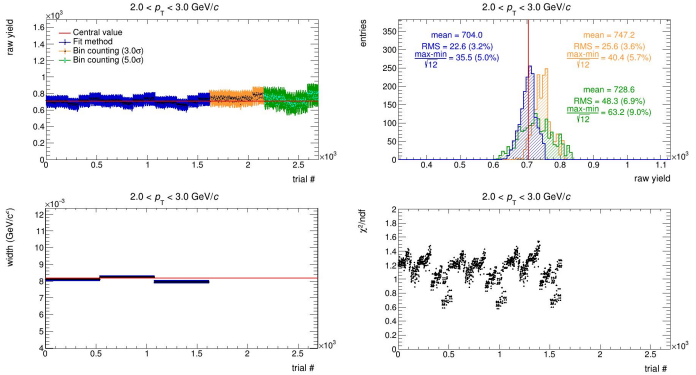


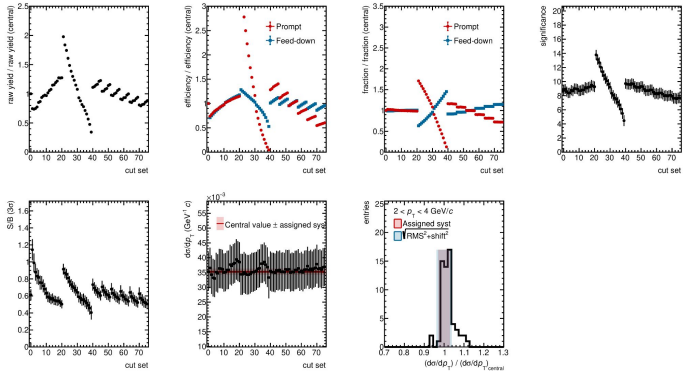


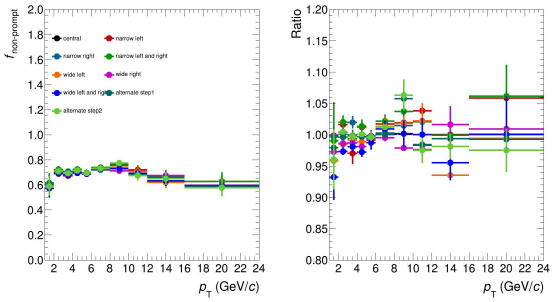
- Central cutset: $f_{np} \sim 70\%$ in full p_T
- With looser selection $\varepsilon_{FD} \approx 5\varepsilon_{prompt}$
- With tighter selection $\varepsilon_{FD} \approx 70\varepsilon_{prompt}$

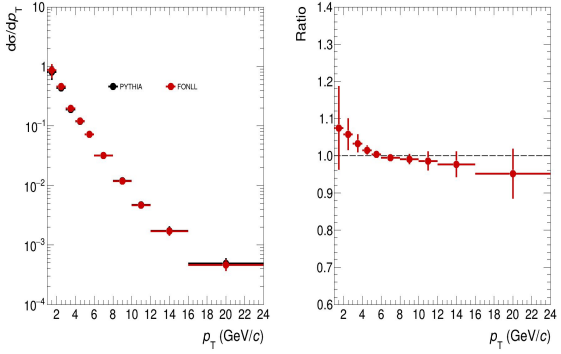
Estimation of systematic uncertainties

- Systematic source: (more details in backup)
  - Raw yield extraction: multi-trial approach
  - Selection efficiency: cut-variations on ML-output score
  - Non-prompt fraction estimation ($f_{\text{non-prompt}}$): data-driven method
  - MC p_{T} shape: repeat full analyses applying p_{T} weights from FONLL p_{T} shapes for generated signal in MC simulations
  - PID: inherited from prompt D^{+} meson analyses
  - Tracking: consider single-track systematic uncertainty and ITS-TPC matching efficiency using the D-meson decay kinematics

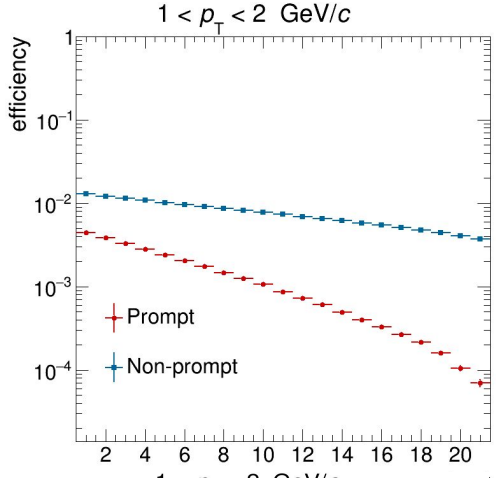


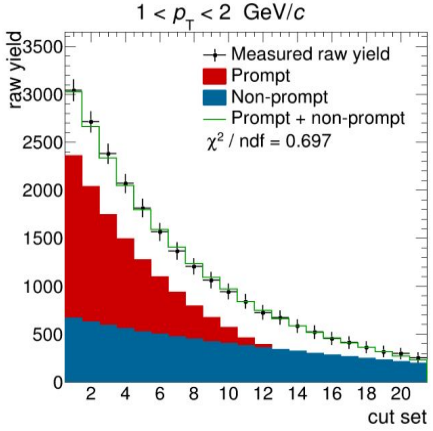


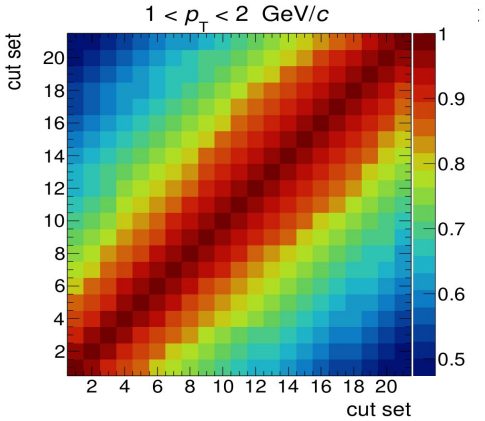


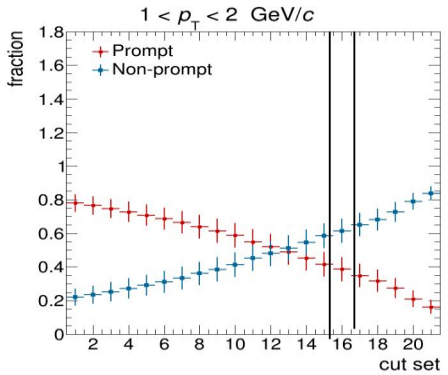


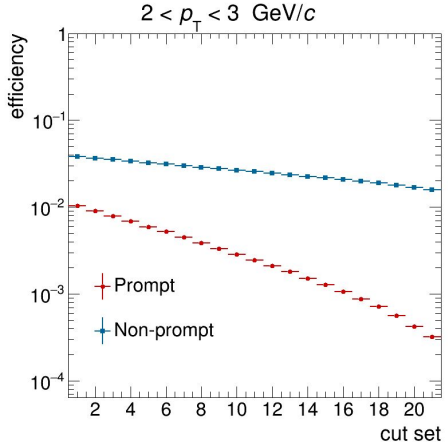
D^+ f_{FD} estimation ([1-2], [2-3])

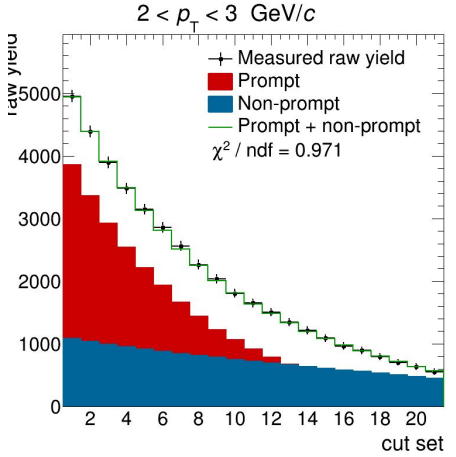


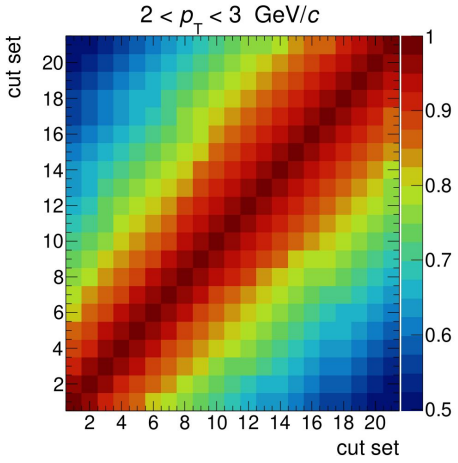


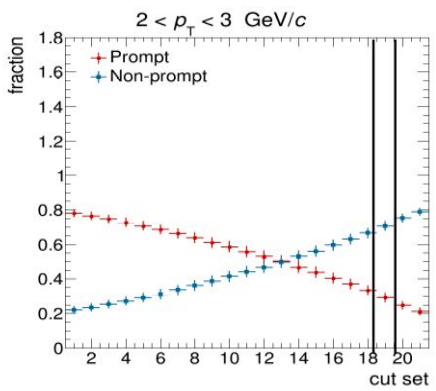




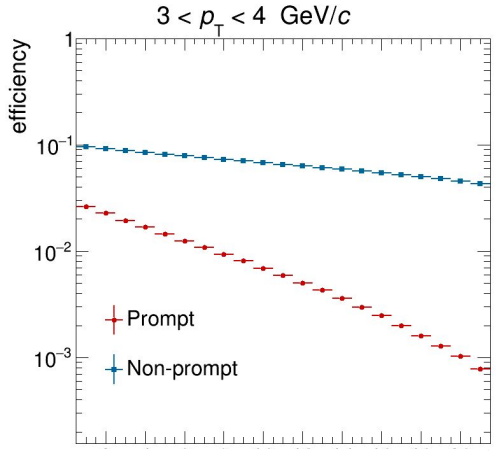


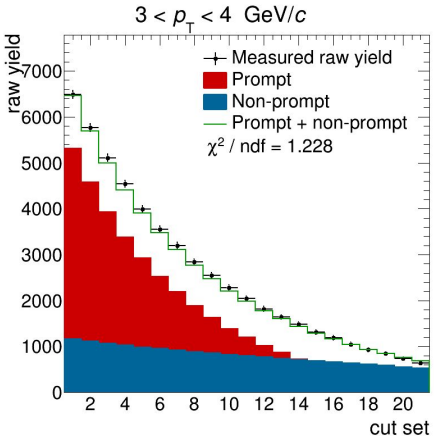


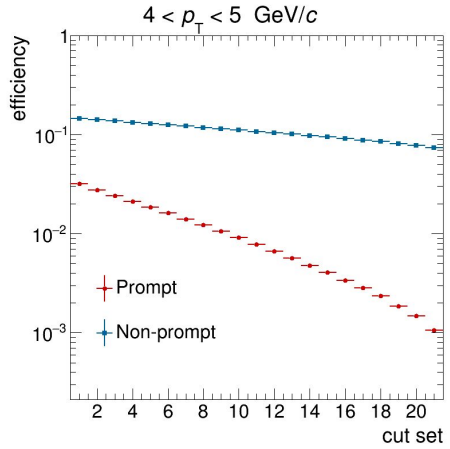


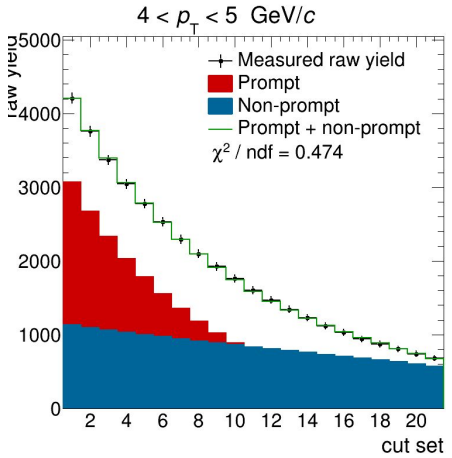


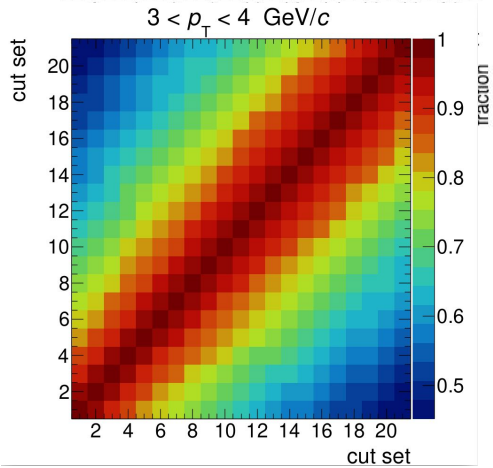
D^+ f_{FD} estimation ([3-4], [4-5])

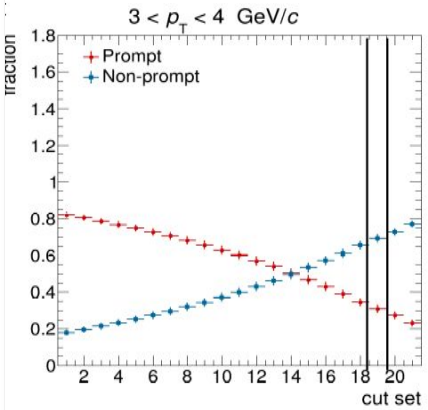


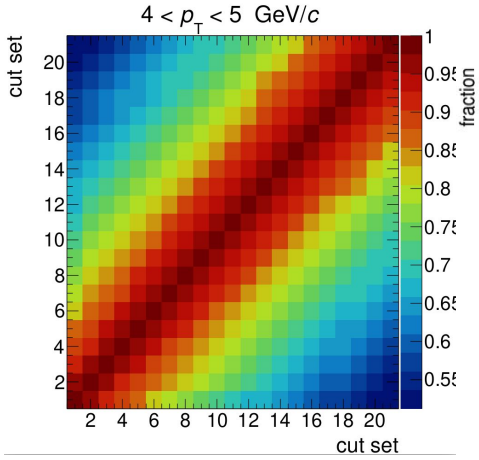


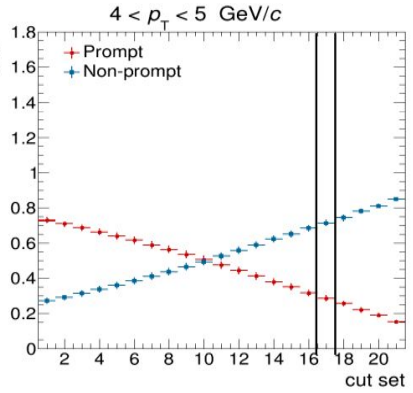




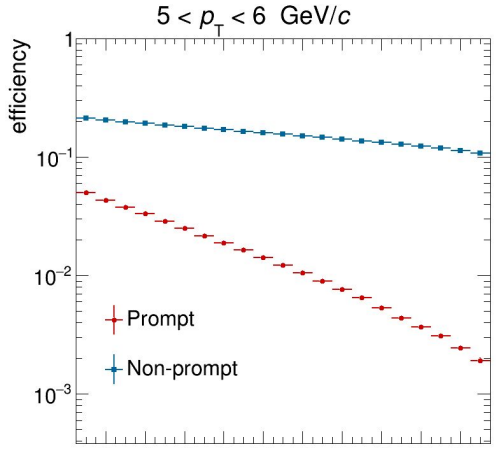


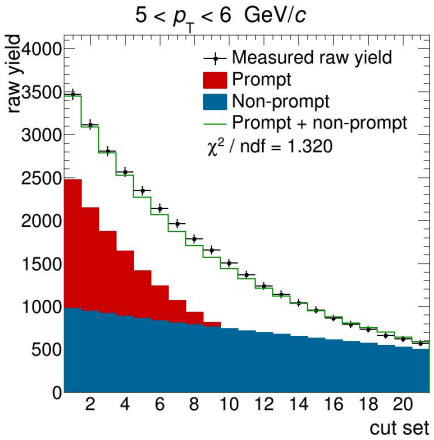


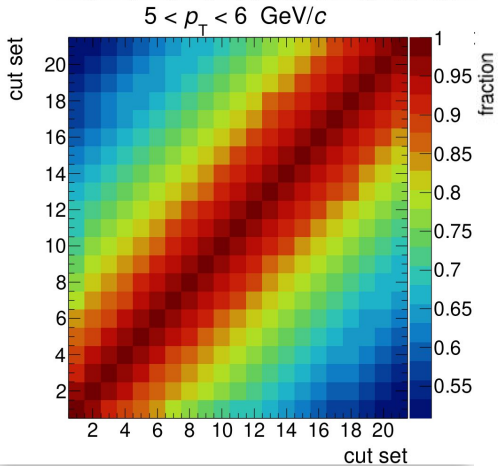


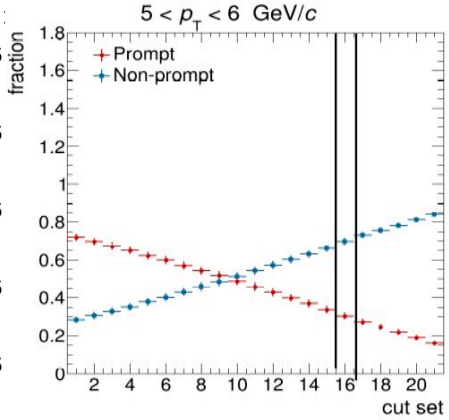


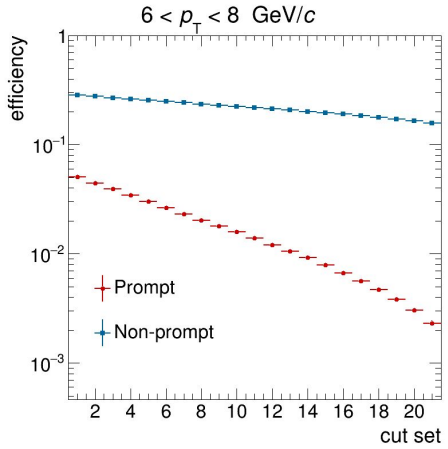
D^+ f_{FD} estimation ([5-6], [6-8])

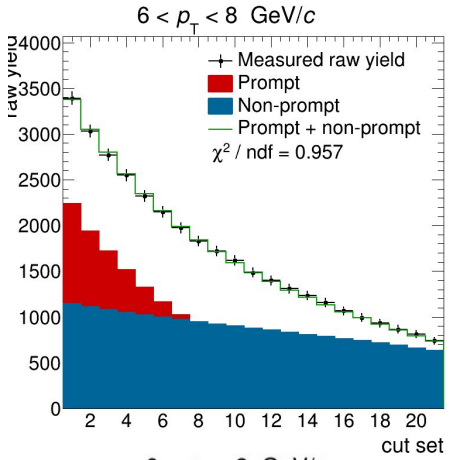


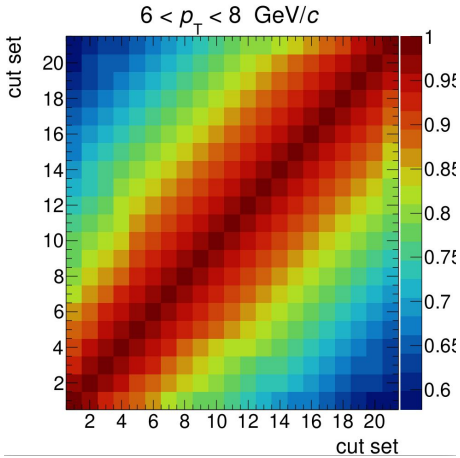


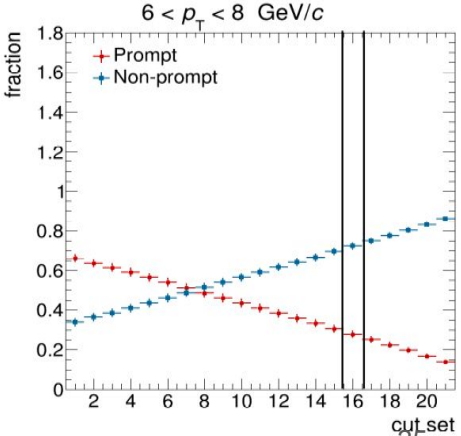




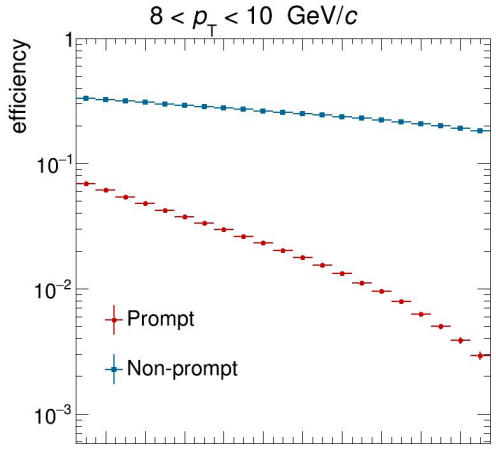


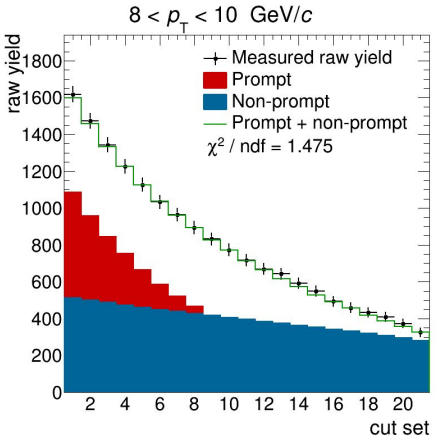


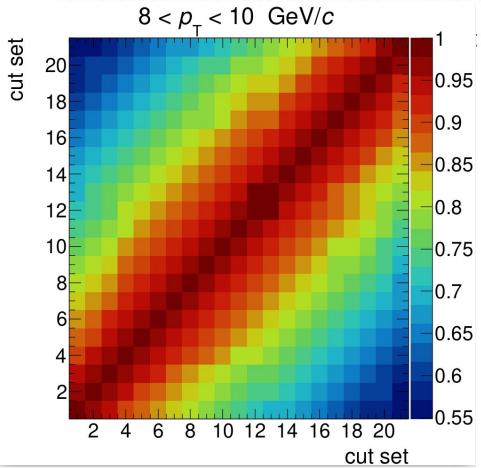


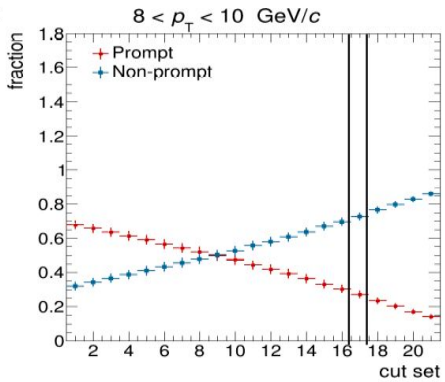


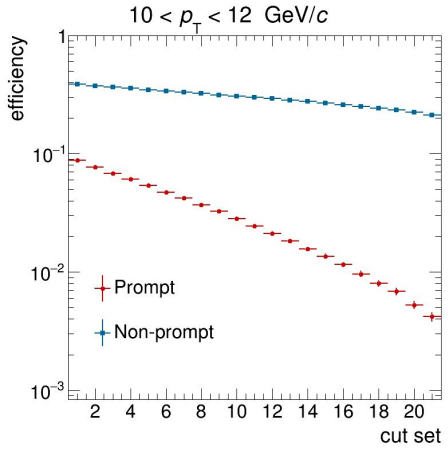
D^+ f_{FD} estimation ([8-10], [10-12])

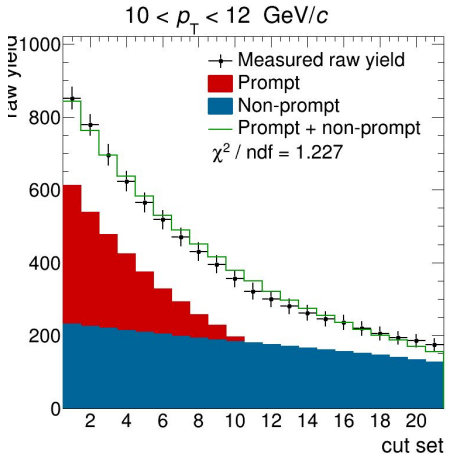


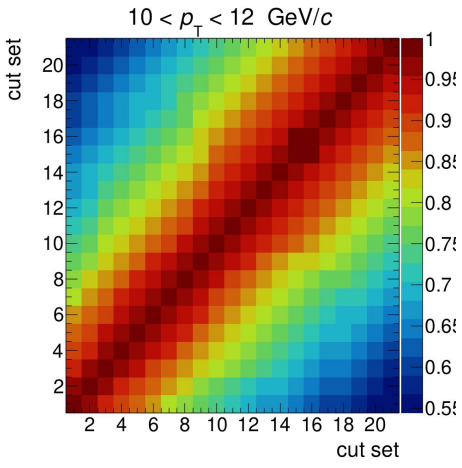


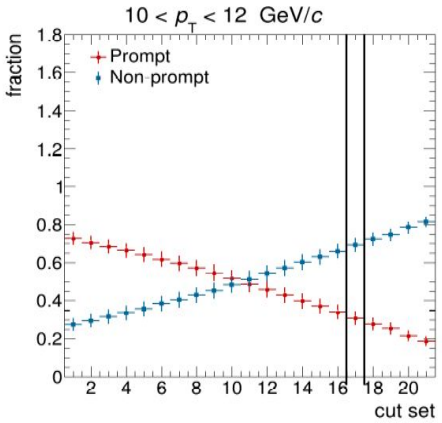




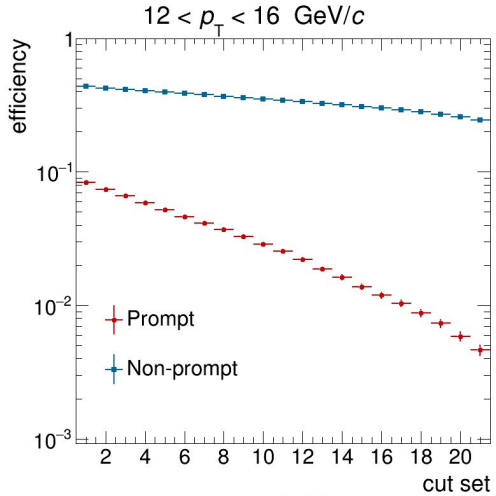


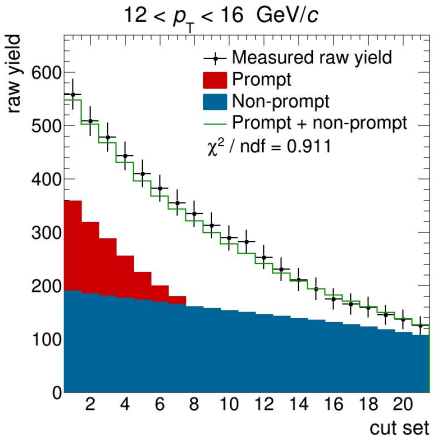


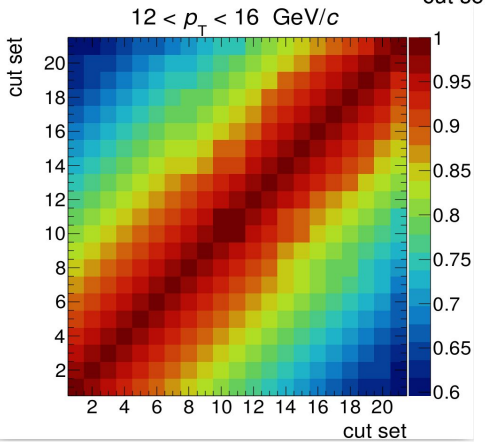


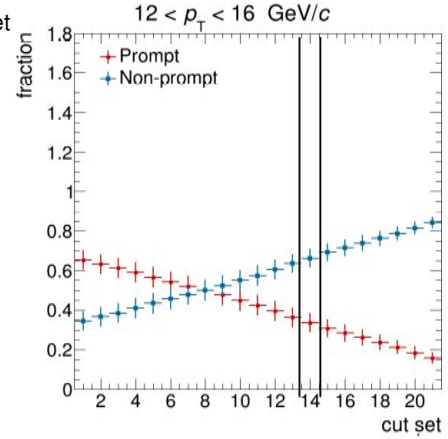


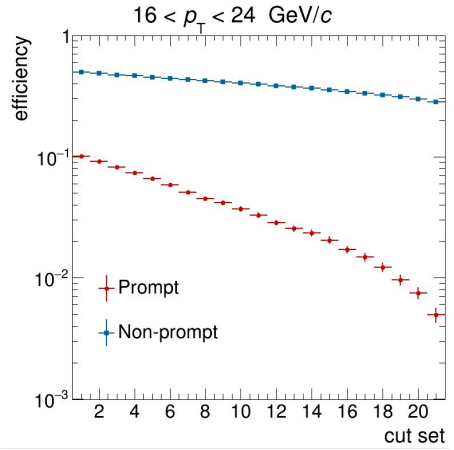
D^+ f_{FD} estimation ([12-16], [16-24])

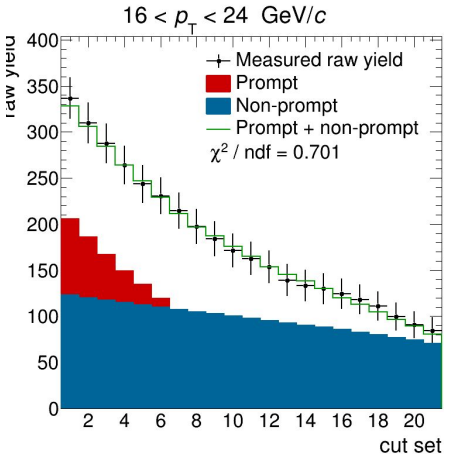


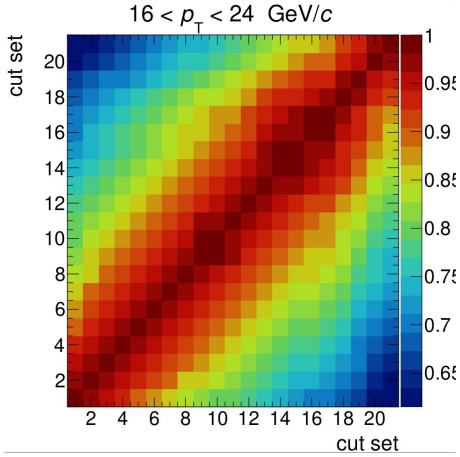


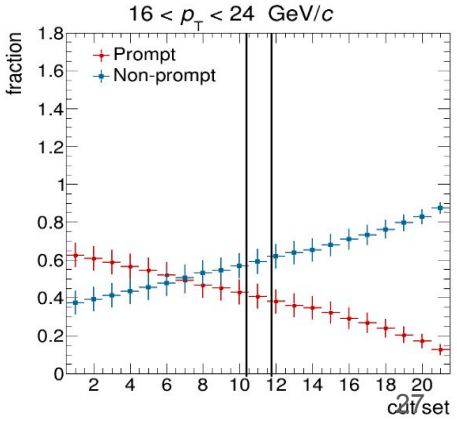




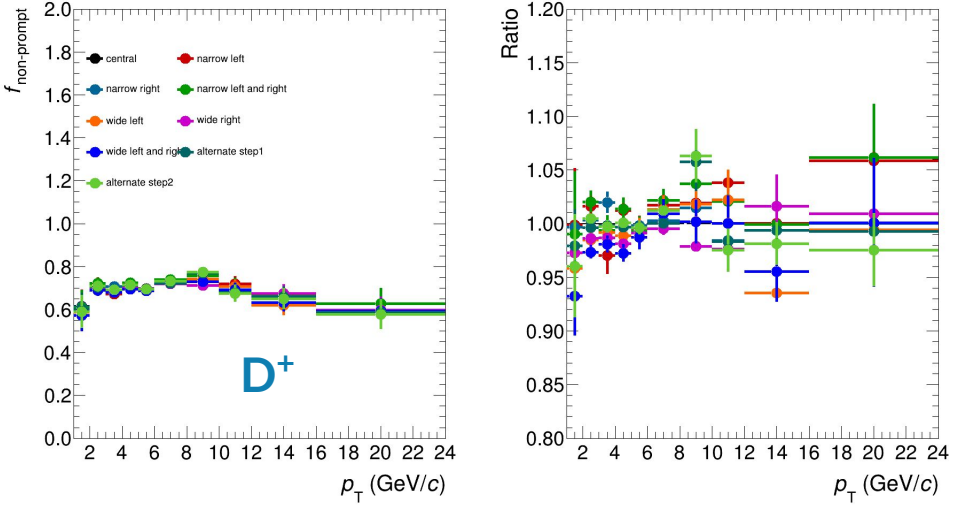








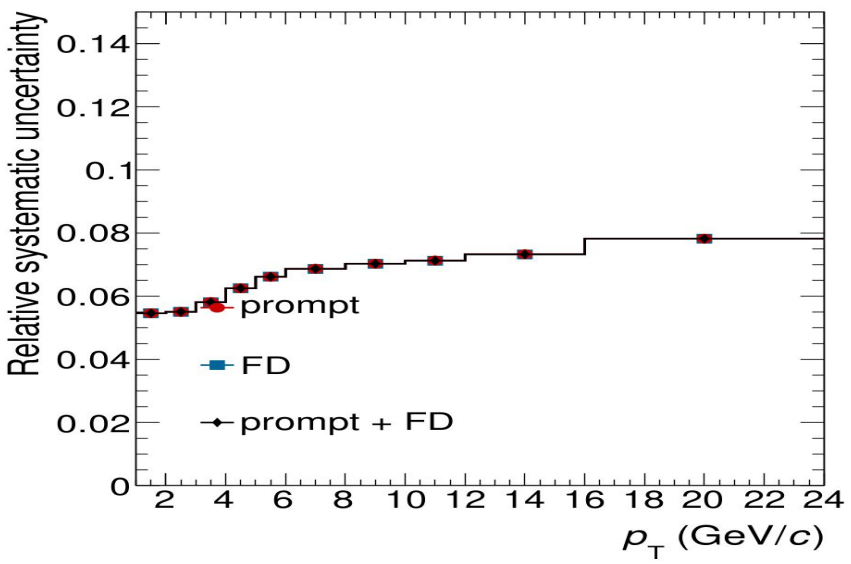
Systematic uncertainties: non-prompt fraction



Configuration	Meaning
Narrow	tightest (“*right*”) and loosest (“*left*”) cut sets are **removed** from the minimisation
Wide	tighter (“*right*”) and looser (“*left*”) cut sets are **added** in the minimisation
alt step	**different step sizes** are considered among the cut sets

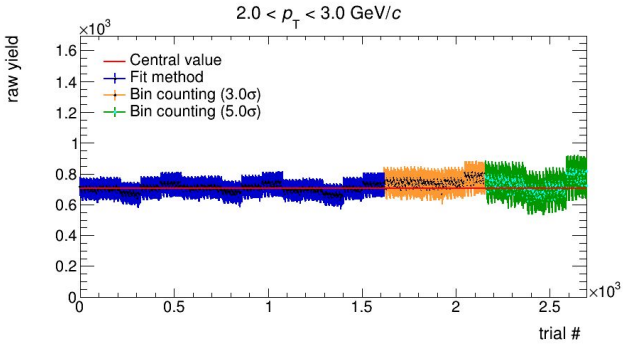
- The systematic uncertainty on the non-prompt fraction is evaluated by varying the sets of cuts considered in the system minimisation
- Assigned uncertainty range from 3% to 5% for D^+ and 2% to 10% for D_s^+.

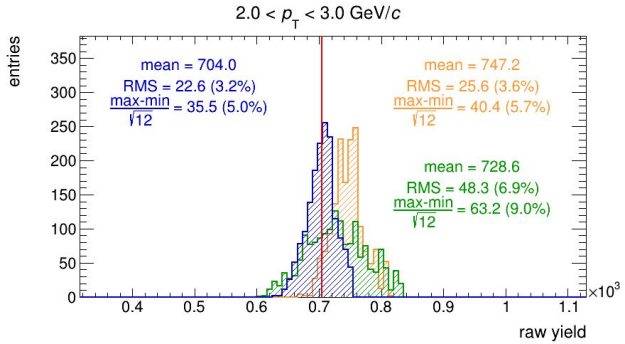
Systematic uncertainties: tracking efficiency for D^+

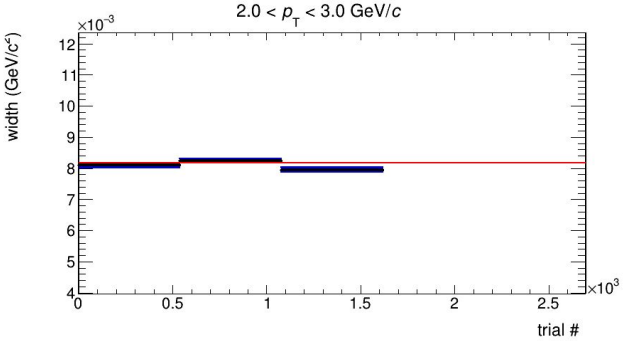


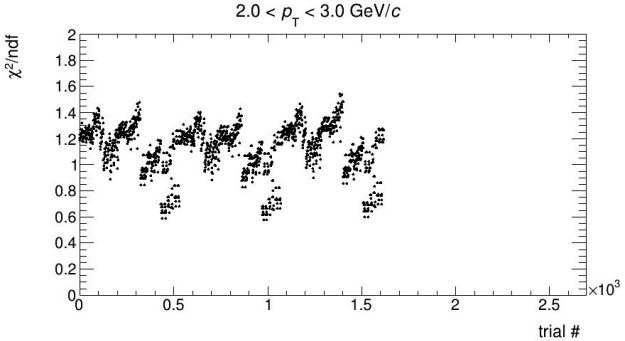
- Propagate single-track systematic uncertainty on ITS-TPC matching efficiency using decay kinematics of non-prompt D^+.
  - tracking selection efficiency systematic for non-prompt D are taken from DPG
- Assigned uncertainty range from 6% to 8% for non-prompt D^+

Systematic uncertainties: D^+ raw-yield - $2 < p_T < 3$ GeV/c



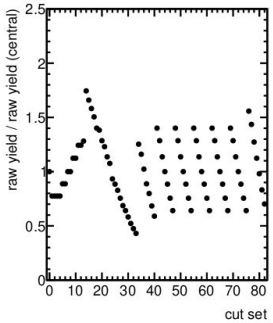


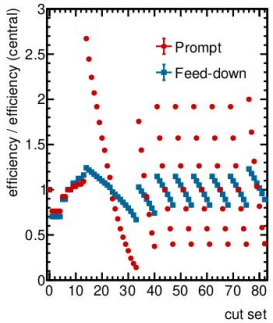


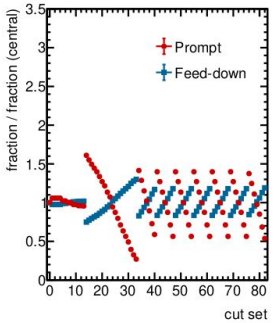


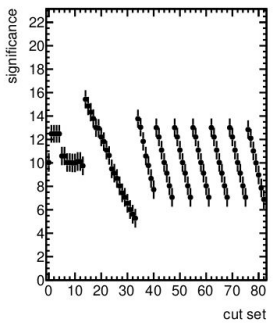
- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
- Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%

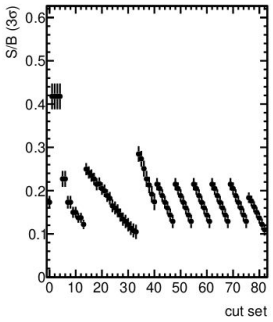
Systematic uncertainties: ML sel. eff. - D^+, $2 < p_T < 3$ GeV/c

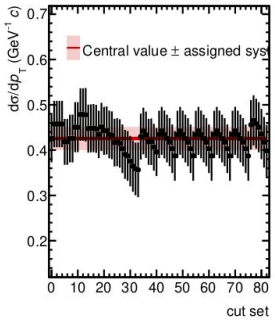


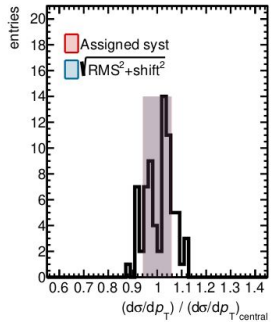






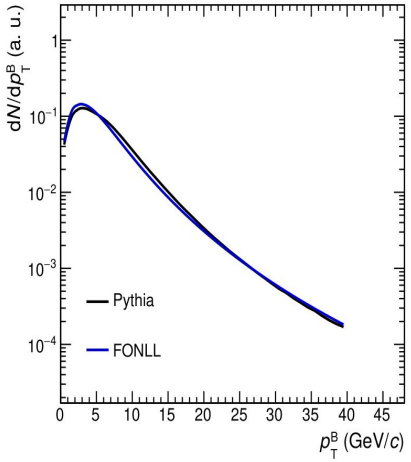


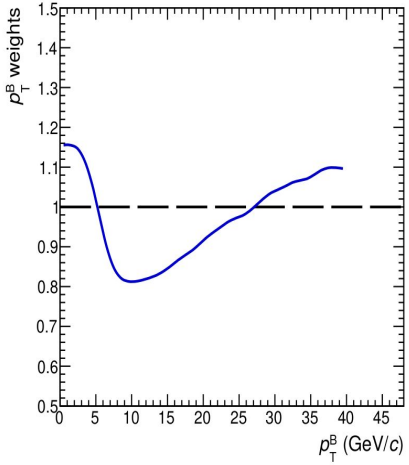


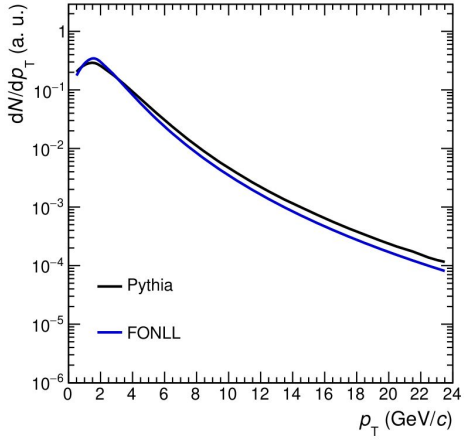


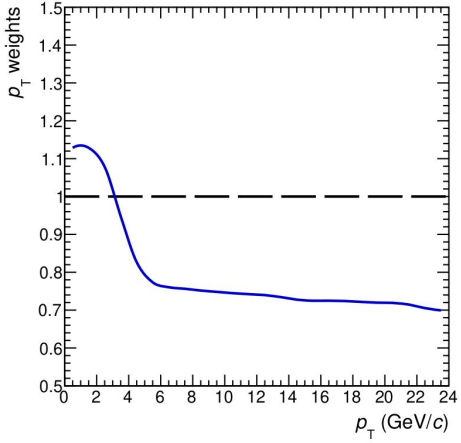
- 12 different Bkg_score variations
- 20 different FD_score variations
- 48 simultaneous Bkg_score & FD_score variations
- Quality check:
  - signif. > 3, $\chi^2 < 2$
  - 0.5 < rel. eff. variation < 2.5
- Systematic uncertainties assigned range from 4% to 10%

Systematic uncertainties: MC p_T shape - D^+





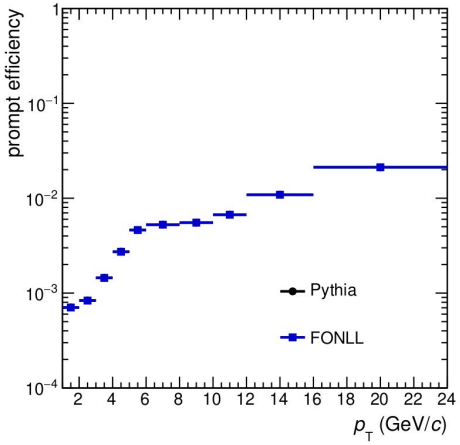


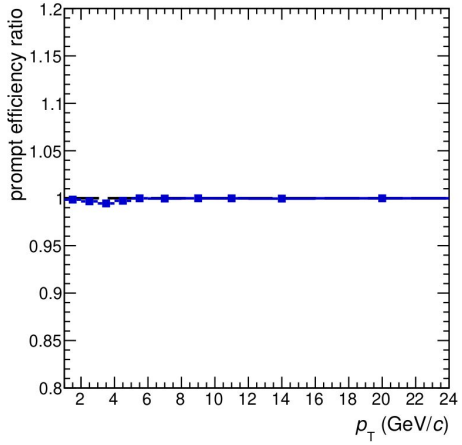


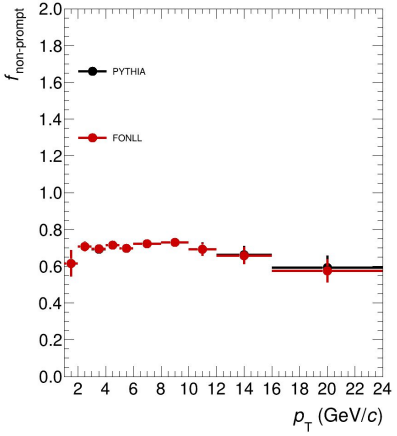
p_T shape in the MC re-weighted in order to reproduce a realistic distribution:

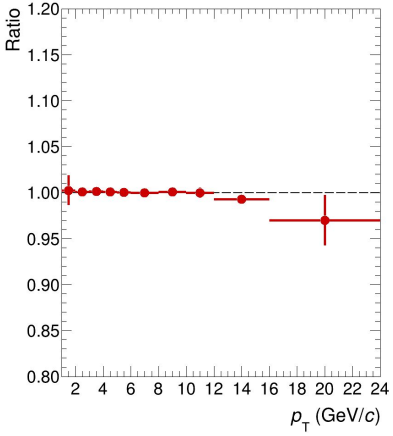
- reference case: PYTHIA
- p_T weights computed using FONLL shape for prompt D
- p_T weights computed using FONLL shape for B (mixture)

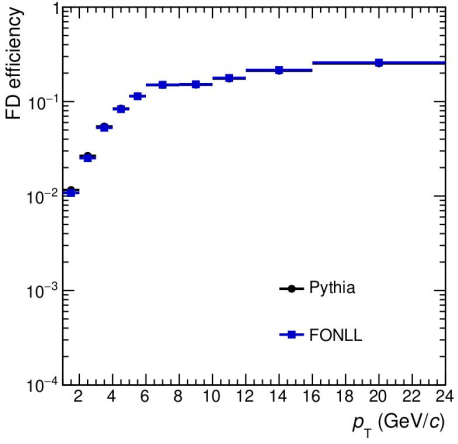
Systematics: MC p_T shape D^+

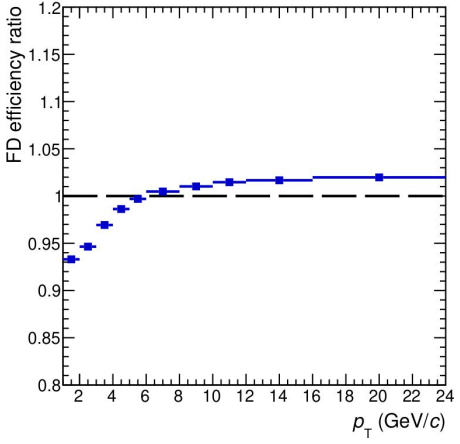


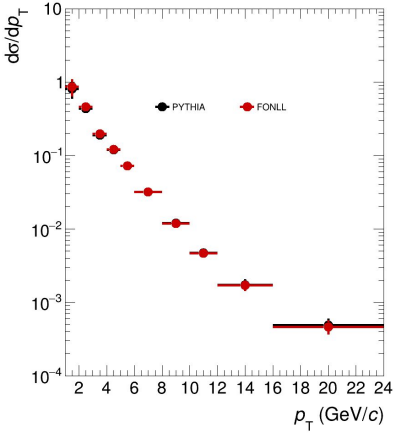


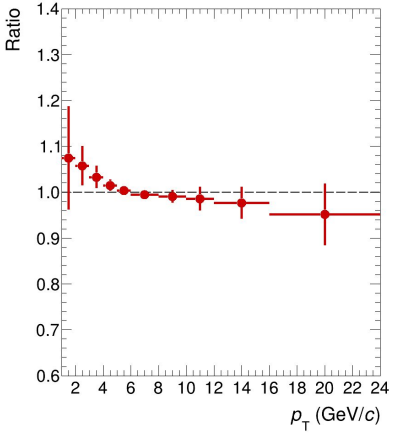




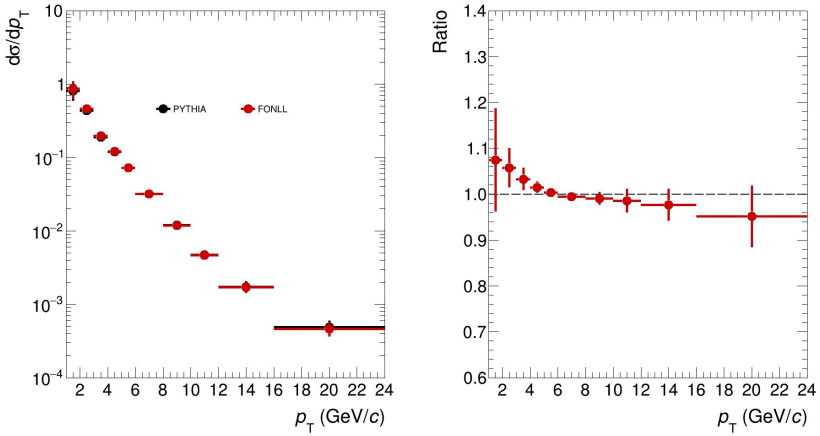








Systematic uncertainties: MC p_T shape - D^+

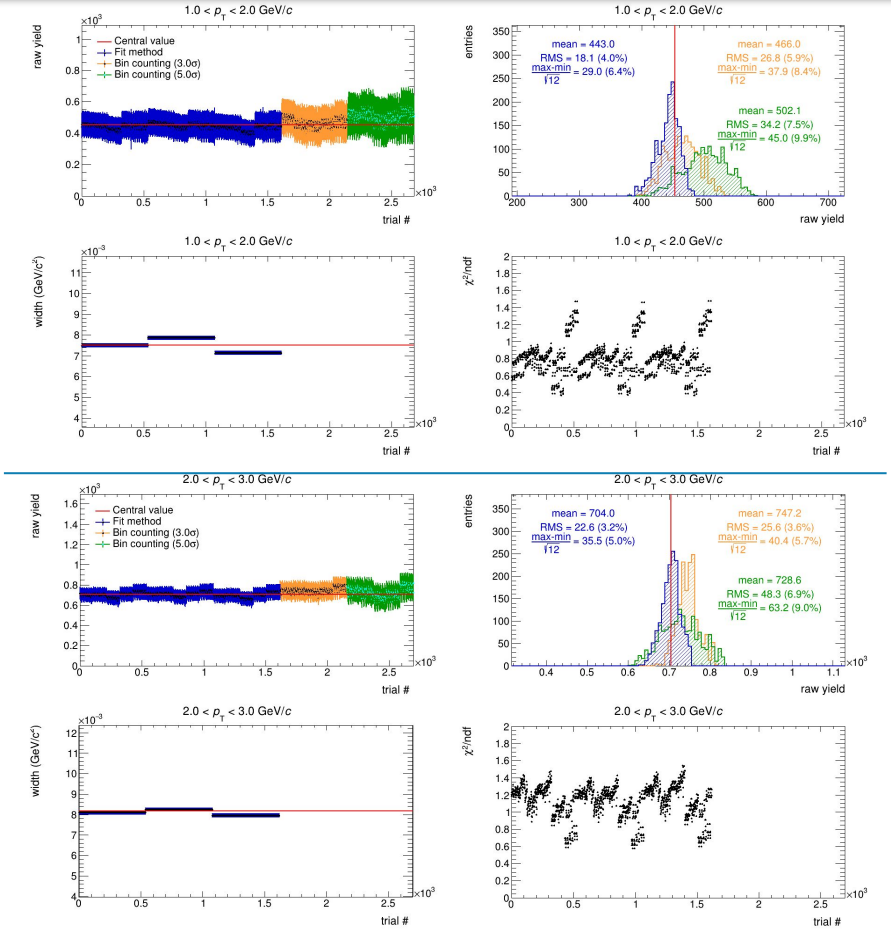


Repeated full analyses with and w/o MC p_T weights:

- FONLL
- PYTHIA (reference)
- Syst. unc. estimated considering the effect on $f_{\text{non-prompt}}$ and cross-section

– Assigned uncertainty ranges from 2% to 6% for D^+

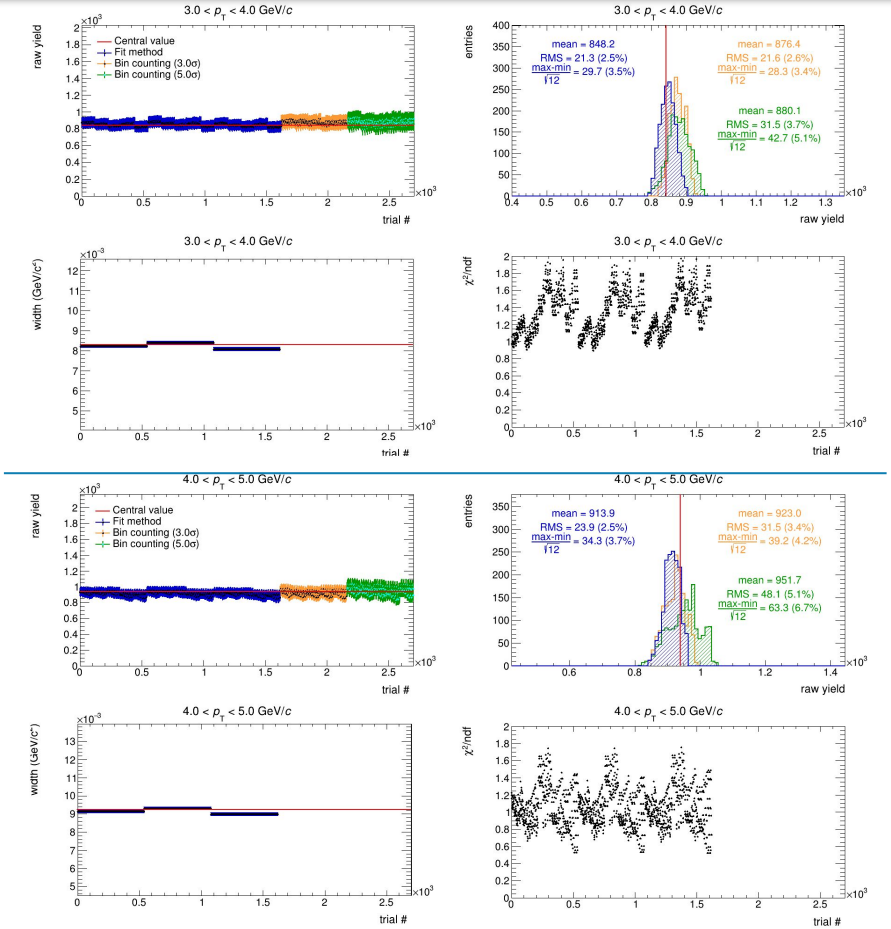
Systematics: D^+ raw yields extraction ([1-2], [2-3])



Multi-trial fit

- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
  - Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%
    - ➡ [5%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 5%, 5%]

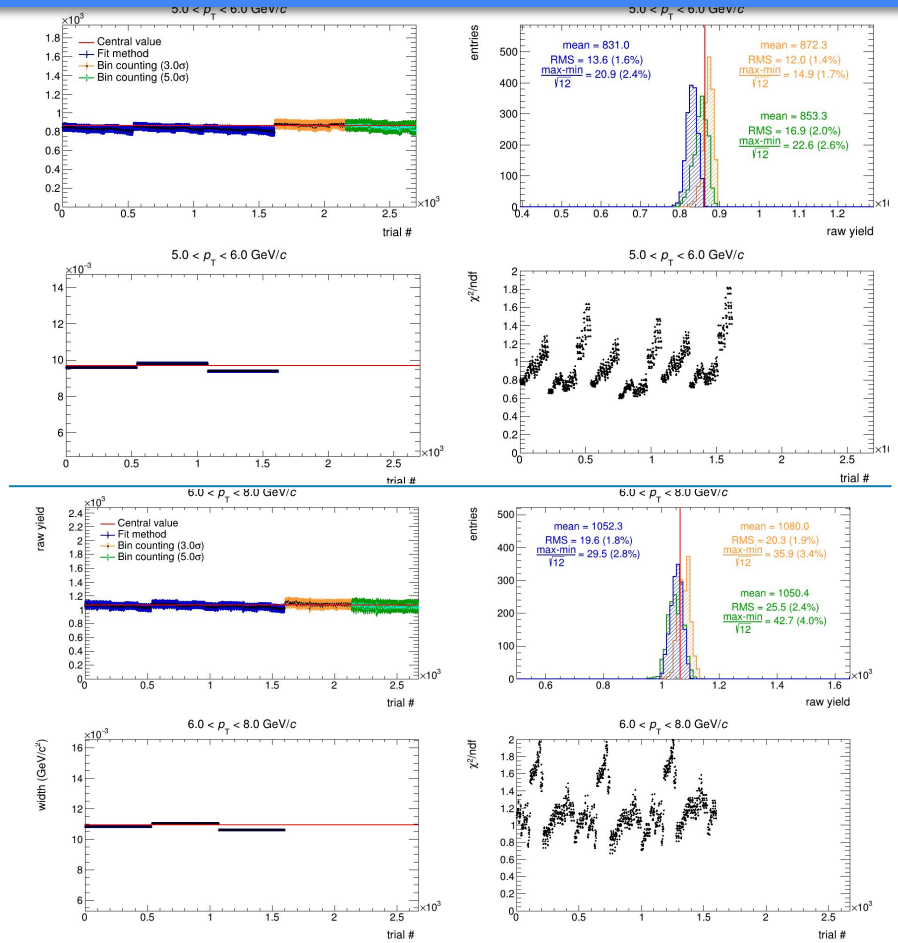
Systematics: D^+ raw yields extraction ([3-4], [4-5])



Multi-trial fit

- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
  - Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%
    - ➡ [5%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 5%, 5%]

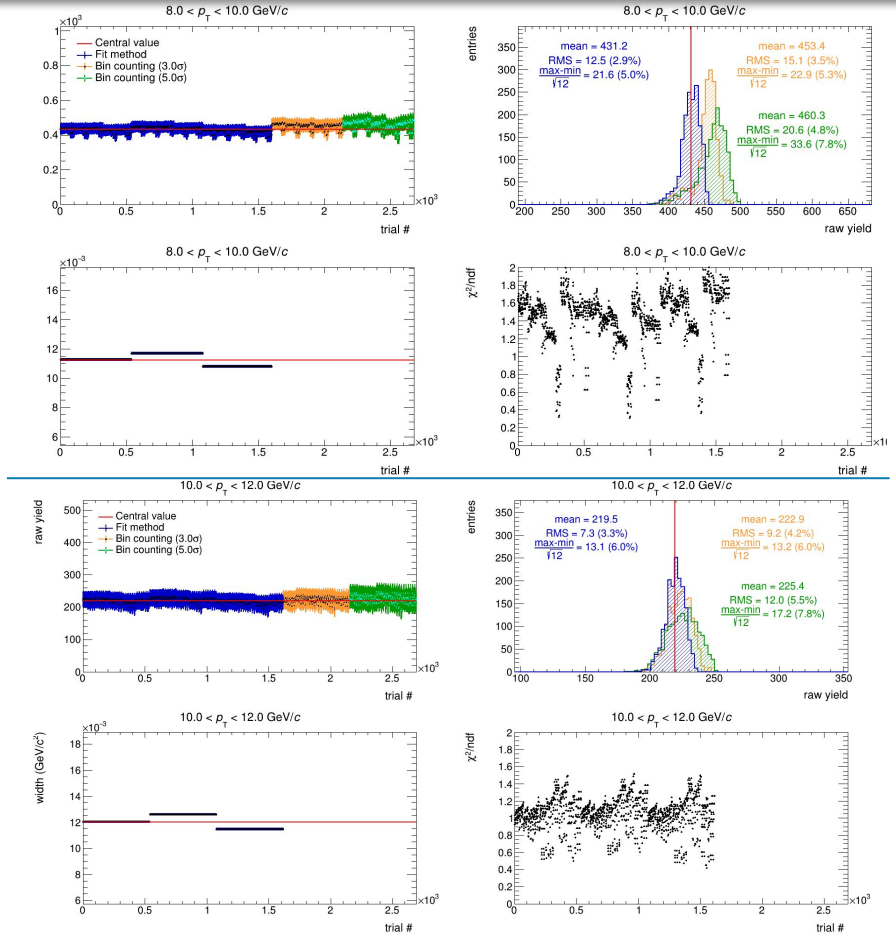
Systematics: D^+ raw yields extraction ([5-6], [6-8])



Multi-trial fit

- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
  - Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%
    - ➡ [5%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 5%, 5%]

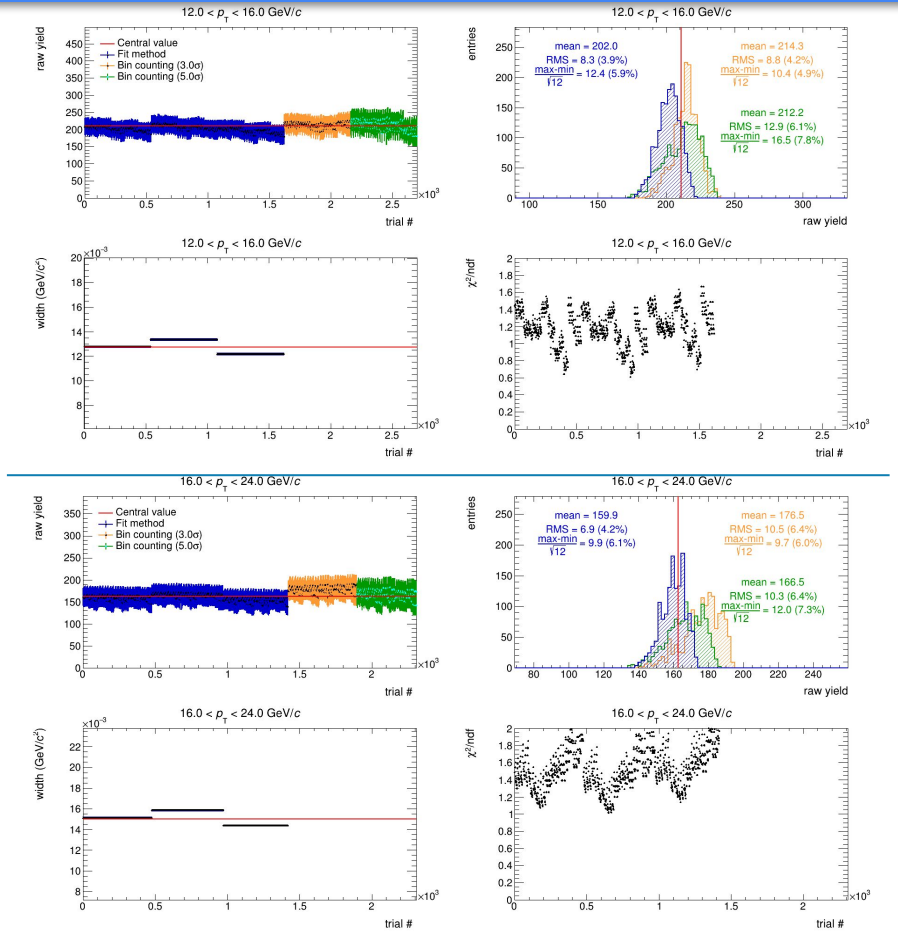
Systematics: D^+ raw yields extraction ([8-10], [10-12])



Multi-trial fit

- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
  - Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%
    - ➡ [5%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 5%, 5%]

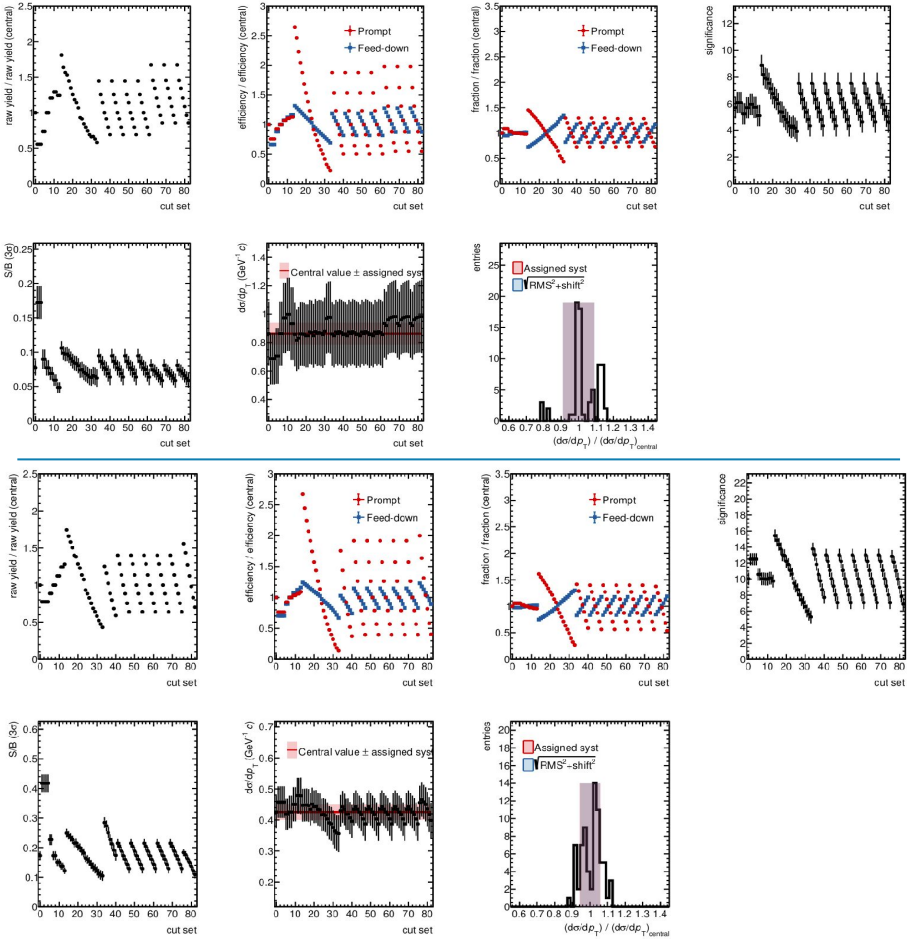
Systematics: D^+ raw yields extraction ([12-16], [16-24])



Multi-trial fit

- 3 bkg functions (lin, pol2, expo)
- different upper/lower limits
- 5 different rebin
- mean: free
- sigma: fixed to prompt-en. ± unc
  - Syst. unc. estimated as sum in quadrature of RMS and shift w.r.t. the trial distribution
  - Assigned uncertainty range from 4% to 5%
    - ➡ [5%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 4%, 5%, 5%]

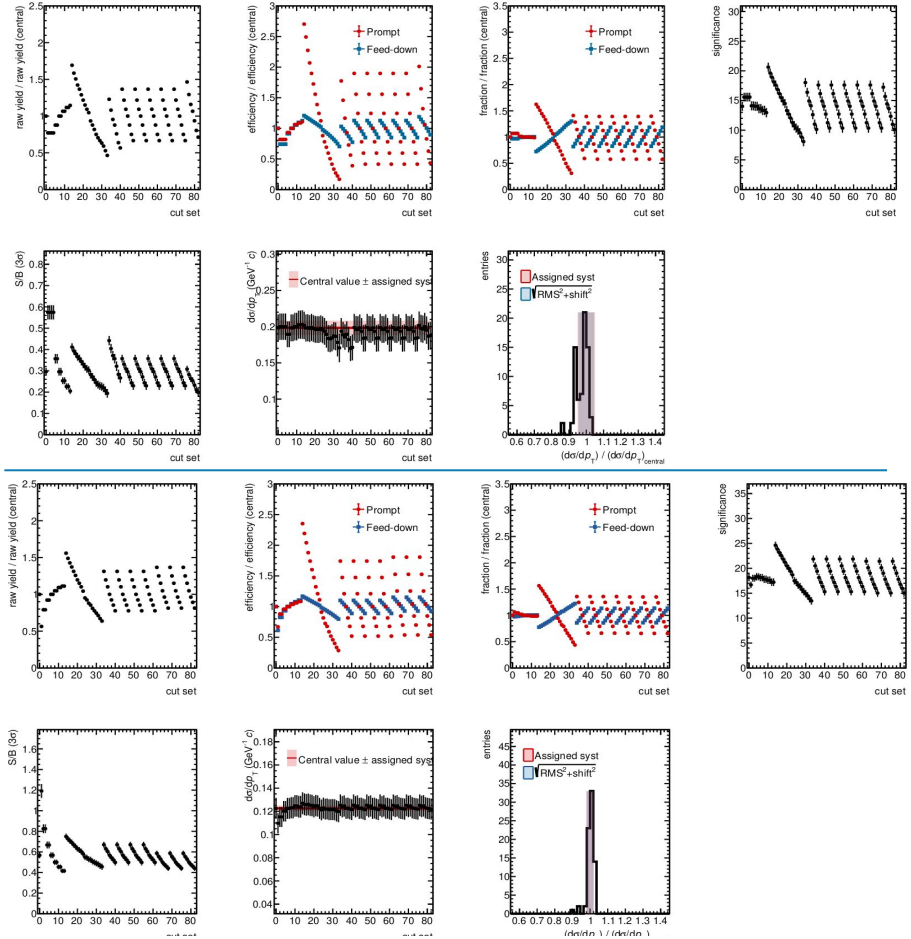
Systematics: D^+ selection efficiency ([1-2],[2-3])



Full analysis repeated with different central selections:

- 12 different Bkg_score selections (6 tighter, 6 looser)
- 20 different FD_score selections (10 tighter, 10 looser)
- around 45 different Bkg_score & FD_score selections
- Quality check:
  - signif. > 3
  - 0.5 < rel. eff. variation < 2.5
- Systematic evaluated as the sum in quadrature of RMS and shift on the relative variation of the corrected yield
  - Assigned uncertainty range from 4% to 10%
    - ➡ [10%, 6%, 5%, 4%, 4%, 4%, 4%, 4%, 4%. 4%]

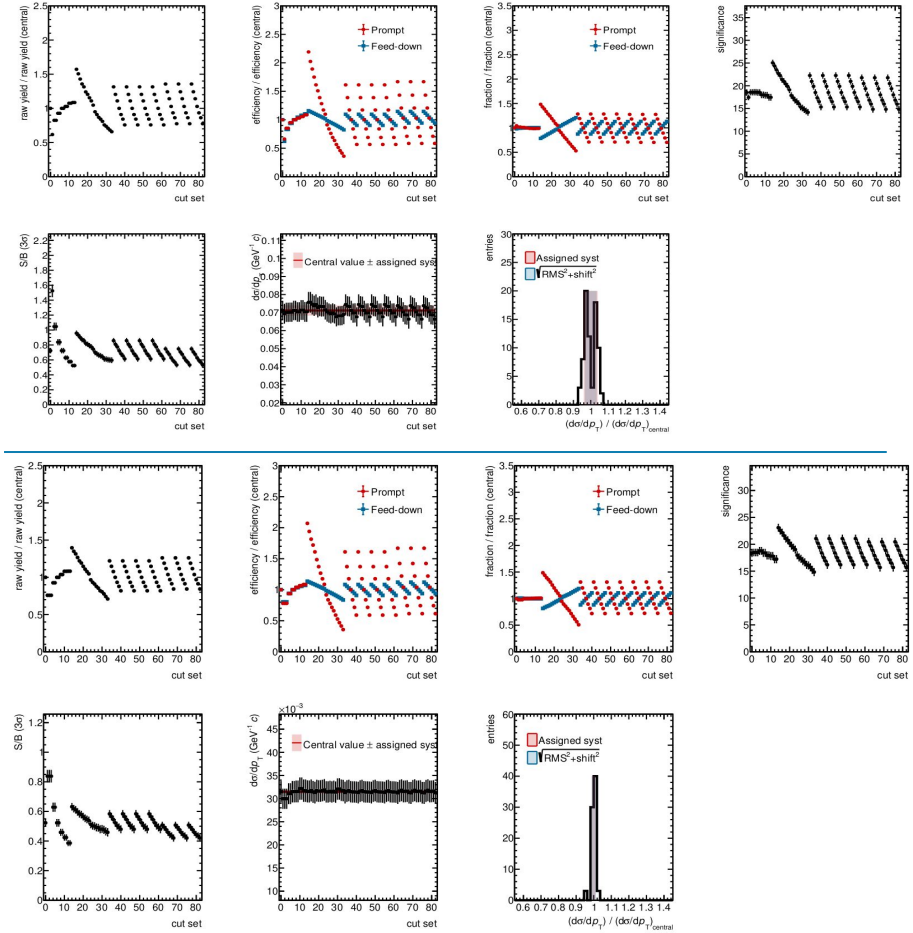
Systematics: D^+ selection efficiency ([3-4],[4-5])



Full analysis repeated with different central selections:

- 12 different Bkg_score selections (6 tighter, 6 looser)
- 20 different FD_score selections (10 tighter, 10 looser)
- around 45 different Bkg_score & FD_score selections
- Quality check:
  - signif. > 3
  - 0.5 < rel. eff. variation < 2.5
- Systematic evaluated as the sum in quadrature of RMS and shift on the relative variation of the corrected yield
  - Assigned uncertainty range from 4% to 10%
    - ➡ [10%, 6%, 5%, 4%, 4%, 4%, 4%, 4%, 4%. 4%]

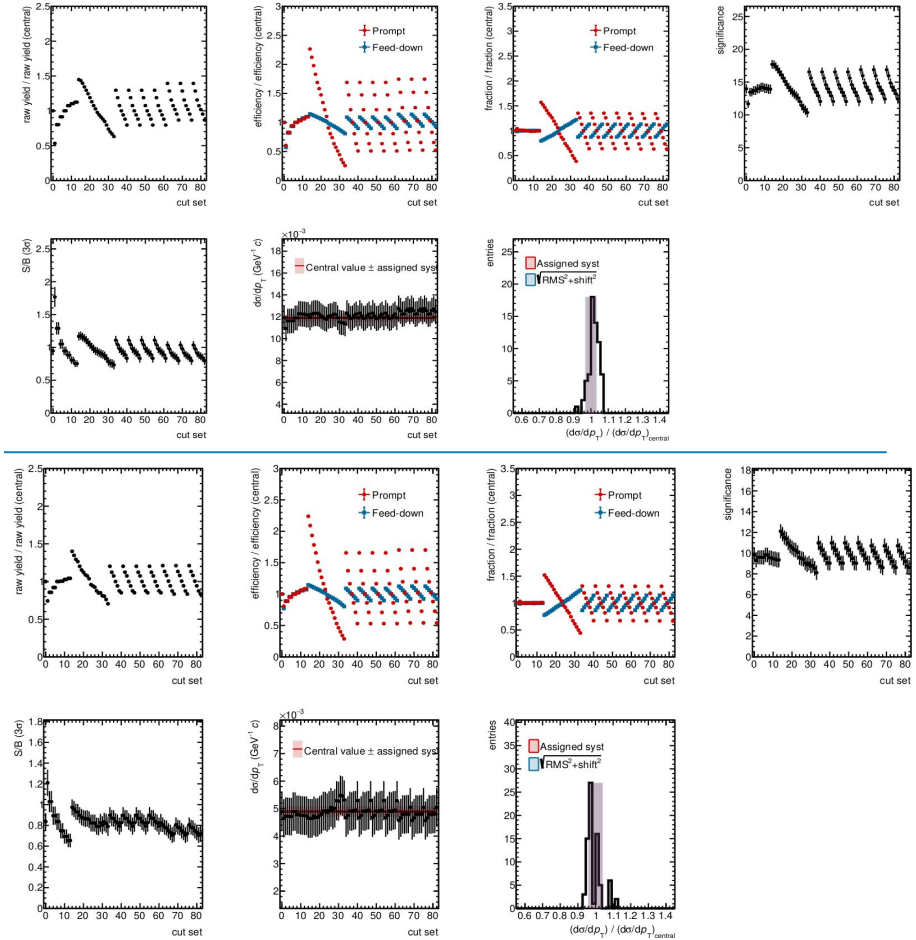
Systematics: D^+ selection efficiency ([5-6],[6-8])



Full analysis repeated with different central selections:

- 12 different Bkg_score selections (6 tighter, 6 looser)
- 20 different FD_score selections (10 tighter, 10 looser)
- around 45 different Bkg_score & FD_score selections
- Quality check:
  - signif. > 3
  - 0.5 < rel. eff. variation < 2.5
- Systematic evaluated as the sum in quadrature of RMS and shift on the relative variation of the corrected yield
  - Assigned uncertainty range from 4% to 10%
    - ➡ [10%, 6%, 5%, 4%, 4%, 4%, 4%, 4%, 4%. 4%]

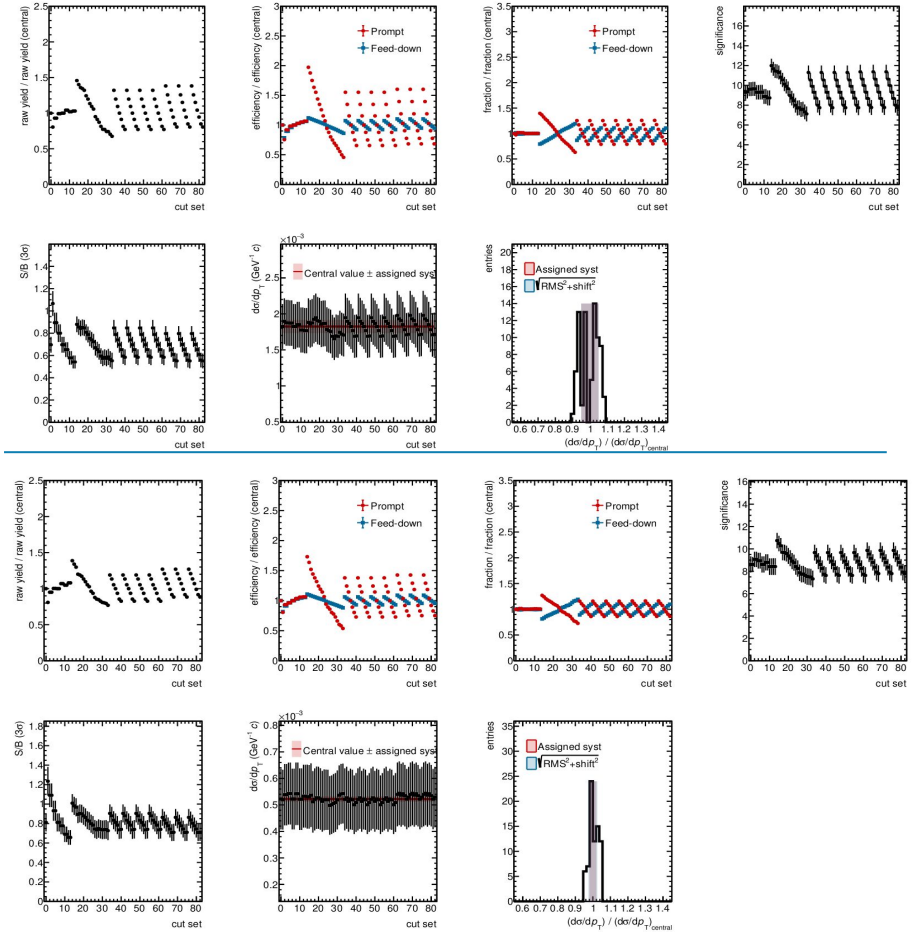
Systematics: D^+ selection efficiency ([8-10],[10-12])



Full analysis repeated with different central selections:

- 12 different Bkg_score selections (6 tighter, 6 looser)
- 20 different FD_score selections (10 tighter, 10 looser)
- around 45 different Bkg_score & FD_score selections
- Quality check:
  - signif. > 3
  - 0.5 < rel. eff. variation < 2.5
- Systematic evaluated as the sum in quadrature of RMS and shift on the relative variation of the corrected yield
  - Assigned uncertainty range from 4% to 10%
    - ➡ [10%, 6%, 5%, 4%, 4%, 4%, 4%, 4%, 4%. 4%]

Systematics: D^+ selection efficiency ([12-16],[16-24])

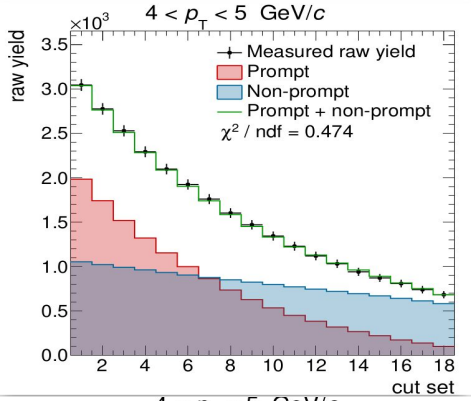


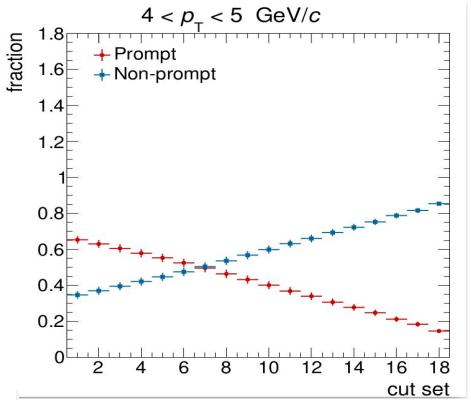
Full analysis repeated with different central selections:

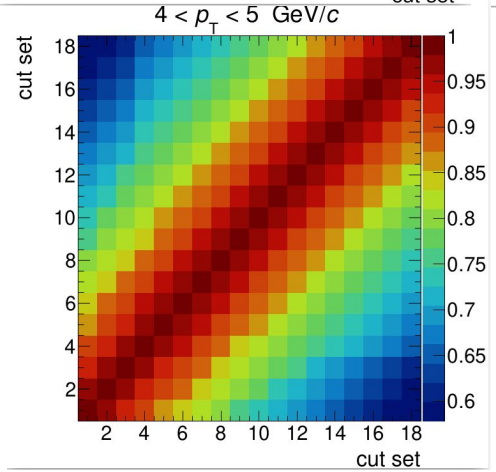
- 12 different Bkg_score selections (6 tighter, 6 looser)
- 20 different FD_score selections (10 tighter, 10 looser)
- around 45 different Bkg_score & FD_score selections
- Quality check:
  - signif. > 3
  - 0.5 < rel. eff. variation < 2.5
- Systematic evaluated as the sum in quadrature of RMS and shift on the relative variation of the corrected yield
  - Assigned uncertainty range from 4% to 10%
    - ➡ [10%, 6%, 5%, 4%, 4%, 4%, 4%, 4%, 4%. 4%]

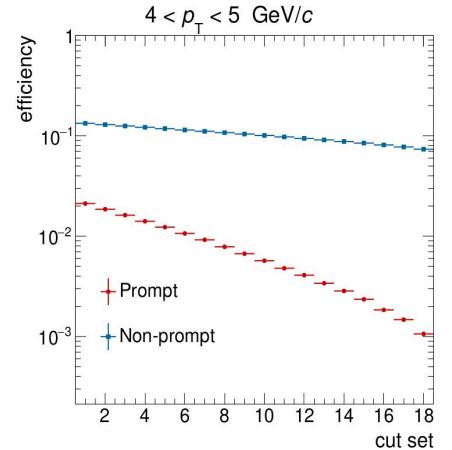
Systematics: D^+ f_{FD} [4-5] -narrow right & left

narrow left

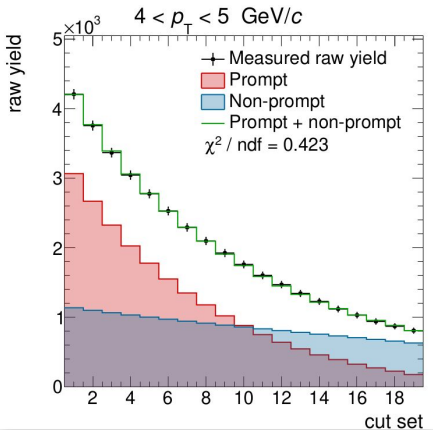


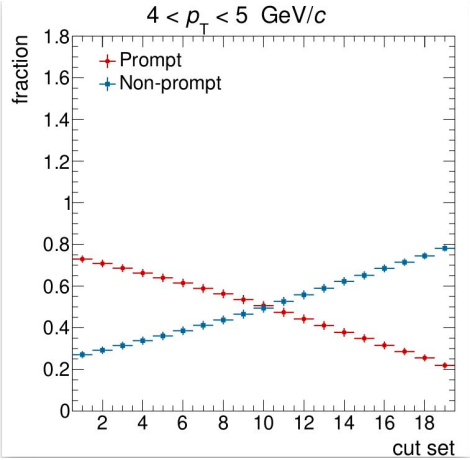


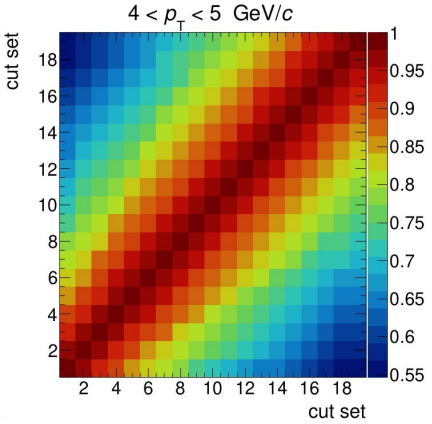


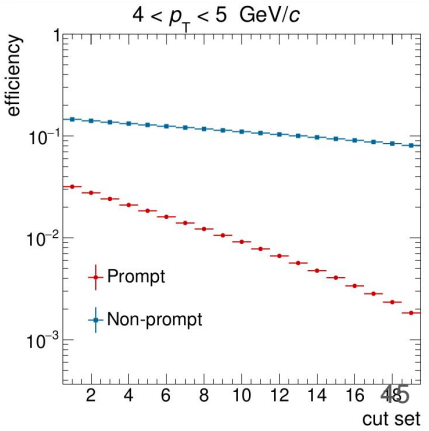


narrow right



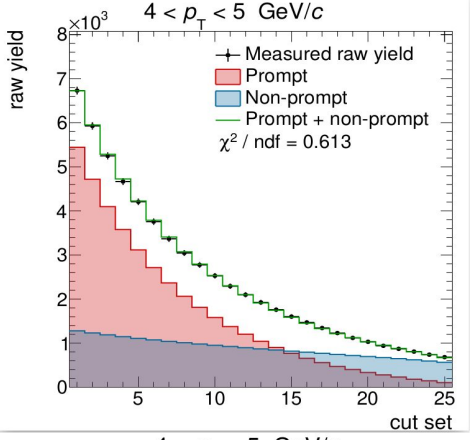


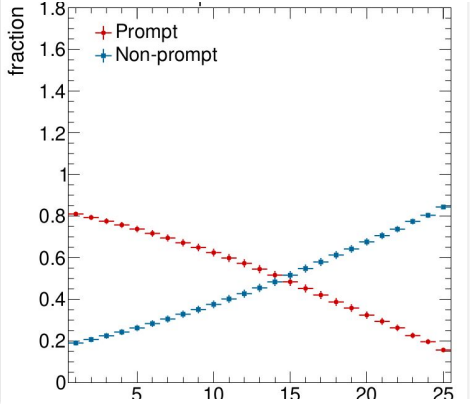


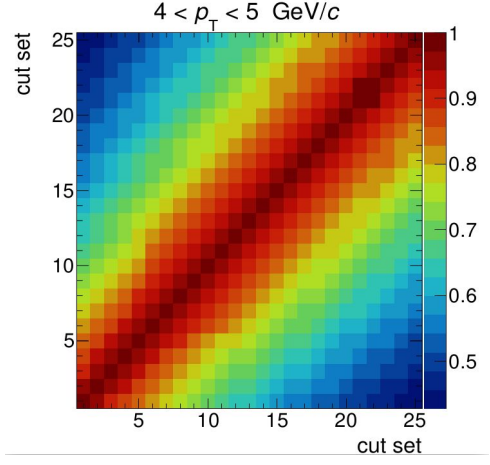


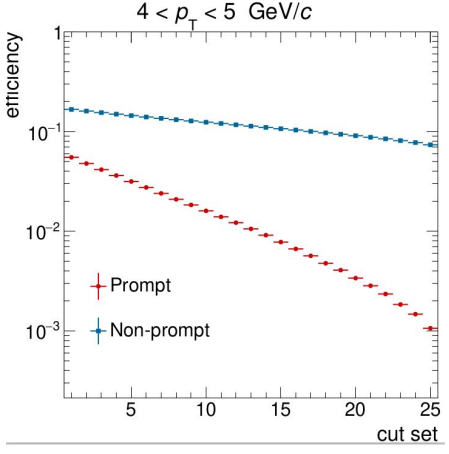
Systematics: D^+ f_{FD} [4-5] -narrow right & left

Wide left

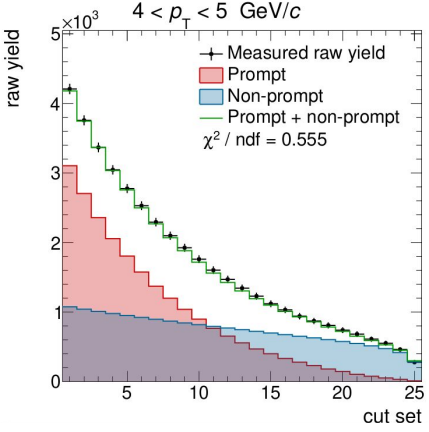


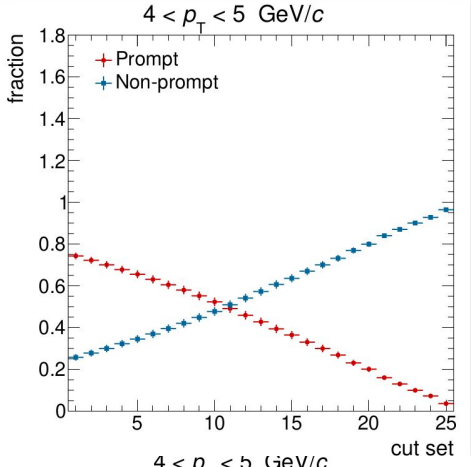


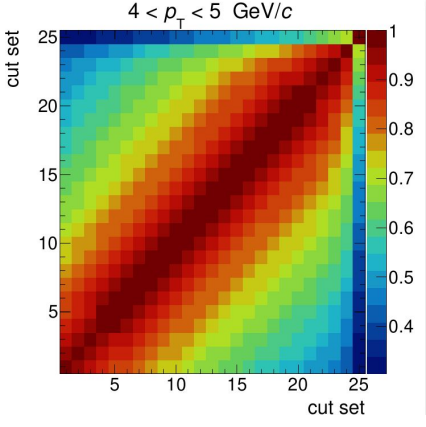


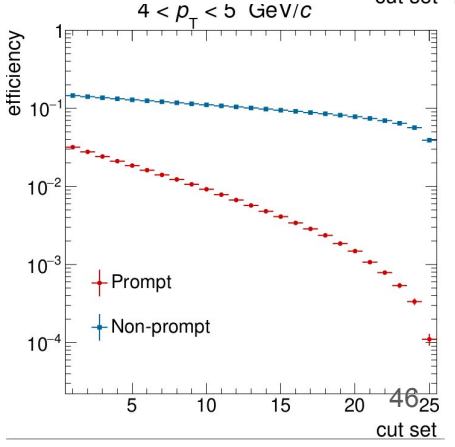


Wide right



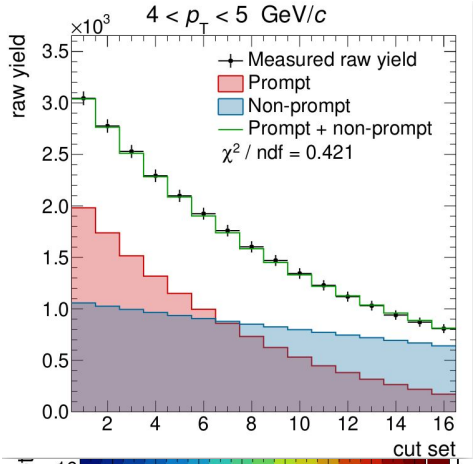


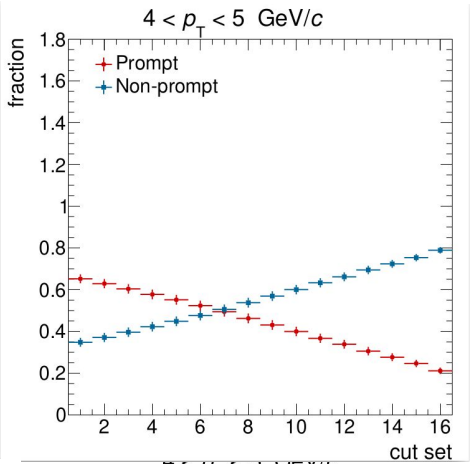


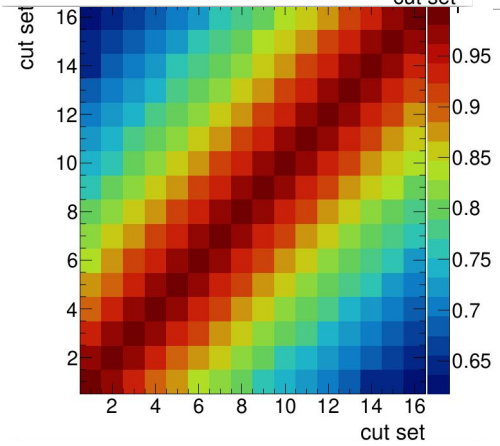


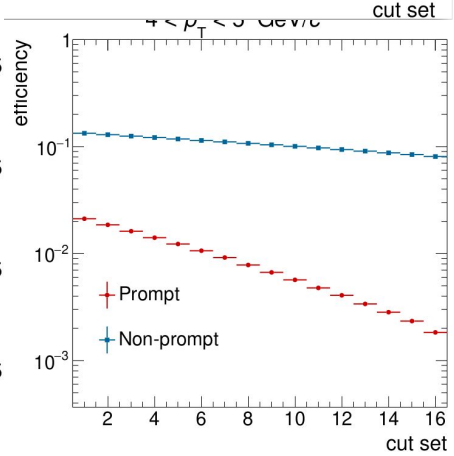
Systematics: D^+ f_{FD} [4-5] -narrow right & left

Narrow left & right

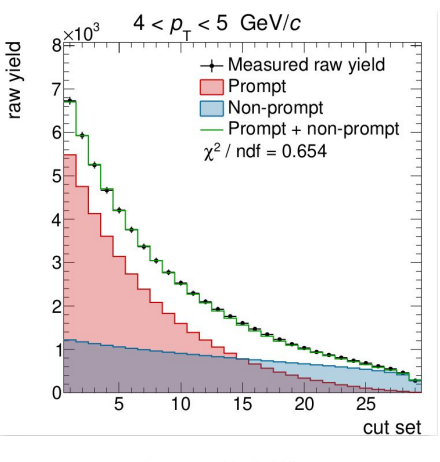


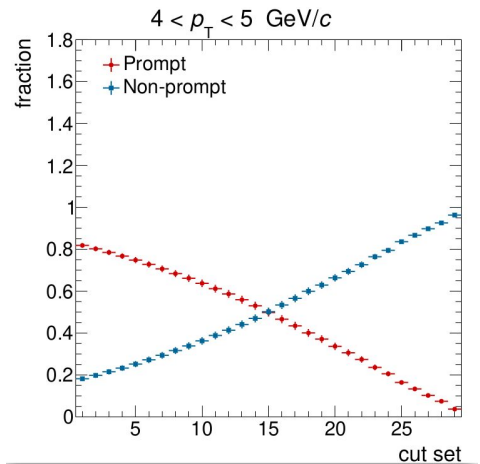


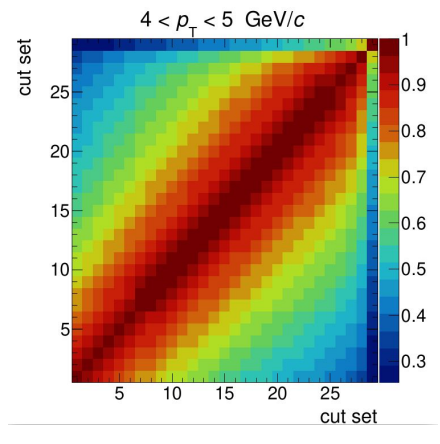


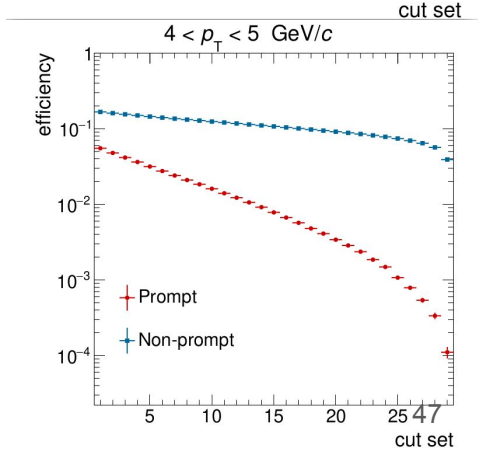


Wide left & right



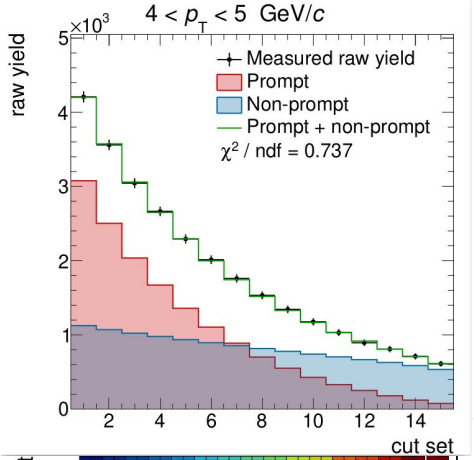


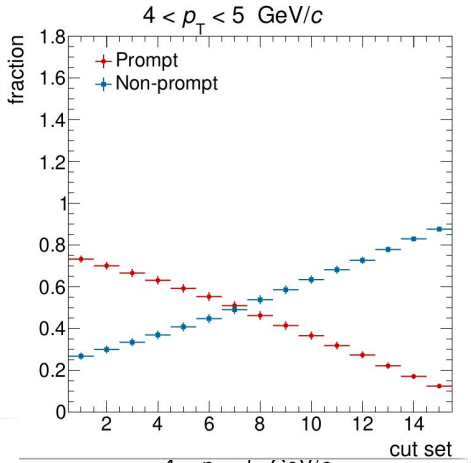


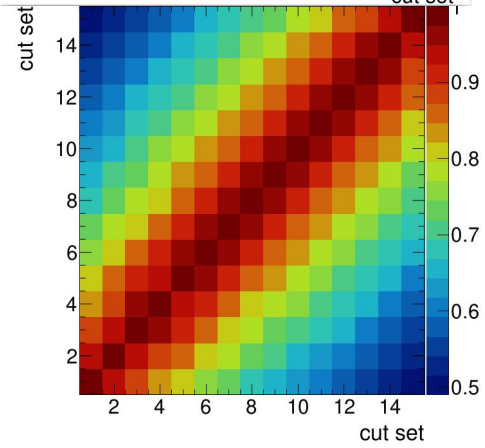


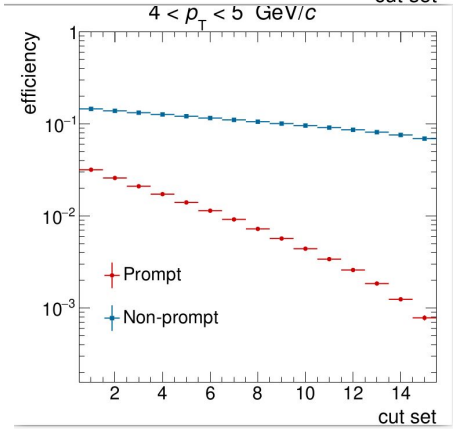
Systematics: D^+ f_{FD} [4-5] -narrow right & left

Alternate1









Alternate2