Chapter 6

Evaluation of raw yield corrections

To evaluate the ratio between the production yield ratio of the two D-meson species, a number of corrections must be applied to the extracted raw yields. These corrections are necessary to account for the acceptance of the detector, the selection efficiency of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons, the branching ratio of the decay channels, and the feed-down from beauty-hadron decays. The prompt $\mathrm{D_s^+}/\mathrm{D^+}$ production yield ratio is then obtained by dividing the corrected yield of $\mathrm{D_s^+}$ by the corrected yield of $\mathrm{D^+}$:

$$\mathrm{D_s^+}/\mathrm{D^+} = \frac{N_{\mathrm{raw}}^{\mathrm{D_s^+}} \cdot f_{\mathrm{prompt}}^{\mathrm{D_s^+}}}{(\mathrm{Acc} \times \varepsilon)_{\mathrm{prompt}}^{\mathrm{D_s^+}} \cdot \mathrm{BR}^{\mathrm{D_s^+}}} \cdot \left(\frac{N_{\mathrm{raw}}^{\mathrm{D^+}} \cdot f_{\mathrm{prompt}}^{\mathrm{D^+}}}{(\mathrm{Acc} \times \varepsilon)_{\mathrm{prompt}}^{\mathrm{D^+}} \cdot \mathrm{BR}^{\mathrm{D^+}}} \right)^{-1} \quad , \tag{6.1}$$

where $(\mathrm{Acc} \times \varepsilon)$ is the product of the detector acceptance and the efficiency of D-meson selection, f_{prompt} is the fraction of prompt D mesons in the extracted raw yield, and BR is the branching ratio of the considered decay channel.

In the following sections, a detailed description of the corrections applied to the raw yields is given.

6.1 Acceptance and efficiency correction

The first correction takes into account that of the D mesons produced at midrapidity ($|y| < 0.5$), only a fraction Acc can be detected by the ALICE apparatus, due to the geometry of the detector, and that only a fraction ε of the detected D mesons passes the selection criteria described in Chapters **??** and **??**.

These corrections are evaluated using a pure sample of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons from Monte Carlo (MC) simulations. To avoid the introduction of biases, a different data sample is used for the evaluation of the acceptance and efficiency corrections compared to the one used for the ML model training and performance evaluation. Proton-proton collisions are simulated using the PYTHIA 8 event generator [1] with colour-reconnection Mode 2 [2], and the generated particles are propagated through the ALICE experimental apparatus using the GEANT4 transport simulation toolkit [3]. Due to the displaced topology of heavy-flavour decays and the continuous readout employed by the ALICE detector, the selection of events with charm or beauty hadrons produces “fake” vertices arising from the association of displaced decay tracks, affecting the reconstruction of heavy-flavour hadrons. To overcome this problem, minimum bias events are generated between charm- or beauty-injected ones (*gap-triggered* approach). Studies performed using different gap sizes have shown that a gap of 5 minimum bias events between heavy-flavour-injected events reduces the number of fake vertices to an acceptable level, while keeping the simulation time reasonable.

Special care is taken to ensure that the MC simulation reproduces the experimental conditions and the reconstruction configuration used for data. To improve the description of the spatial resolution in data, a smearing of the track impact parameters is applied via a dedicated workflow (`track-tuner`), to reproduce that observed in data. The workflow has been tuned by comparing the impact-parameter resolution in data and MC obtained by fitting their distributions for primary particles. It is needed as a slightly worse impact-parameter resolution is obtained in data, probably due to a residual misalignment of the ITS, as shown in Fig. 6.1, where the impact parameter resolutions in the transverse plane and along the beam direction are shown for both data and MC simulations.

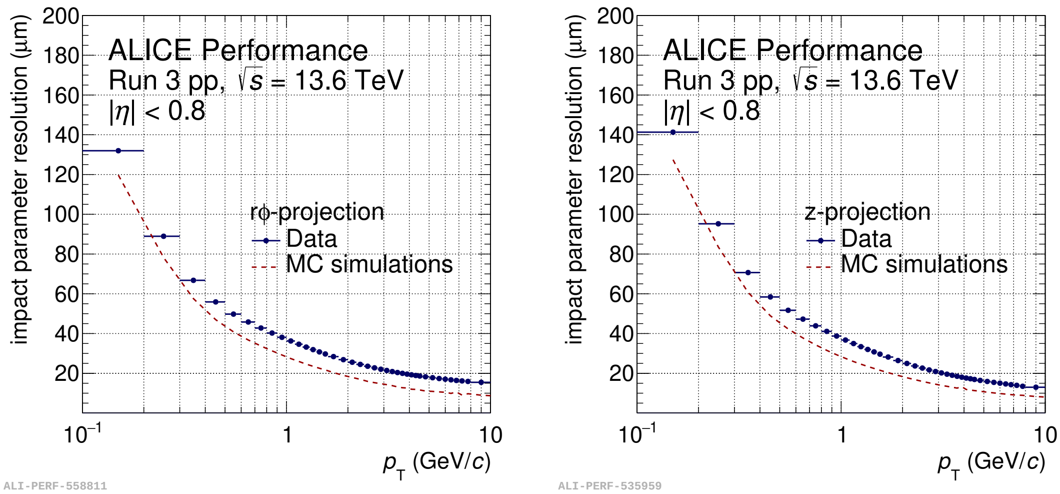

Figure 6.1: Impact-parameter resolution in the transverse plane (left panel) and beam direction (right panel) of primary particles in data and MC. Figure from ALICE figure repository [4].

The $(\mathrm{Acc}\times\varepsilon)$ factor is defined as

$$(\mathrm{Acc}\times\varepsilon) = \frac{N_{\mathrm{sel.}}^{|y|<0.5}}{N_{\mathrm{gen.}}^{|y|<0.8}} \quad ,$$

where $N_{\mathrm{gen.}}^{|y|<0.8}$ is the number of generated D mesons with $|y| < 0.8$, and $N_{\mathrm{sel.}}^{|y|<0.5}$ is the number of D mesons with $|y| < 0.5$ that pass the applied selection criteria. The $(\mathrm{Acc}\times\varepsilon)$ factor is illustrated in Fig. 6.2 for both $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons as a function of p_{T}. Due to the different decay topologies of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons, with the second decaying on average at a larger distance from the primary vertex, the acceptance times efficiency is lower for $\mathrm{D_s^+}$ mesons than for $\mathrm{D^+}$ mesons. For the same reason, the (Acc$\times\varepsilon$) factor is smaller for promptly-produced D mesons than for non-prompt ones. The correction presents a strong dependence on p_{T}. The larger Lorentz boost of D mesons at high p_{T} results in a more displaced decay vertex, facilitating the identification of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons. To extract a significant signal at low p_{T}, where a larger number of background tracks (and therefore combinatorial background candidates) are present, the selection criteria are tightened, reducing the efficiency of the selection. The (Acc$\times\varepsilon$) correction ranges from about 0.3 at the highest p_{T} interval down to about 7×10^{-4} at the lowest considered p_{T} interval.

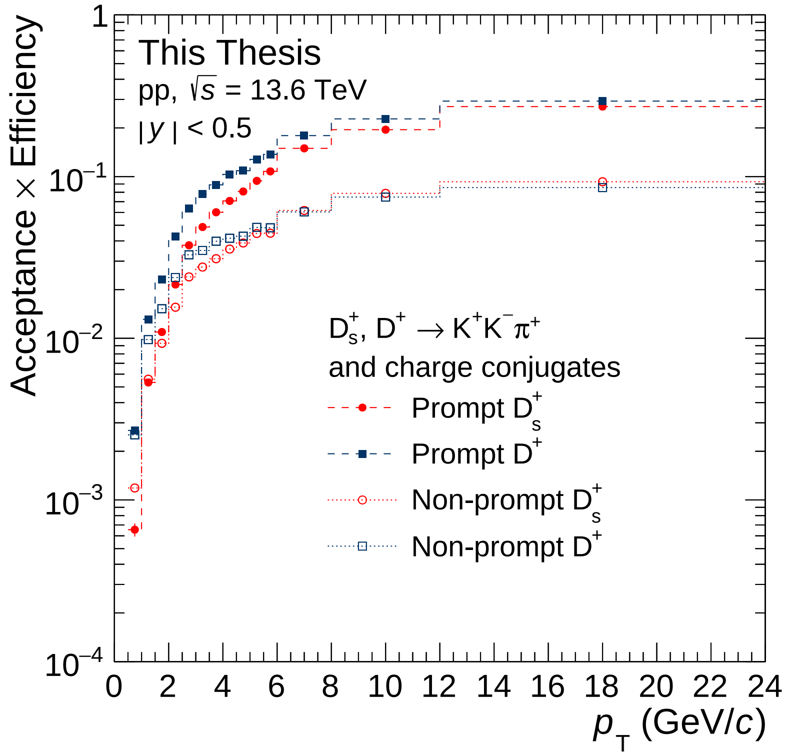

Figure 6.2: Acceptance-times-efficiency correction factor as a function of p_{T} for prompt and non-prompt $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons.

6.2 Prompt fraction correction

The fraction of prompt $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons in the measured raw yield ($f_{\mathrm{prompt}}^{\mathrm{D_s^+}}$ and $f_{\mathrm{prompt}}^{\mathrm{D^+}}$) is estimated using a data-driven method that involves varying the BDT probability threshold for candidate selection. This variation changes the balance between prompt and non-prompt contributions to the signal, exploiting their different trends with the BDT output score to disentangle their relative abundances in the extracted raw yield. Introduced in Run 2 in Ref. [5], this approach leverages the relationship between the raw yield and the number of produced prompt (N_{prompt}) and non-prompt ($N_{\mathrm{non\text{-}prompt}}$) D mesons decaying into the considered decay channel:

$$N_{\mathrm{raw}} = N_{\mathrm{prompt}} \times (\mathrm{Acc}\times\varepsilon)_{\mathrm{prompt}} + N_{\mathrm{non\text{-}prompt}} \times (\mathrm{Acc}\times\varepsilon)_{\mathrm{non\text{-}prompt}} \quad . \tag{6.2}$$

Eq. 6.2 holds for every chosen BDT selection criterion, defining the $(\mathrm{Acc}\times\varepsilon)_{\mathrm{prompt}}$ and $(\mathrm{Acc}\times\varepsilon)_{\mathrm{non\text{-}prompt}}$ correction factors. Since these factors can be determined from MC simulations (as described in the previous section), Eq. 6.2 contains only two unknown variables, N_{prompt} and $N_{\mathrm{non\text{-}prompt}}$. They can be evaluated by varying the BDT working point, extracting the corresponding raw yield, and estimating the $(\mathrm{Acc}\times\varepsilon)_{\mathrm{prompt}}$ and $(\mathrm{Acc}\times\varepsilon)_{\mathrm{non\text{-}prompt}}$ factors. However, a single variation of the selection criteria may lead to large uncertainties, as a small change in the $(\mathrm{Acc}\times\varepsilon)$ correction adds limited information on N_{prompt} and $N_{\mathrm{non\text{-}prompt}}$. To overcome this issue, the BDT threshold values are varied several times to span a wide range of $(\mathrm{Acc}\times\varepsilon)_{\mathrm{prompt}}$ and significantly change the f_{prompt} in the extracted raw yield. For each selection criterion i, the acceptance-times-efficiency factor is calculated for prompt $(\mathrm{Acc}\times\varepsilon)^{\mathrm{i}}_{\mathrm{prompt}}$ and non-prompt $(\mathrm{Acc}\times\varepsilon)^{\mathrm{i}}_{\mathrm{non\text{-}prompt}}$ D mesons, and the raw yields $N^{\mathrm{i}}_{\mathrm{raw}}$ are extracted by fitting the invariant mass distribution of candidates passing the selections. By including the entire set of selection criteria, a system of equations can be defined:

$$\begin{cases} N^{1}_{\mathrm{raw}} = N_{\mathrm{prompt}} \times (\mathrm{Acc}\times\varepsilon)^{1}_{\mathrm{prompt}} + N_{\mathrm{non\text{-}prompt}} \times (\mathrm{Acc}\times\varepsilon)^{1}_{\mathrm{non\text{-}prompt}} \\ N^{2}_{\mathrm{raw}} = N_{\mathrm{prompt}} \times (\mathrm{Acc}\times\varepsilon)^{2}_{\mathrm{prompt}} + N_{\mathrm{non\text{-}prompt}} \times (\mathrm{Acc}\times\varepsilon)^{2}_{\mathrm{non\text{-}prompt}} \\ \vdots \\ N^{\mathrm{n}}_{\mathrm{raw}} = N_{\mathrm{prompt}} \times (\mathrm{Acc}\times\varepsilon)^{\mathrm{n}}_{\mathrm{prompt}} + N_{\mathrm{non\text{-}prompt}} \times (\mathrm{Acc}\times\varepsilon)^{\mathrm{n}}_{\mathrm{non\text{-}prompt}} \end{cases} ,$$

For n > 2, the system of equations is overconstrained, and the solution can be found through a minimisation procedure. The system of equations can be written in matrix notation as

$$\begin{pmatrix} (\mathrm{Acc}\times\varepsilon)^{1}_{\mathrm{prompt}} & (\mathrm{Acc}\times\varepsilon)^{1}_{\mathrm{non\text{-}prompt}} \\ (\mathrm{Acc}\times\varepsilon)^{2}_{\mathrm{prompt}} & (\mathrm{Acc}\times\varepsilon)^{2}_{\mathrm{non\text{-}prompt}} \\ \vdots & \vdots \\ (\mathrm{Acc}\times\varepsilon)^{\mathrm{n}}_{\mathrm{prompt}} & (\mathrm{Acc}\times\varepsilon)^{\mathrm{n}}_{\mathrm{non\text{-}prompt}} \end{pmatrix} \begin{pmatrix} N_{\mathrm{prompt}} \\ N_{\mathrm{non\text{-}prompt}} \end{pmatrix} - \begin{pmatrix} N^{1}_{\mathrm{raw}} \\ N^{2}_{\mathrm{raw}} \\ \vdots \\ N^{\mathrm{n}}_{\mathrm{raw}} \end{pmatrix} = \begin{pmatrix} \delta^{1} \\ \delta^{2} \\ \vdots \\ \delta^{\mathrm{n}} \end{pmatrix} ,$$

where δ^{i} are the residuals due to the uncertainty in the raw yield and $(\mathrm{Acc}\times\varepsilon)$ correction factors for the i-th selection. For each selection criterion, the uncertainty on δ^{i} is estimated by propagating the uncertainties on the raw yield and $(\mathrm{Acc}\times\varepsilon)$ factors as

$$\sigma^{2}_{\mathrm{i}} = \sigma^{2}_{N^{\mathrm{i}}_{\mathrm{raw}}} + N_{\mathrm{prompt}} \times \sigma^{2}_{(\mathrm{Acc}\times\varepsilon)^{\mathrm{i}}_{\mathrm{prompt}}} + N_{\mathrm{non\text{-}prompt}} \times \sigma^{2}_{(\mathrm{Acc}\times\varepsilon)^{\mathrm{i}}_{\mathrm{non\text{-}prompt}}} .$$

Given that the corrected yields are unknown variables, an iterative procedure is used to define the total uncertainty: initially, N_{prompt} and $N_{\mathrm{non\text{-}prompt}}$ are set to zero and only the uncertainty on the raw yields is taken into account. From the second iteration, the corrected yields N_{prompt} and $N_{\mathrm{non\text{-}prompt}}$ obtained in the previous step are also used. This iteration is repeated until the difference between the corrected yields in successive iterations falls below a predefined threshold.

The solution of the system of equations is found by minimising the residuals with a least-squares method, with a χ^2 defined as

$$\chi^{2} = \boldsymbol{\delta}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{\delta} ,$$

where $\boldsymbol{\delta}$ is the vector of residuals and $\mathbf{C}$ is the covariance matrix of the residuals.

In this analysis, the BDT threshold is fixed for the background score, at a tighter value than the one used to extract the central raw yields to ensure the convergence of the invariant mass fits. The minimum probability for being a non-prompt D meson is varied in different ranges for the two D mesons, to guarantee that a large enough variation in the $f_{\text{non-prompt}}$ is achieved. Because the selection criterion is increasingly tightened, the $\mathrm{i}+1$-th selected sample will be entirely contained in the i-th one. The residuals δ^{i} will therefore exhibit a degree of correlation. The off-diagonal elements $\sigma_{\mathrm{i,j}}$ of the covariance matrix $\mathbf{C}$, which are the covariance between the residuals of the i-th and j-th selection criteria, can be estimated as

$$\sigma_{\mathrm{i,j}} = \rho_{\mathrm{i,j}}\sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \quad ,$$

where it can be demonstrated [6] that the correlation coefficient $\rho_{\mathrm{i,j}}$ is given by

$$\rho_{\mathrm{i,j}} = \frac{\sigma_{\mathrm{i}}}{\sigma_{\mathrm{j}}}$$

if the measurement i is made on a dataset that is fully included in the one used for the measurement j.

The minimisation procedure described above leads to the determination of the true number of prompt and non-prompt D mesons decaying into the considered decay channel (N_{prompt} and $N_{\text{non-prompt}}$, respectively), which are independent of the applied selection criteria. The results for the evaluation of the $\mathrm{D_s^+}$ f_{prompt} correction factor in the $1.5 < p_{\mathrm{T}} < 2.0$ GeV/c interval are shown in Fig. 6.3. In the top-left panel, the correlation factor ρ between the residuals δ^{i} is shown for the different selection criteria; in the top-right panel, the $(\mathrm{Acc}\times\varepsilon)$ factors are shown as a function of the BDT selection criterion for both prompt and non-prompt $\mathrm{D_s^+}$ mesons; in the bottom-left panel, the relative contribution of the prompt and non-prompt components to the extracted $\mathrm{D_s^+}$ raw yields is shown as a function of the ML selections. Lastly, in the bottom-right panel, the extracted yields are fitted using a template fit, utilising the $(\mathrm{Acc}\times\varepsilon)^{\mathrm{D_s^+}}_{\text{prompt}}$ and $(\mathrm{Acc}\times\varepsilon)^{\mathrm{D_s^+}}_{\text{non-prompt}}$ evolution as a function of the BDT selection criterion as input. The minimisation procedure described above can in fact be interpreted as a template fit to the extracted raw yields, where the templates are the $(\mathrm{Acc}\times\varepsilon)$ factors. The prompt and non-prompt components of the raw yield, obtained for each BDT-based selection from the minimisation procedure as $(\mathrm{Acc}\times\varepsilon)_{\text{prompt}}\times N_{\text{prompt}}$ and $(\mathrm{Acc}\times\varepsilon)_{\text{non-prompt}}\times N_{\text{non-prompt}}$ respectively, are represented with red and blue filled histograms. Their sum is reported by the green histogram. Given the small χ^2/ndf value obtained from the fit, the f_{prompt} factor is considered to be well determined. The leftmost data point of each distribution corresponds to the looser selection on the BDT non-prompt score, while the rightmost one corresponds to the strictest selection, which is expected to preferentially select non-prompt D mesons.

The f_{prompt} factor is then calculated for each D-meson species and for any given selection criterion with efficiencies $(\mathrm{Acc}\times\varepsilon)_{\text{prompt}}$ and $(\mathrm{Acc}\times\varepsilon)_{\text{non-prompt}}$ for prompt and non-prompt D mesons respectively, as

$$f_{\text{prompt}} = \frac{N_{\text{prompt}}\times(\mathrm{Acc}\times\varepsilon)_{\text{prompt}}}{N_{\text{prompt}}\times(\mathrm{Acc}\times\varepsilon)_{\text{prompt}} + N_{\text{non-prompt}}\times(\mathrm{Acc}\times\varepsilon)_{\text{non-prompt}}} \quad .$$

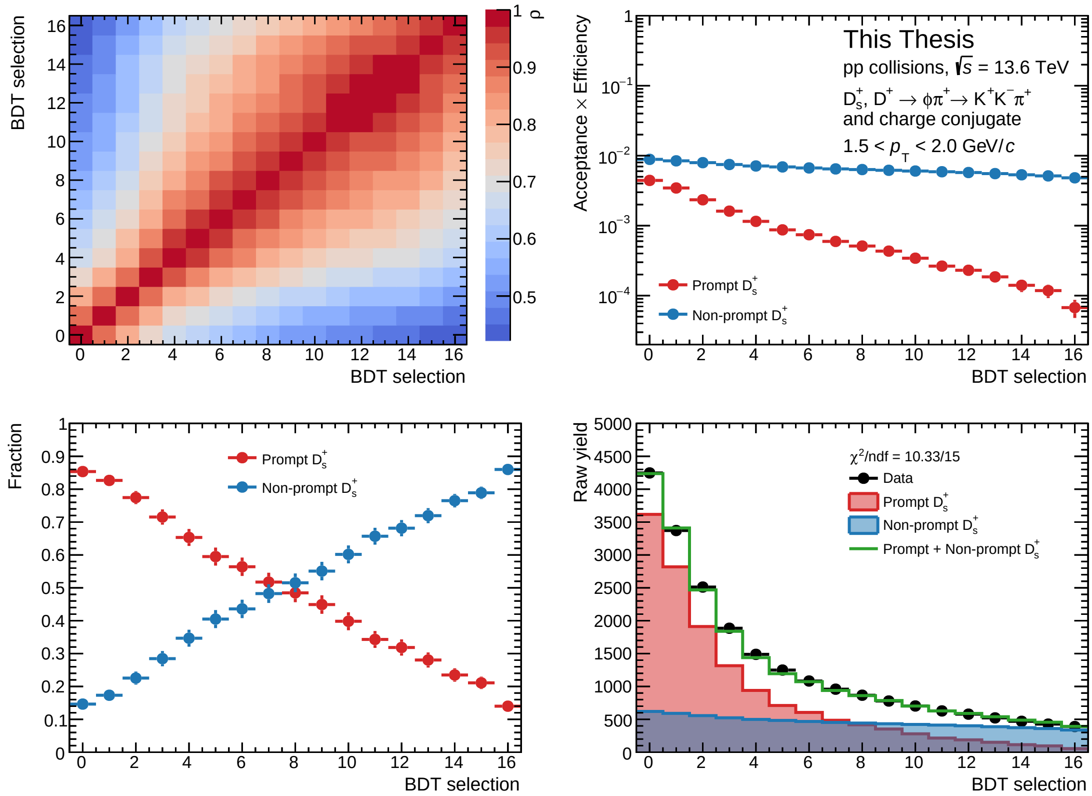

Figure 6.3: Results for the evaluation of the D_s^+ f_{prompt} correction factor in the $1.5 < p_T < 2.0$ GeV/c interval. The correlation factor ρ (top-left panel), $(\text{Acc}\times\varepsilon)$ factors for both prompt and non-prompt D_s^+ mesons (top-right panel), prompt and non-prompt fraction in the extracted D_s^+ raw yields (bottom-left panel), and contribution of prompt and non-prompt to the extracted raw yields (bottom-right panel) are shown as a function of the BDT selection criterion.

As a cross-check, the f_{prompt} factor is also estimated using a theory-driven approach [7], which relies on FONLL [8] predictions for beauty-hadron production and the decay kinematic description of PYTHIA 8 [1] for estimating the p_T differential cross section of non-prompt D mesons. The f_{prompt} factor is then calculated for each D-meson species as

$$f_{\text{prompt}} = 1 - \left.\frac{d^2\sigma}{dp_T dy}\right|_{\text{FONLL+PYTHIA 8}}^{\text{non-prompt}} \times \frac{(\text{Acc}\times\varepsilon)_{\text{non-prompt}}\Delta y \Delta p_T \text{BR}\mathcal{L}_{\text{int}}}{\frac{1}{2}\times N_{\text{raw}}} \quad ,$$

where Δy and Δp_T are the rapidity and p_T intervals, respectively, $\mathcal{L}_{\text{int}}$ is the integrated luminosity, corresponding to $1\,\text{pb}^{-1}$, N_{raw} is the extracted raw yield of the considered D meson, and the factor 1/2 takes into account that both particle and antiparticles are selected.

A comparison between the f_{prompt} correction factors for D_s^+ and D^+ mesons obtained with the two methods is illustrated in Fig. 6.4. A good level of agreement is observed between the two results, which result to be compatible within their uncertainties across the studied p_T range, with the exception of the highest p_T

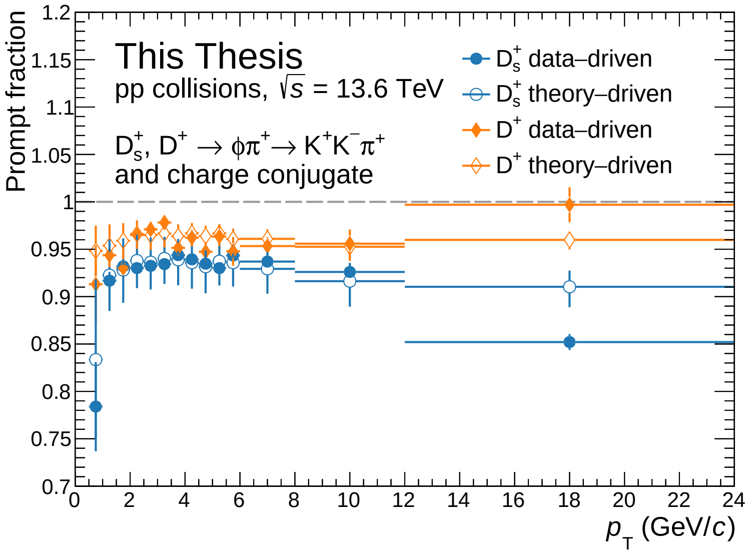

Figure 6.4: f_{prompt} correction factor for D_s^+ (blue) and D^+ (orange) mesons as a function of p_T. The results obtained with the data-driven (filled markers) method are compared with those exploiting a theory-driven approach (void markers).

interval, where fluctuations in the yield extraction with the data-driven method lead to a small discrepancy between the two methods. For the evaluation of the central f_{prompt} values, the data-driven method is used, as it is not sensitive to possible shortcomings in the theoretical description of $b\bar{b}$ production from FONLL which may affect the theory-driven approach.

6.3 Branching ratio correction

The branching ratio correction is applied to account for the fact that of the produced D_s^+ and D^+ mesons, only those decaying into a $KK\pi$ final state are reconstructed. The values of the branching ratios for the considered decays are taken from Ref. [9], and correspond to $BR(D_s^+) = (2.21 \pm 0.06) \times 10^{-2}$ and $BR(D^+) = \big((2.69^{+0.07}_{-0.08}) \times 10^{-3}\big)$. Since these values are not evaluated through the analysis carried out in this Thesis, but rather taken from a global fit reported in a different publication, a systematic uncertainty is assigned to account for that provided by the PDG. The uncertainty is propagated to the D_s^+/D^+ production yield ratio with a gaussian approach, yielding an asymmetric $^{+3.7}_{-4.0}\%$ uncertainty on the ratio.

6.4 Systematic uncertainties

Measurements of ratios between different particles' production yields allow for the cancellation of some of the systematic uncertainties which have to be taken into account for the measurement of production cross sections, such as that related to luminosity. In addition, the reconstruction of D_s^+ and D^+ mesons through the same decay channel allows for the cancellation of supplementary sources of systematic uncertainties, such as those related to tracking and PID efficiency, thereby enhancing the precision of the results. Nonetheless, the measurement is still affected by several sources of systematic uncertainties, due to arbitrary choices made in the analysis, or the need to rely on MC simulations, which could not perfectly reproduce the data.

6.4.1 BDT selection efficiency

The choice of the set of BDT threshold values used to extract the raw yields is arbitrary, albeit driven by a defined criterion through the maximisation of the significance on a subsample of the analysed dataset. Variations in the ML selection criteria may lead to differences in the extracted raw yields, and therefore in the D_s^+/D^+ production yield ratio, due to possible imperfections in the MC description of the topological, kinematic, and PID variables used in the training of the BDT model. An imperfect MC description could in fact cause a bias in the final ratio when correcting for the efficiency terms of D_s^+ and D^+ mesons. The systematic uncertainty on the BDT selection efficiency is assessed by repeating the analysis varying the BDT threshold values on prompt and background probabilities, independently and then simultaneously, for a total of 48 variations. Three tighter and three looser BDT thresholds on both the background and prompt D meson scores are used on top of those employed for the central case. To avoid extreme variations in the selection criteria, the $(\mathrm{Acc} \times \varepsilon)_{\mathrm{prompt}}$ for both D-meson species is required to be within 30% of that estimated for the central selections. In addition, to ensure a good fit quality, the χ^2/ndf of the invariant mass fits is required to be below 2, and the significance of the extracted signal to be above 3 and to be larger than half of the central value, for both D_s^+ and D^+ mesons.

The results for the estimation of the systematic uncertainty on the BDT selection efficiency are shown in Fig. 6.5 for the $3.0 < p_T < 3.5$ GeV/c p_T interval. In the top row, the raw yields of D_s^+ and D^+ mesons, the prompt and non-prompt D_s^+ selection efficiency, the prompt and non-prompt D^+ selection efficiency, and the prompt fraction of D_s^+ and D^+ mesons are shown for the different BDT selections used to estimate the systematic uncertainty, normalised to their respective central values. In the bottom row, the statistical significance of the D_s^+ and D^+ extracted signals, the signal-over-background ratio for the two meson species, and the D_s^+/D^+ production yield ratio are reported for the considered BDT selections. The different results are ordered such that the central value is the leftmost one (BDT selection 0). In the following seven variations, the background score threshold is fixed to the central case, while the prompt score threshold is varied through the whole set of considered values. Then, in the next 21 variations, three different tighter thresholds are applied to the background score, while the prompt score is varied. The last 21 variations are obtained by applying three looser thresholds to the background score and varying the prompt score. In the rightmost bottom panel, the distribution of the D_s^+/D^+ production yield ratio for the different BDT selections divided by the central value is shown. The systematic uncertainty on the BDT selection efficiency is estimated as the sum in quadrature of the standard deviation of the D_s^+/D^+

distribution and of the difference (Δ) between its mean and the central value. To avoid the inclusion of statistical fluctuations in the systematic uncertainty, the $p_{\rm T}$ dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty ranges from 1% to 10% depending on $p_{\rm T}$.

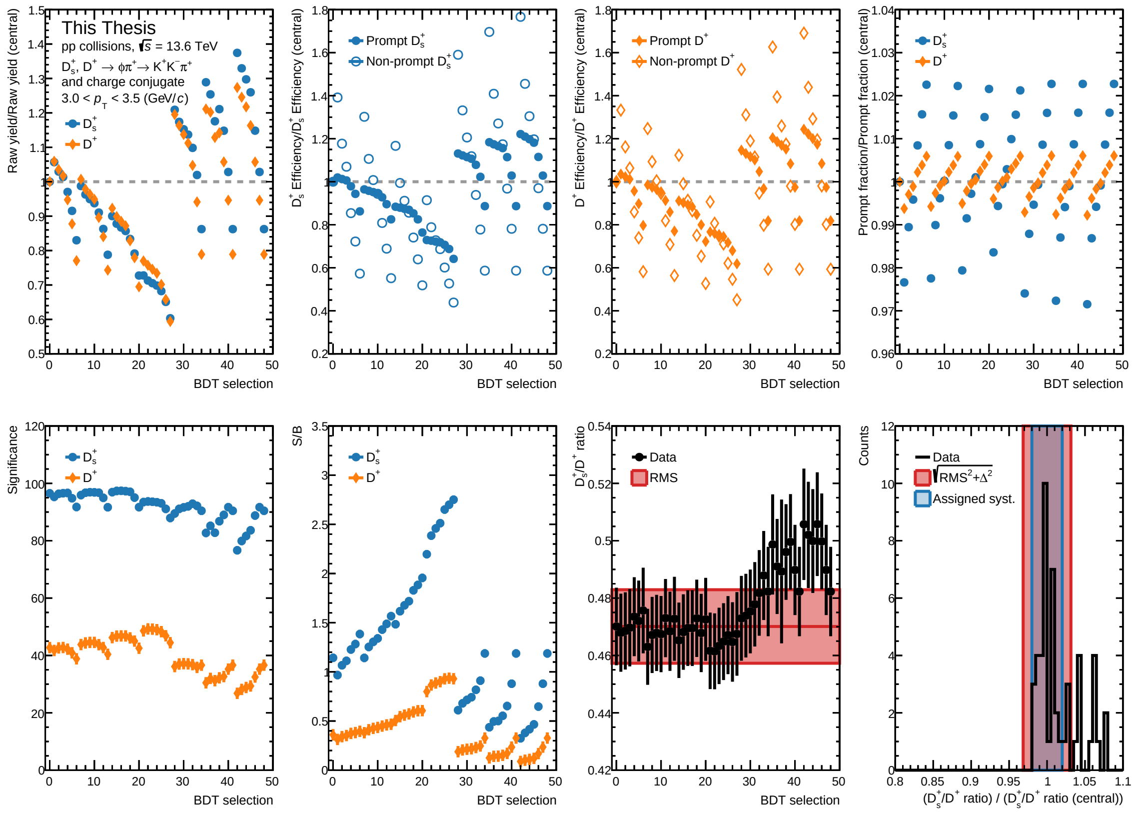

Figure 6.5: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency for the $3.0 < p_{\rm T} < 3.5$ GeV/c $p_{\rm T}$ interval. In the top row, the raw yields of $\rm D_s^+$ (blue) and $\rm D^+$ (orange) mesons, the prompt (filled markers) and non-prompt (void markers) $\rm D_s^+$ selection efficiency, the prompt and non-prompt $\rm D^+$ selection efficiency, and the prompt fraction of $\rm D_s^+$ and $\rm D^+$ mesons are shown for the different BDT selections used to estimate the systematic uncertainty, normalised to their respective central values. In the bottom row, the statistical significance of the $\rm D_s^+$ and $\rm D^+$ extracted signals, the signal-over-background ratio for the two meson species, and the $\rm D_s^+$/dpl production yield ratio are reported for the different BDT selections. A red band is shown in this last panel around the central value, and represents the RMS of the measurements. The rightmost bottom panel shows the distribution of the $\rm D_s^+/D^+$ production yield ratio for the different BDT selections divided by the central value. A red (blue) band is shown around the central value, representing the sum in quadrature of the standard deviation of the $\rm D_s^+/D^+$ distribution and of the difference between its mean and the central value (the assigned systematic uncertainty).

6.4.2 Prompt fraction

The choice of the set of selection criteria employed for evaluating the f_{prompt} correction may introduce a systematic uncertainty on the $\mathrm{D_s^+}/\mathrm{D^+}$ production yield ratio, due to possible imperfections in the description of the efficiency evolution as a function of the applied selections. To estimate the magnitude of this uncertainty, the f_{prompt} correction is evaluated using different sets of BDT threshold values. Of the 17 variations on the non-prompt score threshold used to estimate the prompt fraction correction as described in Sec. 6.2, a subsample of 13 is used to estimate the systematic uncertainty. The f_{prompt} estimation is repeated using only the first, last, and central selections, corresponding to the loosest, tightest, and intermediate selection criteria. For each set of BDT threshold values, the $f_{\mathrm{prompt}}^{\mathrm{D_s^+}}$ and $f_{\mathrm{prompt}}^{\mathrm{D^+}}$ are extracted using the same procedure as in Sec. 6.2. The correction factor applied to Eq. 6.1 is then calculated for each p_{T} interval as the ratio between the $f_{\mathrm{prompt}}^{\mathrm{D_s^+}}$ and $f_{\mathrm{prompt}}^{\mathrm{D^+}}$ correction factors, considering all the possible combinations of results from the four different sets of BDT selections (also the central values are considered). A total of 16 values of the $f_{\mathrm{prompt}}^{\mathrm{D_s^+}}/f_{\mathrm{prompt}}^{\mathrm{D^+}}$ ratio are obtained for each p_{T} interval, as shown in the left panel of Fig. 6.6. On the right panel, the ratio of the different $f_{\mathrm{prompt}}^{\mathrm{D_s^+}}/f_{\mathrm{prompt}}^{\mathrm{D^+}}$ factors to that obtained in the central case is reported. The systematic uncertainty on the f_{prompt} correction is then estimated as the standard deviation of the $f_{\mathrm{prompt}}^{\mathrm{D_s^+}}/f_{\mathrm{prompt}}^{\mathrm{D^+}}$ distribution. To avoid the inclusion of statistical fluctuations in the systematic uncertainty, the p_{T} dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty ranges from 1% to 4% depending on p_{T}.

6.4.3 Imperfections in the description of the tracking performance

An imperfect description of the topological variables in the MC simulations may imply a systematic error in the determination of the selection efficiencies of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons. Such imperfections might be related to the description of the decay topology and decay kinematics at generated level (i.e. $c\tau$ implemented in the adopted MC generator) as well as to the reproduction of the detector resolution. This source of systematic uncertainty is typically tested by varying the topological selections applied and evaluating the stability of the result against the variations, as was done in Sec. 6.4.1. In addition to this, the sensitivity of the measurement to the possible discrepancies in the impact parameter resolution between data and MC was evaluated by using different configurations of the track-tuner workflow. In particular, a configuration with a 10% worse resolution than the default one was considered. Moreover, a version of the track-tuner implementing a smearing of the track p_{T} at the innermost update point was also used. In this case, a factor 2 worse p_{T} resolution was used to account for the discrepancy between the widths of the mass peaks in data and MC. The analysis was repeated using the two different configurations of the track-tuner for estimating the correction factors, and the measured $\mathrm{D_s^+}/\mathrm{D^+}$ production yield ratios are shown in Fig. 6.7, together with the ratio to the central result. The systematic uncertainty is estimated for each p_{T} interval as the semi-dispersion of the

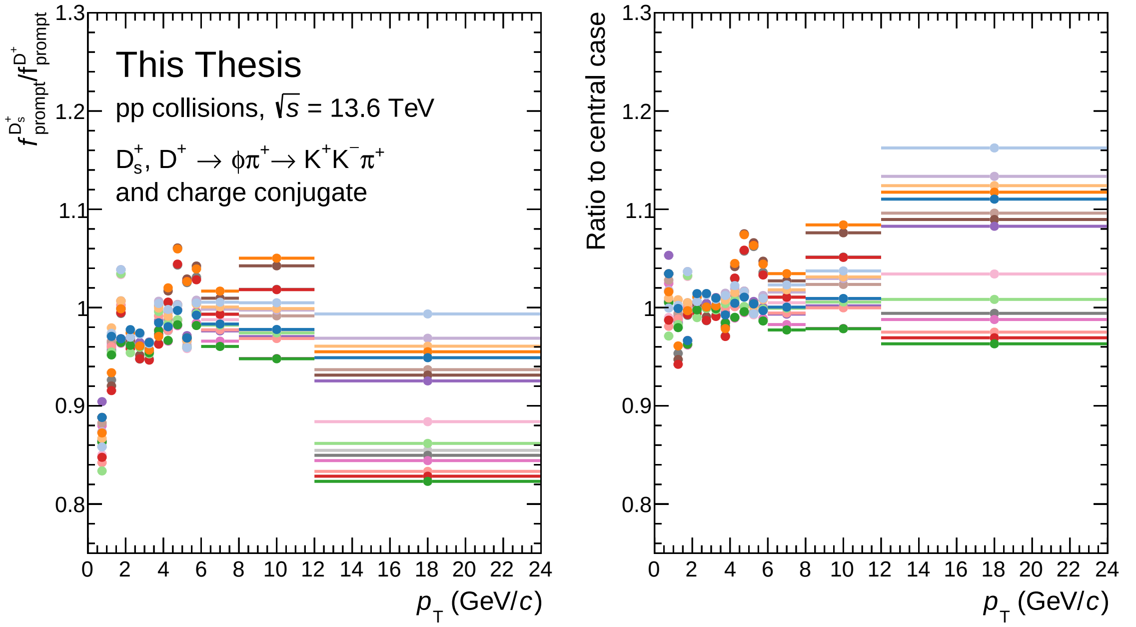

Figure 6.6: Results for the evaluation of the systematic uncertainty on the $f_{\rm prompt}$ correction. In the left panel, the different $f_{\rm prompt}^{\rm D_s^+}/f_{\rm prompt}^{\rm D^+}$ corrections obtained with different sets of BDT selections are shown. In the right panel, their ratio to that obtained in the central case is reported.

three results. As for the other sources of systematic uncertainty, the $p_{\rm T}$ dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty ranges from 1% to 4% depending on $p_{\rm T}$.

6.4.4 Generated Monte Carlo $p_{\rm T}$ shape

A non-realistic $p_{\rm T}$ shape of generated D mesons in the MC simulations could lead to a bias in the determination of the efficiency because of the efficiency variation within the $p_{\rm T}$ intervals considered in the analysis. This source of systematic uncertainty is expected to be negligible, as: i. it was observed to have almost no impact for D mesons in previous analyses(e.g., in Refs. [5, 10]), ii. because possible biases in the $p_{\rm T}$ shape of the generated D mesons would affect both $\rm D_s^+$ and $\rm D^+$ mesons in a similar way, canceling out in the ratio and iii. because very narrow $p_{\rm T}$ intervals are considered in the analysis. Nonetheless, the efficiency steeply decreases at low $p_{\rm T}$, as shown in Fig. 6.2. Therefore, it was checked that changes in the $p_{\rm T}$ distribution of generated D mesons in the MC simulations do not affect the results. The systematic uncertainty is estimated by changing the $p_{\rm T}$ distribution of the generated D mesons in the MC simulations used for evaluating the $(\mathrm{Acc} \times \varepsilon)$ corrections, which are based on the PYTHIA 8 [1] with colour-reconnection Mode 2 [2] event generator, as described in Sec. 6.1. The $p_{\rm T}$ distribution of the generated D mesons is modified to that predicted by FONLL [8]. The $p_{\rm T}$ distributions of $\rm D_s^+$ ($\rm D^+$) mesons are shown in the top-left (top-right) panel of Fig. 6.8 for PYTHIA 8 (filled markers) and

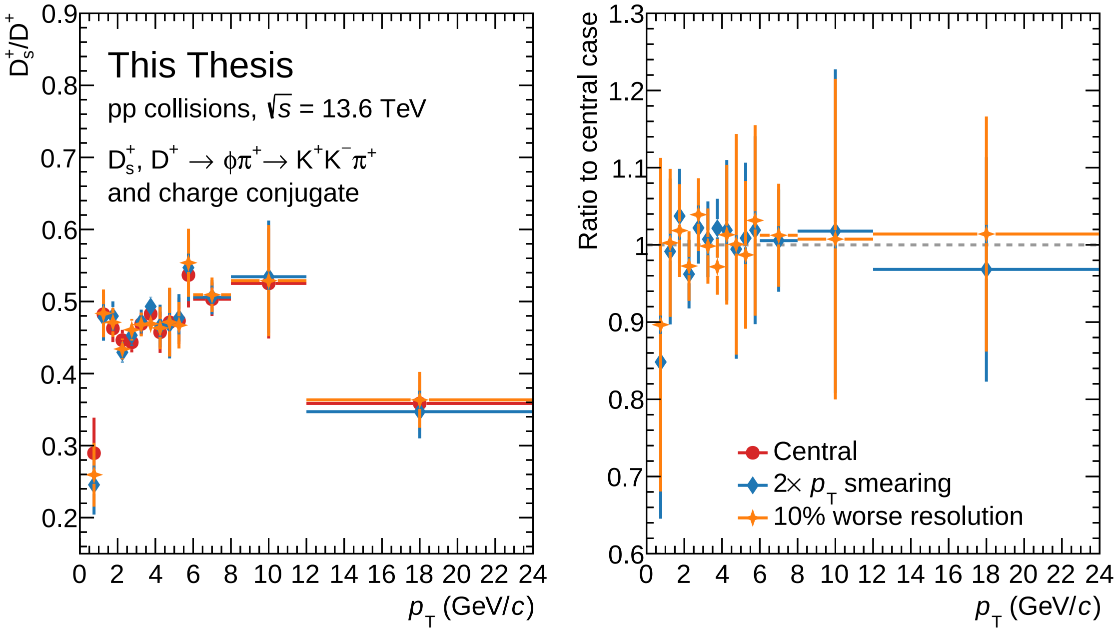

Figure 6.7: Results for the evaluation of the systematic uncertainty on a possible imperfect description of the topological variables in MC simulations. The D_s^+/D^+ production yield ratio is shown in the left panel for the central result (red) and for the two different configurations of the track-tuner: with a factor 2 worse p_T resolution (blue) and a 10% worse resolution. In the right panel, the ratio of the different D_s^+/D^+ factors to that obtained in the central case is reported.

FONLL (void markers) predictions. In the bottom-left panel, the ratio between the FONLL and PYTHIA p_T distributions is shown for D_s^+ and D^+ mesons. Lastly, the bottom-right panel shows the ratio between the FONLL- and PYTHIA 8-based $(\mathrm{Acc}\times\varepsilon)^{D_s^+}_{\mathrm{prompt}}/(\mathrm{Acc}\times\varepsilon)^{D^+}_{\mathrm{prompt}}$ factors, which is the quantity that is used to correct the D_s^+/D^+ production yield ratio in Eq. 6.1. The systematic uncertainty is estimated as the difference between the central result and the one obtained with the FONLL p_T shape. It resulted to be $< 1\%$ for the considered p_T intervals, with the exception of the lowest, where a 1% effect is present. However, given that this uncertainty is largely covered by the other sources of systematic uncertainty, it is not considered in the final result.

6.4.5 Summary of systematic uncertainties

The total systematic uncertainty assigned to the different p_T intervals considered in the analysis is evaluated by summing in quadrature the contributions from the different sources of systematic uncertainty, as the sources are considered to be uncorrelated. The assigned systematic uncertainty ranges from 4% to 13% depending on p_T. A summary of the considered systematic uncertainties, together with the total systematic uncertainty on the D_s^+/D^+ production yield ratio, is reported in Table 6.1.

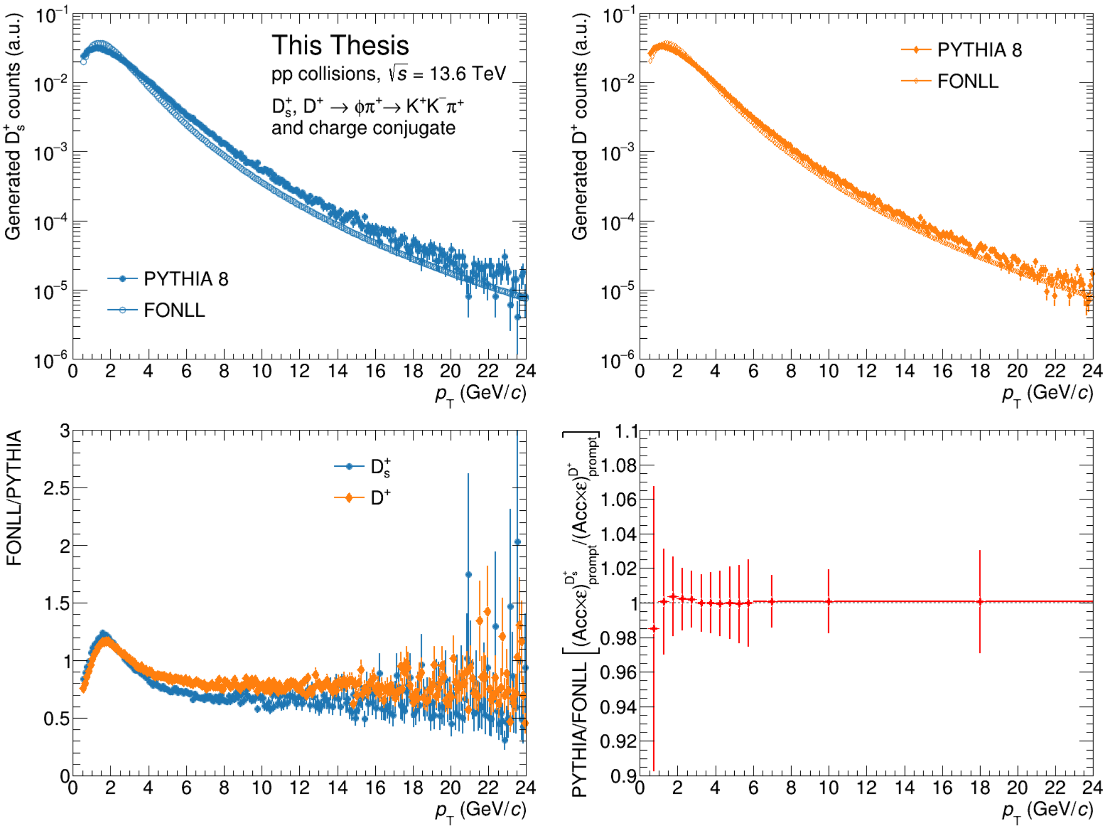

Figure 6.8: Comparison of the p_{T} distributions of generated D_s^+ (left panel) and D^+ (right panel) mesons in the MC simulations used for evaluating the $(\mathrm{Acc} \times \varepsilon)$ corrections. The distributions obtained with the PYTHIA 8 event generator are shown with filled markers, while those obtained with FONLL are shown with void markers.

p_{T} (GeV/c)	0.5–1	1–1.5	1.5–2	2–2.5	2.5–3	3–3.5	3.5–4	4–4.5	4.5–5	5–5.5	5.5–6	6–8	8–12	12–24
Raw yield extraction (%)	3	3	3	3	3	3	3	5	5	5	5	8	9	10
BDT selection efficiency (%)	10	5	3	3	2	2	1	1	1	2	2	2	2	2
FD fraction (%)	2	2	2	1	1	1	1	2	2	2	2	2	3	4
Topological variables (%)	8	2	2	2	2	1	1	1	1	1	1	1	1	2
Tot. systematic uncertainty (%)	13	7	5	5	4	4	4	6	6	6	6	8	10	13
BR	$^{+3.7}_{-4.0}$													

Table 6.1: Summary of the assigned systematic uncertainties on the D_s^+/D^+ production yield ratio.

Bibliography

[1] 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 [hep-ph]`.

[2] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP* **08** (2015) 003, `arXiv:1505.01681 [hep-ph]`.

[3] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[4] **ALICE** Collaboration, "ALICE Figure." `https://alice-figure.web.cern.ch/`.

[5] **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* **05** (2021) 220, `arXiv:2102.13601 [nucl-ex]`.

[6] G. Cowan, *Statistical data analysis.* Oxford university press, 1998.

[7] **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. C* **77** (2017) 550, `arXiv:1702.00766 [hep-ex]`.

[8] M. Cacciari, M. Greco, and P. Nason, "The p_T spectrum in heavy-flavour hadroproduction.", *JHEP* **05** (1998) 007, `arXiv:hep-ph/9803400`.

[9] **Particle Data Group** Collaboration, R. L. Workman and Others, "Review of Particle Physics", *PTEP* **2022** (2022) 083C01.

[10] **ALICE** Collaboration, S. Acharya *et al.*, "Charm production and fragmentation fractions at midrapidity in pp collisions at $\sqrt{\mathrm{s}} = 13$ TeV", *JHEP* **12** (2023) 086, `arXiv:2308.04877 [hep-ex]`.