6

Data-driven cut-variation method for the D^+ measurement

In this Chapter I will present an alternative data-driven method for the measurement of the p_T-differential production cross section of prompt and feed-down D^+-mesons in the decay channel $D^+ \to K^-\pi^+\pi^+$, which is based on the different decay topologies of the two types of D^+. As the impact-parameter fit method presented in Chapter 5, this approach has the advantage of not depending on theoretical calculations and on any assumption on the feed-down D^+-meson R_{pPb}. The first section of this Chapter is dedicated to the description of the method and the startegy adopted, while in the second part the analysis of the Minimum Bias p-Pb data sample described in 4.1.1 is provided. Finally, in the last section of this Chapter, the results obtained are compared with the standard analysis.

6.1 Cut-variation feed-down subtraction method

As explained in section 4, the D^+-meson candidates reconstructed in the decay channel $D^+ \to K^-\pi^+\pi^+$ include prompt D^+, feed-down D^+ and the combinatorial background. Therefore, for each set of topological cuts, a different amount of prompt and feed-down D^+ is selected, as illustrated in figure 6.1. This sketch shows how the extracted raw yield in the same p_T interval changes for two different set of topological cuts. Since a different number of candidates is selected, also the

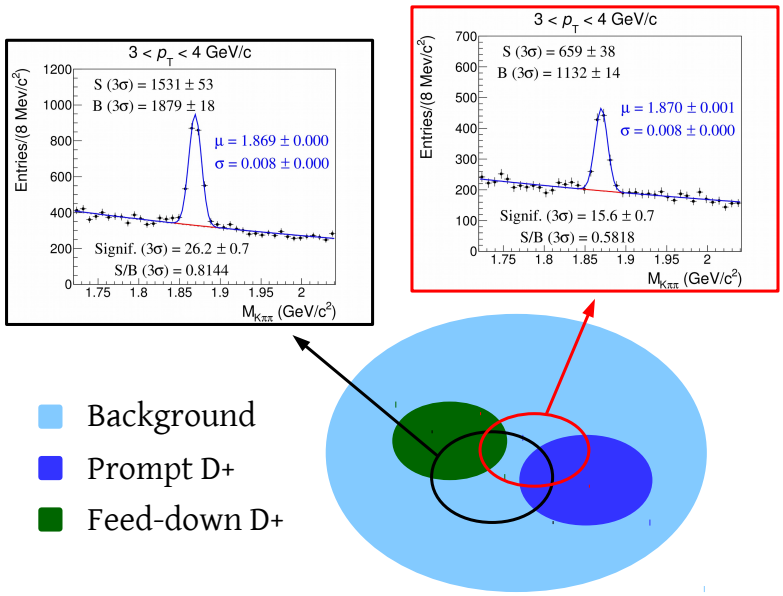

Figure 6.1. Schematic representation of the selection of the D^+ candidates for two different sets of topological cuts.

reconstruction efficiencies for prompt and feed-down D^+ depends on the topological cuts, as explained in section 4.3.1. Hence we can define a equation, which relates the extracted raw yield Y_i and the reconstruction efficiencies for prompt D^+ (ϵ_i^{prompt}) and feed-down D^+ ($\epsilon_i^{feed-down}$), for each given set of topological cuts i:

$$\epsilon_i^{prompt} \cdot N_{prompt} + \epsilon_i^{feed-down} \cdot N_{feed-down} = Y_i, \tag{6.1}$$

where N_{prompt} and $N_{feed-down}$ are the corrected yields for prompt and feed-down D^+ respectively, which we want to determine in order to compute the cross section. However, it is not possible to obtain the corrected yields from a single equation, while at least two are needed. The basic idea of the cut-variation method is therefore to define n different sets of topological cuts in order to obtain a system of n

equations

$$\begin{cases} \epsilon_1^{prompt} \cdot N_{prompt} + \epsilon_1^{feed-down} \cdot N_{feed-down} = Y_1 \\ .. \\ .. \\ \epsilon_n^{prompt} \cdot N_{prompt} + \epsilon_n^{feed-down} \cdot N_{feed-down} = Y_n \end{cases} \tag{6.2}$$

and determine N_{prompt} and $N_{feed-down}$. The system can be exactly solved only in the case of two equations, otherwise it is likely to be overdetermined. Furthermore, since the extracted raw yields and the efficiencies are affected by the experimental uncertainty, a number of sets of topological cuts larger than two would lead to a better estimation of the corrected yields. For this reason, a minimisation procedure is needed in order to obtain an approximated solution of the system. In particular for this analysis two different approaches have been used.

6.1.1 Analytic minimisation

The first method used in the analysis to get an approximated solution of the system 6.2 is based on the minimisation of the χ^2. In order to give a general description of this method, the matrix formalism for n generic sets of topological cuts will be used. The quantities involved in the minimisation are the following:

- The column vector of raw yields Y_i

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ .. \\ Y_n \end{pmatrix}. \tag{6.3}$$

- The efficiency matrix for prompt and feed-down D^+

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1^{prompt} & \epsilon_1^{feed-down} \\ \epsilon_2^{prompt} & \epsilon_2^{feed-down} \\ .. & .. \\ \epsilon_n^{prompt} & \epsilon_n^{feed-down} \end{pmatrix}. \tag{6.4}$$

- The corrected yields vector

$$\boldsymbol{N} = \begin{pmatrix} N_{prompt} \\ N_{feed-down} \end{pmatrix}. \tag{6.5}$$

Hence, the system of equation 6.2, can be rewritten in a more compact way:

$$\boldsymbol{\epsilon} \times \boldsymbol{N} - \boldsymbol{Y} = \boldsymbol{0}, \tag{6.6}$$

where $\boldsymbol{0}$ is the n-dimensional null vector. As already mentioned, for $n > 2$ the system is likely to be overdetermined, and therefore an additional quantity $\boldsymbol{\delta}$ is needed. In particular, $\boldsymbol{\delta}$ is the *residuals vector*, which coincides with the null vector $\boldsymbol{0}$ in case of a determined system, while is different otherwise. Therefore, the latter equation becomes

$$\begin{pmatrix} \epsilon_1^{prompt} & \epsilon_1^{feed-down} \\ \epsilon_2^{prompt} & \epsilon_2^{feed-down} \\ .. & .. \\ \epsilon_n^{prompt} & \epsilon_n^{feed-down} \end{pmatrix} \times \begin{pmatrix} N_{prompt} \\ N_{feed-down} \end{pmatrix} - \begin{pmatrix} Y_1 \\ Y_2 \\ .. \\ Y_n \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ .. \\ \delta_n \end{pmatrix}. \tag{6.7}$$

The best estimation of the corrected yields is obtained by minimising of the χ^2, defined as

$$\chi^2 = \boldsymbol{\delta^T C^{-1} \delta}, \tag{6.8}$$

where $\boldsymbol{\delta^T}$ is the row vector of residuals and $\boldsymbol{C}$ the weighting matrix

$$\boldsymbol{C} = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix}, \tag{6.9}$$

which is considered diagonal. This is clearly an approximation, however the correlation among the n sets of topological cuts is hard to quantify. For this reason only three as much as possible different sets of cuts will be considered for the analysis, as will be described in section 6.2. The weights σ_i^2 are calculated starting from the statistical errors on the raw yields and the efficiencies, according to the following relation:

$$\sigma_i^2 = \sigma_{Y_i}^2 + N_{prompt} \cdot \sigma_{\epsilon_i^{prompt}}^2 + N_{feed-down} \cdot \sigma_{\epsilon_i^{feed-down}}^2. \tag{6.10}$$

Since the corrected yields are not yet determined at this stage, the weights are evaluated interactively: in the first step N_{prompt} and $N_{feed-down}$ are set to zero and therefore only the uncertainty on the raw yields is taken into account, while from the second iteration the weights are computed using the value of the corrected yields obtained in the previous step. Only few iterations are needed to obtain a stable result, since the statistical error on the efficiencies is typically much smaller than the one on the raw yield. Then, the minimisation of the χ^2 with respect to the corrected yields leads to

$$\boldsymbol{N} = \boldsymbol{Cov(N)\epsilon^T C^{-1} Y}, \tag{6.11}$$

where $\boldsymbol{Cov(N)} = (\boldsymbol{\epsilon^T C^{-1} \epsilon})^{-1}$ is the covariance matrix which contains the statistical uncertainties on the corrected yields.

6.1.2 Incentre minimisation

Beside the analytic minimisation, another method based on the geometrical meaning of the set of topological cuts has been used for the analysis. Considering the equation 6.1, each set of cuts can be eventually interpreted as a line

$$N_{prompt} = -\frac{\epsilon_i^{feed-down}}{\epsilon_i^{prompt}} \cdot N_{feed-down} + \frac{Y_i}{\epsilon_i^{prompt}}, \tag{6.12}$$

where the slope is given by the ratio between the efficiencies, while the constant term is given by the ratio of the raw yield and the efficiency. Considering now three different sets of cuts, the three corresponding lines should intersect in one single point, which coordinates correspond to the corrected yields. However, because of the experimental uncertainties on the raw yield extraction and on the efficiency determination, the number of intersections becomes equal to the number of lines. Since the considered sets of cuts is three, the intersections define a triangle. The corrected yields are then evaluated by the coordinates of the incentre of the triangle, which is defined as the crossing point of the internal angle bisectors or equivalently as the point equidistant from the triangle's sides.

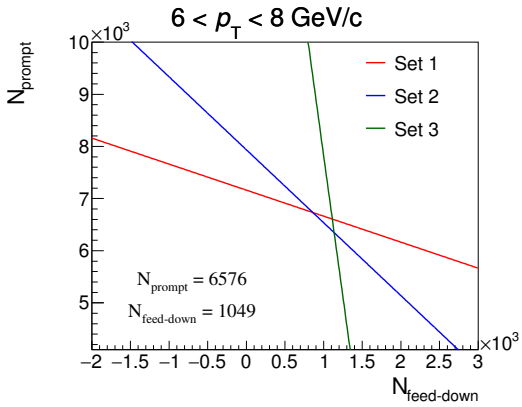

Figure 6.2. Three lines corresponding to different sets of topological cuts in the interval $6 < p_{\mathrm{T}} < 8$ GeV/c. Their intersections define a triangle, which incentre is used to evaluate the corrected yields for prompt and feed-down D^+-mesons.

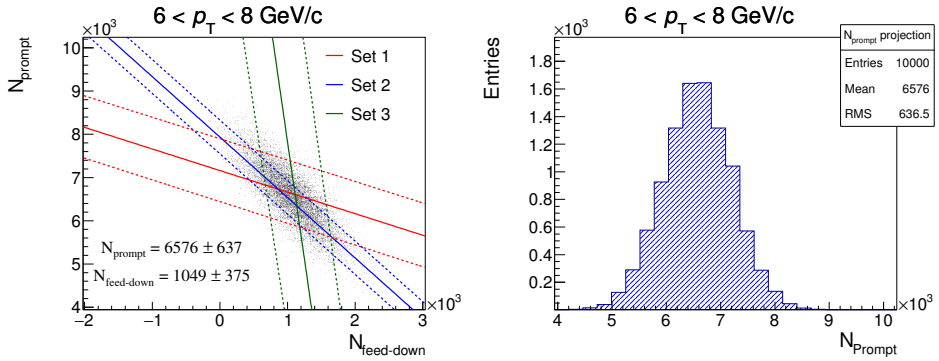

Figure 6.3. Left: distribution of the generated incentres with the toy MC simulation superimposed to the lines corresponding to three different sets of cuts in the interval $6 < p_{\mathrm{T}} < 8$ GeV/c. The dotted lines represent the maximum effect of the error on the line parameters within 1σ. Right: projection of the distribution of the generated incentres on the N_{prompt}-axis.

Estimation of the statistical error on the incentre coordinates

The error on the slope of the lines depends on the uncertainty on the efficiencies for prompt and feed-down D^+-mesons, while the error on the constant term, which implies a shift of the line, is dominated by the raw yield uncertainty, since is typically larger than the one on the efficiency. The statistical uncertainty on the coordinates of the incentre, which represent the corrected yields, is propagated from the errors on the line parameters with a toy Monte Carlo simulation. It consists in generating several times the parameters of the lines from a Gaussian distribution, with mean equal to the value of the considered parameter and sigma equal to its error. For each generation then the coordinates of the incentre are computed and a two-dimentional histrogram is filled. Finally, the uncertainty on the single coordinate is evaluated as the RMS of the one-dimensional distribution, obtained projecting the two-dimentional distribution of the incentres to the corresponding axis. In the right panel of figure 6.3 is shown an example of incentre distribution obtained with the toy MC simulation in the transverse momentum inerval $6 < p_{\mathrm{T}} < 8$ GeV/c, together with the lines used for the incentre determination, which correspond to three different sets of topological cuts. The dotted lines are obtained setting the line parameters to their value plus or minus their error, in order to give a visual representaition of the uncertainty on the determination of the lines. In the right panel of the same figure is shown the projection on the N_{prompt}-axis of the distribution of the incentre coordinates, which was used to assign the statistical uncertainty on the corrected yield for prompt D^+.

6.2 Strategy for the selection of topological cuts

As described in the previous section, the weighting matrix used for the analytic minimisation is diagonal (see equation 6.9) and therefore it neglects possible correlations among the sets of topological cuts considered. Furthermore, considering the incentre method, if two or more sets are similar (and therefore correlated), the corresponding lines would be almost parallel, resulting in a poor precision of the incentre calculation, because of the large overlap of the regions defined by the dotted lines in figure 6.3. Concerning that, the sets of topological cuts have to be chosen in order to generate sub-samples of D^+ candidates as much as possible disjoined. Moreover, a large number of sets would reduce significantly the statistics in each sub-sample, then only three cut sets were defined for the analysis:

- **Maximised prompt contribution**: this first set of topological cuts is used to maximise the prompt D^+ contribution. However, since the efficiency is typically higher for feed-down D^+, the aim for this achieve a similar or slightly higher efficiency for the prompt contribution with respect to the feed-down one.

- **Maximised significance**: the goal of this cut set is to extract the best signal from the invariant-mass spetra. For this reason this set results similar to the one defined for the standard analysis, which was optimised to obtain the best signal over background ratio and the best significance possible.

- **Maximised feed-down contribution**: the third cut set aim to maximise the feed-down D^+ contribution. Even if the maximisation of the efficiency for feed-down D^+ can be easily achieved, the complication for this set of topological cuts is to obtain a reasonable yield extraction, since the feed-down D^+ are the minority of the candidates.

6.2.1 Topological variables

In order to obtain the three different cut sets described above, the topological varibles used must have a significant separation power between the prompt and the feed-down contributions. Concerning that, the following quantities are used.

Decay length in the xy-plane

As already discussed in section 4.1.4, the decay length and the normalised decay length can discriminate between the signal and background candidates (see right panel of figure 4.3). However this quantity can also be used to separate the prompt from the feed-down contribution, since the D^+ which decay from B-hadrons are on average more displaced. Furthermore, the decay length in the transverse plane

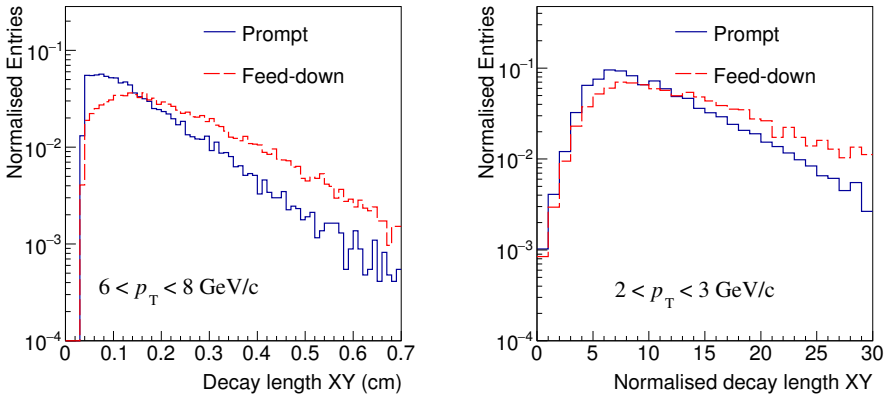

Figure 6.4. Comparisons between the prompt and feed-down D^+ distributions of the decay length in the xy-plane (left) in the interval $6 < p_T < 8$ GeV/c and the normalised decay length in the xy-plane (right) for $4 < p_T < 5$ GeV/c from MC simulations.

would allow to achieve a better separation, because of the better resolution with respect to the z-coordinate. An example of this propriety is illustrated in figure 6.4, where is shown the comparison between the distributions of prompt and feed-down D^+ of the decay length in the xy-plane for $6 < p_T < 8$ GeV/c (left) and th normalised decay length in the xy-plane in the interval $2 < p_T < 3$ GeV/c (right) obtained from MC simulations.

Cosine of the pointing angle in the xy-plane

The momentum vector of the feed-down D^+-mesons in general is not parallel to the line which connect the primary and the decay vertex of the D^+, but their relative orientation depends on the direction of the D^+-meson with respect to the B-hadron. As a result, the distribution of the cosine of the pointing angle for feed-down D^+ is less peaked at 1 than the one of prompt D^+. In the left panel of figure 6.5 the MC distributions for prompt and feed-down D^+ in the transverse momentum interval $12 < p_T < 16$ GeV/c are compared. In order to achieve a better separation of the two contributions, only the xy-components of the cosine of the pointing angle are exploited, because of the better resolution with respect to the z-coordinate.

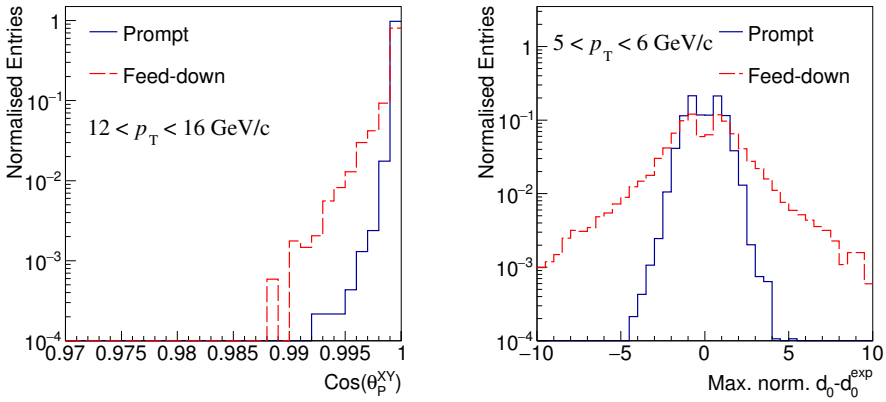

Figure 6.5. Comparisons between distributions of the cosine of the pointing angle in the xy-plane in the interval $12 < p_{\mathrm{T}} < 16$ GeV/c (left) and the topomatic variable (right) in the interval $5 < p_{\mathrm{T}} < 6$ GeV/c for prompt and feed-down D^+ from MC simulations.

Normalised residual of the daughter impact parameter

Another quantity which can discriminate the prompt from the feed-down contribution is the so called *topomatic* variable, defined as the maximum of the normalised residuals of the daughter impact parameter in the transverse plane:

$$max\ n\sigma_{res} = max\left\{\frac{d_{0,xy}^{exp} - d_{0,xy}^{meas}}{\sqrt{\sigma_{d_{0,xy}^{exp}}^2 + \sigma_{d_{0,XY}^{meas}}^2}}\right\}, \tag{6.13}$$

Where $d_{0,xy}^{meas}$ is the measured value of the daughter impact parameter in the xy-plane and $\sigma_{d_{0,XY}^{meas}}$ its error. The expected value of the impact parameter for each daughter is computed as

$$d_{0,xy}^{exp} \simeq L_{xy}\sin(\theta_{xy}^{tr}), \tag{6.14}$$

where L_{xy} is the D^+ decay length and θ_{xy}^{tr} is the angle between the reconstructed momenta of the considered daughter and the D^+-meson in the xy-plane, as shown in the sketch in figure 6.6. Furthermore, the error on $d_{0,xy}^{exp}$ is calculated negleting the error on $\sin(\theta_{xy}^{tr})$, and therefore as

$$\sigma_{d_{0,xy}^{exp}} = \sigma_{L_{xy}}\sin(\theta_{xy}^{tr}). \tag{6.15}$$

The topomatic variable allows to reach a significant separation between the two contributions, since it is on average larger for feed-down D^+, as shown in the right

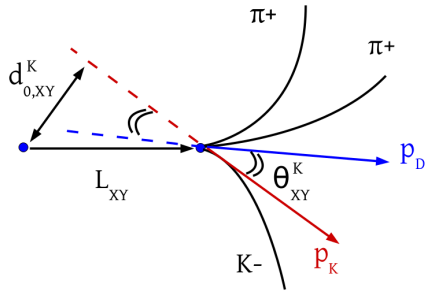

Figure 6.6. Sketch of the D^+ decay topology with the quantites used to compute the topomatic variable.

panel of figure 6.5, where the distributions of prompt and feed-down D^+ in the p_T interval $5 < p_T < 6$ GeV/c from MC simulations are compared.

6.2.2 Sets of topological cuts

For all the sets, in order to remove as much as possible the combinatorial background and extract the signal, some common topological cuts are applied. These cuts are the same as those in the standard analysis, which are summarised in table

	p_T (GeV/c)	[2,4]	[4,6]	[6,8]	[8,12]	[12,16]
Set 1	L_{xy} (cm)	[0.02,0.14]	[0.02,0.15]	[0.02,0.15]	[0.04,0.15]	[0.04,0.25]
	Norm. L_{xy}	[7,14]	[4,16]	[4,16]	[4,16]	[4,22]
	$Cos(\theta_P^{xy})$	>0.998	>0.998	>0.998	>0.998	>0.998
	Topomatic	[−1.5,1.5]	[−1.5,1.5]	[−1.5,1.5]	[−1.5,1.5]	[−2.0,2.0]
Set 2	L_{xy} (cm)	>0.05	>0.05	>0.05	>0.10	>0.10
	Norm. L_{XY}	>9	>9	>9	>6	>6
	$Cos(\theta_P^{xy})$	>0.990	>0.990	>0.990	>0.990	>0.990
Set 3	L_{xy} (cm)	>0.08	>0.12	>0.20	>0.20	>0.20
	Norm. L_{xy}	>9	>12	>20	>20	>20
	$Cos(\theta_P^{xy})$	>0.990	>0.990	>0.990	>0.990	>0.990
	Topomatic	<-1.5	<-1.5	<-1.5	<-1.5	<-1.0

Table 6.1. Summary table of the three sets of topological cuts used for the D^+ cut-variation analysis.

4.2. Furthermore, all the event, track quality and PID selections described in chapter 4 are used. The only difference concerns the cuts applied on the variables listed above, which are not the same as in the standard analysis, but they are different in order to obtain the three sets of topological cuts described in section 6.2. The values of the cuts used for this purpose are reported in table 6.1.
For the first set an upper cut to the decay length and the normalised decay length in the transverse plane has been applied to reject the feed-down contribution. However, in order to extract the raw yields, also a lower cut has been applied to these variables. Furthermore, a lower cut to the cosine of the pointing angle in the xy-plane and a symetric cut with respect to the zero to the topomatic variable, has been applied to remove the tails, where the feed-down contribution is dominant.
The second set of topological cuts, as already mentioned, was chosen similar to the one of the standard analysis, therefore lower cuts to the decay length, the normalised decay length and the cosine of the pointing angle in the transverse plane have been applied in order to reject the combinatorial background.
For the third set, tighter lower cuts to the decay lenght and the normalised decay length has been applied to reduce the prompt contribution. For the same purpose only the left tail of the distribution of the topomatic variable was kept. Finally, the cut on the cosine of the pointing angle was set only to reject the background, since an upper cut to reduce the prompt contribution implies a significant loss of signal, which does not allow to extract the raw yields.

The p_{T} range considered for this analysis is $2 < p_{\mathrm{T}} < 16$ GeV/c, since for the first and the last p_{T} bins of the standard analysis was not possible to extract the signal in all the three sub-samples, due to the high combinatorial bakground at low p_{T} and the low statistics at high p_{T}. The p_{T} bins are the same as those in the impact-parameter fit analysis.

6.3 Efficiency

The efficiencies for prompt and feed-down D^+-mesons are obtained from the MC simulation, as explained in 4.3.1. However an additional consideration must be taken into account, since the efficiency is computed in p_{T} intervals of finite size and therefore it depends on the p_{T} spectra of D^+-mesons and B-hadrons which decay into D^+. Hence, in order to avoid a bias due to the different p_{T} shape of generated D^+ in data and MC, the D^+-meson spectrum was re-weighted to reproduce the FONLL p_{T} shape at $\sqrt{s_{NN}} = 5.02$ TeV. For the feed-down D^+ the weights have been computed in order to match the FONLL p_{T} spectrum of the B-hadrons which decay into D^+. The difference between the PYTHIA and FONLL p_{T} shape of D^+-mesons and B-hadrons and the ratio between them, which was used to re-weight the efficiencies, is shown in figure 5.14.

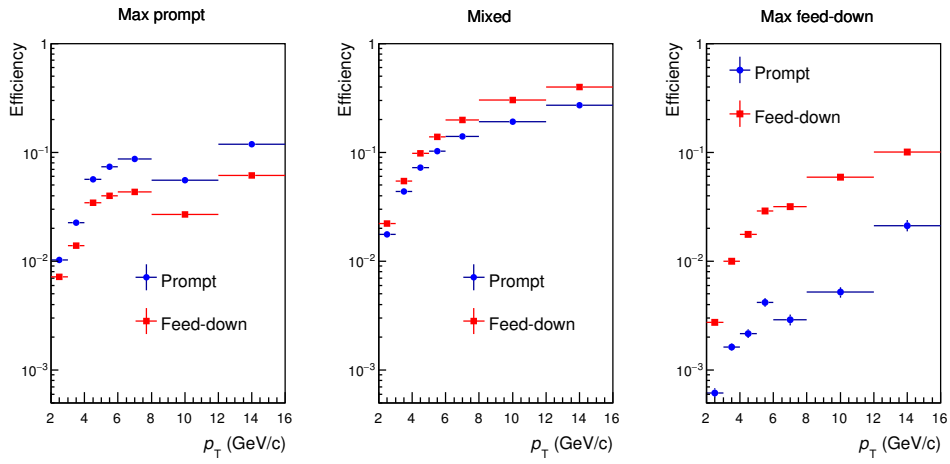

Figure 6.7. Comparison between the efficiencies for prompt and feed-down D^+ for the three sets of topological cuts used in the cut-variation analysis.

The resulting efficiencies for the maximised prompt cut set are shown in the left panel of figure 6.7. The efficiency for prompt D^+ is higher than the corresponding efficiency for feed-down D^+ in the full p_T range, by a factor $\sim 1.5-2$. For this set of topological cuts the values of the efficiency for prompt (feed-down) D^+ increases from about 1% (0.7%) up to a maximum of about 12% (6%), with increasing transverse momentum.
The values of the efficiency for the second set, shown in the middle panel of figure 6.7, are close to the ones of the standard analysis (see section 4.3.1), since the topological cuts applied are similar. In particular, the efficiency for feed-down D^+ varies from $\sim 2\%$ for $2 < p_T < 3$ GeV/c up to about 40% in the highest p_T bin of the analysis. It results to be slightly higher than the prompt D^+ efficiency, which reaches $\sim 30\%$ for $12 < p_T < 16$ GeV/c, since at the same transverse momentum the feed-down D^+ are on average more displaced from the primary vertex.
Finally, in the right panel of figure 6.7 are shown the resulting efficiencies for the cut set used to maximise the feed-down contribution. For this set of topological cuts the prompt contribution is strongly suppressed, as confirmed by its efficiency, which varies from about 0.06% to 2%. The feed-down D^+ efficiency is also lower than the one in the second cut set, however it is higher than the one for prompt D^+-mesons by up to a factor ~ 10 and therefore it results the dominant contribute.

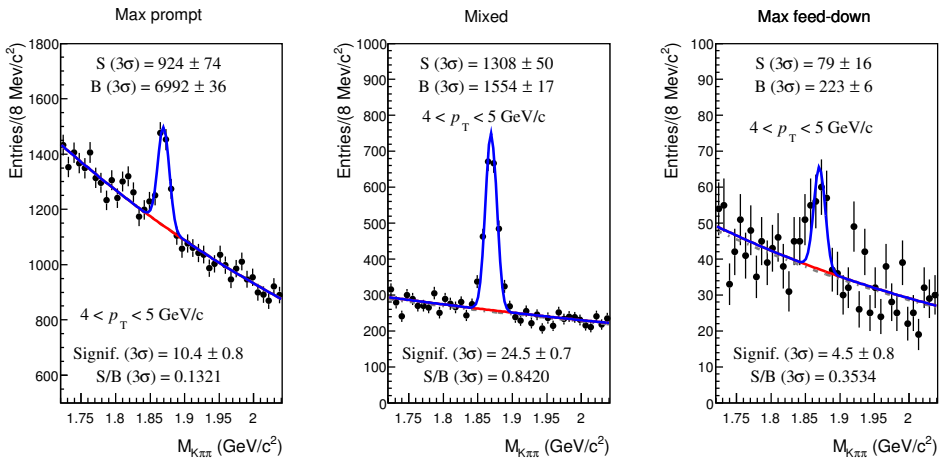

Figure 6.8. Example of invariant-mass fits for the three different sets of topological cuts in the transverse momentum interval $4 < p_{\mathrm{T}} < 5$ GeV/c.

6.4 Raw yield extraction

The raw yields were obtained by fitting the invariant-mass spectra of each subsample obtained applying the three different sets of topological cuts summarised in table 6.1, in addition to all the common selections (event, track quality, PID and common topological cuts), as explained in section 6.2.2. The fitting function is the sum of an exponential function which describes the background and a Gaussian function for the signal. In figure 6.8 is shown an example of invariant-mass fits in the interval $4 < p_{\mathrm{T}} < 5$ GeV/c for the prompt-enhanced set (left), mixed set (middle) and feed-down-maximised set (right). The most critical invariant-mass fits are those obtained applying the third cut set, used to maximise the feed-down contribution, because of the low statistics. In particular, the statistical flactuations for this set lead to a very unstable result of the fit in the full p_{T} range letting the Gaussian width as a free parameter. For this reason, the Gaussian width was fixed in each set to the values of the MC simulation obtained with the mixed cut set, which is the most stable since it was optimised to obtain a good significance. This was justified by comparing the Gaussian width obtained by fitting the invariant-mass spectra in the data to those in the MC simulation. The peak width was found to be compatible within two sigmas with the MC values in all the considered p_{T} intervals, except for the bin $4 < p_{\mathrm{T}} < 5$ GeV/c, as shown in the left panel of figure 6.9. This discrepancy was taken into account in the evaluation of the systematic errors, as will be described in section 6.5.2. Also the peak position was found to

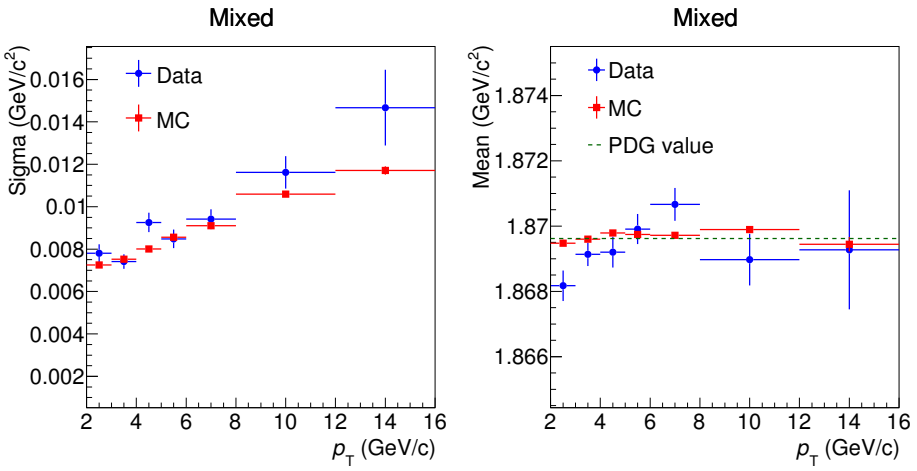

Figure 6.9. Comparison between the peak width (left) and mean (right) obtained by fitting the invariant-mass spectra in data and MC for the mixed cut set. The peak position is also compared to the PDG value of the mass of the D^+-meson.

be compatible within two sigmas with the MC and with the PDG values, except for the first p_T bin, as shown in the right plot of figure 6.9 for the second set of topological cuts. Then, for each cut set, the D^+ raw yields are obtained integrating the signal fit function over the mass range within 3σ. The extracted raw yields together with the significance for each cut set are reported in table 6.2, while all the invariant-mass fits can be found in the appendix A.5.1.

	S (3σ)			*significance* (3σ)		
p_T (GeV/c)	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
[2,3]	479 ± 36	897 ± 43	40 ± 14	11.8 ± 0.8	18.7 ± 0.7	2.2 ± 0.7
[3,4]	659 ± 38	1531 ± 53	74 ± 20	15.6 ± 0.7	26.2 ± 0.7	3.1 ± 0.8
[4,5]	924 ± 74	1308 ± 50	79 ± 16	10.4 ± 0.8	24.5 ± 0.7	4.5 ± 0.8
[5,6]	588 ± 57	999 ± 41	68 ± 14	8.7 ± 0.8	21.8 ± 0.7	4.3 ± 0.8
[6,8]	633 ± 56	1126 ± 44	42 ± 11	9.9 ± 0.8	23.3 ± 0.7	4.1 ± 0.7
[8,12]	211 ± 29	787 ± 29	53 ± 12	6.7 ± 0.8	18.6 ± 0.7	4.0 ± 0.8
[12,16]	95 ± 17	242 ± 24	27 ± 8	5.3 ± 0.8	9.6 ± 0.8	2.8 ± 0.7

Table 6.2. Summary table of the raw yields and the significance for the three sets of topological cuts.

6.5 Systematic uncertainties

In this section, the sources of systematic uncertainties on the determination of the corrected yields with the cut-variation analysis are discussed. The systematics were studied for both the analytic and the incentre minimisation methods, evaluating the influence of each source via the variation of the corrected yields. The three contributions, which have been tested, are:

1. Systematic uncertainty due to the efficiency of the selection criteria for the sets of topological cuts.
2. Systematic uncertainty due to the signal extraction.
3. Systematic uncertainty due to the p_{T} shape of generated D^{+} and B mesons.

Furthermore, two additional sources have been taken into account for the evaluation of the total systematic uncertainty on the cross section measurement:

4. Systematic uncertainty due to the tracking efficiency.
5. Systematic uncertainty due to the PID selection efficiency.

The latter two contributions were assumed to be equal to those in the standard analysis. In particular, the uncertainty on the Particle Identification was found to be negligible, while the one on the tracking efficiency was of the order of 3% for each measured track, resulting in a 9% total uncertainty in the full p_{T} range considered.

6.5.1 Systematic uncertainty due to the topological selection efficiency

If the distributions of the topological variables were exactly reproduced in the Monte Carlo simulation, the result would be independent from the cut values used for the analysis. However, a possible difference between the distributions in data and MC could lead to a bias in the result. For this reason, the minimisation procedure was repeated varying the selection criteria for each cut set. In particular, the corrected yields have been re-computed, adopting a lower and a tighter cut value for each topological variable used, separately for each set. For example, the minimisation procedure was repeated setting the cut on the cosine of the pointing angle in the xy-plane for the prompt-enhanced set to 0.995 and 0.999, while the central value was 0.998. It is important to notice that the cut values adopted for the variations cannot be too different from the central values (see table 6.1), in order to keep the three sets of topological cuts as much as possible uncorrelated. In the left panel of figure 6.10 is shown the ratio between the corrected yields for

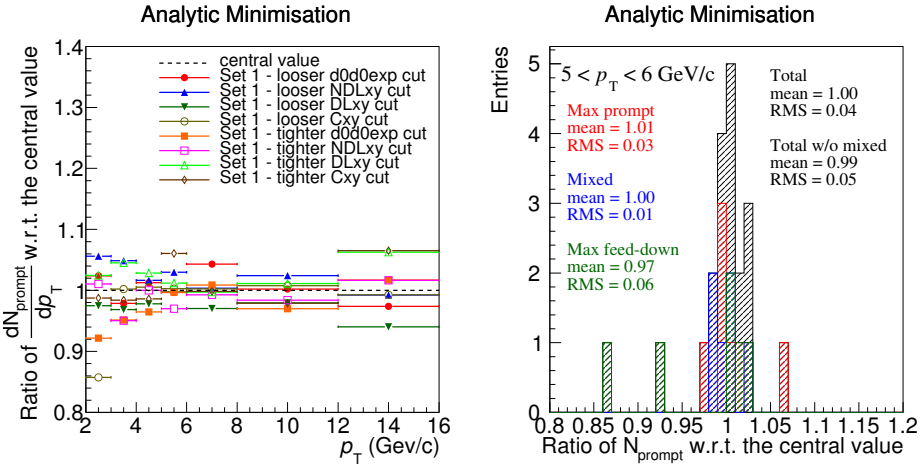

Figure 6.10. Left: ratio of the corrected yields for prompt D^+ as a function of p_T obtained with the analytic minimisation with respect to the central value, for the prompt-enhanced set cut variations. Right: distributions of the corrected yield variations in the interval $5 < p_T < 6$ GeV/c for the three cut sets separately (red, blue, green) and for all the trials together (black).

prompt D^+-mesons obtained with the analytic minimisation for each trial with respect to central value, for the variations applied to the maximised prompt set. The procedure was repeated for the other two sets of topological cuts used in the analysis and for both the minimisation methods. The resulting variations have been treated with two different approaches:

- Considering the variation of the three sets of topological cuts as uncorrelated.
- Considering all the variations together.

In the first case the systematic uncertainty was assigned adding in quadrature the three contributes, evaluated with the sum of the RMS and the shift of the mean, with respect to the unity, of the distribution of the variations for each set of topological cuts separately. In the second case, the uncertainty is estimated considering the RMS and the shift of the distributions of all the variations together. The resulting values are summarised in table 6.3, where the column entitled *Sum* corresponds to the uncorrelated case, while the one entitled *Tot* corresponds to the second case. It is possible to notice how the values obtained considering all the trials together are systematically smaller than the ones evaluated considering the variations of three cut sets as uncorrelated. This is due to a partial compensation

p_T (GeV/c)	Prompt D^+ corrected yields									
	Analytic					Incentre				
	Set 1	Set 2	Set 3	Sum	Tot	Set 1	Set 2	Set 3	Sum	Tot
[2,3]	8%	2%	9%	12%	7%	5%	2%	11%	12%	7%
[3,4]	5%	2%	5%	7%	4%	7%	2%	5%	9%	6%
[4,5]	2%	1%	3%	4%	3%	3%	0%	4%	5%	2%
[5,6]	4%	1%	9%	10%	4%	6%	0%	6%	8%	4%
[6,8]	2%	2%	8%	8%	5%	4%	0%	4%	6%	4%
[8,12]	3%	2%	6%	7%	3%	6%	0%	3%	7%	3%
[12,16]	5%	1%	4%	6%	3%	6%	1%	2%	6%	4%

Table 6.3. Variations of the prompt D^+ corrected yields for the single cut sets, the sum in quadrature of the three contributions (*Sum*) and the total variation induced by all the trials together (*Tot*).

of the variations induced by the three cut sets, which leads to an overestimation of the uncertainty if the three variations are considered uncorrelated. In the right panel of figure 6.10 is shown an example of this effect in the transverse momentum interval $5 < p_T < 6$ GeV/c: the distribution of the variations for the first cut set (red) is shifted toward values higher than unity, while the one for the third cut set (green) toward lower values. Therefore, the total distribution results to be centred at one. Furthermore, from the values reported in table 6.3, it is possible to notice how the mixed cut set is much more stable than other two in the full p_T range, for both the minimisation methods. Then, considering the distribution of all the variations together, it implies an artificial reduction of the RMS and therefore of the uncertainty. Concerning that, the systematic uncertainty on the prompt D^+ corrected yields was assigned considering the RMS plus the shift of the distributions without considering the variations of mixed cut set. The comparison between the

	p_T (GeV/c)	[2,3]	[3,4]	[4,5]	[5,6]	[6,8]	[8,12]	[12,16]
Tot	analytic	7%	4%	3%	4%	5%	3%	3%
	incentre	7%	6%	2%	4%	4%	3%	4%
W/o mixed	analytic	9%	4%	4%	6%	6%	4%	4%
	incentre	8%	7%	4%	5%	4%	4%	5%

Table 6.4. Comparison between the systematic uncertainty due to the topological cut efficiency for prompt D^+ estimated considering all the variations and excluding the variations of the mixed cut set.

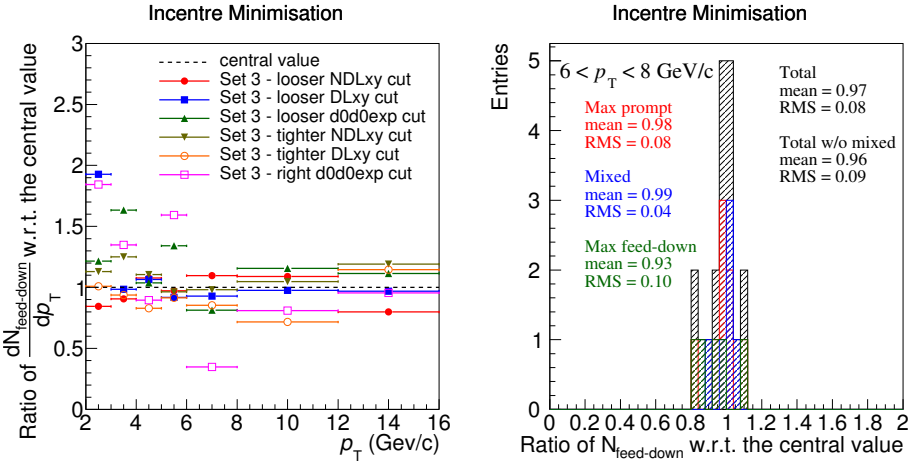

Figure 6.11. Left: ratio of the corrected yields for feed-down D^+ as a function of p_T obtained with the incentre minimisation with respect to the central value, for the feed-down-enhanced set variations. Right: distributions of the corrected yield variations in the interval $6 < p_T < 8$ GeV/c for the three cut sets separately (red, blue, green) and for all the trials together (black), obtained removing the bad trial, corresponding to the variation of the topomatic variable cut.

values obtained considering all the three sets and those obtained removing the mixed one are reported in table 6.4.

The same procedure was done for the corrected yields for feed-down D^+-mesons. In this case the result was found to be more unstable, since the determination of the feed-down contribution mostly depends on the third cut set, which it is the most critical in terms of signal extraction. Therefore, the choice of tighter cut values could lead to a bad raw yield extraction, which is reflected in the determination of the corrected yield. For this reason, the trials corresponding to a bad mass-fit result were excluded, in order not to double count the systematic uncertainty on the raw yield extraction and to limit as much as possible the contribution due to the statistical fluctuations. In the right panel of figure 6.11 is shown the ratio of the corrected yields for the feed-down contribution with respect to the central value for the cut variations of the maximised feed-down set. In the right plot of the same figure are represented the distributions of the variations in the interval $6 < p_T < 8$ GeV/c, where the bad trial corresponding to the variation of the topomatic variable for the third cut set has been excluded. All the values obtained considering the three sets of topological cuts separately and together are reported in table 6.5.

	Feed-down D^+ corrected yields									
p_T	Analytic					Incentre				
(GeV/c)	Set 1	Set 2	Set 3	Sum	Tot	Set 1	Set 2	Set 3	Sum	Tot
[2,3]	14%	10%	29%	34%	26%	27%	27%	36%	52%	25%
[3,4]	26%	21%	34%	48%	28%	19%	28%	35%	49%	23%
[4,5]	10%	5%	10%	15%	10%	7%	6%	11%	14%	9%
[5,6]	12%	2%	38%	40%	20%	9%	5%	32%	34%	21%
[6,8]	10%	5%	17%	20%	11%	7%	4%	18%	20%	10%
[8,12]	10%	6%	19%	22%	11%	5%	4%	22%	23%	11%
[12,16]	15%	7%	16%	23%	12%	12%	8%	17%	22%	12%

Table 6.5. Variations of the feed-down D^+ corrected yields for the single cut sets, the sum in quadrature of the three contributions (*Sum*) and the total variation induced by all the trials together (*Tot*).

As expected, the largest contribution is given by the variations of the feed-down-enhanced set but, unlike the values obtained for the prompt D^+ corrected yields, the uncertainty induced by the mixed set is not negligible. Therefore, in case of the corrected yields for the feed-down D^+-mesons, the uncertainty is assigned considering the distributions of all the variations, without excluding the second set of topological cuts.

6.5.2 Systematic uncertainty due to the yield extraction

The systematic uncertainty on the yield extraction was evaluated by changing the parameters of the invariant-mass fits for all the three sets of topolgical cuts and performing the minimisation procedure for each trial. In particular, the fits have been repeated setting all the possible combinations of the following variations:

- The Gaussian width of the signal was fixed to the value obtained from the MC and varying it within ± 15%.
- The Gaussian mean of the signal was first let free and then fixed to the PDG value of the D^+-meson mass.
- The bin width of the invariant-mass distributions was changed from 4 to 16 MeV/c^2.
- The range of the invariant-mass fit has been varied using 4 different lower and upper limits.

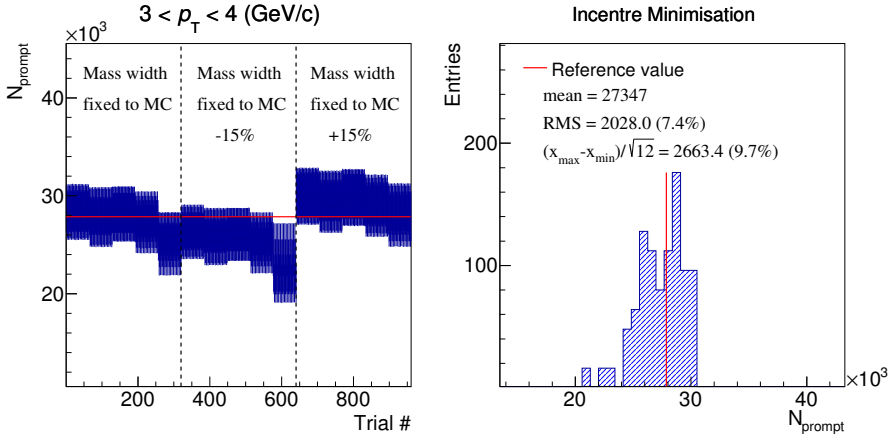

Figure 6.12. Prompt D^+ corrected yield obtain with the incentre minimisation method as a function of the trial number (left) and the resulting distribution (right) for the p_T interval $3 < p_T < 4$ GeV/c.

- Two different functions have been used for the parametrisation of the background (exponential and parabolic).

The systematic uncertainty was then estimated considering the RMS of the distributions of the corrected yields computed using both the minimisation procedures. Since the resulting distributions are not Gaussian for all the p_T intervals considered in the analysis, also the variance assuming a uniform distribution was evaluated. In figures 6.12 and 6.13 are shown two examples of variation of the prompt and the feed-down corrected yields in the interval $3 < p_T < 4$ GeV/c and $5 < p_T < 6$ GeV/c respectively. In the left panel is shown the variation of the corrected yields as a function of the trial number, while in the right plot is shown the resulting distributions. For the prompt D^+ case the maximum variation is obtained by

	p_T (GeV/c)	[2,3]	[3,4]	[4,5]	[5,6]	[6,8]	[8,12]	[12,16]
Prompt D^+	analytic	12%	8%	8%	8%	9%	9%	15%
	incentre	12%	8%	8%	8%	10%	12%	15%
Feed-down D^+	analytic	26%	20%	20%	16%	10%	11%	20%
	incentre	30%	20%	20%	18%	11%	11%	20%

Table 6.6. Systematic uncertainty on the corrected yields due to the raw yield extraction.

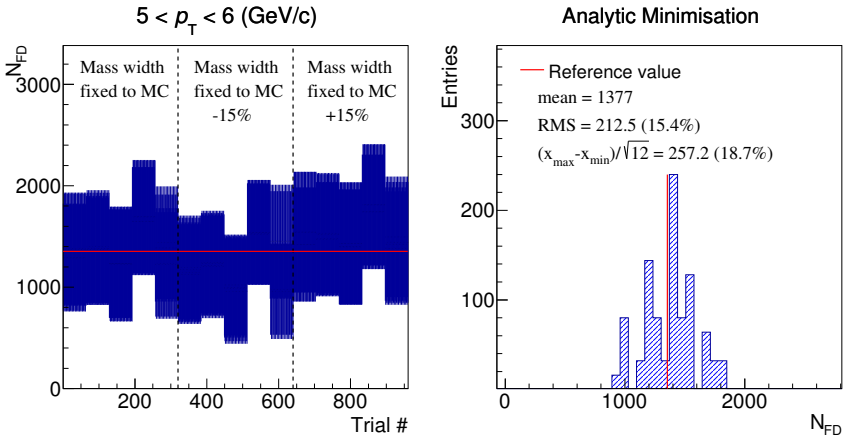

Figure 6.13. Feed-down D^+ corrected yield obtained with the analytic minimisation method as a function of the trial number (left) and the resulting distribution (right) for the p_T interval $5 < p_T < 6$ GeV/c.

changing the Gaussian width, however all the trials are compatible among each other within the statistical uncertainty. In table 6.6 are reported the values of systematic uncertainty assigned to the corrected yields for both the contributions and the minimisation methods, in the full p_T range of the analysis.

6.5.3 Systematic uncertainty due to the p_T shape

The shape of the p_T distribution of the generated D^+ affects the efficiency, as discussed in section 6.3. Therefore, the systematic uncertainty on the determination of the efficiency was estimated from the relative variation of the corrected yields obtained using the efficiencies computed with the FONLL and MC p_T shapes. In the left panel of figure 6.14 is shown the ratio of the efficiencies for prompt and

	p_T (GeV/c)	[2,3]	[3,4]	[4,5]	[5,6]	[6,8]	[8,12]	[12,16]
Prompt D^+	analytic	1%	1%	1%	1%	1%	2%	2%
	incentre	1%	1%	1%	1%	1%	4%	2%
Feed-down D^+	analytic	1%	1%	1%	1%	1%	5%	5%
	incentre	1%	1%	1%	1%	1%	5%	5%

Table 6.7. Systematic uncertainty on the corrected yields due to the p_T shape of the generated D^+-mesons.

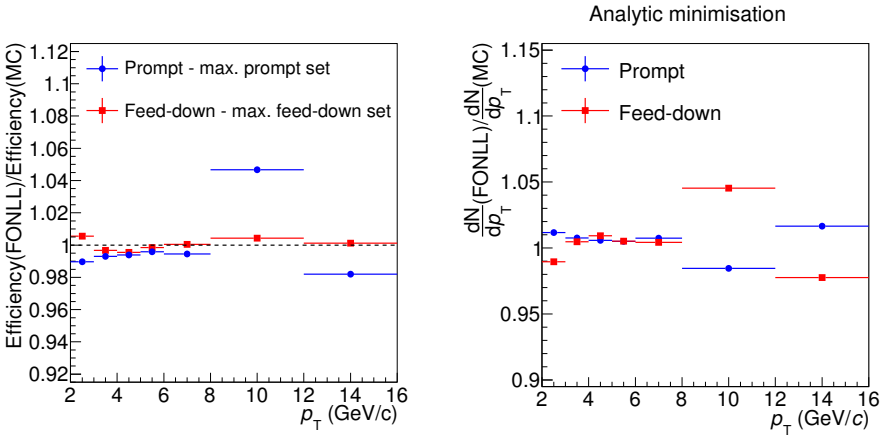

Figure 6.14. Left: ratio of the efficiency for prompt and feed-down D^+-mesons obtained using the FONLL and MC p_T shapes, for the prompt-enhanced and the maximised feed-down cut set respectively. Right: ratio of the corrected yields obtained with the analytic minimisation, using the efficiencies computed with the FONLL and MC p_T shapes.

feed-down D^+-mesons obtained with the two different p_T shapes for the prompt-enhanced and the maximised feed-down cut sets respectively, while in the right panel the ratio of the corrected yields computed with the analytic minimisation, using the efficiencies obtained with the two different p_T shapes. In table 6.7 are reported the values assigned as systematic uncertainty on the basis of the corrected yield variations.

6.6 Results

After the minimisation procedure, the p_T-differential cross section was calculated as:

$$\left.\frac{d\sigma_X^{D^+}}{dp_T}\right|_{y<0.5} = \frac{1}{2} \cdot \frac{N_X^{D^\pm}|_{y<y_{fid}}}{\Delta p_T \Delta y \cdot (Acc)_X} \cdot \frac{\sigma_{MB}^{pPb}}{BR \cdot N_{ev}}, \tag{6.16}$$

where X indicates the prompt or the feed-down production, and $N_X^{D^\pm}$ are the corresponding corrected yields computed with the minimisation methods explained in 6.1. Like in the formula used in the standard analysis (see equation 4.8), the factor 1/2 takes into account the fact that the corrected yields include both the particles and the anti-particles, Δy the rapidity coverage, which is common for all

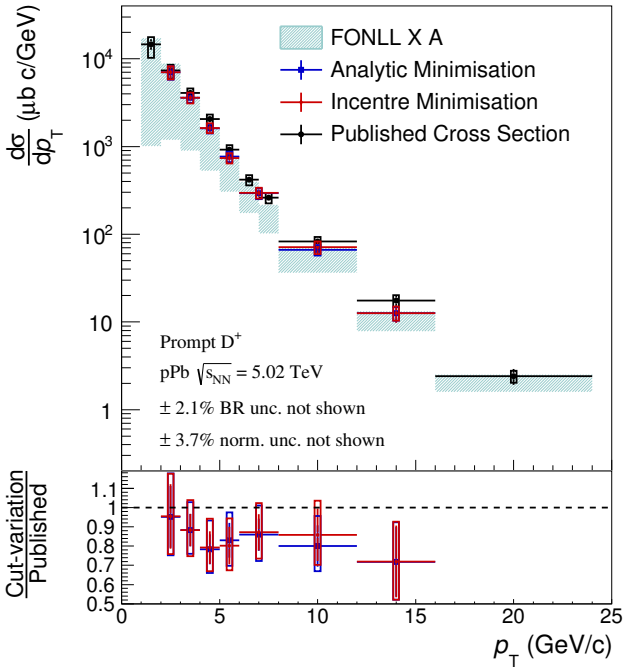

Figure 6.15. Top: prompt D^+ production cross section measured in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV using the cut-variation method, compared to the result obtained with the standard approach published by the ALICE Collaboration [36]. The vertical bars represent the statistical errors, while the boxes the systematic uncertainties. Bottom: ratio of the measured cross section with the cut-variation approach and the published result.

the transverse momentum intervals Δp_{T}, σ_{MB}^{pPb} is the total inelastic Minimum Bias cross section and N_{ev} is the normalisation factor computed as reported in equation 4.9. On the contrary, in equation 6.16 does not compare the efficiency factor, since $N_X^{D^\pm}$ are already corrected for the efficiency, and the fraction of prompt D^+, which is not needed in this analysis. The resulting p_{T}-differential cross section of prompt D^+-mesons obtained with the cut-variation method is shown in figure 6.15, compared with the published result obtained with the standard approach and to the FONLL prediction at $\sqrt{s_{NN}} = 5.02$ TeV rescaled for the lead mass number, which should describe the data in case of binary scaling. This is not the case, since we

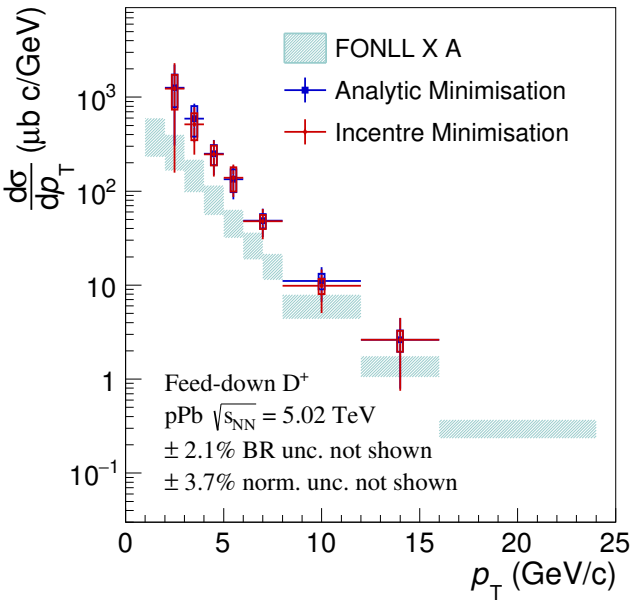

Figure 6.16. Feed-down D^+ production cross section measured in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV using the cut-variation method. The vertical bars represent the statistical errors, while the boxes the systematic uncertainties.

know that the binary scaling is broken, however the measured nuclear modification factor R_{pPb} of D-mesons shows that the Cold Nuclear Matter effcts are small (see section 2.2.1). In the bottom panel it is possible to see the ratio between the result obtained with the cut-variation method and the standard approach. The statistical and the systematic uncertainties have been treated as uncorrelated in the ratio, except for the systematic error on the tracking efficiency, which has been considered as fully correlated. The results obtained with the analytic and the incentre minimisation methods are very similar and they are both compatible within the uncertainties with the published one, however systematically lower of about 20%. This effect could be due to the tight selection criteria needed to obtain three different sets of topological cuts, which can create a bias in case of a not perfect description of the distributions of the topological variables in the MC simulation. Finally, in figure 6.16 is shown the p_{T}-differential cross section of feed-down D^+-mesons obtained with the cut-variation analysis compared to the FONLL prediction at $\sqrt{s_{NN}} = 5.02$ TeV rescaled for the lead mass number. As already

mentioned the FONLL prediction does not describe the data in p-Pb collisions, however with the current level of uncertainty for feed-dwon D^+ the difference should be negligible. On the contrary, the result obtained with the cut-variation analysis is significantly higher than the FONLL prediction, especially at low p_T. This could be a consequence of the effect observed in the prompt D^+ differential cross section, since a small relative deficit in the prompt yields can imply a large relative increase in the feed-down yields because the feed-down contribution is the smallest fraction of total D^+-mesons.