5

Impact-parameter fit method for the prompt D^+ fraction determination

In this Chapter I will present the main data-driven method that I developed for the evaluation of D^+ prompt fraction. The main advantage of this technique is that it does not rely on any theoretical assumption, unlike the FONLL-based methods presented in section 4.7. It also aims to reduce the large systematic uncertainty of the theory-driven approach, especially at low transverse momentum. This first data-driven method exploits the different shapes of the impact-parameter

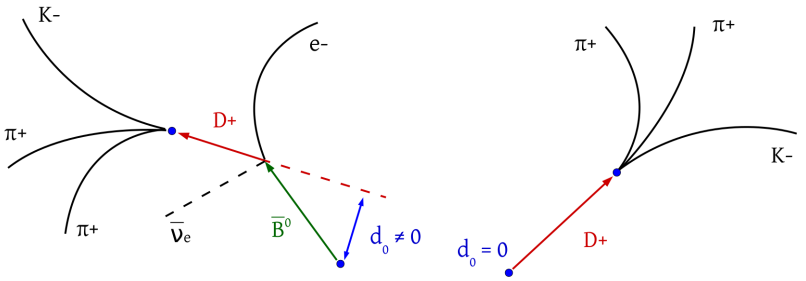

Figure 5.1. Sketch of the decay topology for prompt (right) and feed-down (left) D^+ with particular emphasis on the true impact parameter.

distributions for prompt and feed-down D^+. The impact parameter is defined as the distance of closest approach between the prolonged momentum direction of the D^+-meson and the primary vertex. Since the prompt D^+-mesons come from the primary vertex, their real impact parameter is by definition exactly zero. Therefore, we can expect the measured distribution to be centred at zero and smeared for resolution effects. On the other hand, the impact parameter for feed-down D^+ is in general different from zero and it depends on the decay length of the B-hadron and its relative direction with respect to the D^+-meson. In figure 5.1 is illustrated an example of decay and the corrispective true impact parameter for both prompt and feed-down D^+-mesons. Beside that, there is an alternative data-driven method, that will be presented in Chapter 6.

5.1 Analysis strategy

Before to start with the description of the analysis, I would like to introduce the strategy adopted. First and foremost, all the selections applied (event, track quality, topological cuts and PID) are the same as those used for the standard analysis (see Chapter 4). This is very important in order to be able to compare the two results, since the prompt fraction depends on the topological cuts applied.
Then, the procedure is divided in two main steps:

- A prefit phase, which consists in the parametrisation of the impact-parameter distributions in the transverse plane (d_0^{xy}) for the three different types of candidates that contribute to the measured distribution: prompt D^+, feed-down D^+ and background. For the prompt and feed-down D^+ the distributions are obtained from the Monte Carlo sample described in section 4.3. The distribution for the combinatorial background is directly taken from the data sample (see section 4.1.1), in the invariant-mass region aside the D^+ peak (see section 5.2.3). Each contribution is fitted with a specific function, which is called *template*.

- An unbinned log-likelihood fit to the impact-parameter distributions in the data using the templates obtained in the previous step.

All the parameters obtained in the prefit phase are fixed in the final fit on data, except for the detector-resolution term which is kept free in order to compensate a possible not perfect description of the impact-parameter resolution in the MC simulation (a detailed description will be provide in section 5.2.1).
This method requires a certain amount of statistics in order to obtain a stable fit result, therefore the transverse momentum range considered for this analysis is narrower ($2 < p_{\mathrm{T}} < 16$ GeV/c instead of $1 < p_{\mathrm{T}} < 24$ GeV/c) with respect to the one used for the standard analysis obtained with the theory-driven approach, presented in section 4.9.

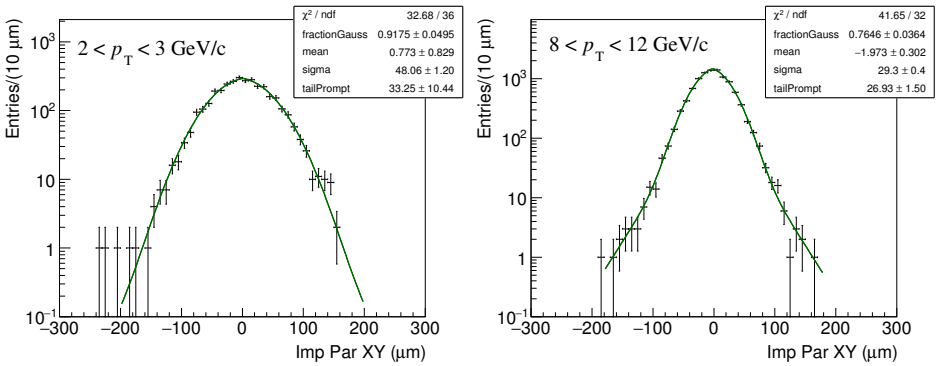

Figure 5.2. Examples of fit to the simulated impact-parameter distributions for prompt D^+-mesons in two different p_T intervals.

5.2 Prefit on the impact-parameter distributions

In this section I will present the prefits performed to the impact-parameter distributions in the transverse plane for prompt and feed-down D^+-mesons and for the combinatorial background.

5.2.1 MC impact-parameter distributions for prompt D^+

The impact-parameter distributions in the transvere plane for prompt D^+ were obtained from the Monte Carlo simulation described in section 4.3, in seven p_T intervals in the range $2 < p_T < 16$ GeV/c. Since the generated impact parameter for prompt D^+-mesons is by definition zero, the distributions from the reconstruction of the MC data (*simulated* distributions) are approximately Gaussian due to the finite track and primary vertex resolution of the detector. However, as shown in figure 5.2, they clearly present non-Gaussian tails, which are more prominent at higher D^+ transverse momentum. For this reason, a simple Gaussian function would not give a proper description of the distribution and therefore the fit function used is the sum of a Gaussian and a symmetric exponential function

$$F^{prompt}(d_0^{xy}) = F^{g+exp}(d_0^{xy}) = \\ = N \cdot \left\{ \frac{f_{gaus}^{prompt}}{\sqrt{2\pi}\sigma_{prompt}} e^{-\frac{(d_0^{xy}-\mu_{prompt})^2}{2\sigma_{prompt}^2}} + \frac{1-f_{gaus}^{prompt}}{2\lambda_{prompt}} e^{-\frac{|d_0^{xy}-\mu_{prompt}|}{\lambda_{prompt}}} \right\}, \tag{5.1}$$

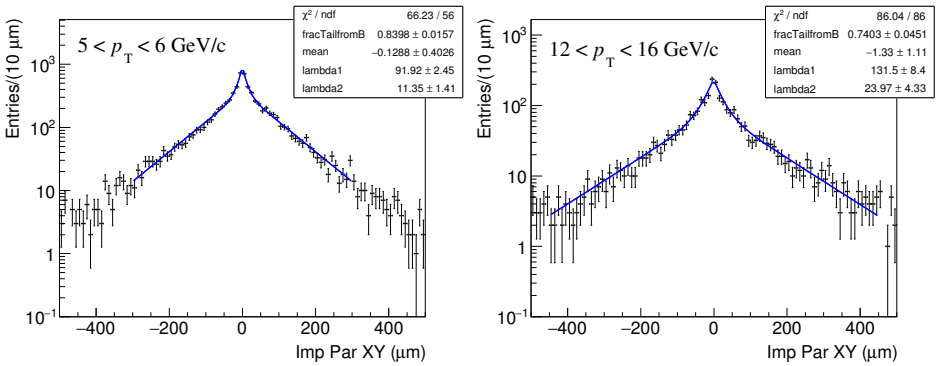

Figure 5.3. Examples of fit to the true generated impact-parameter distributions for feed-down D^+-mesons in two different p_T intervals.

where N is a normalisation factor which is fixed to the integral of the histogram, f_{gaus}^{prompt} represents the fraction of the integral contained in the Gaussian function, μ_{prompt} is the common mean value for the two contributions, λ_{prompt} is the slope of the exponential function and σ_{prompt} is the width of the Gaussian function, which represents the resolution of the detector. As shown in the two examples in figure 5.2, the χ^2 values show that the chosen functional form gives a reasonably good description of the distributions. Furthermore, it is possible to notice how the Gaussian width decreases with increasing D^+ transverse momentum, due to the better impact-parameter resolution on high p_T tracks (right panel of figure 3.6). Also the f_{gaus}^{prompt} parameter decreases with increasing p_T, because of the higher contribution in the exponential. All the other fits to the impact-parameter distributions for prompt D^+ in the p_T intervals considered for the analysis can be found in the appendix A.1.

5.2.2 MC impact-parameter distributions for feed-down D^+

The true impact parameter for the feed-down D^+-mesons is different from zero because in general the momentum of the D^+ is not parallel to the one of the B-hadron, as shown in the left panel of figure 5.1. In particular it depends on the decay length of the B-hadron and the angle between the directions of the two particles. For this reason the impact-parameter distribution of feed-down D^+ is more difficult to parametrise with respect to the one of prompt D^+. In order to achieve a satisfactory description of the distribution, two steps were performed. First, the true impact-parameter distribution in the xy-plane for feed-down D^+ from the Monte Carlo simulation in the seven p_T intervals of the analysis, applying

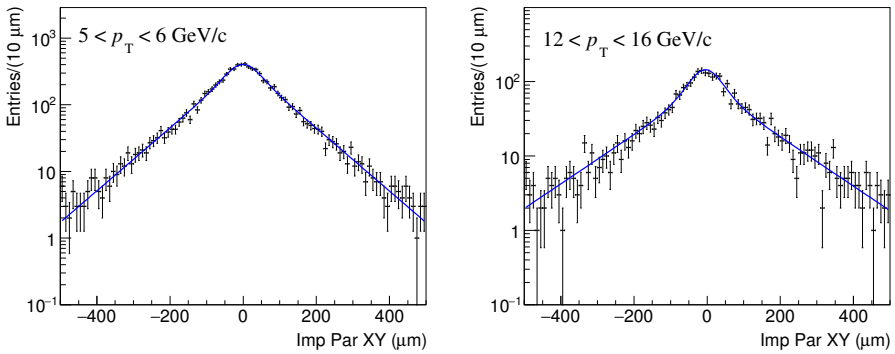

Figure 5.4. Examples of reconstructed impact-parameter distributions for feed–down D^+-mesons in two different p_T intervals. The function superimposed is the convolution between the true impact parameter function for feed-down D^+ (eq. 5.2) and the resolution term (eq. 5.1).

all the selections described in Chapter 4, was fitted with a double exponential function, symmetrised with respect to its mean value

$$F_{true}^{feed-down}(d_0^{xy}) = N \cdot \left\{ \frac{f_{\lambda_1}^{FD}}{2\lambda_1^{FD}} e^{-\frac{|d_0^{xy}-\mu_{FD}|}{\lambda_1^{FD}}} + \frac{1-f_{\lambda_1}^{FD}}{2\lambda_2^{FD}} e^{-\frac{|d_0^{xy}-\mu_{FD}|}{\lambda_2^{FD}}} \right\}, \qquad (5.2)$$

where N is again fixed to the integral of the histogram, λ_1^{FD} and λ_2^{FD} are the slopes of the two symmetrised exponential functions, $f_{\lambda_1}^{FD}$ is the fraction of integral contained in the first function and μ_{FD} is the common mean value. The impact-parameter range in the fit increases with the transverse momentum, starting from $[-250,250]$ μm in the interval $2 < p_T < 3$ GeV/c up to $[-450,450]$ μm in $12 < p_T < 16$ GeV/c, in order to get a good description of the distribution in the steepest region. Two examples of fit to the true impact-parameter distribution for feed-down D^+ are presented in figure 5.3. The reasonable agreement between the functional form chosen and the distributions is confirmed by the χ^2 values obtained. The second step for the parametrisation consisted in a convolution of the function obtained for the feed-down D^+ true impact-parameter distribution $F_{true}^{feed-down}(d_0^{xy})$ and the $F_{prompt}(d_0^{xy})$ function, which represents the effect of the detector resolution, for each p_T interval

$$F_{reco}^{feed-down}(d_0^{xy}) = N \cdot \int_{d_{0,min}^{xy}}^{d_{0,max}^{xy}} F_{true}^{feed-down}(d_0^{xy\prime}) F^{prompt}(d_0^{xy} - d_0^{xy\prime}) dd_0^{xy\prime}, \quad (5.3)$$

where the limits in the convolution were set in order to cover the impact-parameter range between [−1000,1000] μm. In figure 5.4 the simulated impact-parameter distributions in the same p_{T} intervals are shown, with the function obtained with the convolution superimposed. All the other plots can be found in the appendix A.2.

5.2.3 Impact-parameter distributions for the background

Unlike the impact-parameter distributions for prompt and feed-down D^+, which were modelled from the MC simulation, the distributions for the combinatorial background were obtained directly from the Minimum Bias p-Pb data sample described in section 4.1.1, selecting two mass intervals on the two sides of the D^+ peak, where there is no signal. The invariant-mass intervals considered, called *side-band regions*, are defined by:

$$4\sigma < |M - M_{peak}| < 15\sigma, \tag{5.4}$$

where M_{peak} and σ are respectively the mean value and the width of the Gaussian function used to describe the signal in the invariant-mass fit. The lower limit of 4σ allows to remove most of the signal (99.99%) from the impact-parameter distribution and at the same time to select a mass region close to the peak, in order to get a reasonable description of the background under the D^+ peak. The mass interval is symmetric with respect to the mass of the D^+ to keep more statistics for the parametrisation and to average on a possible invariant-mass dependence of the

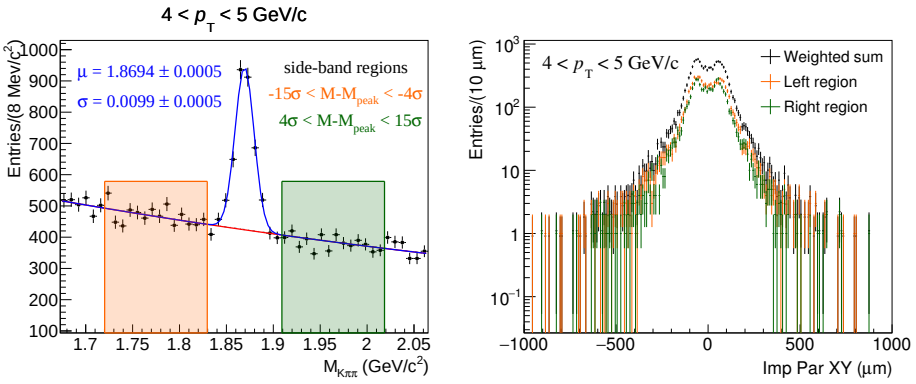

Figure 5.5. Left: invariant-mass intervals considered for the parametrisation of the impact-parameter distribution for the background (*side-band regions*) in the transverse momentum interval $4 < p_{\mathrm{T}} < 5$ GeV/c. Right: resulting background impact-parameter distribution.

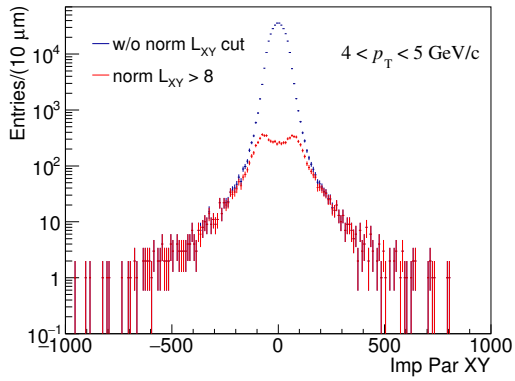

Figure 5.6. Comparison between the impact-parameter distributions for the combinatorial background obtained from the side-bands region, with (red) and without (blue) cut on normalised decay length xy. In both cases all the other topological selections reported in table 4.2 are applied.

impact-parameter distribution. In figure 5.5 the side-band regions selected (left) and the resulting background impact-parameter distributions in the p_{T} interval $4 < p_{\mathrm{T}} < 5$ GeV/c (right) are shown. The distributions obtained from the two side-band regions are combined with a weighted sum, according to their integrals. The double-peak structure is induced by the topological selections applied, especially by the cut on the normalised decay length in the xy-plane, as shown in figure 5.6. The red distribution is obtained from the side-bands with the normalised decay length cut, while the blue one without. Both the distributions are obtained applying all the other topological selections summarised in table 4.2. Because of this double-peak shape, a sum of two Gaussian functions with exponential tails is used for the fit

$$F^{bkg}(d_0^{xy}) = N \cdot \Big\{ F_1^{gaus+exp}(d_0^{xy};\ f_{gaus}^{bkg}, \mu_1^{bkg}, \sigma_{bkg}, \lambda_{bkg}) + \\ + F_2^{gaus+exp}(d_0^{xy};\ f_{gaus}^{bkg}, \mu_2^{bkg}, \sigma_{bkg}, \lambda_{bkg}) \Big\}. \tag{5.5}$$

In order to avoid as much as possible statistical fluctuations, the two $F^{gaus+exp}$ functions were forced to have all the parameters equal, except for the two mean values, μ_1 and μ_2. The common parameters are hence the following: f_{gaus}^{bkg} is the fraction of integral contained in each Gaussian function, σ_{bkg} is the width of the Gaussians and λ_{bkg} the slope of the exponential constributions. The χ^2 values

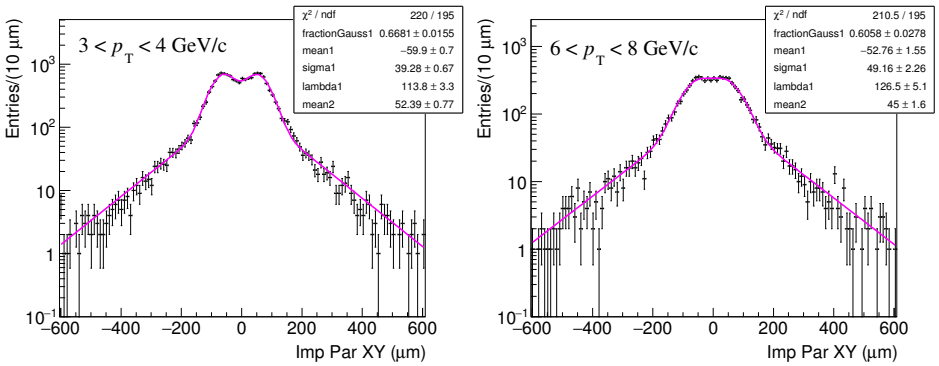

Figure 5.7. Examples of fit to the impact-parameter distributions for the combinatorial background obtained from the side-band regions, in two different p_T intervals.

obtained confirm that this functional form gives an acceptable description of the distributions. From the two fit examples shown in figure 5.7 is it possible to notice how the double-peak sructure becomes less pronounced with increasing p_T. All the other fits to the impact-parameter distributions in the transverse plane for the background are reported in the appendix A.4.

5.3 Unbinned log-likelihood fits on data

In this section I will describe the procedure used to fit the impact-parameter distributions in the transverse plane for D^+ candidates in the Minimum Bias data sample for p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, to extract the prompt fraction. First a brief motivation for using the log-likelihood unbinned fit technique is provided, then the fitting procedure is described.

5.3.1 Binned vs. unbinned fit

Given a set of experimental observables $\vec{x} = (x_1, ..., x_n)$ described by the probability density function $f(\vec{x}; \vec{\theta})$, which depends on the set of parameters $\vec{\theta} = (\theta_1, ..., \theta_m)$, the likelihood function is equal to the probability function evaluated with fixed $\vec{x}$, depending on $\vec{\theta}$

$$\mathcal{L}(\vec{\theta}; \vec{x}) = f(\vec{x}; \vec{\theta}). \tag{5.6}$$

Both the *binned* and *unbinned* fit techniques are performed finding the set of parameters $\vec{\theta}$ which maximise the likelihood function, or more conveniently the logarithm

of $\mathcal{L}$. The difference between the two fit methods is in the treatment of the data, which leads a different definition of the likelihood function itself.
Let assume x to be a single variable and $\boldsymbol{x} = (x_1, ..., x_N)$ a sample of N independent measurements of x. In case of unbinned fit the log-likelihood is defined as

$$\ln\mathcal{L}(\vec{\theta}; \boldsymbol{x}) = \sum_{i=1}^{N} \ln f(x_i; \vec{\theta}), \tag{5.7}$$

where $f(x_i, \vec{\theta})$ is the probability density function evaluated in the point x_i and depending on the parameters $\vec{\theta}$. On the other hand, the binned fit is performed by first filling an histogram with the measurements $\boldsymbol{x}$, and therefore the log-likelihood is modified as follow

$$\ln\mathcal{L}(\vec{\theta}) = \sum_{i=1}^{N_{bins}} n_i \ln\nu_i(\vec{\theta}) \tag{5.8}$$

where N_{bins} is the number of bins of the histogram and n_i is the number of entries in the i-th bin. In this case the expected probability for the value x_i does not enter directly in the formula, which instead contains the expected number of entries in the i-th bin, given by

$$\nu_i(\vec{\theta}) = n \int_{x_i^{min}}^{x_i^{max}} f(x, \vec{\theta}) dx. \tag{5.9}$$

Therefore the binned fit becomes convenient in terms of computing time when the sample of measurement $\boldsymbol{x}$ is large, while it results less precise in case of low statistics, because of the few entries in the bins of the histogram.

The unbinned technique was chosen for the fits on data because the impact-parameter distributions in the transverse plane show a small number of entries in the tails, where the shape for prompt and feed-down D^+-mesons are more different playing a crucial role for the proper separation of the two contributions.

5.3.2 Impact-parameter fits

The impact-parameter distributions for the final fit were obtained from the data, applying all the selections described in Chapter 4, for the seven p_{T} intervals of the analysis. Furthermore, the invariant-mass range considered for the D^+ candidates is

$$|M - M_{peak}| < 2\sigma \tag{5.10}$$

in order to keep as much high as possibile the signal-to-background ratio, while keeping the majority of the signal (95.5%). In figure 5.8 the selected invariant-mass range for the $4 < p_{\mathrm{T}} < 5$ GeV/c interval (left) and the corresponding impact-parameter distribution obtained (right) are shown. As already mentioned the fit

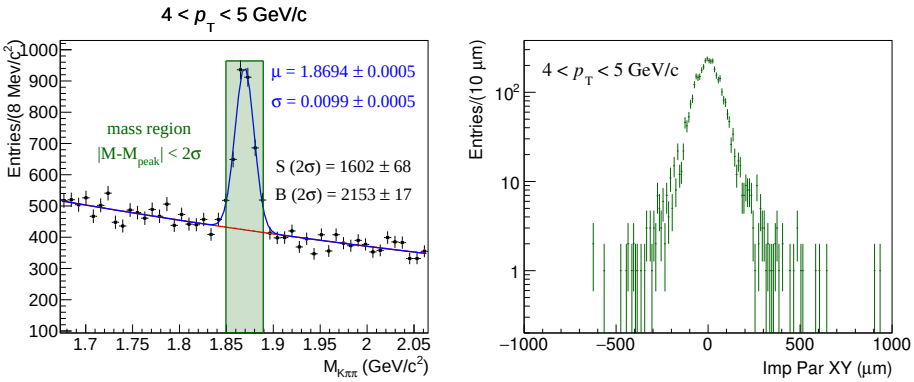

Figure 5.8. Left: invariant-mass range considered for the impact-parameter distribution final fit in the transverse momentum interval $4 < p_{\mathrm{T}} < 5$ GeV/c. Right: resulting impact-parameter distribution.

was not performed to the histogram, but with the unbinned technique, and therefore to a ROOT structure called *tree*, which allows to store the data without binning them. The fitting function used is

$$F(d_0^{xy}) = N \cdot \left\{ \frac{S}{N} \left[f_{prompt} F^{prompt}(d_0^{xy};\ \sigma_{prompt}) + \right.\right. \\ \left.\left. + (1 - f_{prompt}) F_{reco}^{feed-down}(d_0^{xy};\ \sigma_{prompt}) \right] + \frac{N - S}{N} F^{bkg}(d_0^{xy}) \right\}, \tag{5.11}$$

where the normalisation factor N is fixed to the number of entries in the tree and S is the parameter which represent the amount of signal, fixed to the value extracted with the fit to the invariant-mass distribution, integrating the signal fit function over the selected invariant-mass range (see figure 5.8). The values S obtained with this procedure are reported in table 5.1. $F^{prompt}(d_0^{xy})$, $F_{reco}^{feed-down}(d_0^{xy})$ and $F^{bkg}(d_0^{xy})$ are the functions for prompt D^+-mesons, feed-down D^+-mesons and the combinatorial background respectively, obtained in the prefit phase (see section 5.2). All the parameters for each contribution are fixed to the values obtained from the prefits, except for the Gaussian width of the detector-resolution term σ_{prompt}, which is kept free to vary within 20% around the value obtained by fitting the MC distribution for prompt D^+ to compensate a possible not perfect description of the impact-parameter resolution in the Monte Carlo simulation. None of the values obtained from the fit reaches the 20% limit. In section 5.4.1 the values obtained from the data and the MC simulation are compared in order to evaluate a possible systematic uncertainty. The only other free parameter in the fit function is the fraction of prompt D^+-mesons f_{prompt}, which is the quantity that we want

p_{T} (GeV/c)	$S(2\sigma)$	$\sigma_S/S(2\sigma)$ stat.	σ_S/S syst.
[2,3]	975 ± 60	0.06	0.08
[3,4]	1734 ± 72	0.04	0.05
[4,5]	1602 ± 68	0.04	0.05
[5,6]	1122 ± 52	0.05	0.05
[6,8]	1297 ± 55	0.04	0.05
[8,12]	985 ± 53	0.05	0.05
[12,16]	263 ± 27	0.10	0.08

Table 5.1. Raw yields in the invariant-mass range $|M - M_{peak}| < 2\sigma$ with the statistical error in the seven p_{T} intervals of the analysis. The systematic uncertainties on the yield extraction are taken from the values estimated for the standard analysis [59].

to measure. The impact-parameter fit range gradually increases with increasing transverse momentum to better account for the broadening of the distributions. Figure 5.9 shows the resulting fits in the transverse momentum intervals of the analysis. In this figure the histogram and the curves, rescaled according to the integral of the histogram, are depicted in order to provide a visual representation of the unbinned fit.

5.4 Systematic uncertainties

In this section the various sources of systematic uncertainty that affect the fraction of prompt D^+ extracted from the fit of the impact-parameter distributions in the xy-plane are presented. They include:

1. Systematic uncertainty due to the shapes assumed in the MC simulation:
   a. Assumed shapes for prompt D^+, feed-down D^+ and background impact-parameter distributions.
   b. Assumed p_{T} shape of the generated prompt D^+ and B mesons decaying to D^+.
2. Systematic uncertainty induced by the uncertainty on the signal parameter S.
3. Systematic uncertainty evaluated by checking the consistency of the procedure.

Several tests have been performed in order to quantify these sources of systematics. Concerning the first category, tests on the degree of agreement of the resolution

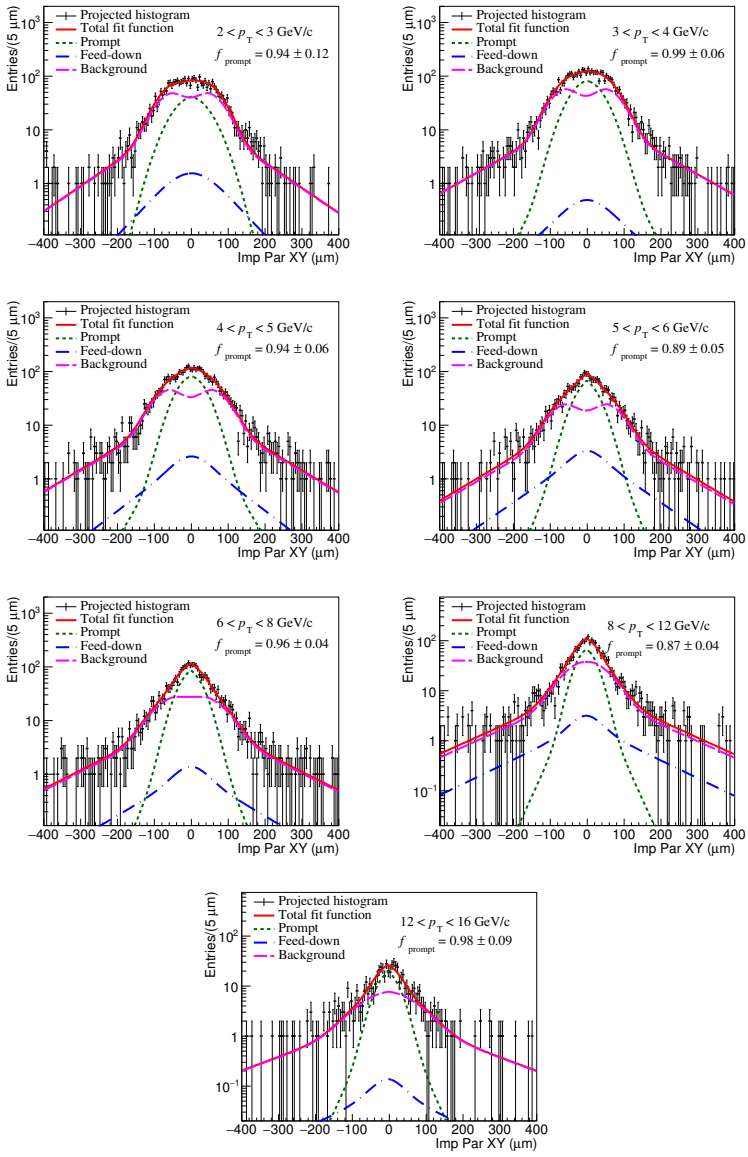

Figure 5.9. Fits to the D^+ impact-parameter distributions in the xy-plane in the seven transverse momentum intervals of the analysis. The curves are the fit functions for prompt D^+ (green), feed-down D^+ (blue), background (magenta) and the sum of all contributions (red).

term in data and MC, the impact-parameter fitting range and the parametrisation of the background distribution from the side-band regions have been performed. To take into account the discrepancy between the p_{T}-shape in data and MC, the impact-parameter distributions of prompt and feed-down D^+ used in the prefit phase have been reweighted in order to match the FONLL p_{T}-shape. The error induced by the signal yield was quantified by changing the S parameter within 1σ in the fitting function. Finally, to check the consistency of the procedure the invariant-mass window used to obtain the impact-parameter distributions was changed and a MC closure test was performed. All these tests will be described in more details in the following subsections.

5.4.1 Systematic uncertainty due to the assumed impact-parameter shapes

Test on the degree of agreement between the resolution term in data and MC

In order to test the degree of agreement between the detector resolution in data and MC, all the fits have been repeated fixing σ_{prompt} to the values obtained from

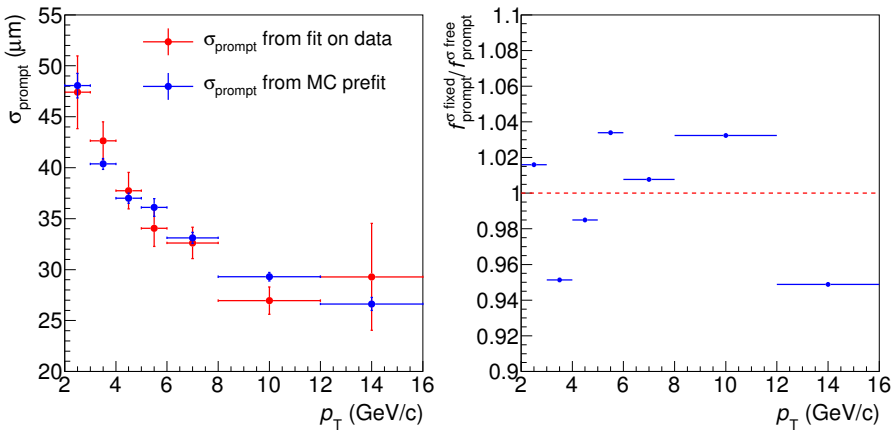

Figure 5.10. Left: comparison between the σ_{prompt} parameter as a function of p_{T} obtained from the impact-parameter fit on data and the one obtained by fitting the MC distributions for prompt D^+ in the prefit phase. Right: ratio between the D^+ prompt fraction f_{prompt} obtained by fixing and letting free σ_{prompt} in the fit on data as a function of p_{T}.

the MC impact-parameter distributions for prompt D^+ in the prefit phase. To give a quantitative evaluation of the uncertainty, the ratio of f_{prompt} obtained with and without the σ_{prompt} parameter fixed was computed. As shown in the right panel of figure 5.10 this ratio flactuates around unity without any particular trend, therefore no systematic uncertainty due to the detector-resolution term is assigned. This observation is also confirmed by the comparison between the values of σ_{prompt} obtained by fitting the data and in the prefit phase on the simulated sample, which shows that all the values agree within the statistical uncertainty.

Variation of the fitting range

Another test performed to quantify the systematic uncertainty due to the assumed shape of the impact-parameter distributions consisted in varying the impact-parameter fitting range. The prefit and the fit phases were repeated varying the impact-parameter ranges from [−200,200] μm, up to [−1000,1000] μm coherently. Then, in order to disentangle the effect of the variation in the bakground distribution from the the others, the same procedure was repeated fixing the fitting range of the bakground in the prefit phase to the value of [−1000,1000] μm. Figure 5.11 shows the ratio of the prompt fraction obatained for each trial with respect to the

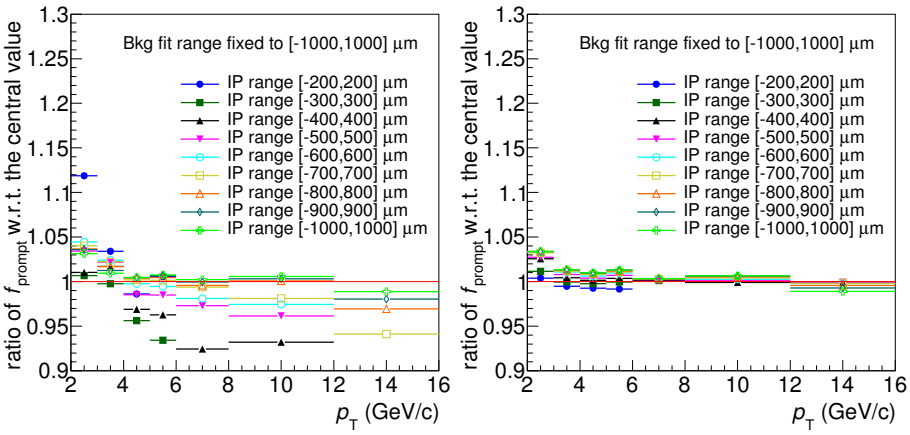

Figure 5.11. Ratio of f_{prompt} as a function of p_T obtained varying the fit range with respect to the central value. In the left plot all the fit ranges both in the fit and the prefit phase are changed coherently, while in the right one the impact-parameter range for the background is fixed to [1000,1000] μm.

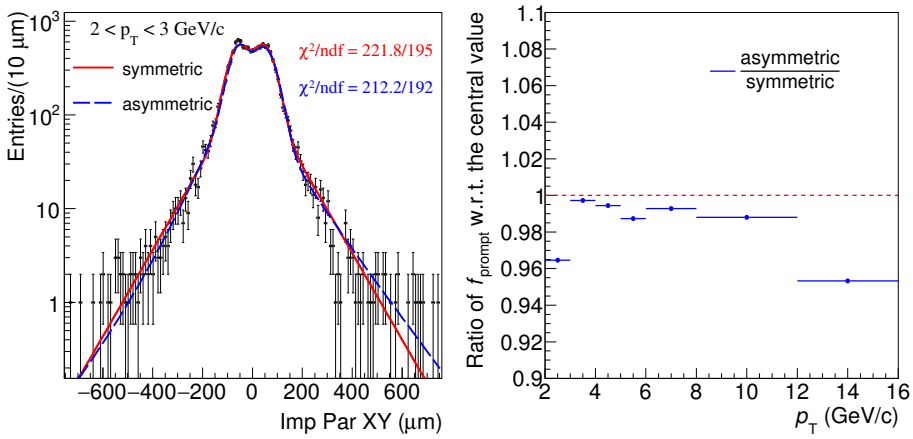

Figure 5.12. Left: example of fit to the background distribution in the p_{T} interval $2 < p_{\mathrm{T}} < 3$ GeV/c with the symmetric and asymmetric functions. Right: the ratio of f_{prompt} obtained with the symmetric and the asymmetric fit function for the background.

central value, in the two cases. From these plots it is clear that there is a systematic trend and that the larger source of uncertainty comes from the variation of the fit range of the background distribution in the prefit phase. It should be also pointed out that the narrower impact-parameter ranges tested do not allow to describe properly the background distribution, which is much broader, especially at high p_{T}. For this reason, the first two fitting ranges were not considered for the transverse momentum interval $6 < p_{\mathrm{T}} < 12$ GeV/c, and the first three for $12 < p_{\mathrm{T}} < 16$ GeV/c. The systematic uncertainties were assigned by evaluating the RMS of the trials shown in the left panel of figure 5.11 for each p_{T} bin, which results to be of about 2% in the full transverse momentum range.

Uncertainty on the determination of the background distribution

For the evaluation of the systematic uncertainty due to the parametrisation of the combinatorial bakground distributions, two different sources have been taken into account:

- Systematic uncertainty due to the fit function chosen.
- Systematic uncertainty due to the invariant-mass range in the side-bands region taken into account to obtain the impact-parameter distribution for the background.

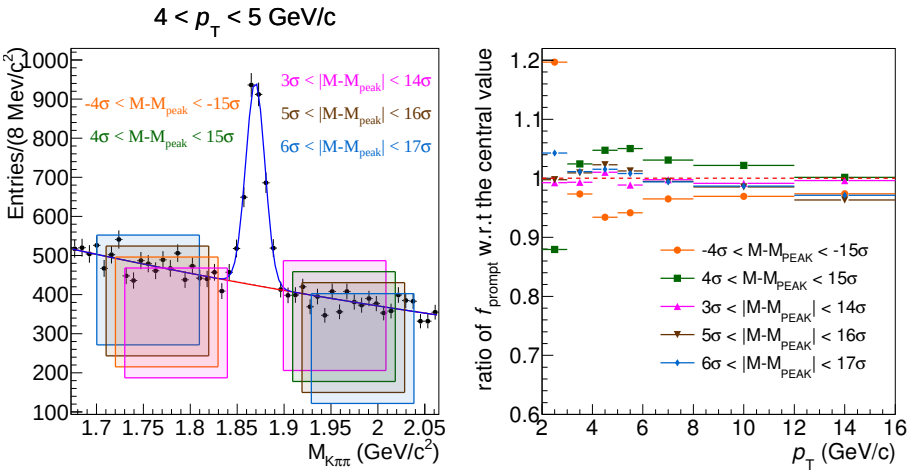

Figure 5.13. Left: the five invariant-mass intervals in the side-bands region tested to evaluate the systematic uncertainty for the combinatorial background parametrisation in the $p_{\rm T}$ interval $4 < p_{\rm T} < 5$ GeV/c. Right: The ratio of f_{prompt} with respect to the reference value obtained with the mass region $4\sigma < |M - M_{peak}| < 15\sigma$ for the background parametrisation as a function of $p_{\rm T}$.

In order to take into account the possibility that the background distribution is not symmetric with respect to its central value, not only for statistical flactuations, the prefit phase and the final fits have been repeated with a slightly different and asymmetric fit function for the background contribution. This alternative function is indeed the same used as default (see equation 5.5), but with all the parameters of the two Gaussian functions with exponential tails free and independent. The left plot of figure 5.12 shows the comparison between the two different fit functions obtained for the $p_{\rm T}$ interval $2 < p_{\rm T} < 3$ GeV/c. The two curves result to be similar and both give an acceptable value of the χ^2. In the right panel of the same plot is shown the relative variation between the prompt fraction of $\rm D^+$-mesons obtained in the two cases, which was used to assign a 4% uncertainty for $2 < p_{\rm T} < 3$ GeV/c and 1% in the remaining $p_{\rm T}$ interval. In particular in the $p_{\rm T}$ range $12 < p_{\rm T} < 16$ GeV/c the uncertainty was reduced because of the bad description of the distribution obtained with the asymmetric function.

As mentioned in section 5.2.3 the impact-parameter distribution for the combinatorial background was obtained from the $\rm D^+$ candidates in the side-bands region. However the choice of the invariant-mass range is arbitrary and therefore can be

a source of systematic uncertainty. Concerning that, the fitting procedure was repeated changing the limits of the side-band regions for the parametrisation of the background. In particular two tests have been performed:

- Selecting the same invariant-mass window used for the central value, but only the left or the right side from the D^+ peak.

- Selecting different invariant-mass ranges symmetric with respect to the D^+ peak.

In the right panel of figure 5.13 are illustrated all the different ranges considered in the side-bands region, while in the right plot is shown the ratio of f_{prompt} with respect to the central value, as a function of the transverse momentum. The larger variation is observed for the two trials performed selecting only the left or the right invariant-mass interval. However these two tests are extreme, since the choice of a symmetric range was done to averege on a possible mass dependence of the impact-parameter and therefore to get a better estimation of the background distribution under the D^+ peak. For this reason the assigned systematic uncertainty is 6% in the first p_T interval and 3% for $3 < p_T < 16$ GeV/c, instead of the full spread.

5.4.2 Systematic uncertainty due to the assumed p_T shape

Since the shape of the impact-parameter distributions in the xy-plane depends on the transverse momentum, and the distributions for prompt and feed-down are modelled starting from the MC distributions in p_T intervals of finite size, it is important to have a good description in the Monte Carlo simulation of the p_T spectrum of D^+-mesons and B-mesons which decay to D^+. However, as shown in figure 5.14, the shape of the generated p_T spectrum in the MC is slightly different with respect to the FONLL one, which describes reasonably well the measurements (see section 2.1). This discrepancy could generate a bias, especially in the transverse momentum intervals where the p_T shape in FONLL and MC is more different and in the larger p_T bins of the analysis. Concerning that, the MC impact-parameter distributions for prompt (feed-down) D^+ have been re-weighted, according to the ratio between the p_T spectrum of D^+-mesons (B-mesons) predicted by FONLL and in the MC simulation. Figure 5.15 shows that the difference between the impact-parameter distributions obtained with and without applying the weights, for prompt and feed-down D^+, is very small. As a consequence, the variation of f_{prompt} obtained with the re-weighted impact-parameter distributions results to be of the order of $\sim 0.1\%$, which is completely negligible with respect to the other sources of systematic uncertainty.

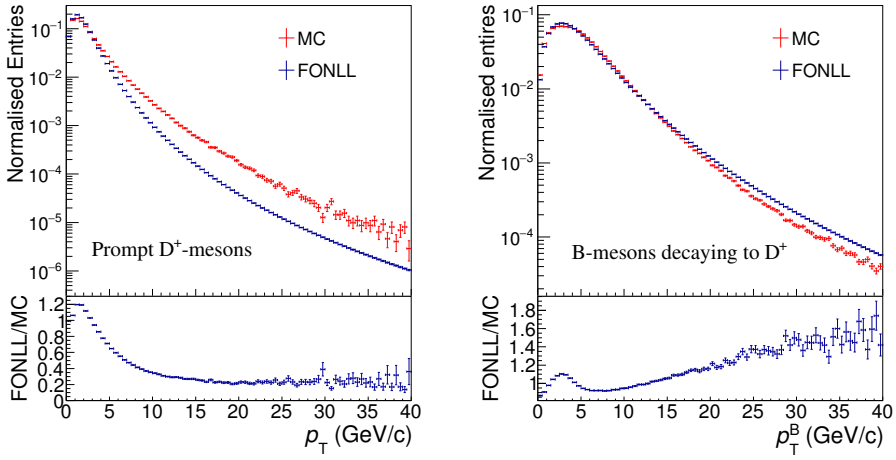

Figure 5.14. p_{T}-spectrum of generated D^+-mesons (left) and B-mesons decaying to D^+ (right) from Monte Carlo simulation compared to the FONLL p_{T}-spectrum.

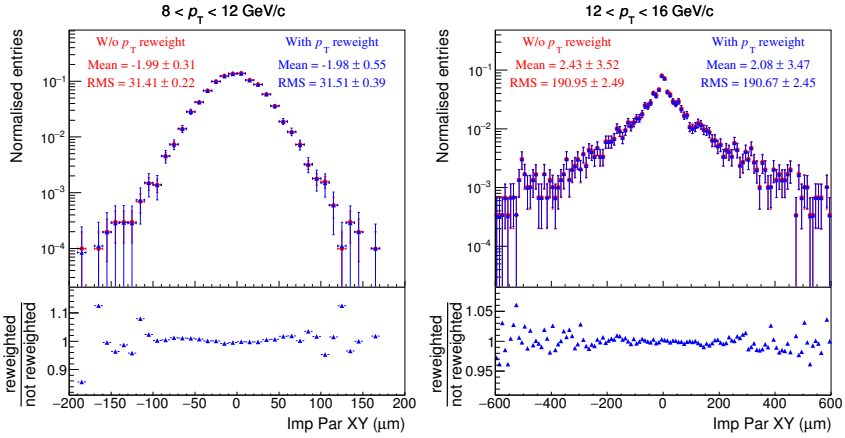

Figure 5.15. Top: comparison between the prompt and true feed-down D^+ impact-parameter distributions with and without applying the weight factor according to the FONLL p_{T} spectrum for prompt D^+-mesons and B-mesons which decay to D^+ respectively, in two p_{T} intervals of the analsysis. Bottom: ratio between the reweighted and not reweighted impact-parameter distributions.

5.4.3 Systematic uncertainty induced by the uncertainty on the signal parameter S

In the impact-parameter fit on data, the amount of signal S, which is a parameter of the fit function 5.11, is fixed to the value of the extracted raw yield within 2σ. However this value is affected by a statistic and systematic error, which has to be taken into account in the evaluation of the D^+ prompt fraction. Therefore, the fits were repeated fixing the S parameter to the value of the raw yield plus or minus its error, which is reported in table 5.1. The systematic uncertainty for the raw yield extraction is taken from the values of the standard analysis published by ALICE [36], which were evaluated in [59], while the total uncertainty is computed considering the statistical and systematic errors as uncorrelated

$$\sigma_S = \sqrt{\sigma_S^2(stat) + \sigma_S^2(syst)}. \tag{5.12}$$

As shown in the left panel of figure 5.16, to an increase of the signal corresponds a decrease of the D^+ prompt fraction. Since f_{prompt} is computed for the calculation of the p_T-differential cross section, which is proportional to both the prompt fraction and the raw yield (see equation 4.8), we can expect a partial compensation of this uncertainty in the cross section evaluation. This is the case, as shown in the right panel of figure 5.16, where is reported the relative variation of the p_T-differential

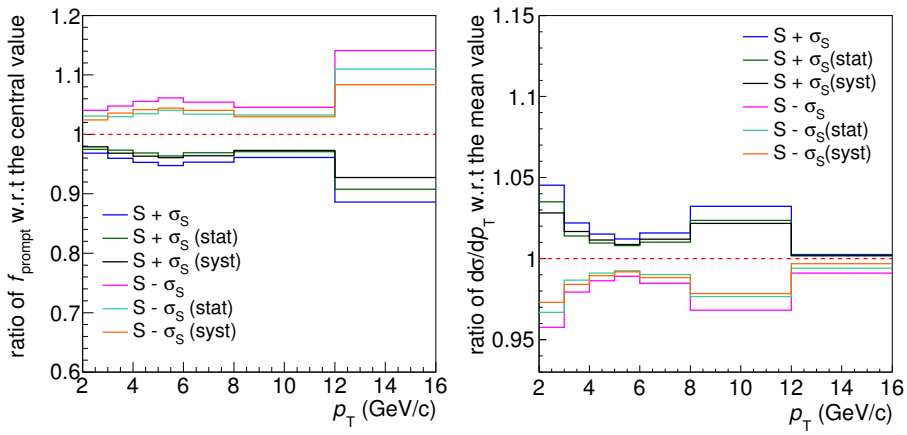

Figure 5.16. Left: ratio of f_{prompt} obtained fixing the signal parameter S to its value plus or minus its statistical and systematic error, with respect to the central value. Right: relative variation of the p_T-differential cross section due to the variation of the signal within its statistical and systematic error.

cross section calculated using the formula 4.8, where the f_{prompt} is evaluated with the impact-parameter method, and setting the raw yields and the parameter S in the impact-parameter fits to $S \pm \sigma_S$. Not to double count the error the signal extraction for the computation of the cross section, the uncertainty is assigned on the basis of the $d\sigma/dpt$ variation, and therefore 4% in the transverse momentum interval $2 < p_{\rm T} < 3$ GeV/c and 2% in the remaining $p_{\rm T}$ range.

5.4.4 Consistency of the fitting procedure

Variation of the invariant-mass window

The stability of the result obtained selecting an invariant-mass region, from which the candidate impact-parameter distribution is obtained and the amount of signal S is extracted from the fit to the invariant-mass distributions (see figure 5.8), was evaluated changing the number of σ of the $\rm D^+$ peak. In particular the impact-parameter fits have been repeated on the distributions obtained with four different invariant-mass windows, from a minimum of 1σ to a maximum of 3σ, as shown in the left plot of figure 5.17 for the $p_{\rm T}$ interval $4 < p_{\rm T} < 5$ GeV/c. The estimation of the systematic uncertainty was then performed computing the ratio between each value of f_{prompt} and the reference value obtained using the mass window $|M - M_{peak}| < 2\sigma$. The right plot of figure 5.17 shows that this ratio flactuates around unity without any clear trend, therefore no systematic uncertainty has been assigned based on this test.

Monte Carlo closure test

In this section I present the fast Monte Carlo closure test, which was performed to verify the consistency and stability of the fitting procedure. The aim of this study is to check if the fits can reproduce well the input values of f_{prompt}, to be sure not to have biases in the measurements. The closure is then divided in two steps:

- Simulation of the entries in the tree in the same $p_{\rm T}$ intervals of the analysis, starting from the MC impact-parameter distributions for prompt and feeed-down $\rm D^+$ and from the distributions obtained from the sideband region for the background, which are the same used in the prefit phase.
- Unbinned fit of the simulated impact-parameter distributions, applying the procedure described in section 5.3.

The number of entries for each contribution is given by three input parameters (N, S and f^{gen}_{prompt}), as follow:

- $N_{bkg} = N - S$, the number of entries extracted from the background distribution.

- $N_{prompt} = f^{gen}_{prompt} \times S$, the number of entries extracted from the prompt D^+ distribution.
- $N_{feed-down} = (1 - f^{gen}_{prompt}) \times S$, the number of entries extracted from the feed-down D^+ distribution.

The N and S parameters were fixed to the values in the data, in order to simulate the same statistics that is available for the analysis, while for the input prompt fraction four different values have been tested: $f^{gen}_{prompt} = 0.85$, 0.90, 0.95 and finally f^{gen}_{prompt} was fixed to the values obtained with the theory-driven N_b method (see section 4.2.1). The resulting distributions were then fitted following the same procedure and using the same fit function (see equation 5.11) with f_{prompt} and the detector-resolution term σ_{prompt} as free parameters. The simulation and the fitting procedure were repeated 50 times. In figure 5.18 are shown the mean value (left) and the RMS (right) of the residuals distributions for f_{prompt} as a function of the transverse momentum, for all the four different input values of the prompt fraction. The mean value is compatible with 0 within the statistical uncertainty in the almost the full $p_{\rm T}$ range, which means that the measured value of f_{prompt} reproduce reasonably well the value set as input. However,it seems to be slightly stematically higher, therefore a 1% systematics is added to the measurement, as reported in table 5.2. In order to test whether the error on the single measurement

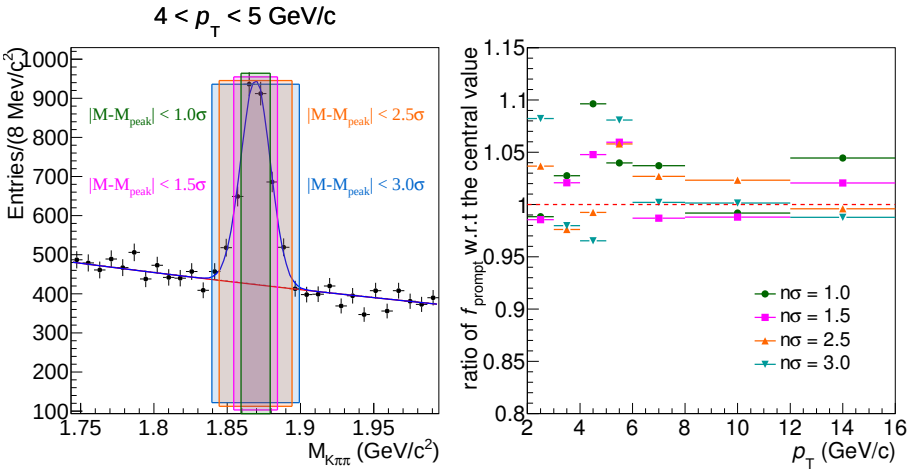

Figure 5.17. Left: the four invariant-mass windows tested for the systematic uncertainty estimation in the $p_{\rm T}$ interval $4 < p_{\rm T} < 5$ GeV/c. Right: The ratio of f_{prompt} with respect to the reference value obtained with $|M - M_{peak}| < 2\sigma$ as a function of $p_{\rm T}$.

is well estimated, the pull distributions are also checked in the p_{T} intervals of the analysis. The pull is defined as the Gaussian width of the D^+ prompt fraction obtained from the fit minus the one set as input, divided by the error on the measured prompt fraction:

$$Pull(p_{\mathrm{T}}) = \sigma\left(\frac{f^{meas}_{prompt} - f^{gen}_{prompt}}{\sigma_{f^{meas}_{prompt}}}\right)(p_{\mathrm{T}}). \tag{5.13}$$

Since the difference between the input and measured f_{prompt} should be covered by the statistical errors, the pull distribution is expected to be a normal Gaussian distribution with zero mean and width equal to unity. As shown in the left panel of figure 5.19 the pulls obtained stay between 0.8 and 1.1, indicating that the statistical error is reasonably well estimated. Finally, the right panel of the same figure shows the mean value of σ_{prompt} obtained for each input prompt fraction tested, compared with the value obtained from the MC prefit phase. Since the fitted distributions are simulated from the MC distributions for prompt and feed-down D^+-mesons, the two values of this parameter should be the equal, in fact, they result to be well in agreement within the uncertainties.

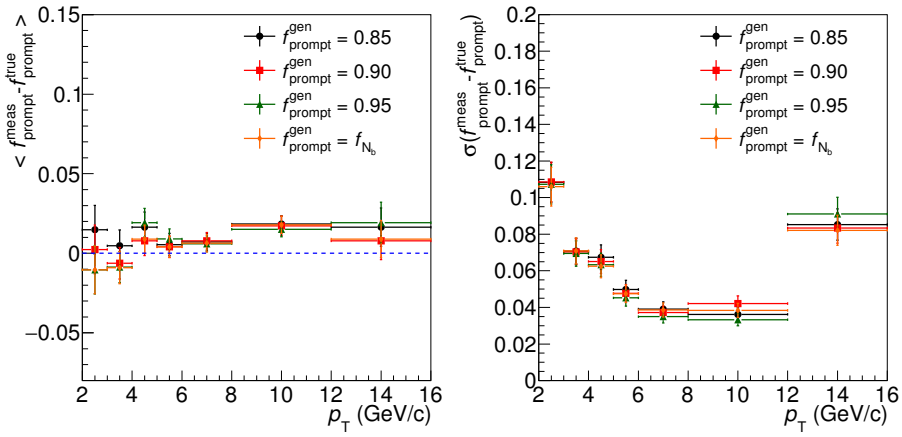

Figure 5.18. Mean (left) and RMS (right) of the residuals distributions for the D^+ prompt fraction evaluated with the MC closure test, for four different input values of f_{prompt}.

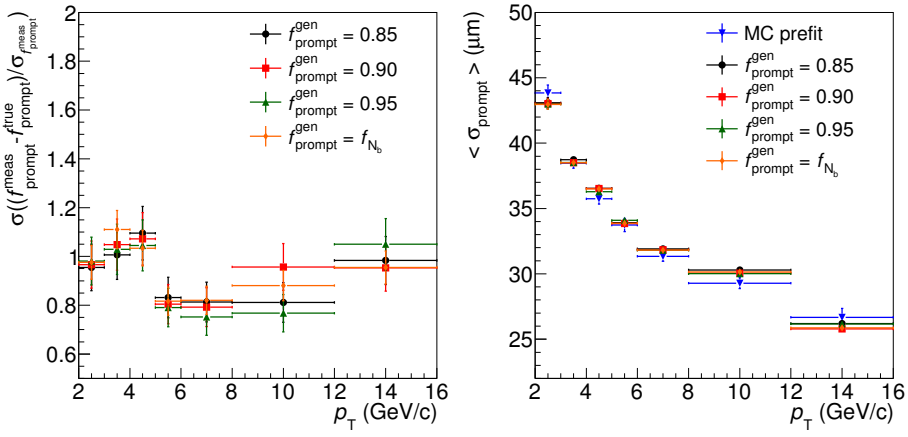

Figure 5.19. Left: pulls of f_{prompt} (left) and the mean value of σ_{prompt} (right) evaluated with the MC closure test, for four different input values of f_{prompt}.

5.4.5 Summary of the systematic uncertainties

The total systematic uncertainty was evaluated by considering all the tests described above. The systematic uncertainty due to the assumed shapes of the impact-parameter distributions was found to be the dominant source, especially

p_T (GeV/c)	[2,3]	[3,4]	[4,5]	[5,6]	[6,8]	[8,12]	[12,16]
σ_{prompt}	negl.	negl.	negl.	negl.	negl.	negl.	negl.
Mass window	negl.	negl.	negl.	negl.	negl.	negl.	negl.
Fit range	2%	2%	2%	2%	2%	2%	2%
Bkg fit function	4%	1%	1%	1%	1%	1%	1%
Sidebands region	6%	3%	3%	3%	3%	3%	3%
S parameter	4%	2%	2%	2%	2%	2%	2%
p_T shape	negl.	negl.	negl.	negl.	negl.	negl.	negl.
MC closure test	1%	1%	1%	1%	1%	1%	1%
total syst.	9%	4%	4%	4%	4%	4%	4%

Table 5.2. Summary of the systematic uncertainty for the data-driven impact–parameter analysis for the evaluation of the fraction of prompt D^+-mesons.

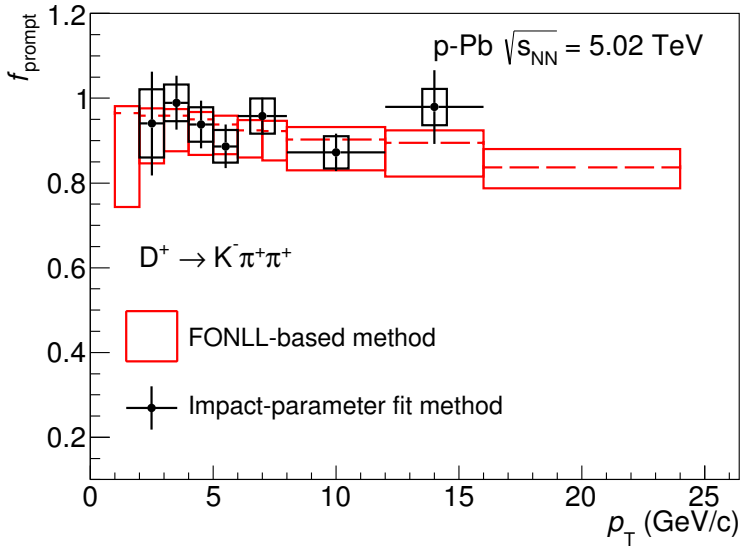

Figure 5.20. D^+ prompt fraction obtained with the impact-parameter fit method, compared to the FONLL-based method (N_b and f_c combined). The vertical bars represent the statistical uncertainty, while the boxes the systematic one.

considering the tests performed on the impact-parameter distributions of the background. To compute the total systematic uncertainty, all the variations observed performing the tests were treated as uncorrelated and therefore summed in quadrature, as reported in table 5.2.

5.5 Results

The results for f_{prompt} obtained with the impact-parameter fit method in the seven p_T intervals are found to be compatible with the one obtained with the FONLL-based method within the uncertainties in the full p_T range, as shown in figure 5.20. The vertical bars represent the statistical error, while the boxes the systematic one.

In some p_T bins the error bars exceed the unity, because the statistical uncertainty obtained by the fitting procedure is symmetric with respect to the central value. This is due to the relative amount of prompt D^+, feed-down D^+ and background estimated with the fit. Furthermore, the systematic studies were performed without limiting f_{prompt} at 1, since a bias in the result was observed if this limit was set.

However, a fraction of prompt D^+ in the raw yield larger than 1 has not a physical meaning, and therefore the total uncertainty should be truncated at one. The reason why it was chosen not to represent the total uncertainty in the picture and truncate it at 1, was to provide information about the different contribution of the statistical and systematic uncertainty to the total uncertainty. It is indeed possible to notice that the statistical uncertainty is larger in each p_T interval, which means that this analysis will greatly benefit starting from the Run II, because of the larger data samples expected. The total uncertainty reported in table 5.3 was computed treating the statistical and systematic uncertainties as uncorrelated and therefore summed in quadrature. The fraction of prompt D^+ evaluated by fitting the impact-parameter distributions has similar uncertainties in the intermediate p_T intervals ($3 < p_T < 12$ GeV/c) and larger uncertainties in the first and in the last analysed p_T intervals, with respect to the theory-driven method. Even if this method does not allow to improve significantly the measurement with the current statistics, it provides an important confirmation of the result obtained with the FONLL-based method, which depends on theoretical pQCD calculations and the assumption on the R_{pPb} of feed-down D^+-meson. Furthermore, the f_{prompt} measurement with this method will improve in the future, because of the larger data samples expected in the LHC Run II, which has already started in 2015, and even more with the upgrade of the Inner Tracking System planned for 2019, as will be briefly discussed in the next section. The results of this work have been published as a cross-check of the standard approach by the ALICE Collaboration [1]. Figure 5.21, taken from [1], shows an example of fit (left column) and the resulting f_{prompt} as a function of p_T (right column) for the three D-meson species analysed: D^0 (top), D^+ (middle) and D^{+*} (bottom). The prompt fraction evaluated with the impact-parameter fit method was found to be compatible with the FONLL-based estimation within the uncertainties for the other D-meson species as well. In particular, for the D^0 the data-driven approach provides a more precise estimation of f_{prompt} in the transverse momentum range $2 < p_T < 12$ GeV/c as compared to the FONLL-based approach, while for $1 < p_T < 2$ GeV/c only a lower limit at a 95% confidence level

p_T (GeV/c)	[2,3]	[3,4]	[4,5]	[5,6]	[6,7]	[7,8]	[8,12]	[12,16]
Theory-driven (%)	+2 -12	+2 -9	+2 -9	+2 -8	+3 -7	+3 -7	+3 -8	+3 -9
Imp. par. fit (%)	±16	±7	±7	±7	±6		±6	±10

Table 5.3. Comparison between the total uncertainty on the D^+ prompt fraction evaluated with the impact-parameter fit method and the FONLL-based method.

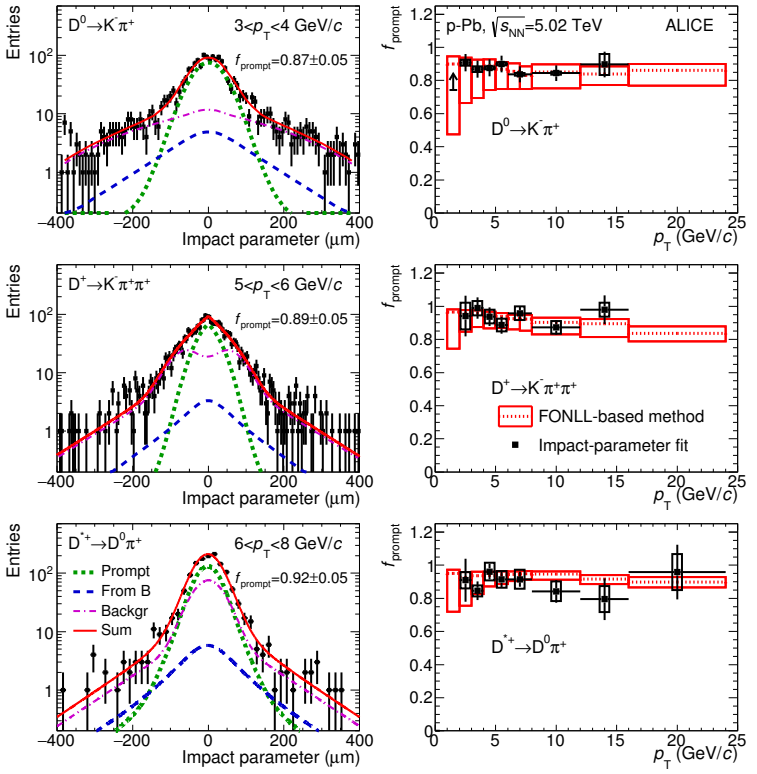

Figure 5.21. Examples of impact-parameter fits (left column) and resulting prompt fraction as a function of p_{T} (right column) for the three different D-meson species analysed: D^0 (top), D^+ (middle), D^{+*} (bottom) [1]. The arrow in the interval $1 < p_{\mathrm{T}} < 2$ GeV/c for the D^0-meson represents the minimum value within a 95% confidence level.

could be estimated, given the poor accuracy of the impact-parameter fit.

5.6 Perspective of an improved measurement with the ITS upgrade

The ALICE Collaboration is preparing a major upgrade of the experimental apparatus, planned for installation in the second long LHC shutdown in 2019, which will also involve the Inner Tracking System. The main goals of the upgrade of the

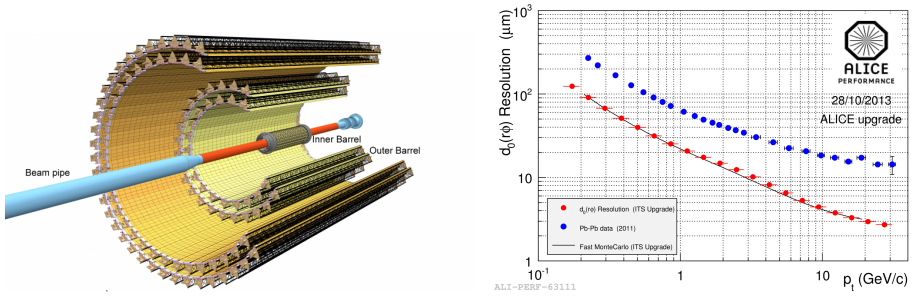

Figure 5.22. Left: Layout of the upgraded ALICE ITS. Right: improvement of the impact-parameter resolution in the transverse plane expected with the upgraded ITS [66].

ALICE ITS are an improved reconstruction of primary and secondary vertices and an improved performance in the tracking of low-momentum particles. In order to achieve these targets, the following changes will be applied:

- Replace the six ITS layers of the current setup with seven concentric barrel layers of silicon pixel detectors with radii between 22 mm for the innermost and 400 mm for the outermost layer, as shown in the left panel of figure 5.22.
- Reduce the material budget down to 0.3% X_0 in the innermost and $\sim$ 0.9% X_0 in the outer and middle layers compared to 1.14% X_0 per layer in the current ALICE Silicon Pixel Detector.
- Use sensors with a pixel size of 30 μm x 30 μm, compared to 50 μm x 425 μm in the current pixel detector.

The central part of the upgraded ALICE ITS is the pixel chip. The most challenging requirements for the detector design are the high spatial resolution and the extremely low material budget, in particular for the inner layers. These requirements, together with the modest radiation environment compared to the other LHC trackers, led to the choice of Monolithic Active Pixel Sensors (MAPS) [65]. As already mentioned, the impact-parameter resolution will greatly benefit from this upgrade. The right panel of figure 5.22 shows the comparison between the current impact-parameter resolution in the xy-plane (blue points) and the expected one (red points) with the upgraded ITS, obtained with a MC simulation, in the 10% most central Pb-Pb collisions [66]. The improvement of the impact-parameter resolution in the xy-plane, which will reach a factor $\sim$ 3, will allow to achieve a better separation between the prompt and feed-down D^+ contributions, which, together with the higher statistics expected, will result in a more precise estimation of the prompt fraction with the impact-parameter fit method. Tests of the

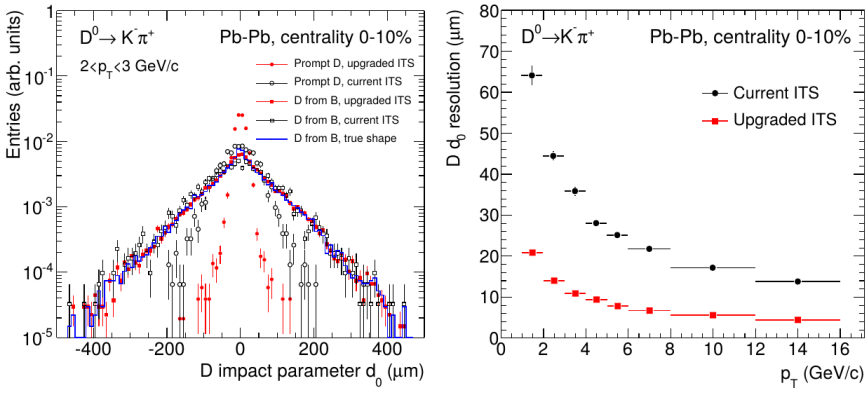

Figure 5.23. Left: simulated impact-parameter distributions for prompt and feed-down D^0 obtained with the current and upgraded ITS configurations in the interval $2 < p_T < 3$ GeV/c [67]. Right: D^0 impact-parameter resolution as a function of p_T

performances of the upgraded ITS for the measurement of the prompt fraction with the impact-parameter fit method have been carried out for D^0 mesons [67]. In the left panel of figure 5.23 the simulated impact-parameter distributions for prompt and feed-down D^0 with $2 < p_T < 3$ GeV/c, obtained with the current and upgraded ITS configurations, are compared. In the right panel of the same figure the comparison between the impact-parameter resolution of D^0 with the two different configurations is shown; as already mentioned, with the upgraded ITS the resolution will be improved up to a factor 3.