

# 7

## Prompt $D^+$ reconstruction with the Kalman filter vertexing package

In this Chapter I will present the measurement of the  $p_T$ -differential production cross section of prompt  $D^+$ -mesons in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV, performed using an alternative vertex-reconstruction approach with respect to the one implemented in the ALICE analysis framework, based on the Kalman filter algorithm. The software used for this analysis (*KFParticle* package [68]) is under development for the CBM experiment, which will be one of the major scientific pillars of the future Facility for Antiproton and Ion Research (FAIR) in Darmstadt [69]. However, because of the possibility of an improvement in the analyses which involve decayed particles, such as open heavy-flavoured or strange hadrons, it is also planned to include this software in the ALICE framework. At the moment it is only possible to use it to analyse the ALICE data at the GSI (*Gesellschaft für Schwerionenforschung*) laboratory, where the *KFParticle* package has been included as an external library in the software.

In the first section of this Chapter a brief introduction to the Kalman filter algorithm and, in particular, to the application for the decayed-particle reconstruction is provided. In the second section I will present a test of the performances of the *KFParticle* package for the prompt  $D^+$ -meson reconstruction in the MC simulation described in section 4.3. Finally, the analysis performed reconstructing the  $D^+$ -mesons with the *KFParticle* package will be presented and the measured  $p_T$ -differential cross section of prompt  $D^+$ -mesons compared to the published result.

## 7.1 The Kalman filter algorithm

The Kalman filter is an algorithm which provides a precise estimation  $\mathbf{r}$  and its covariance matrix  $\mathbf{C}$  of an unknown state vector  $\mathbf{r}^t$ , using a series of  $k = 1, \dots, n$  measurements  $\mathbf{m}_k$ , containing statistical noise. The algorithm starts with an approximation of the vector, which is subsequently updated adding one measurement after another. In general the vector  $\mathbf{r}^t$  is changed after each iteration. However, in order to provide a simplified description of the algorithm, I assume  $\mathbf{r}^t$  identical for each determination and the measurement  $\mathbf{m}_k$  linearly dependent on  $\mathbf{r}^t$ :

$$\mathbf{m}_k = H_k \mathbf{r}^t + \boldsymbol{\eta}_k, \quad (7.1)$$

where the matrix  $H_k$  is the model of measurement and  $\boldsymbol{\eta}_k$  is the statistical error on the  $k$ -th measurement  $\mathbf{m}_k$ , which is assumed to be unbiased and its covariance matrix  $V_k$  to be note:

$$\begin{aligned} \langle \boldsymbol{\eta}_k \rangle &= 0, \\ \langle \boldsymbol{\eta}_k \cdot \boldsymbol{\eta}_k^T \rangle &= V_k. \end{aligned} \quad (7.2)$$

Then the fitting algorithm is divided in two steps:

- **Initialisation:** the state vector is approximated to  $\mathbf{r}_0$  and its covariance matrix is set to  $\mathbf{C} = \mathbf{1} \cdot N$ , where  $N$  is a large number.
- **Filtration:** for each  $k$ -th measurement  $\mathbf{m}_k$ , starting from the values obtained in the  $(k-1)$ -th step, the following quantities are calculated:

$$\begin{aligned} S_k &= (V_k + H_k \mathbf{C}_k H_k^T)^{-1}, \\ K_k &= \mathbf{C}_k H_k^T S_k, \\ \zeta_k &= \mathbf{m}_k - H_k \mathbf{r}_{k-1}, \\ \mathbf{r}_k &= \mathbf{r}_{k-1} + K_k \zeta_k, \\ \mathbf{C}_k &= \mathbf{C}_{k-1} - K_k H_k \mathbf{C}_{k-1}, \\ \chi_k^2 &= \chi_{k-1}^2 + \zeta_k^T S_k \zeta_k. \end{aligned} \quad (7.3)$$

In the above formula  $K_k$  is the gain matrix,  $\zeta_k$  the residuals vector and  $\chi_k^2$  the total chi square of the obtained estimation  $\mathbf{r}_k$  from the  $k$ -th measurement.

The best estimation  $\mathbf{r}_n$  of the state vector  $\mathbf{r}^t$  is obtained after the filtration of the last measurement, together with the covariance matrix  $\mathbf{C}$ . The same procedure can be performed of measurements non-linearly dependent on  $\mathbf{r}^t$ ; in this case one speaks of *extended* Kalman filter.

### 7.1.1 Reconstruction of decayed particles with the KFParticle package

The KFParticle package provides an estimation of the parameters of the decayed particles [70], starting from the daughter tracks and their covariance matrices. Furthermore, the algorithm is improved with respect to the standard Kalman filter described in the section above and it takes into account the correlation among the  $k$ -th measurement  $\mathbf{m}_k$  and the previous measurements  $\mathbf{m}_{i < k}$ . The state vector used for the parametrisation of the particle is

$$\mathbf{x}^T = (x, y, z, p_x, p_y, p_z, E, s), \quad (7.4)$$

where  $\mathbf{v} = (x, y, z)$  are the spatial coordinates of the point where the particle is build (the decay vertex),  $\mathbf{p} = (p_x, p_y, p_z)$  is the particle momentum in the considered spatial point,  $E$  is the energy and  $s$  is a parameter called *signed path*, which is defined as the length of the particle trajectory normalised to the momentum. Therefore,  $s$  is not zero only in case of the particle is computed using the production vertex information (see next section). By parametrising the trajectory in this way, the particle with the corresponding covariance matrix can be transported to each point of its trajectory.

### 7.1.2 Fit constraints

The precision of the secondary vertex parameters obtained in the fit can be improved by taking into account additional information known about the decayed particle, for example the production vertex or the mass. These assumptions can be used to constraint the parameters of the state vector. In particular I focus on the *topological constraint*, which will be used in the analysis presented in the next sections.

#### Topological constraint

The topological constraint consists in the allignement of a decayed particle with the already known production vertex  $\mathbf{v}_{pv}$ . Considering the simplified case of a neutral particle, the trajectory which connect the primary to the decay vertex is a straight line:

$$\mathbf{v} - s \cdot \mathbf{p} = \mathbf{v}_{pv}. \quad (7.5)$$

This latter equation is used to fix a constraint to the parameters of the state vector (see equation 7.4) and force the reconstructed particle to point to the production vertex. The same procedure can be repeated for charged particles in a magnetic field, but in this case the algorithm takes into account the curvature of the particle trajectory.

A possible application of the topological constraint can be the separation of the prompt from the feed-down contribution of  $D^+$ -mesons by evaluating the  $\chi^2$  of the particle after the constraint. If the  $D$ -meson is prompt, it is produced in the primary vertex by definition, and then the resulting  $\chi^2$  value divided by the number of degree of freedom should be on average close to 1. Since the feed-down  $D^+$ -mesons do not originate from the primary vertex, fixing the topological constraint at the primary vertex for them would imply on average a larger value of the  $\chi^2$ . This is the basic idea of the analysis presented in this Chapter. In the next section the performances of the reconstruction of the prompt  $D^+$ -mesons using the Kalman filter algorithm are tested for the MC sample described in section 4.3, while in the following section this analysis method is applied for the measurement of the  $p_T$ -differential production cross section of prompt  $D^+$ -mesons.

## 7.2 Reconstruction performances of the KFParticle package for prompt $D^+$ -mesons

The precision and the quality of the reconstructed quantities of the prompt  $D^+$ -meson in the decay channel  $D^+ \rightarrow K^- \pi^+ \pi^+$  using the KFParticle package was tested in the MC sample presented in section 4.3, applying the event and track quality selections described in Chapter 4. The checks have been performed selecting only prompt  $D^+$  in the MC simulation, in order to test the topological constraint at the primary vertex, which, in case of prompt  $D^+$ , coincides with the production vertex. The results were also compared to the values obtained with the vertex-reconstruction algorithm used as default in the ALICE Collaboration. In

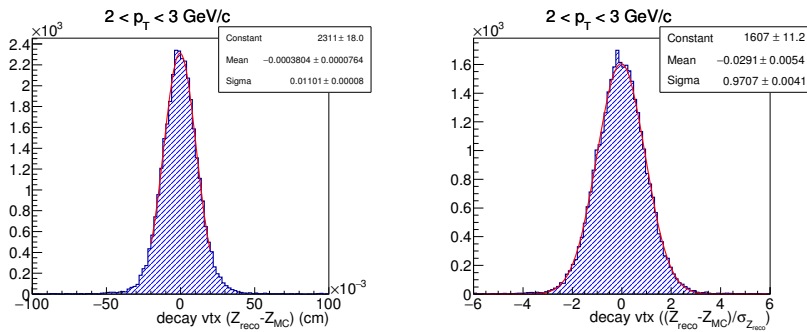


Figure 7.1. Example of Gaussian fit to the residual (left) and the pull (right) distributions of the  $z$  coordinate of the  $D^+$  decay vertex in the interval  $2 < p_T < 3$  GeV/c obtained with the KFParticle package.

particular, the precision was quantified by computing the resolution, defined as the Gaussian width obtained by fitting the residual distributions of a given observable  $X$  in a  $p_T$  interval:

$$Res(p_T) = \sigma(X_{reco} - X_{MC})(p_T), \quad (7.6)$$

where  $X_{reco}$  is the reconstructed value of the variable  $X$ , while  $X_{MC}$  is the value known from the MC truth. To check whether the uncertainty on the variable  $X$  is well estimated, the pulls are evaluated:

$$Pull(p_T) = \sigma \left( \frac{X_{reco} - X_{MC}}{\sigma_{X_{reco}}} \right) (p_T). \quad (7.7)$$

The pull distribution is expected to be Gaussian with zero mean and width equal to unity, if the difference between the reconstructed and the MC values is covered by the statistical error  $\sigma_{X_{reco}}$  on the reconstructed observable  $X$ . In figure 7.1 is depicted an example of Gaussian fit to the residual (left) and the pull (right) distributions of the  $z$  coordinate of the  $D^+$  decay vertex, in the transverse momentum interval  $2 < p_T < 3$  GeV/c.

### 7.2.1 Prompt $D^+$ -mesons at the decay vertex

The left panel of figure 7.2 shows the resolution on the  $D^+$  decay vertex position (only  $x$  and  $z$  coordinates, since the  $y$  coordinate is equivalent to the  $x$  one) evaluated as a function of  $p_T$ , using the KFParticle package (*KF*), with and without the topological constraint at the reconstructed primary vertex, and the standard

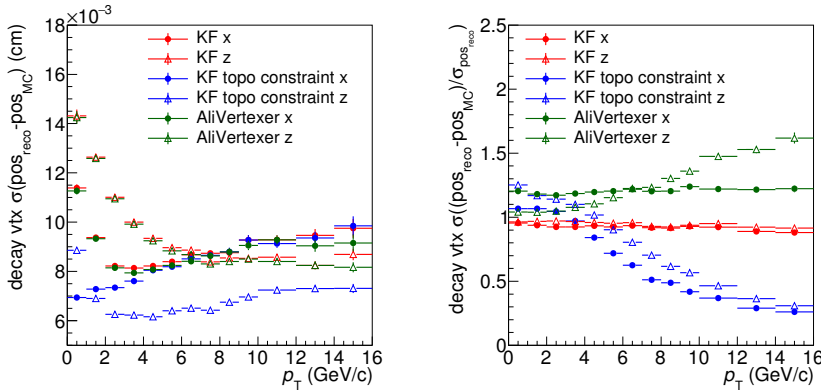


Figure 7.2. Resolution (left) and pulls (right) of the  $x$  and  $z$  coordinates of the  $D^+$  decay vertex position as a function of  $p_T$ .

algorithm used in ALICE (*AliVertexer*). In case of topological constraint, the particle trajectory is recomputed and the parameters of the reconstructed particle are stored at the production vertex. Therefore it is necessary to transport the reconstructed particle to the decay vertex. Since in the analysis of the data it is only possible to use reconstructed points, the particle has been transported to the point on its trajectory closest to the reconstructed decay vertex, stored before applying the constraint. As expected, the resolution is worse for the  $x$  coordinate rather than the  $z$  coordinate, for all the reconstruction algorithms tested. The resolution obtained with the Kalman filter without constraint is compatible with the one obtained using the standard vertex algorithm, while applying the topological constraint an improvement in the low (full)  $p_T$  region is observed for the  $x$  ( $z$ ) coordinate. This observation confirms the improved reconstruction of prompt  $D^+$  achieved by using the topological constraint, however it cannot be used in the analysis, since it is not possible to separate a priori the prompt from the feed-down  $D^+$ -mesons. The values of the pulls obtained with the standard vertexer are slightly larger than one, indicating an underestimation of the errors of about 20% (up to 50% for the  $z$  coordinate). The pulls obtained with the Kalman filter are close to unity in case of no constraint, while they decrease up to 0.2 if the topological constraint is applied. This effect is due to the fact that the errors on the reconstructed trajectory of the particle and the secondary vertex previously stored are strongly correlated. However, as already mentioned, the reconstructed position of the secondary vertex and its error should never be used after applying the topological constraint.

### 7.2.2 Prompt $D^+$ -mesons at the production vertex

The resolution and the pulls have been tested also for the position of the prompt  $D^+$  at the production vertex. In case of topological constraint the particle parameters are already stored at the production vertex, while, if the constraint is not applied, the particle must be transported to the point of closest approach to the reconstructed primary vertex (which in this case corresponds to the production vertex). In the latter case the errors on the reconstructed primary vertex are taken into account. Figure 7.3 shows the resolution (left) and the pulls (right) of the  $x$  and  $z$  coordinates of the production vertex position in both cases. The resolution is better for the  $x$  coordinate and it is improved for both the coordinates after applying the topological constraint due to the fact that the particle trajectory is forced to pass closer to the primary vertex. The pulls are close to unity for the  $z$  coordinate, while 10%-20% lower for the  $x$  coordinate. A quantity related to the position at the production vertex is the impact parameter, which is computed as the distance of closest approach of the  $D^+$  trajectory to the reconstructed primary vertex. The true value of the impact parameter is zero, since the test is performed only on prompt  $D^+$ -mesons. In figure 7.4 is shown the resolution (left) and the

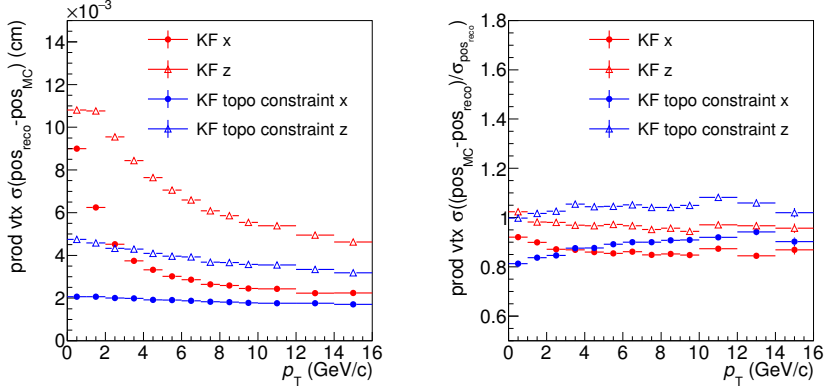


Figure 7.3. Resolution (left) and pulls (right) of the  $x$  and  $z$  coordinates of the  $D^+$  production vertex position as a function of  $p_T$ .

pulls (right) for the impact parameter in the transverse plane obtained with the Kalman filter compared to that achieved with the standard algorithm. Even if, with the standard algorithm, the impact parameter is computed by approximating the prolongation of the  $D^+$  trajectory with a straight line, the resolution (left panel of figure 7.4) is similar to the one achieved with the Kalman filter without

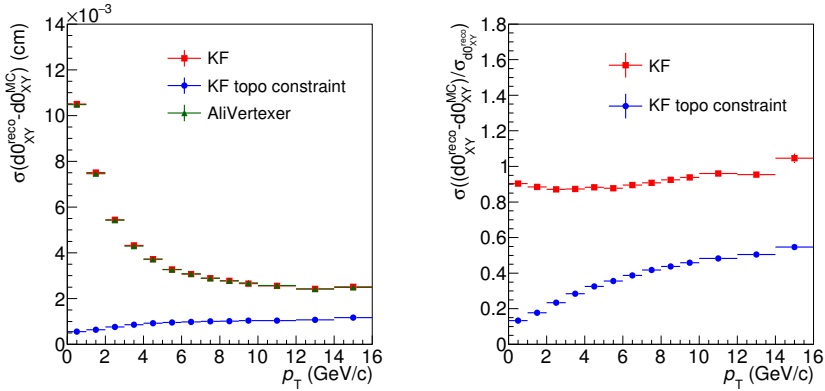


Figure 7.4. Resolution (left) and pulls (right) of the impact parameter of the  $D^+$ -mesons as a function of  $p_T$ .

constraints, which takes into account the curvature of the  $D^+$  track induced by the magnetic field. This is a consequence of the short decay length of the  $D^+$ -mesons, for which the curvature becomes negligible. The impact parameter resolution results to be improved by construction applying the topological constraint. The right panel of figure 7.4 shows that the pulls of the impact parameter in the transverse plane obtained with the Kalman filter without constraints are close to unity, while they indicate a significant overestimation of the errors in case of topological constraint. However, the impact parameter obtained applying the constraint cannot be used in the analysis of the data since it is not possible to discriminate a priori prompt from feed-down  $D^+$ .

### 7.2.3 Prompt $D^+$ -mesons decay length

The decay length is computed differently by the standard algorithm of ALICE and the Kalman filter. In the first case it is calculated as the distance between the reconstructed decay vertex of the  $D^+$ -meson and the reconstructed primary vertex, and therefore it is approximated as a straight line, while in the second case it is given by the  $s$  parameter, multiplied by the momentum of the particle (see section 7.1.1). In the panel of figure 7.5 is shown the resolution of the decay length in the transverse momentum calculated with the two different approaches. The resolution seems to be worse considering the reconstruction of Kalman filter, however the MC value is computed as the distance between the MC primary and secondary vertices, since the information of the decay length is not stored in the MC truth. In the right panel of the same figure the pulls are plotted. For the standard algorithm

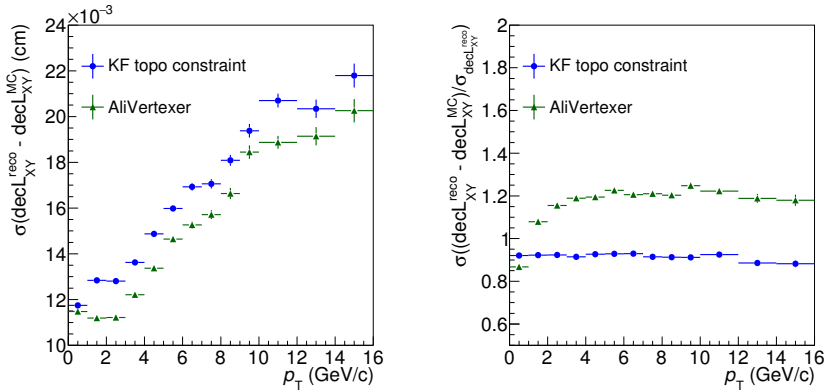


Figure 7.5. Resolution (left) and pulls (right) of the decay length in the transverse plane of the  $D^+$ -mesons as a function of  $p_T$ .



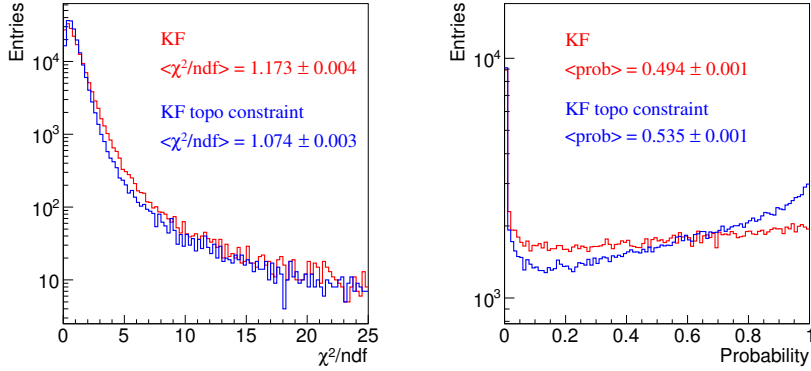


Figure 7.6.  $\chi^2$  (left) and probability (right) distributions for the  $D^+$ -mesons reconstructed with the Kalman filter algorithm, with and without fixing the topological constraint.

they suggest an underestimation of the error of about 20%, while for the Kalman filter an overestimation of about 10%. This feature will result in a shift to lower values of the normalised decay length distribution reconstructed with the Kalman filter algorithm with respect to the standard one.

#### 7.2.4 Prompt $D^+$ -mesons $\chi^2$ and probability

The possibility to evaluate the  $\chi^2$  of the reconstructed decayed particle is the feature of the Kalman filter that, in the next section, will be exploited for the subtraction of the feed-down distribution. However, it is important to check whether the  $\chi^2$  is well evaluated for prompt  $D^+$ , especially in the case of topological constraint, before applying the selection on this variable. Figure 7.6 shows the distribution of the  $\chi^2$  divided by the number of degrees of freedom (left) and the associated probability (right). As expected in case of a good estimation of the errors, the mean value of the  $\chi^2/ndf$  distribution is close to unity, while the probability distribution is more or less flat, both with and without applying the topological constraint.

#### 7.2.5 Primary vertex manipulations

Another feature of the Kalman filter algorithm is the possibility to add or remove measurements to the evaluation of the state vector. In the case of the primary vertex it results in adding or removing tracks from the fit. Therefore, not only the tracks of the  $D^+$  daughters (which do not originate from the primary vertex) can be

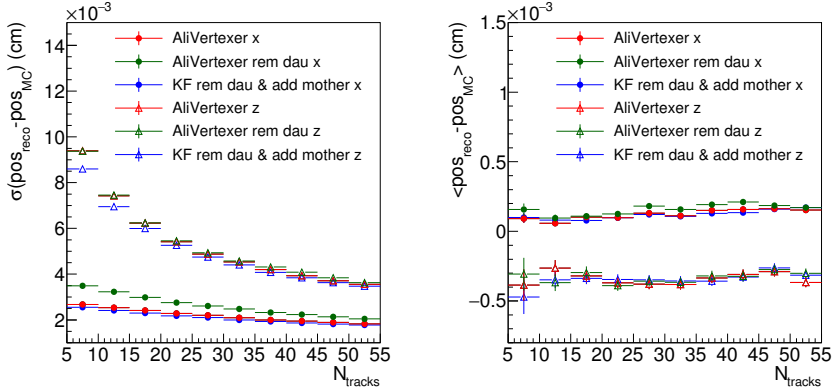


Figure 7.7. Resolution (left) and bias (right) of the  $x$  and  $z$  coordinates of the reconstructed primary vertex position after several manipulations.

removed from the primary vertex evaluation, however it is also possible to add the reconstructed  $D^+$  itself, as an additional measurement. In figure 7.7 are shown the Gaussian widths (resolution) and the means (bias) of the residuals distributions of the  $x$  and  $z$  coordinates of the primary vertex position, as a function of the number of tracks used to evaluate the primary vertex position, for events with at least a  $D^+$ -meson. The resolution (left panel) results to be better for the  $x$  coordinate with respect to the  $z$  coordinate. Furthermore, it seems to be worse if the daughter tracks are removed, but it is only due to the fact that the resolution worsens with a decreasing number of the tracks that contribute to its determination. On the contrary, the resolution is slightly improved if the reconstructed track of the  $D^+$  is added to the determination of the primary vertex using the Kalman filter algorithm, since the prompt  $D^+$ -mesons originate from the primary interaction vertex. However, like the evaluation of the observables applying the topological constraint, it cannot be used in the analysis, since it is not possible to identify which  $D^+$ -mesons are prompt or feed-down. Despite that, the  $\chi^2$  of the primary vertex, after adding the  $D^+$  to the fit, can be used as an additional information to discriminate between the two contributions. In fact, since the feed-down  $D^+$  trajectories in general do not point at the primary vertex, adding them in the fit would imply a worse value of the  $\chi^2$ . Including the daughter tracks in the primary vertex determination should create a bias in the  $D^+$  flight direction. This is the reason to exclude them from the fit. However, the  $D^+$  direction is random with respect to the single coordinates and therefore the mean values of the residuals distributions for the single coordinates do not change after removing the daughter tracks or adding the prompt  $D^+$ , as illustrated in the right plot of figure 7.7. In

particular, for all the tests performed, a small bias is observed, which is of the order of  $\sim 1 - 2 \mu\text{m}$  in the  $x$  direction, while of about  $\sim 3 - 4 \mu\text{m}$  for the  $z$  coordinate.

## 7.3 Prompt $D^+$ measurement with the KFPARTICLE package

The  $p_{\text{T}}$ -differential production cross section of prompt  $D^+$  was measured in the p-Pb data sample at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$  presented in Chapter 4, using the KFPARTICLE package for the reconstruction of the  $D^+$  candidates. The transverse momentum range covered is  $1 < p_{\text{T}} < 24 \text{ GeV}/c$  and the  $p_{\text{T}}$ -intervals are the same as in the standard analysis, published by the ALICE Collaboration [36]. For this analysis, all the event, track quality and PID selections used in the standard analysis were applied. As will be explained in the next section, the selection criteria on the topological variables were slightly changed with respect to those reported in table 4.2 and the  $\chi^2/\text{ndf}$  of the reconstructed  $D^+$ -mesons with Kalman filter algorithm, evaluated after applying the topological constraint, added, in order to separate the prompt from the feed-down contribution. As already anticipated in Chapter 4, the code for the latter selections and the  $D^+$  candidate reconstruction has been re-implemented, since it was necessary to use the ESDs instead of the AODs. Furthermore, the available statistics at the GSI farm was slightly lower, therefore the resulting number of analysed events is of the order of  $80 \cdot 10^6$ .

### 7.3.1 Topological selections

All the quantities related to the  $D^+$  decay topology exploited in the analysis were computed by reconstructing the  $D^+$  candidates with the KFPARTICLE package, without applying any constraint, except for the decay length and the  $\chi^2$  of the candidate, for which the topological constraint at the primary vertex was used. In the

$p_{\text{T}}$ (GeV/c)	[1,2]	[2,3]	[3,8]	[8,12]	[12,16]	[16,24]
$p_{\text{T}}^{\text{dau}}$ (GeV/c) >	0.3	0.3	0.3	0.3	0.3	0.3
Norm. $L_{xy}$ >	6	5	5	5	6	6
$\cos(\theta_P)$ >	0.985	0.985	0.985	0.985	0.985	0.985
$\sigma_{\text{vertex}}$ (cm) <	0.030	0.030	0.034	0.050	0.050	0.050
$\chi^2/\text{ndf}$ <	2.5	3.0	3.0	3.5	3.5	4.0

Table 7.1. Summary table of the  $D^+$  topological cuts optimised for the analysis performed using the Kalman filter algorithm for the reconstruction of the  $D^+$ -mesons.

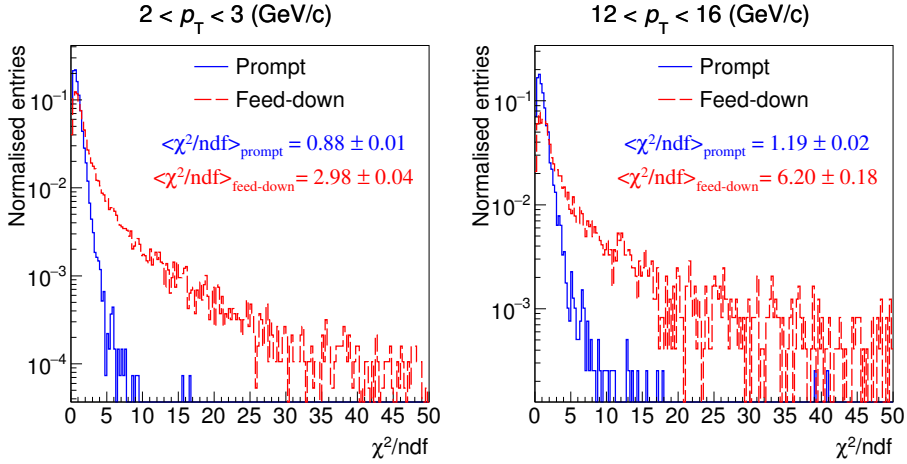


Figure 7.8. Comparison between the distributions of  $\chi^2/ndf$  of prompt and feed-down  $D^+$ -mesons after applying the topological constraint at the primary vertex from MC simulations, in two different  $p_T$  intervals of the analysis.

first case the topological constraint was applied because the Kalman filter algorithm needs the information of the production vertex to compute the decay length, while in the second it was applied to separate the prompt from the feed-down  $D^+$  contribution. Figure 7.8 shows the comparison between the distributions of the  $\chi^2/ndf$  of prompt and feed-down  $D^+$  obtained from MC simulations in two different  $p_T$  intervals of the analysis. From this figure it is possible to note that the mean value of  $\chi^2/ndf$  is close to unity for prompt  $D^+$ , while it is wider for feed-down  $D^+$ , and that the difference between the two distributions increases with increasing  $D^+$  transverse momentum. In order to exploit as much as possible the selection on the  $D^+$  candidates  $\chi^2/ndf$  after applying the topological constraint at the primary vertex, the other selections have been slightly released, as reported in table 7.1. In particular, the cuts on the cosine of the pointing angle and on the normalised decay length in the transverse plane were set to lower values with respect to those reported in table 4.2, while the selections on  $\cos(\theta_P^{xy})$  and the decay length were not applied. Furthermore, the cut values of the  $\chi^2/ndf$  were chosen taking into account the amount of signal rejected, the efficiency of prompt and feed-down  $D^+$ , the fraction of prompt  $D^+$  and the possible bias on the  $p_T$ -differential cross section induced by the discrepancy between the distributions of  $\chi^2/ndf$  in data and MC.

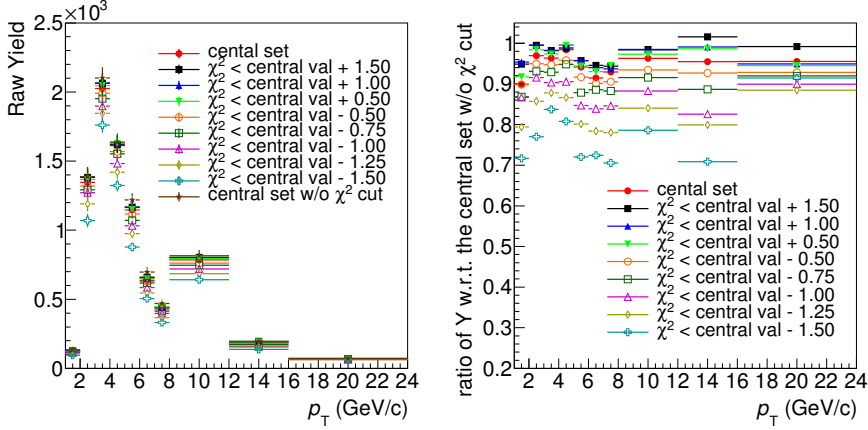


Figure 7.9. Values of the extracted raw yield in the  $p_T$  bins of the analysis for each  $\chi^2/ndf$  cut value tested (left) and relative variation with respect to the values obtained without applying the  $\chi^2/ndf$  cut (right).

### 7.3.2 Variation of the $D^+$ -meson $\chi^2/ndf$ selection

In this section I present all the variations of the cut on the  $D^+$  candidate  $\chi^2/ndf$  after applying the topological constraint, performed in order to choose the values reported in table 7.1 and to test the stability of the result. In particular, several cut values have been tested, varying those reported in table 7.1 from a minimum of  $-1.5$  to a maximum of  $+1.5$ .

#### Raw yield variation

The raw yield was extracted in the  $p_T$  intervals of the analysis by fitting the invariant-mass distributions with the sum of a Gaussian function for the signal (prompt and feed-down  $D^+$ ) and an exponential function for the combinatorial background. For these tests the Gaussian width was fixed to the values obtained from the MC simulations in order to avoid as much as possible the statistical fluctuations. The absolute value of the extracted raw yields in the  $p_T$  intervals of the analysis for each trial is shown in the left panel of figure 7.9, while in the right panel is shown the relative variation with respect to the values obtained without applying the cut on  $\chi^2/ndf$ . The amount of signal (prompt and feed-down  $D^+$ ) loss applying the central value of  $\chi^2/ndf$  cut is of the order of about 5-10% (red points), while for the tightest cut tested the raw yields decrease up to 30%. Therefore, the central cut values of  $\chi^2/ndf$  reported in table 7.1 allows one to keep the majority

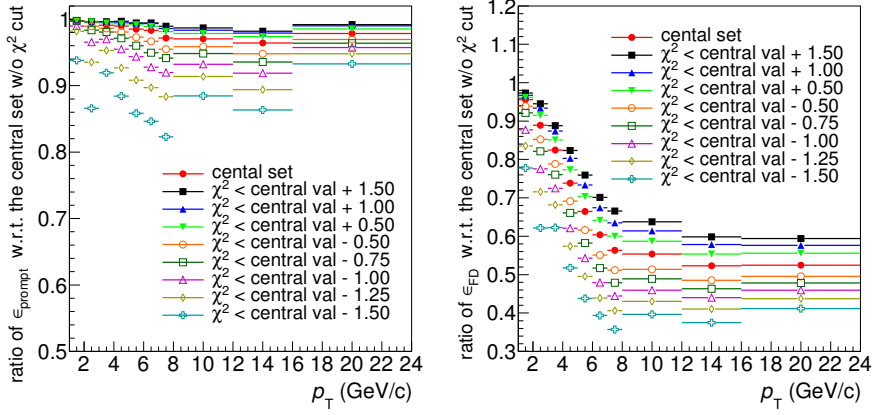


Figure 7.10. Relative variations of the efficiencies of prompt (left) and feed-down (right)  $D^+$  obtained for each  $\chi^2/ndf$  cut value tested with respect to the efficiencies obtained without applying the  $\chi^2/ndf$  cut.

of the raw yields and hence not to increase significantly the statistical uncertainty.

### Efficiency variation

In figure 7.10 are shown the relative variations of the efficiencies for prompt (left) and feed-down (right)  $D^+$  for each cut value of  $\chi^2/ndf$  tested, with respect to the efficiency obtained without applying the  $\chi^2/ndf$  cut. The decrease of the efficiency is lower than 5% (20%) for prompt  $D^+$  and up to 45% (60%) for feed-down  $D^+$  mesons for the central (tightest) value of the  $\chi^2/ndf$  cut. These variations show that the rejection of the feed-down contribution achieved with the central value of the  $\chi^2/ndf$  cut is similar to that obtained with the tightest cut value tested, but it allows to keep more signal as discussed in the previous section. Furthermore, the relative variation of the efficiency for feed-down  $D^+$  is strongly dependent on the  $D^+$  transverse momentum, because of the shape of the  $\chi^2$  distribution, which becomes wider with increasing  $D^+$  transverse momentum.

### $D^+$ prompt fraction variation

The  $D^+$  prompt fraction has been evaluated with the FONLL-based methods,  $N_b$  and  $f_c$ . The central value is given by the  $N_b$  method, while the  $f_c$  method is used to evaluate the systematic uncertainty, as explained in section 4.2.1. The resulting prompt fraction is larger than 0.9 for each  $\chi^2/ndf$  cut value tested and

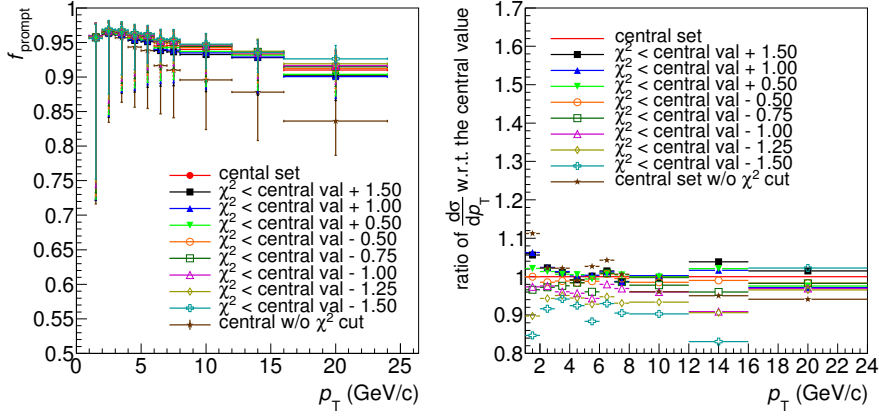


Figure 7.11. Left:  $D^+$  prompt fraction evaluated with the FONLL-based methods for each  $\chi^2/ndf$  cut value tested. Right: relative variation of the  $p_T$ -differential cross section obtained for each  $\chi^2/ndf$  cut value tested with respect to the central value.

slightly dependent on the  $D^+$   $p_T$ . As illustrated in the left panel of figure 7.11, the subtraction of the feed-down contribution achieved with the cut on the  $\chi^2/ndf$  is significant at high  $p_T$ , while negligible at low  $p_T$ . The variation of the  $D^+$  prompt fraction among the different cut values tested is small. Furthermore, the improved feed-down subtraction provides also a reduction of the uncertainties which affect the  $D^+$  prompt fraction.

### $p_T$ -differential cross section variation

If the distributions of  $\chi^2/ndf$  were equal in data and MC, then the  $p_T$ -differential cross section should not vary changing the value of the  $\chi^2/ndf$  cut. In the right panel of figure 7.11 is shown the ratio of the  $p_T$ -differential production cross section of prompt  $D^+$  for each trial, computed using equation 4.8, with respect to the one obtained with the central  $\chi^2/ndf$  cut value. From this plot it is possible to see that the  $p_T$ -differential cross section decreases with tightening of the  $\chi^2/ndf$  cut, indicating a possible discrepancy between in data and MC in the region where the distributions are steeper ( $\chi^2/ndf \lesssim 2$ ). This effect was taken into account in the evaluation of the systematic uncertainties (see section 7.4.1). However, for the central values chosen, this bias is not observed or, in any case, very small.

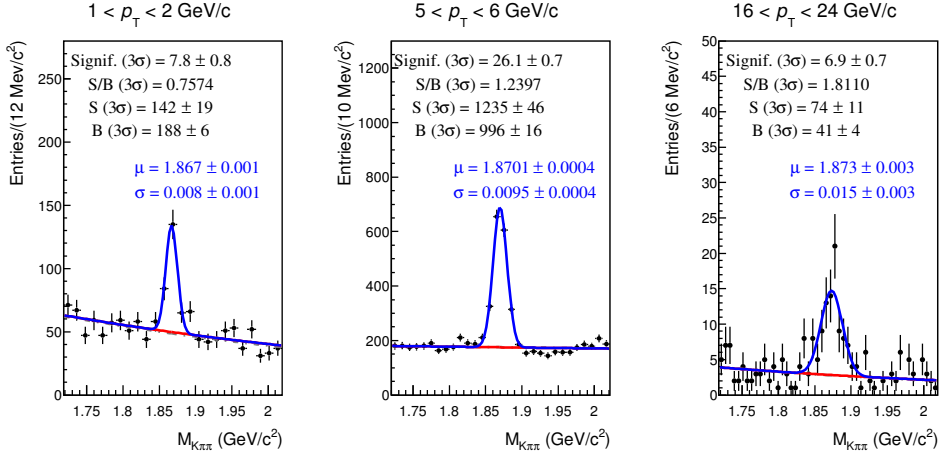


Figure 7.12. Invariant-mass fits in the transverse momentum intervals  $1 < p_T < 2$  GeV/c,  $5 < p_T < 6$  GeV/c and  $16 < p_T < 24$  GeV/c obtained applying the topological cuts reported in table 7.1 and the PID selection.

### 7.3.3 Raw yield extraction

In order to compute the central value of the  $p_T$ -differential cross section, the raw yields have been extracted from the invariant-mass spectra obtained applying the topological cuts summarised in table 7.1 and the PID selection on the daughter tracks. The fitting function is composed by a Gaussian function for the signal (prompt and feed-down  $D^+$ ) and an exponential function, for the parametrisation of the combinatorial background. Both the Gaussian mean and width are left free in the fit. In figure 7.12 three examples of invariant-mass fits for different  $p_T$  intervals are shown, while the others can be found in the appendix A.6. At low  $p_T$  this is due to the strong PID strategy applied necessary to extract the signal because of the large combinatorial background, while at high  $p_T$  because of the low statistics. In figure 7.13 the values of the Gaussian width (left) and mean (right) are compared to those obtained fitting the invariant-mass distributions from Monte Carlo simulations. The peak mean is also compared to the PDG value of the  $D^+$  mass. Concerning the peak width, the values are all compatible within 2 sigmas, except for the  $5 < p_T < 6$  GeV/c and  $8 < p_T < 12$  GeV/c, while for the mean only the value obtained in the interval  $4 < p_T < 5$  GeV/c is not in agreement within the uncertainty. The discrepancy between the values of the parameters extracted from the data and MC simulations will be taken into account in the evaluation of the systematic uncertainties. In the lowest and the highest  $p_T$  intervals the statistical



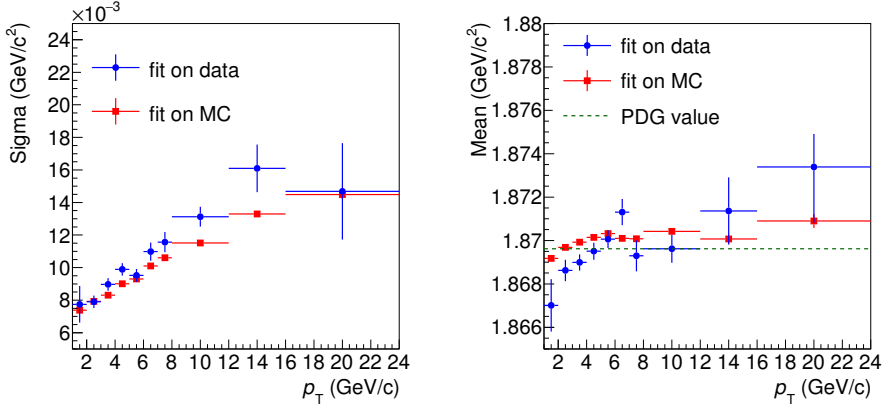


Figure 7.13. Comparison between the peak width (left) and mean (right) obtained by fitting the invariant-mass spectra obtained from the data and the MC simulation. The peak position is also compared to the PDG value of the  $D^+$  mass.

significance is lower with respect to the other  $p_T$  bins. The extracted raw yield, the signal-to-background ratio and the significance within  $3\sigma$  obtained in all the  $p_T$  intervals of the analysis are reported in table 7.2.

$p_T$ (GeV/c)	<i>Signal</i>	<i>S/B</i> ( $3\sigma$ )	<i>Significance</i> ( $3\sigma$ )
[1,2]	$142 \pm 19$	0.7574	$7.8 \pm 0.8$
[2,3]	$1432 \pm 76$	0.3376	$19.0 \pm 0.9$
[3,4]	$2249 \pm 85$	0.4970	$27.3 \pm 0.9$
[4,5]	$1842 \pm 65$	0.7851	$28.5 \pm 0.8$
[5,6]	$1235 \pm 46$	1.2397	$26.1 \pm 0.7$
[6,7]	$735 \pm 35$	1.3602	$20.6 \pm 0.7$
[7,8]	$503 \pm 26$	1.7449	$17.9 \pm 0.6$
[8,12]	$910 \pm 38$	1.4563	$23.2 \pm 0.7$
[12,16]	$201 \pm 16$	2.1726	$11.7 \pm 0.6$
[16,24]	$74 \pm 11$	1.8110	$6.9 \pm 0.7$

Table 7.2. Raw yield, signal-to-background ratio and significance obtained from the fit to the invariant-mass distributions obtained by reconstructing the  $D^+$  candidates with the KFParticle package.

## 7.4 Systematic uncertainties

The sources of systematic uncertainties evaluated for this analysis are the following:

1. Systematic uncertainty due to the topological selection efficiency.
2. Systematic uncertainty due to the raw yield extraction.
3. Systematic uncertainty due to the generated  $p_T$  shape of  $D^+$  and B mesons.
4. Systematic uncertainty due to the PID selection efficiency.

Moreover, for the computation of the  $p_T$ -differential production cross section of prompt  $D^+$ , other four sources have been considered:

1. Systematic uncertainty due to the beauty feed-down subtraction.
2. Systematic uncertainty due to the tracking efficiency.
3. Systematic uncertainty on the integrated luminosity.
4. Systematic uncertainty on the branching ratio.

In particular, the systematic uncertainty due to the beauty feed-down subtraction was reduced with respect to the standard analysis, thanks to the selection on the  $\chi^2/ndf$  of the reconstructed  $D^+$  candidates, evaluated after applying the topological constraint at the primary vertex. The comparison between the two results will be presented in section 7.5. For the latter three sources the assigned systematic uncertainties are same as those in the data-driven cut-variation analysis: 9% systematic uncertainty on the tracking efficiency, 3.7% on the integrated luminosity and 2.1% on the branching ratio.

### 7.4.1 Systematic uncertainty due to the topological selection efficiency

As already mentioned, an imperfect description of the distributions of the variables used for the topological selections in the Monte Carlo simulation would induce a bias in the computation of the  $p_T$ -differential cross section. To evaluate a possible systematic effect, the analysis was repeated varying the cut values of the variables used. In particular, for each variable, several looser and tighter selections have been applied, leaving the others unchanged. In the left panel of figure 7.14 the ratio of the  $p_T$ -differential cross section with respect to the central value, obtained varying the cut value on the normalised decay length is shown. Also the variations of the cut value on the  $\chi^2/ndf$  of the reconstructed  $D^+$  candidates after applying the topological constraint (see figure 7.11), performed in order to optimise that

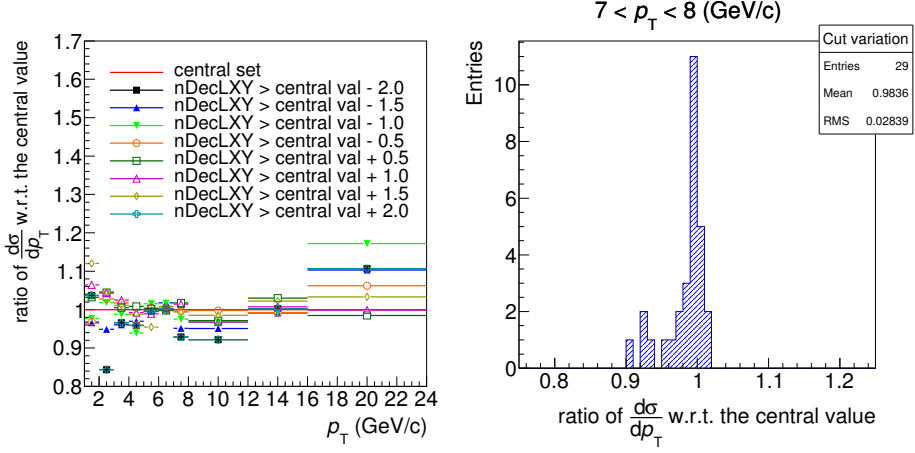


Figure 7.14. Left: relative variation of the  $p_T$ -differential cross section obtained varying the cut on the normalised decay length  $xy$  as a function of  $p_T$ . Right: distribution of the ratio of the  $p_T$ -differential cross section varying all the topological cuts in the interval  $7 < p_T < 8$  GeV/c.

selection, were considered for the evaluation of the systematic uncertainty. The systematic uncertainty was assigned considering the RMS and shifts with respect to unity of the mean of the distributions of all the variations. In the right panel of figure 7.14 an example for  $7 < p_T < 8$  GeV/c is shown. The systematic uncertainty assigned decreases from 10% at low  $p_T$  up to 5% at high  $p_T$ . All the values are summarised in table 7.3.

#### 7.4.2 Systematic uncertainty due to the raw yield extraction

The systematic uncertainty due to the raw yield extraction was evaluated with a multi-trial approach, repeating the fits to the invariant-mass distributions, considering all the possible combinations of the following fit configurations:

- The Gaussian width was first left free and then fixed to the value obtained from the MC simulation.
- The bin width of the invariant-mass distributions was changed from 4 to 16 MeV/c<sup>2</sup>.
- The range of the invariant-mass fit has been varied using 6 different lower and upper limits.

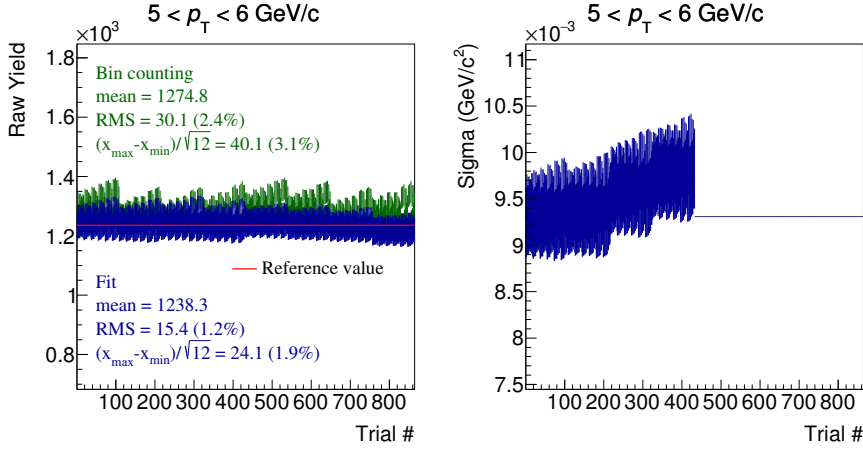


Figure 7.15. Raw yield obtained via the fit to the invariant-mass distributions and the bin counting method (left) and the Gaussian width (right) as a function of the trial number for  $D^+$  candidates reconstructed with the KFParticle package with  $5 < p_T < 6$  GeV/c.

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

For each trial, the amount of signal was extracted from the invariant-mass distributions using an alternative method called *bin counting*, which consists in counting the entries in the invariant-mass histogram within  $3\sigma$  around the peak centre after subtracting the background estimated via the fit. In figure 7.15 the raw yield extracted by integrating the Gaussian function and with the bin counting method (left) and the Gaussian width (right) are shown as a function of the trial number in the interval  $5 < p_T < 6$  GeV/c. The values of Gaussian width without statistical error in the right plot correspond to the trials in which the sigma was fixed to the value obtained by fitting the invariant-mass distributions from the MC simulation. The systematic uncertainty has been assigned considering the RMS of the distributions of the raw yield extracted with the two methods. In case of a not Gaussian distribution, also the variance assuming a flat distribution was evaluated. Finally, since the values obtained with the bin counting are systematically slightly higher than those extracted from the fit, also the difference between the mean values of the two distributions was taken into account. The values assigned as systematic uncertainty are reported in table 7.3.

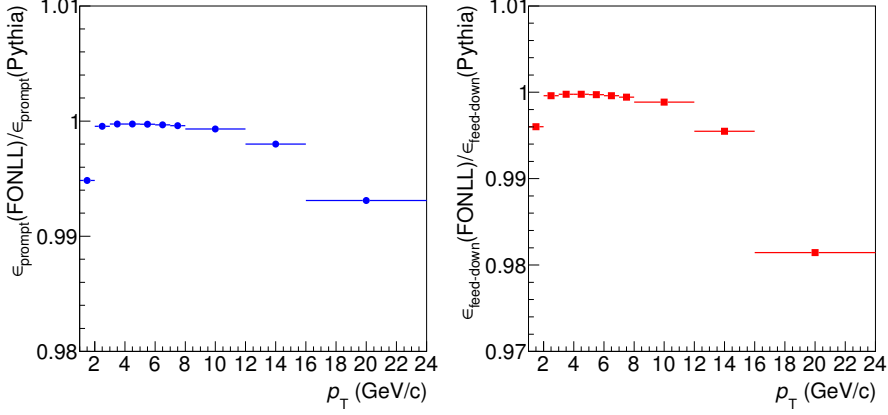


Figure 7.16. Ratio of the efficiencies for prompt (left) and feed-down (right)  $D^+$  obtained with the FONLL and the Pythia  $p_T$ -shapes.

### 7.4.3 Systematic uncertainty due to the $p_T$ shape

To compute the  $p_T$ -differential production cross section of prompt  $D^+$ -mesons, the efficiency of prompt  $D^+$  has been reweighted in order to reproduce the shape of the FONLL  $p_T$  spectrum. That was done also for the efficiency of feed-down  $D^+$ , since it is needed for the computation of the fraction of prompt  $D^+$  with the  $N_b$  and  $f_c$  methods (see section 4.2.1). The weights used to reproduce the FONLL  $p_T$  spectrum for feed-down  $D^+$  have been obtained considering the B-mesons which decay to  $D^+$ , as explained in section 5.4.2. On the other hand, the efficiencies obtained with the Pythia  $p_T$  shape of generated prompt and feed-down  $D^+$  has been considered for the evaluation of the systematic uncertainties. In the left panel of figure 7.16 the ratio between the efficiencies obtained with the two different  $p_T$  shapes is shown. It has been used to assign the systematic uncertainty, which reported in table 7.3.

### 7.4.4 Systematic uncertainty due to the PID efficiency

Since the code for the PID selection was re-implemented for this analysis, the possible systematic uncertainty induced by the PID selection efficiency was re-evaluated for cross-checking. In the left panel of figure 7.17 an example of fit to the invariant-mass distribution for  $D^+$  candidates with  $4 < p_T < 5$  GeV/c obtained with and without applying PID selection is shown. All the topological selections reported in table 7.1 are applied in both cases. In order to avoid as much as

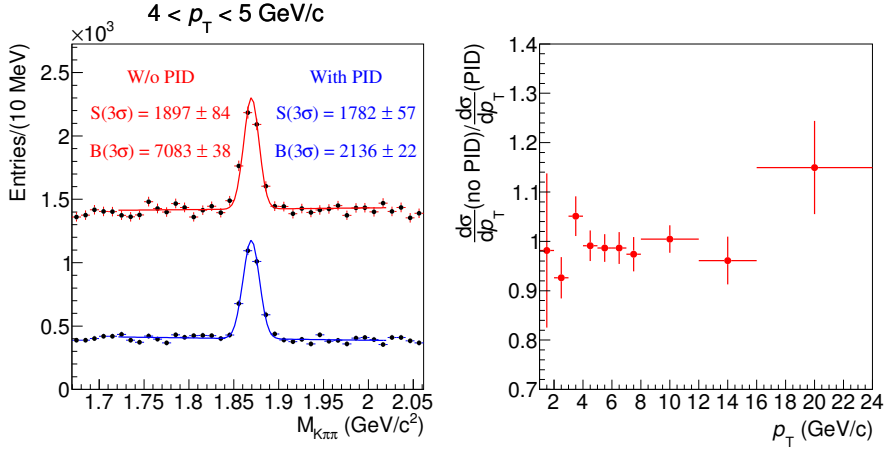


Figure 7.17. Left: comparison between the fit to the invariant-mass distribution obtained with (blue) and without (red) applying the PID selection in the interval  $4 < p_T < 5$  GeV/c. The Gaussian width is fixed to the value obtained from the MC simulation. Right: ratio of the  $p_T$ -differential cross section of prompt  $D^+$  obtained with and without applying the PID selection.

possible the statistical fluctuations for this test the Gaussian width in the fit has been fixed to the value obtained by fitting the invariant-mass distribution from the MC simulation. The amount of signal extracted with and without applying the PID selection is of about 6%, while the amount of compinatorial background is reduced of more than a factor 3 if the PID selection is applied to the tracks of the triplets. The right panel of figure 7.17 shows the ratio between the  $p_T$ -differential cross section obtained with and without applying the PID selection. Since the  $D^+$  candidates obtained with the PID selection are a sub-sample of those obtained without using the PID, the error on the ratio is assumed to be binomial. The ratio fluctuates around unity without any trend and therefore no systematic uncertainty has been assigned for the PID selection efficiency, as in the standard analysis.

## 7.5 Results

The main goal of this study was to achieve an improved subtraction of the feed-down  $D^+$ -mesons in the extracted raw yields. Hence, the result was compared to the one obtained applying the topological selections used in the standard analysis. Furthermore, since the selection of the  $D^+$  candidates was performed with a different vertex-reconstruction algorithm, the  $p_T$ -differential cross section was compared

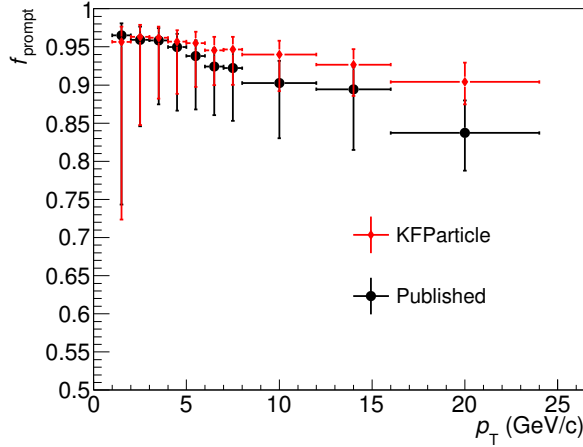


Figure 7.18. Comparison between the fraction of prompt  $D^+$  obtained using the KFParticle package for the reconstruction of the  $D^+$  candidates, applying the topological selections reported in table 7.1, and with the topological selections used in the standard analysis.  $f_{\text{prompt}}$  is computed with the FONLL-based methods in both cases.

to the published result, in order to check the consistency of the analysis.

### 7.5.1 Feed-down $D^+$ -meson subtraction

The fraction of prompt  $D^+$  obtained using the KFParticle for the reconstruction of the  $D^+$  candidates and applying the selections reported in table 7.1 is reported in figure 7.18, compared to the one obtained by applying the cuts used in the published

$p_T$ (GeV/c)	[1,2]	[2,3]	[3,4]	[4,5]	[5,6]	[6,7]	[7,8]	[8,12]	[12,16]	[16,24]
cut var	10%	10%	6%	6%	6%	5%	5%	5%	5%	5%
raw yield	10%	8%	5%	5%	5%	5%	5%	5%	8%	8%
$p_T$ shape	1%	negl.	negl.	negl.	negl.	negl.	negl.	negl.	1%	2%
PID	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.	negl.

Table 7.3. Summary of the systematic uncertainty for the analysis performed by reconstructing the  $D^+$  candidates with the KFParticle package.

$p_T$ (GeV/c)	[1,2]	[2,3]	[3,4]	[4,5]	[5,6]	[6,7]	[7,8]	[8,12]	[12,16]	[16,24]
KFParticle (%)	+2 -22	+2 -12	+2 -9	+2 -9	+2 -8	+3 -7	+3 -7	+3 -8	+3 -9	+5 -6
Published (%)	+2 -24	+2 -12	+2 -8	+2 -7	+2 -6	+2 -5	+2 -5	+2 -5	+2 -5	+3 -3

Table 7.4. Comparison between the total uncertainty on the  $D^+$  prompt fraction obtained using the KFParticle package for the reconstruction of the  $D^+$  candidates and the one obtained in the standard analysis.

result (see table 4.2). Both of them have been computed using the  $N_b$  method for the central value and the  $f_c$  to estimate a systematic uncertainty, as explained in section 4.2.1. The application of the topological constraint at the primary vertex allowed to achieve a larger fraction of prompt  $D^+$  in the raw yield especially at high  $p_T$ , as shown in figure 7.1. The larger value of  $f_{prompt}$  provides also a reduction of the systematic uncertainty, mainly because for a fraction of prompt  $D^+$  close to unity, the values estimated with the  $N_b$  and  $f_c$  methods become more similar. The comparison between the systematic uncertainties on  $f_{prompt}$  for the two different sets of topological selections are reported in table 7.4.

### Comparison to the selection on the maximum difference between the measured and expected impact parameter of the daughter tracks

As discussed in section 6.2.1, another powerful quantity that allows one to discriminate between prompt and feed-down  $D^+$ -mesons is the maximum difference between the measured and expected impact parameter of the daughter tracks. In order to compare the effect of this variable to the one obtained with the  $\chi^2/ndf$  of the reconstructed  $D^+$  candidate after applying the topological constraint, the  $p_T$ -differential cross section has been recomputed replacing the cut on the  $\chi^2/ndf$  with that on the maximum difference between the measured and expected track impact parameter, set to  $[-2.5, 2.5]$  for all the  $p_T$  intervals of the analysis. In figure 7.20 the efficiencies of prompt and feed-down  $D^+$  (left) and the  $D^+$  prompt fraction (right) obtained with the two different cuts are compared. The achieved prompt fraction is very similar in the two cases, while the efficiency is slightly higher for both prompt and feed-down  $D^+$ -mesons if the cut on  $\chi^2/ndf$  is applied. The similar result obtained is justified by the correlation between the two quantities, as confirmed by the scatter plot of the maximum difference between the measured and expected impact parameter of the daughter tracks as a function of the  $\chi^2/ndf$  of the  $D^+$  candidates with  $3 < p_T < 4$  GeV/c, which is reported in the left panel of figure 7.19. The right panel of the same figure shows the ratio between the  $p_T$ -differential cross section of prompt  $D^+$  obtained applying the two different cuts.



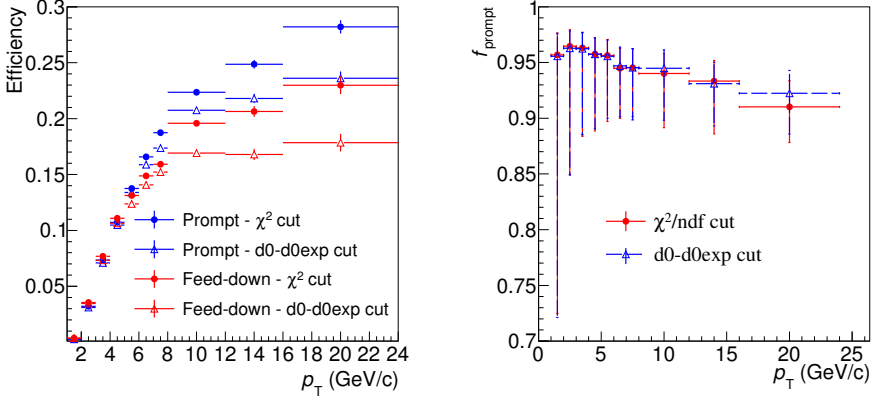


Figure 7.19. Comparison between the efficiencies of prompt and feed-down  $D^+$  (left) and the  $D^+$  prompt fraction (right) obtained applying the selection on the  $\chi^2/\text{ndf}$  and on the maximum difference between the measured and expected impact parameter of the daughter tracks.

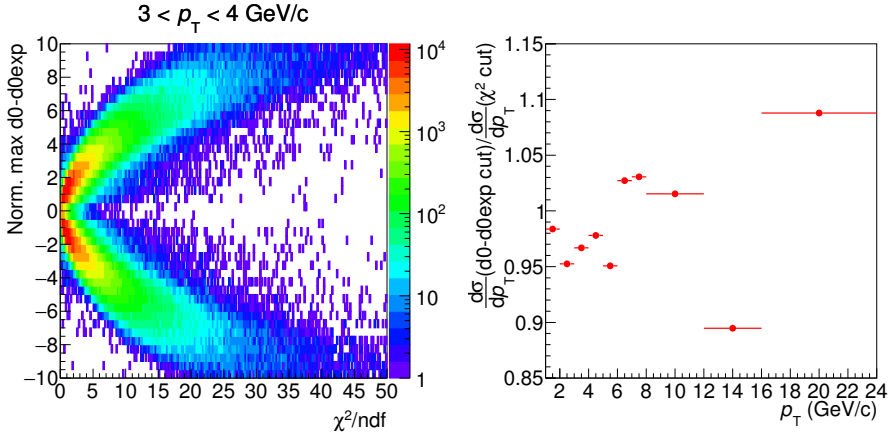


Figure 7.20. Left: scatter plot of the maximum difference between the measured and expected impact parameter of the daughter tracks vs. the  $\chi^2/\text{ndf}$  for the  $D^+$  candidates with  $3 < p_T < 4$  GeV/c. Ratio between the  $p_T$ -differential cross section of prompt  $D^+$  obtained applying the cut on the maximum difference between the measured and expected impact parameter of the daughter tracks and the one obtained applying the  $\chi^2/\text{ndf}$  cut.

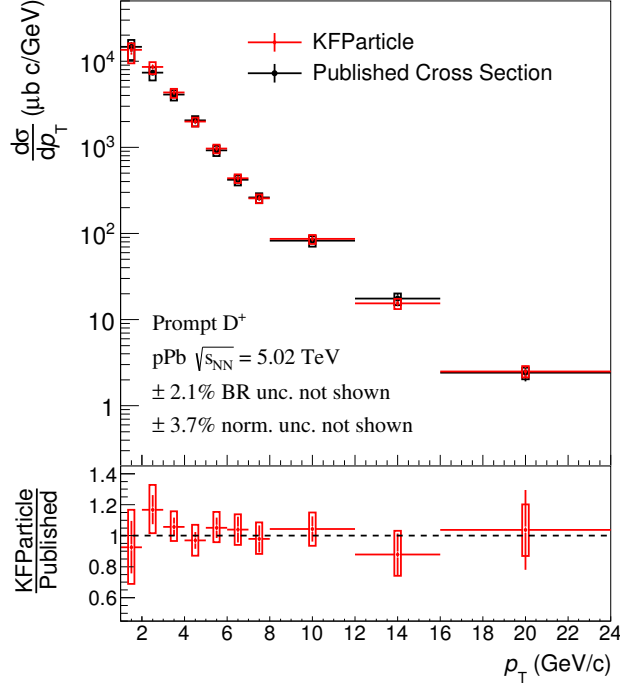


Figure 7.21. Comparison between the  $p_T$ -differential production cross section of prompt  $D^+$ -mesons in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV obtained using the KFPARTICLE package for the reconstruction of the  $D^+$  candidates compared to the published result [36].

This ratio fluctuates around unity without any particular trend, indicating that the application of one selection does not introduce a bias with respect to the result obtained applying the other one.

### 7.5.2 $p_T$ -differential production cross section of prompt $D^+$ -mesons

The  $p_T$ -differential production cross section of  $D^+$  mesons in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV obtained using the KFPARTICLE package for the reconstruction of the  $D^+$  candidates is depicted in figure 7.21, compared to the one obtained with the standard approach, which was published by the ALICE Collaboration [36]. The

central values have been computed using the formula reported in equation 4.8. The vertical bars represent the statistical uncertainty, while the boxes the systematic uncertainty, which was computed considering all the sources presented in section 7.4 as uncorrelated and therefore summed in quadrature. In the bottom panel of figure 7.21 is shown the ratio between the result obtained using the vertex-reconstruction algorithm based on the Kalman filter and the published one. The statistical and systematic uncertainties has been treated as uncorrelated in the ratio, except for the systematic uncertainty due to the tracking efficiency and the beauty feed-down subtraction which have been considered as fully correlated. However, some degree of correlation is also present among the other sources of uncertainties, since the sub-samples of  $D^+$  candidates selected for the two analysis are not completely disjoint. Therefore, the uncertainty on the ratio shown in the bottom panel of figure 7.21 is slightly overestimated. Despite that, the ratio is found to fluctuate around unity, showing that the two results are in agreement between each other and no systematic trend is observed between the two results.



# 8

## Conclusions

This thesis was focused on the measurement of the production of  $D^+$ -mesons in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the ALICE experiment at LHC. All the analyses presented have been performed exploiting the full reconstruction of the  $D^+$ -mesons in the purely hadronic decay channel  $D^+ \rightarrow K^- \pi^+ \pi^+$ . In particular, three different techniques for the measurement of the prompt and feed-down contributions to the  $D^+$ -meson production have been developed. They are respectively based on:

- The unbinned log-likelihood fit to the  $D^+$  candidates impact-parameter distributions.
- The usage of different sets of selection criteria on the  $D^+$  decay topology.
- The reconstruction of the  $D^+$  candidates performed with a different vertex-reconstruction algorithm based on the Kalman Filter.

Figure 8.1 the  $p_T$ -differential production cross section of  $D^+$ -mesons obtained with the three different methods, compared to the published one. The vertical bars represent the statistical uncertainties, while the boxes the systematic ones. All the results are found to be compatible among each other within the experimental uncertainties. The first two methods have the advantage to be fully data-driven and therefore not to rely on perturbative QCD calculations for the determination of the fraction of prompt  $D^+$ . These measurements will be improved starting

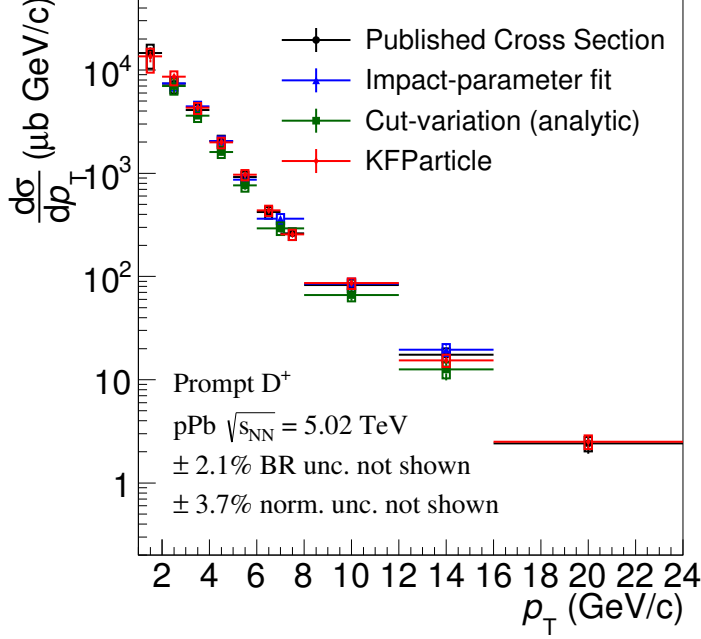


Figure 8.1. Comparison among the  $p_T$ -differential production cross section of prompt  $D^+$ -mesons in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV obtained with all the analysis performed in this thesis and the one published by the ALICE Collaboration [36].

from the Run II at LHC thanks to larger data samples expected and even more with the upgrade of the Inner Tracking System of ALICE, planned for the year 2019, because of the improved vertex-position and impact-parameter resolutions. On the other hand, the Kalman-filter based analysis allowed to subtract significantly the feed-down  $D^+$ . Furthermore, the vertex-reconstruction package based on the Kalman filter could also be used for other analyses, such as the measurement of stranded hadrons (e.g.  $K_s^0$ ,  $\Lambda$  and  $\Xi$ ) or hypernuclei, which have a larger mean proper decay length (e.g.  $(c\tau)_\Lambda = 7.89$  cm [56]) and therefore the curvature of their trajectory due to the magnetic field cannot be neglected.