

Studies of the $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$ decay using 2015 and 2016 LHCb data

Anthony Correia¹

Superviser: Dr. Donal Hill

Professor: Prof. Olivier Schneider

February 21, 2021



Abstract

Comparing the experimental value of $R(D^*) = \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} l^+ \nu_l)}$ to the Standard Model expectation provides a strong test of Lepton Flavour Universality. The branching ratio $\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$ can be determined using the hadronic τ decay modes $\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{\nu}_\tau$ and $\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau$, employing the $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$ decay as a normalisation mode. In this work, the $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$ yield is measured using data collected at $\sqrt{s} = 13$ TeV with the LHCb experiment in 2015 and 2016, corresponding to an integrated luminosity of 2 fb^{-1} .

A fit is performed to the corrected mass $m(D^{*-} \pi^+ \pi^+ \pi^-) - m(D^*) + m(D^*)_{PDG}$, with $m(D^*)_{PDG}$ the world-average value of the D^{*-} meson invariant mass. The normalisation mode yield is determined from this fit. A Gradient Boosted Decision Tree is trained to classify signal and combinatorial background candidates, allowing the relative statistical uncertainty on the normalisation mode yield to be minimised by decreasing the combinatorial background level. The systematic uncertainty due to the use of fixed parameters in the invariant mass fit is also evaluated. The normalisation yield measured is $n(B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-) = 52,322 \pm 279 \text{ (stat.)} \pm 876 \text{ (syst.)}$, where the first uncertainty is statistical and the second systematic.

¹Second year Physics Master, EPFL, Lausanne, Switzerland
anthony.correia@epfl.ch

Contents

1	Introduction	3
1.1	$R(D^*)$ status	3
1.2	Hadronic $R(D^*)$ ratio	3
1.3	The LHCb detector	4
1.4	$B^0 \rightarrow D^*-3\pi$ decay at LHCb	5
2	Datasets and cuts	6
2.1	LHCb and simulated data	6
2.2	$m(D^*)$ -constrained invariant mass	6
3	Fit to the LHCb data	7
3.1	Definitions and notations	8
3.2	Fit to the MC dataset	8
3.3	Fit to $B^0 \rightarrow D^*-K^+\pi^+\pi^-$ RapidSim sample	9
3.4	Fit to $B^0 \rightarrow D^*- (D_s^+ \rightarrow 3\pi)$	9
3.5	Global fit to the $m(D^*)$ -constrained $m(D^*3\pi)$	11
4	Multivariate analysis	13
4.1	Training	13
4.2	Testing	14
4.3	Application to the data	15
4.4	BDT cut optimisation	16
5	Optimised signal yield with systematic uncertainties	18
6	Conclusion	20
7	Annexe	21
7.1	Theory	21
●	<i>Definition of geometrical and kinetic variables</i> - - - - -	21
●	<i>s Weights</i> - - - - -	21
●	<i>Illustration of a multi-variate analysis</i> - - - - -	23
●	<i>Two-sample Kolmogorov-Smirnov test</i> - - - - -	25
7.2	Results	26
●	<i>Distribution of the training variables in the signal and background samples of the BDTs</i> - - - - -	26
●	<i>Rankings by importance of the variables used in the BDT training</i> - - - - -	28
●	<i>Fits to the LHCb data with the optimal BDT > -1.25 cut</i> - - - - -	28
●	<i>Distribution of the parameters fixed in the fit to $m(D^*3\pi)$ in the BDT > -1.25 LHCb data under systematic variations</i> - - - - -	30
7.3	Python codes	32
	References	32

1. Introduction

1.1 $R(D^*)$ status

The $R(D^*)$ branching fraction ratio is defined as

$$R(D^*) = \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} \ell^+ \nu_\ell)} \quad \text{with } \ell \in \{e, \mu\}, \quad (1)$$

where the ratio of semileptonic B^0 decays to τ and light lepton (ℓ) final states is taken. The Feynman diagram for such decays is shown in [Figure 1](#), where a $b \rightarrow c$ quark transition involving the production of a W gauge boson occurs, followed by a leptonic decay of the W boson.

The quantity $R(D^*)$ directly tests the validity of Lepton Flavour Universality (LFU) in the Standard Model (SM), which asserts that decays involving different lepton generations are equally probable (aside from differences in branching fraction due to phase space). The $R(D^*)$ ratio measures the rate at which $b \rightarrow c \tau \nu$ transitions occur relative to $b \rightarrow c \ell \nu$ transitions, and is precisely predicted in the SM to be 0.258 ± 0.005 [12]. Experimental measurements of $R(D^*)$ can be compared to this theoretical prediction to test the validity of LFU in the SM; a significant deviation from the SM expectation would be a clear signature of physics beyond the SM.

$R(D^*)$ has already been measured with the 2011-2012 LHCb data (corresponding to an integrated luminosity of 3 fb^{-1}) using two specific τ decay modes: (1) the muonic decay $\tau^+ \rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau$ [7], and (2) the hadronic mode $\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{\nu}_\tau$ and $\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau$ [8]. Both types of measurement result in higher values of $R(D^*)$ than the SM prediction; the results are summarised in [Table 1](#).

Combining all experimental measurements of $R(D)$ and $R(D^*)$ performed with different experiments (BaBar, Belle, and LHCb), the combined $R(D)$ and $R(D^*)$ discrepancy with the SM expectation is 3.1 standard deviations [12].

This may be a hint of LFU violation and Physics Beyond the Standard Model (BSM), but equally could be due to experimental systematic effects that are not currently understood. Analyses with more data are required to clarify the picture, and either resolve the SM discrepancy or firmly establish the discrepancy as clear evidence of BSM physics. To this end, an analysis using the full 9 fb^{-1} LHCb dataset is strongly motivated.

τ decay mode	Muonic	Hadronic
Decays	$\tau^+ \rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau$	$\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{\nu}_\tau$
Branching ratio (%) [17]	17.39 ± 0.04	9.31 ± 0.05
$R(D^*)$	$0.336 \pm 0.027(\text{stat.}) \pm 0.030(\text{syst.})$	$0.291 \pm 0.019(\text{stat.}) \pm 0.026(\text{syst.}) \pm 0.013$
Agreement with SM prediction	2.1σ	1.1σ

Table 1. $R(D^*)$ measurements with the 2011-2012 LHCb data for the muonic [7] and hadronic [8] τ decay modes. The final uncertainty quoted for the hadronic result is due to the use of external branching fraction information.

1.2 Hadronic $R(D^*)$ ratio

In this project, the hadronic τ decay mode is considered. In order to measure the $R(D^*)$ ratio, the branching ratio $\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)$ can be taken from the Particle Data Group (PDG) [17]. However, the branching ratio $\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$, using the τ^+ lepton decaying to the hadronic $\pi^+ \pi^+ \pi^- \bar{\nu}_\tau$ and $\pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau$ modes, must be measured. To measure this branching fraction with minimal systematic uncertainties, it is sensible to choose a normalisation mode with the same visible final state as the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decay. To this end, the chosen normalisation mode is $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$, denoted $B^0 \rightarrow D^{*-} 3\pi$ for brevity. The Feynman diagram for this decay is shown in [Figure 2](#). The

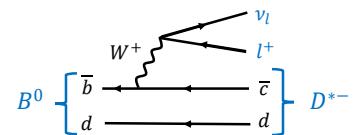


Figure 1. Feynman diagram of $B^0 \rightarrow D^{*-} l^+ \nu_l$, with $l \in \{e, \mu, \tau\}$.

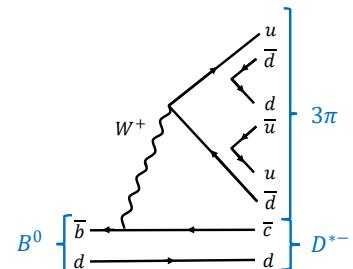


Figure 2. Feynman diagram of the $B^0 \rightarrow D^{*-} 3\pi$ decay.

following ratio must be measured

$$\mathcal{K}(D^*) = \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} 3\pi)} = \frac{\frac{n_{B^0 \rightarrow D^{*-} \tau^+ \nu_\tau}}{\varepsilon_{B^0 \rightarrow D^{*-} \tau^+ \nu_\tau}}}{\frac{n_{B^0 \rightarrow D^{*-} 3\pi}}{\varepsilon_{B^0 \rightarrow D^{*-} 3\pi}}} \times \frac{1}{\mathcal{B}(\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{\nu}_\tau) + \mathcal{B}(\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau)}, \quad (2)$$

where $n_{B^0 \rightarrow D^{*-} \tau^+ \nu_\tau}$ is the yield of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decay with the hadronic τ decay mode, and $n_{B^0 \rightarrow D^{*-} 3\pi}$ is the normalisation mode yield. The terms $\varepsilon_{B^0 \rightarrow D^{*-} \tau^+ \nu_\tau}$ and $\varepsilon_{B^0 \rightarrow D^{*-} 3\pi}$ are efficiency factors computed from Monte-Carlo (MC) simulation, which account for the relative efficiency with which the signal and normalisation modes are reconstructed and selected. The values of $\mathcal{B}(\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{\nu}_\tau)$ and $\mathcal{B}(\tau^+ \rightarrow \pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau)$ are taken from the PDG [17]. The signal branching fraction $\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$ is thus given by

$$\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau) = \mathcal{K}(D^*) \times \mathcal{B}(B^0 \rightarrow D^{*-} 3\pi), \quad (3)$$

where the normalisation mode branching fraction is taken from the PDG [17]. The objective of this project is to measure the yield of the normalisation mode $n_{B^0 \rightarrow D^{*-} 3\pi}$, which will be used as input to the calculation of $\mathcal{K}(D^*)$. The measurement is performed using 2015 and 2016 LHCb data, but in future this will be extended to consider the full LHCb Run 1 + 2 dataset as part of the ongoing $R(D^*)$ analysis.

1.3 The LHCb detector

The LHC (Large Hadron Collider) is the largest and most energetic particle accelerator in the world, with a circumference of 27 km and located near the border between France and Switzerland [2]. Two beams of protons are accelerated to an energy of 13 TeV (in the second run of the LHC between 2015 and 2018), and collided in order to produce large numbers of heavy and short-lived particles.

The LHCb (Large Hadron Collider beauty) detector is one of the four main detectors located around the LHC ring. It specialises in measurements of heavy-flavour hadron decays in the forward pseudorapidity range $2 < \eta < 5$, with a particular focus on hadrons containing beauty (b) quarks. The data analysed in this report were collected with the LHCb experiment during 2015 and 2016, and correspond to an integrated luminosity of 2 fb^{-1} . A schematic illustration of the LHCb detector is shown in [Figure 3](#).

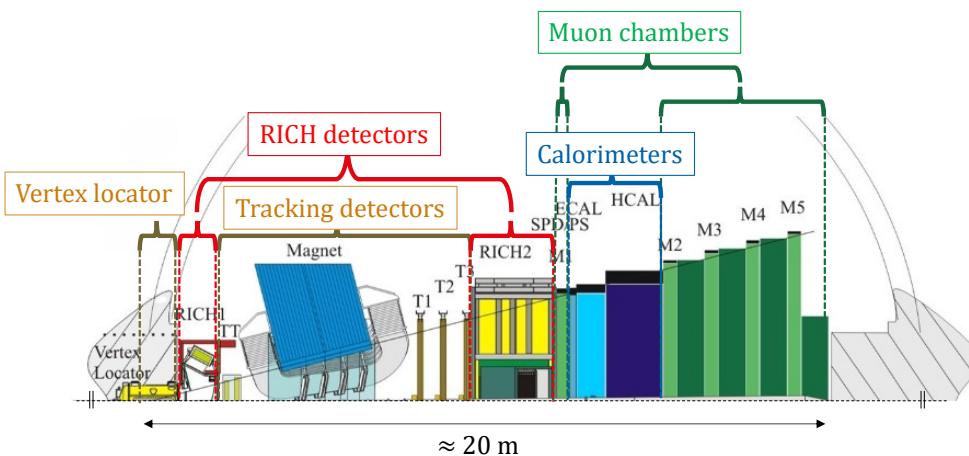


Figure 3. Lateral view of the LHCb detector (adapted from [11]).

The **vertex locator** (VELO) surrounds the proton-proton (pp) collision point, separated from the beam by only 7 mm during data-taking [6]. The VELO is used to measure the decay vertices of particles produced in the pp collisions with a resolution of $O(10 \mu\text{m})$. This detector feature is of particular importance for reconstructing heavy flavour hadron decays, since such particles are long-lived enough to travel a few millimetres before decaying. It is vital to distinguish the decay vertices of these particles from the pp collision point, which is made possible by the excellent VELO resolution.

Stable charged particles produced in heavy flavour decays, such as pions, kaons, and muons, are then tracked by the **Tracker Turicensis** (TT) and the **T-stations** T1, T2 and T3 that “sandwich” a **magnet** to bend their tracks. This allows their momenta to be measured with a relative uncertainty less than 8 per mille [3].

Before and after the tracking system, **Ring-Imaging Cherenkov** (RICH) detectors identify charged hadrons making use of their Cherenkov light. RICH1 (RICH2) is dedicated to low-momentum (high-momentum) particles. After the RICH2 detector, a **Scintillator Pad Detector** (SPD) identifies charged particles and a **Pre-Shower detector** (PS) identifies electromagnetic particles (i.e. electrons and photons) [16]. Then, an **electromagnetic (hadronic) calorimeter** detects and measures the energy of photons and electrons (hadrons) through their interaction with a dense medium. Finally, the outer detector layers form the **muon chambers**, which act to detect muons which predominantly pass through the other detector layers due to their weak interactions with the detector material [23].

1.4 $B^0 \rightarrow D^{*-} 3\pi$ decay at LHCb

As shown in [Figure 4](#), the B^0 particle is produced directly in the pp collision, and decays to a D^{*-} meson and three charged pions inside the vertex locator. The D^{*-} meson decays mainly into a D^0 meson and a π^+ meson (with a branching ratio of 64.7% [17]) via the strong interaction, and hence has a very short lifetime and flight distance that is not resolved by the VELO. The D^0 meson then decays into a kaon K^- and pion π^+ (with a branching ratio of 3.93% [17]).

The six charged hadrons are bent by the magnet and tracked by the tracking detectors, enabling their charge and momentum to be measured. Then, they are absorbed by the HCAL. Thanks to the Particle Identification (PID) RICH detectors, a mass hypothesis m is assigned to each track and the corresponding energy is computed using $E = \sqrt{p^2 + m^2}$. The momenta and energies of the parent particles are computed using four-momentum sums of the reconstructed tracks.

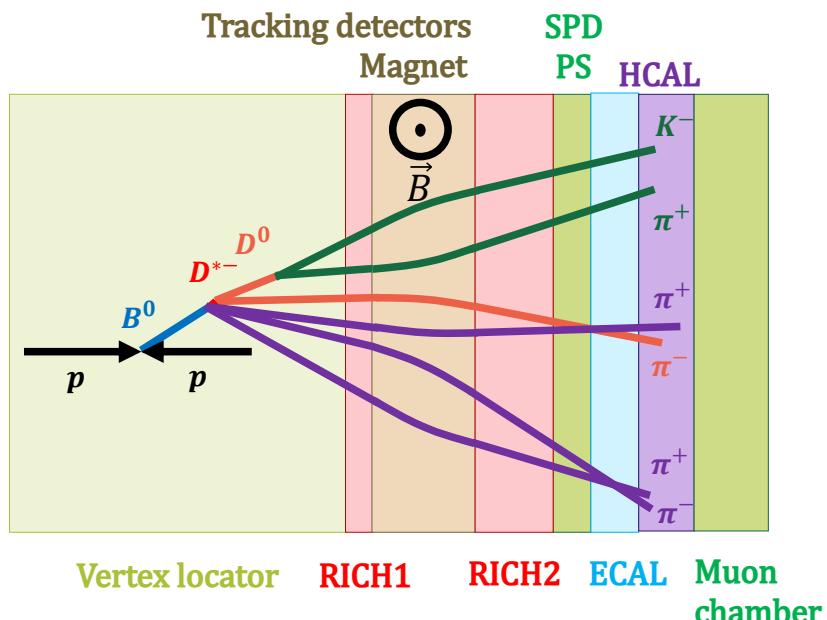


Figure 4. Tracks of the stable particles produced in the $B^0 \rightarrow D^{*-} 3\pi$ decay.

2. Datasets and cuts

2.1 LHCb and simulated data

The 2015–2016 LHCb data has been acquired with a set of online cuts applied by the hardware trigger (L0) and the software high-level triggers (HLT1 and HLT2). These triggers are designed to select the most interesting decays before storing them on disk, which is necessary since the collision rate at LHCb (about 13 MHz within the detector acceptance [4]) is much higher than the rate at which events can be persisted (about 12.5 kHz from 2015 to 2018 [1]). The hardware trigger selects events with high deposited transverse energy E_T in electromagnetic and hadronic calorimeters (ECAL and HCAL). The HLT1 partially reconstructs the events, and the HLT2 performs a more complete event reconstruction to select decays of interest, applying requirements on transverse momenta, impact parameters, and decay topologies [5].

After the online selection, a set of offline cuts on quantities such as invariant mass, vertex and track quality, momentum, and impact parameter, are applied in order to reject background from random combinations of tracks. These cuts improve the purity of the dataset prior to the offline analysis described in this report, and are designed to be highly efficient on genuine signal.

A simulated sample of $B^0 \rightarrow D^{*-} 3\pi$ decays passing the same selection requirements as data is used to study the signal invariant mass distribution and the performance of additional selection cuts. To align the cuts applied to the MC and data, an additional requirement on $\Delta M = m(D^*) - m(D^0)$ must be applied to data, namely $143 \text{ MeV}/c^2 < \Delta M < 148 \text{ MeV}/c^2$, as shown in Figure 5. This cut is used to select a sample that is highly pure in D^* mesons. Because of the small phase space available to pions produced in the $D^{*+} \rightarrow D^0 \pi^+$ decay, genuine D^* decays peak strongly in ΔM . The observed ΔM peak in Figure 5 is narrow because the invariant mass resolution is contributed to only by the pion momentum resolution; the contributions from the D^0 decay product resolutions vanish when subtracting $m(D^0)$ from $m(D^*)$.

2.2 $m(D^*)$ -constrained invariant mass

The $m(D^*)$ distribution of the LHCb data, shown in Figure 6, is clearly dominated by the true D^* contribution; the edge effects visible at low and high $m(D^*)$ values are due to the ΔM window cut described above. The width of the $m(D^*)$ distribution is mainly due to the detector resolution. Hence, it is reasonable to fix $m(D^*)$ to the well-known PDG value $m(D^*)_{PDG} = 2,010.26 \text{ MeV}$ [17], as though the LHCb detectors had perfectly measured $m(D^*)$. The D^* -constrained $m(D^* 3\pi)$ is defined as:

$$m\left(D^* 3\pi \mid m(D^*) = 2,010.26 \text{ MeV}/c^2\right) = m(D^* 3\pi) - m(D^*) + m(D^*)_{PDG} \quad (4)$$

The resulting $m(D^*)$ -constrained signal peak in $m(D^* 3\pi)$ is narrower, as shown in Figure 7. The narrower peak leads to improved purity, as the normalisation peak is now concentrated within a smaller region that includes less background.

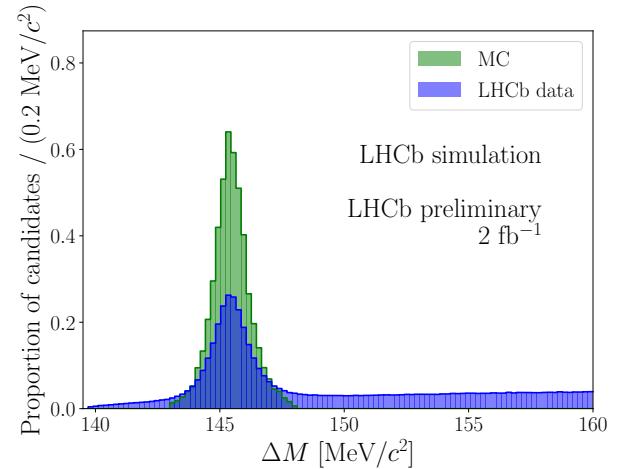


Figure 5. $\Delta M = m(D^*) - m(D^0)$ distribution of the LHCb data (blue) and MC (green). The data distribution includes contributions from fake D^* candidates, which are reduced by applying a cut to the ΔM distribution matching the one already applied to MC.

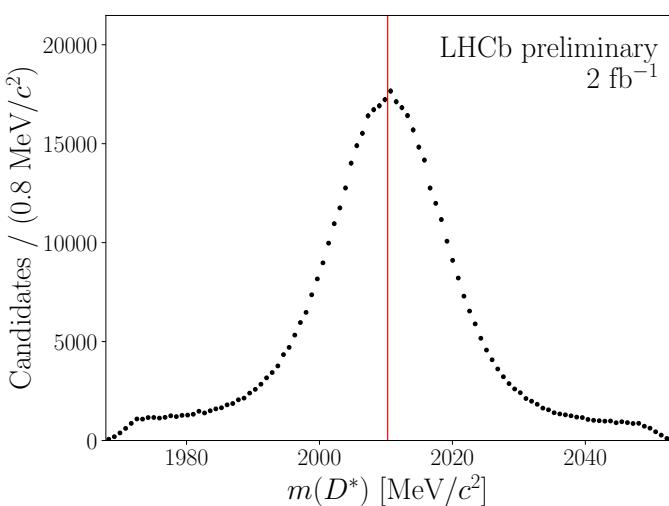


Figure 6. $m(D^*)$ distribution of the LHCb data. The vertical red line marks $m(D^*)_{PDG} = 2,010.26$ MeV

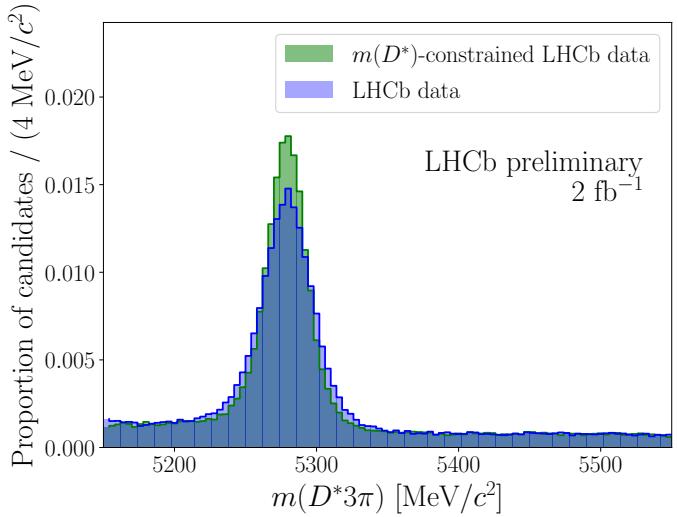


Figure 7. $m(D^*3\pi)$ distribution of the $m(D^*)$ -constrained and unconstrained LHCb data. The $m(D^*)$ constraint is found to improve the resolution of the normalisation peak.

3. Fit to the LHCb data

In order to evaluate the normalisation mode yield in the 2015-2016 LHCb data sample, an unbinned minimum log-likelihood fit is performed to the $m(D^*)$ -constrained invariant mass $m(D^*3\pi)$ in the $5150 - 5545$ MeV/ c^2 range. The data distribution considered in the fit is shown in Figure 8. As the LHCb sample is not a pure signal sample, several components contribute to the total $m(D^*3\pi)$ distribution. The primary component is the $B^0 \rightarrow D^{*-}3\pi$ signal, which contributes the majority of events in the peak around $m(B^0)_{PDG} = 5279.65$ MeV/ c^2 [17]. This peak also has a contribution from the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ decay, which must be taken into account in order not to over-estimate the $B^0 \rightarrow D^{*-}3\pi$ peak yield. At lower mass, around 5215 MeV/ c^2 , sits the $B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$ decay, where the K^+ meson is misidentified as a π^+ meson. Finally, the combinatorial background contributes continuously across to the full invariant-mass range.

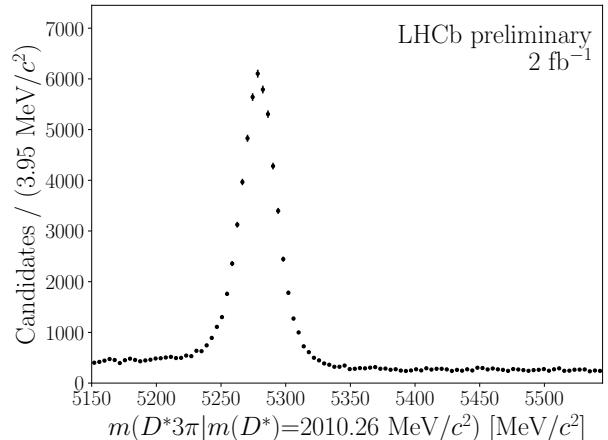


Figure 8. $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in the LHCb data.

At low invariant-mass values, it is possible for partially reconstructed decays such as $B^0 \rightarrow D^{*-}\pi^+\pi^-\mu^+\nu_\mu$ (with a branching ratio $\mathcal{B}(B^0 \rightarrow D^{*-}\pi^+\pi^-\mu^+\nu_\mu) = 0.14\%$ [17]), and $B^0 \rightarrow D^{*-}3\pi\pi^0$ where the ν_μ neutrino and π^0 meson are not reconstructed, to contribute. In order to check whether or not partially reconstructed decays contribute significantly to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution, samples of $B^0 \rightarrow D^{*-}3\pi\pi^0$ and $B^0 \rightarrow D^{*-}\pi^+\pi^-\mu^+\nu_\mu$ decays are simulated using *RapidSim*, “a fast Monte-Carlo generator for simulation of heavy-quark hadron decays” [9].

Out of the full sample of generated $B^0 \rightarrow D^{*-}3\pi\pi^0$ decays, 0.02% are found to contribute in the fit range. This corresponds to a rate of 0.05% compared to the $B^0 \rightarrow D^{*-}3\pi$ signal peak when the relative branching fractions of the decays are considered. Thus, this contribution is deemed to be negligible. In the $B^0 \rightarrow D^{*-}\pi^+\pi^-\mu^+\nu_\mu$ generated sample, 3.9% of events are found to contribute in the fit range. Accounting for relative branching fraction, and allowing for conservatively large $\mu \rightarrow \pi$ misidentification rate of 50%, this mode is expected to contribute below the 0.5% level relative to signal. Thus, this mode is also not considered in the analysis.

3.1 Definitions and notations

Three Probability Density Functions (PDF) are used in this report: the exponential PDF, the Gaussian PDF, and the sum of two Crystal Ball [21] PDFs with tails on opposite sides. Standard notations are used to denote their parameters.

The decay constant of any exponential combinatorial background is denoted $\lambda_{B,c}$, and its yield is denoted $n_{B,c}$. The mean and standard deviation of a Gaussian PDF are denoted μ and σ . The Crystal Ball (CB) PDF p_{CB} is a PDF with a Gaussian core of mean μ and standard deviation σ , where one of the tails follows a $(-n)$ power law for $\frac{x-\mu}{\sigma} \times \text{sign}(\alpha) \leq -|\alpha|$:

$$\forall x \in \mathbb{R}, p_{CB}(x; \mu, \sigma, \alpha, n) = \begin{cases} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) & \text{if } \frac{x-\mu}{\sigma} \times \text{sign}(\alpha) > -|\alpha| \\ A \left(B - \frac{x-\mu}{\sigma}\right)^{-n} & \text{if } \frac{x-\mu}{\sigma} \times \text{sign}(\alpha) \leq -|\alpha| \end{cases}. \quad (5)$$

The parameters A and B are unique and constant real parameters so that the Crystal Ball PDF and its derivative are defined and continuous in \mathbb{R} . If $\alpha > 0$, the PDF has a left power tail whereas if $\alpha_R < 0$, the PDF has a right power tail. In this report, the sum of two CB PDFs with tails on opposite sides is widely used to model signal shapes. It is defined as follows:

$$\forall x \in \mathbb{R}, p(x; \mu_L, \mu_R, \sigma_L, \sigma_R, \alpha_L, \alpha_R, n_L, n_R, f_{\frac{L}{R}}) = f_{\frac{L}{R}} p_{CB}(x; \mu_L, \sigma_L, \alpha_L, n_L) + (1 - f_{\frac{L}{R}}) \times p_{CB}(x; \mu_R, \sigma_R, \alpha_R, n_R), \quad (6)$$

where the parameter $f_{\frac{L}{R}}$ denotes the relative fraction of the left and right Crystal Ball PDFs. α_R is negative so that the corresponding CB PDF has a right power tail. If the two CB PDFs share their mean, the mean will be denoted μ . The fits performed in this report minimise an unbinned negative log-likelihood using the *Minuit* minimiser [14], implemented in *zfit*, a python library [10].

3.2 Fit to the MC dataset

A fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ spectrum in the $B^0 \rightarrow D^{*-}3\pi$ MC allows the expected shape of the normalisation peak in the LHCb data to be studied. Some parameters from this fit can then be used as fixed parameters in the data fit to help constrain the signal PDF. The MC $m(D^*3\pi)$ distribution is modeled using a sum of two CB PDFs with tails on opposite sides and a shared mean. The fit is shown in [Figure 9](#) and the fitted parameters are summarised in [Table 2](#).

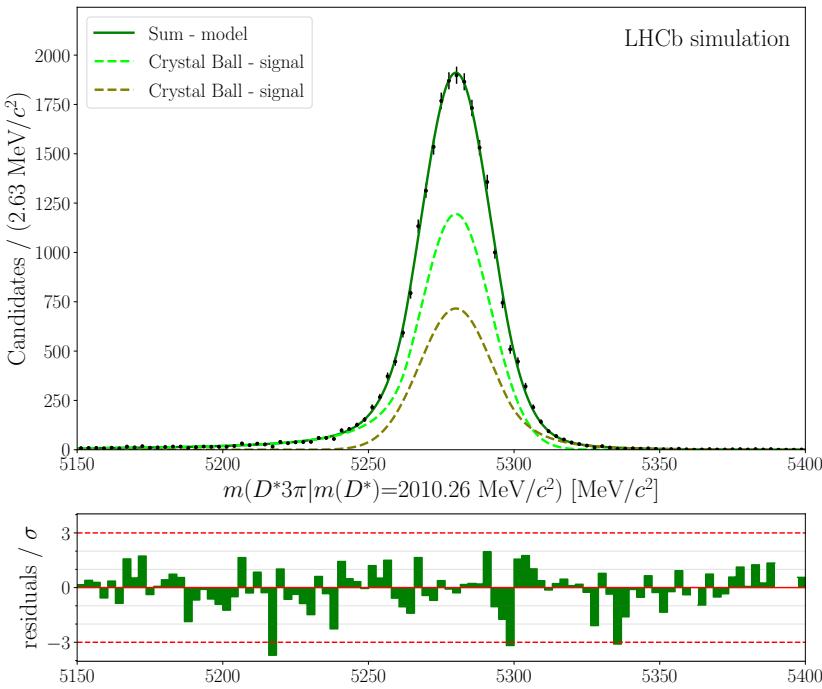


Figure 9. Fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in MC.

Variable	Fitted value
μ_{MC}	5280.0 ± 0.1
$\sigma_{L,MC}$	11.9 ± 0.9
$\sigma_{R,MC}$	12 ± 2
$\alpha_{L,MC}$	1.43 ± 0.21
$\alpha_{R,MC}$	-1.42 ± 0.19
$n_{L,MC}$	1.96 ± 0.35
$n_{R,MC}$	5.8 ± 2.4
$f_{\frac{L}{R},MC}$	0.63 ± 0.15

Table 2. Result of the fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in MC.

3.3 Fit to $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ RapidSim sample

As the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ background component in $m(D^*3\pi)$ is considerably smaller than the $B^0 \rightarrow D^{*-} 3\pi$ normalisation signal and occupies a similar invariant mass range, its shape cannot be directly determined in the data fit. To study the expected shape of this contribution, a $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ simulated sample has been generated using RapidSim. After assigning the π^+ mass hypothesis to the K^+ and constraining $m(D^*)$, the distribution is fitted with the sum of two CB PDFs with tails on opposite sides, as shown in [Figure 10](#). The shape parameters, listed in [Table 3](#), are fixed in the data fit and only the yield varies freely.

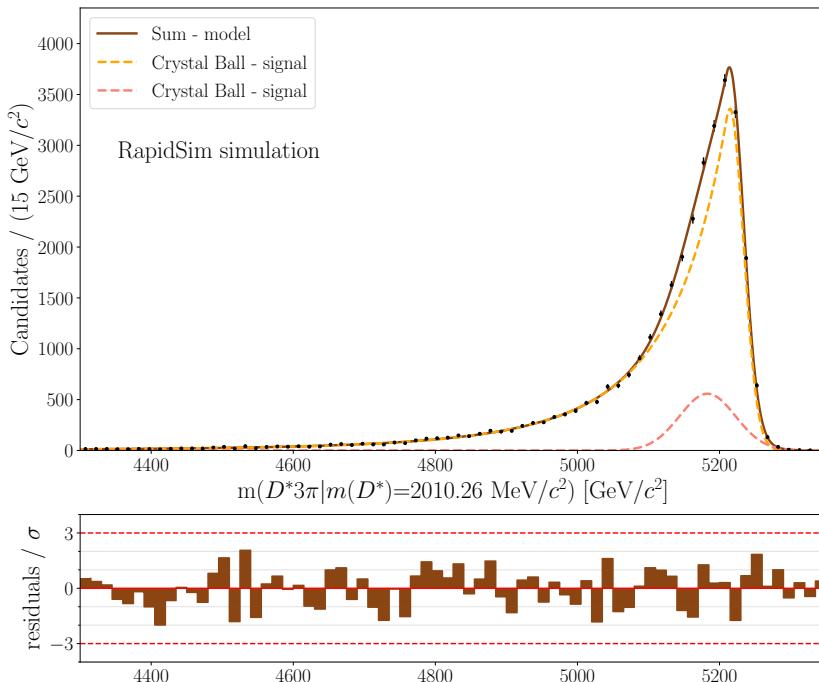


Figure 10. Fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ RapidSim sample.

Variable	Fitted Value
$\mu_{L,D^*K\pi\pi}$	5215 ± 1
$\mu_{R,D^*K\pi\pi}$	5183 ± 3
$\sigma_{L,D^*K\pi\pi}$	19.1 ± 0.6
$\sigma_{R,D^*K\pi\pi}$	39 ± 2
$\alpha_{L,D^*K\pi\pi}$	0.240 ± 0.015
$\alpha_{R,D^*K\pi\pi}$	-3.45 ± 0.21
$n_{L,D^*K\pi\pi}$	4.41 ± 0.24
$n_{R,D^*K\pi\pi}$	$(0.0 \pm 2.2) \times 10^{-4}$
$f_{\frac{L}{R},D^*K\pi\pi}$	0.886 ± 0.015

Table 3. Result of the fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ RapidSim sample.

3.4 Fit to $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$

In data, the $m(3\pi)$ invariant mass distribution exhibits a peak around $m(D_s^+) = 1,968.3$ MeV/ c^2 [17], as shown in [Figure 11](#). Moreover, the two-dimensional histogram of (flight distance, invariant mass) of the 3π system shown in [Figure 12](#) indicates that this peak corresponds to higher flight distance of the 3π system. Thus, this mass peak is a clear contribution from $D_s^+ \rightarrow 3\pi$ decays, where the D_s^+ meson flies before decaying.

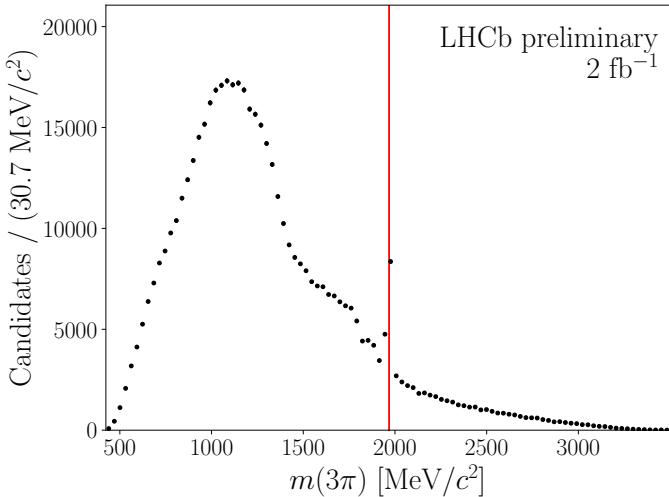


Figure 11. $m(3\pi)$ invariant mass distribution in data. The vertical red line indicates the mass of the D_s^+ meson, around 1,968 MeV.

In order to isolate the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ contribution in data, $m(3\pi)$ is required to be in the region ± 50 MeV/ c^2 around the known D_s^+ mass. Then in order to distinguish events with a genuine $D_s^+ \rightarrow 3\pi$ decay from other decays in this region, a fit is performed to the $m(3\pi)$ spectrum. For this fit, an exponential PDF is used to model the non- D_s^+ contribution, and a Gaussian PDF describes the $D_s^+ \rightarrow 3\pi$ peak; the fit result is shown in [Figure 13](#).

In order to project out the non- D_s^+ component of the $m(D^*3\pi)$ spectrum so that the signal peak in the $m(D^*3\pi)$ spectrum corresponds to $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ decays, sWeights associated with the $D_s^+ \rightarrow 3\pi$ component of the $m(3\pi)$ fit are used. The computation of the sWeights is explained in annex in [7.1](#). The signal and background sWeights are shown in [Figure 14](#); the signal sWeights are higher around $m(D_s)$, where there are more $D_s^+ \rightarrow 3\pi$ events.

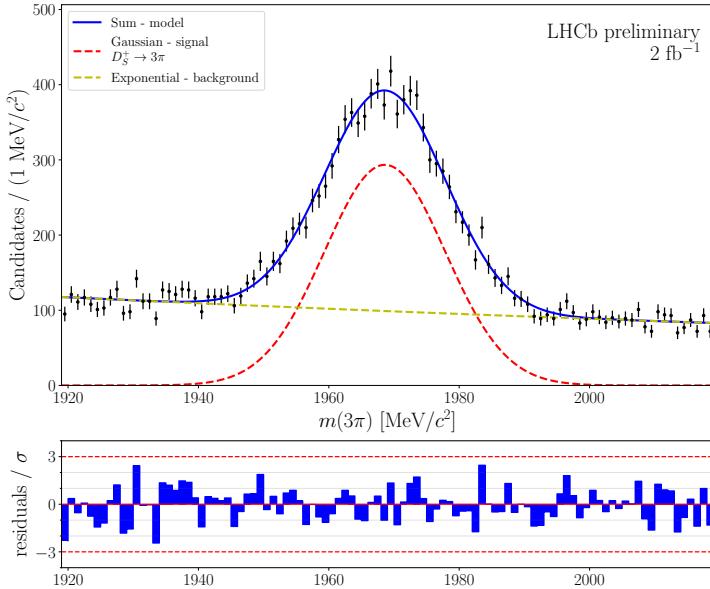


Figure 13. Fit to the $m(3\pi)$ distribution in the LHCb data.

Since the correlation between $m(3\pi)$ and $m(D^*3\pi)$ is about 0.0476, i.e., relatively small, the sWeights can be safely applied to the $m(D^*3\pi)$ distribution where the same cut on $m(3\pi)$ is applied. This enables the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ contribution to the $m(B^0)$ peak to be isolated, and an invariant mass fit performed in order to measure the yield of this component. This yield can then be fixed in the main data fit that measures the $B^0 \rightarrow D^{*-}3\pi$ yield, preventing any over-estimation of the normalisation yield.

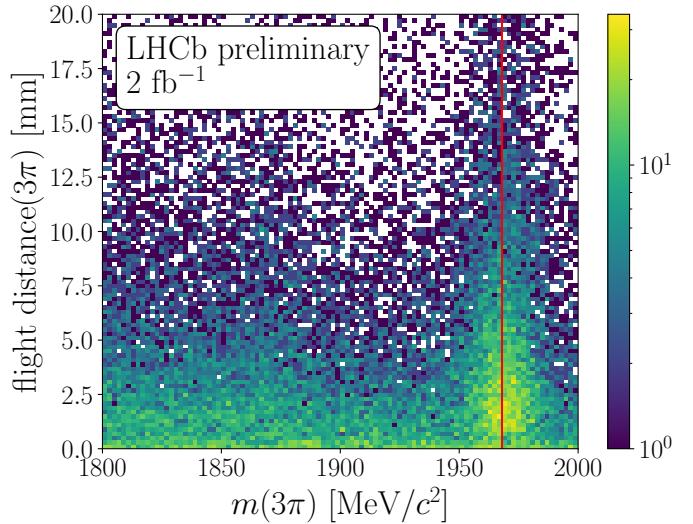


Figure 12. Two-dimensional histogram of flight distance versus invariant mass of the 3π system in data. The vertical red line indicates the mass of the D_s^+ meson.

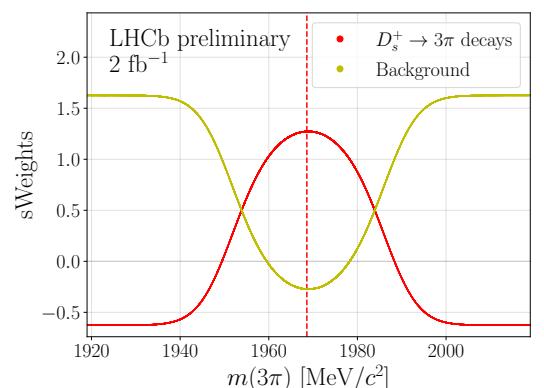


Figure 14. Signal and background sWeights computed from the fit to the $m(3\pi)$ distribution in the LHCb data. The vertical dashed red line indicates the mass of the D_s^+ meson.

The s Weighted $m(D^*3\pi)$ distribution is fitted in the region is performed in the $5150 - 5350$ MeV/ c^2 range by considering three components: (1) The $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ component is modelled by a Gaussian PDF; (2) partially reconstructed decays at lower invariant mass, such as $B^0 \rightarrow D^{*-}(D_s^{*+} \rightarrow D_s^+ \gamma)$, are modelled by a Gaussian PDF; (3) combinatorial background is modelled using an exponential function. The result of the fit is shown in [Figure 15](#) and [Table 4](#). In the global $m(D^*3\pi)$ data fit, the shape and yield of the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ component are fully fixed to the values listed in [Table 4](#).

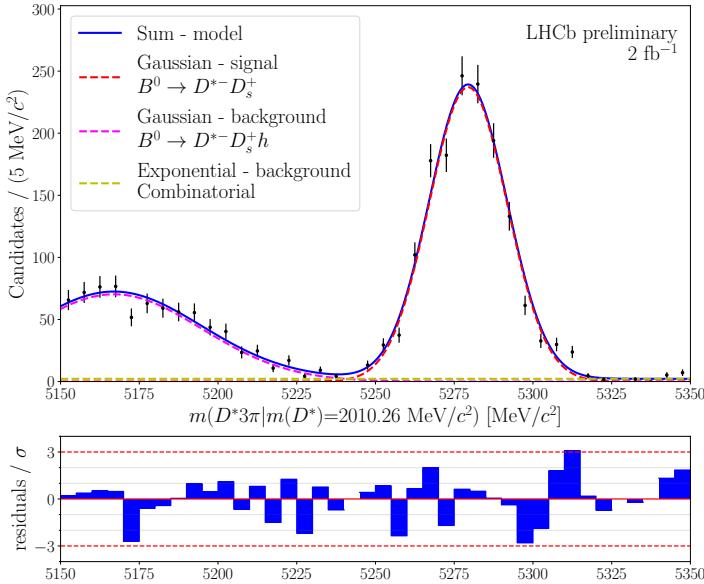


Figure 15. Fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of s Weighted LHCb data, where the s Weights assigned are derived from the $m(3\pi)$ fit shown in [Figure 13](#).

Component	Variable	Fitted value
$B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$	$n_{D^*D_s}$	1468 ± 43
	$\sigma_{D^*D_s}$	12.35 ± 0.47
	$\mu_{D^*D_s}$	5279.40 ± 0.52
Partially reconstructed	$n_{B,D^*D_s h}$	715 ± 42
	$\sigma_{B,D^*D_s h}$	27.9 ± 2.6
	$\mu_{B,D^*D_s h}$	5166.9 ± 3.6
Combinatorial	$n_{B,c'}$	88 ± 60
	$\lambda_{B,c'}$	$(-0.0 \pm 6.3) \times 10^{-10}$

Table 4. Result of the fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of s Weighted LHCb data, where the s Weights assigned are derived from the $m(3\pi)$ fit shown in [Figure 13](#).

3.5 Global fit to the $m(D^*)$ -constrained $m(D^*3\pi)$

The invariant-mass fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in data is performed in the $5150 - 5545$ MeV/ c^2 range. The $B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$ and $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ background contributions are modelled as described in Sections 3.3 and 3.4. The $B^0 \rightarrow D^{*-}3\pi$ signal peak is modelled using the sum of two CB PDFs with tails on opposite sides and shared mean, like in the MC fit described in Section 3.2. In the data fit, the normalisation peak tail parameters $\alpha_{L,S}$, $\alpha_{R,S}$, and $n_{L,S}$ vary freely, while the $n_{R,S}$ parameter is fixed to the value found in the MC fit. The fit results are shown in [Figure 16](#) and [Table 5](#), and the normalisation mode yield is measured to be

$$n_{B^0 \rightarrow D^{*-}3\pi} = 54,484 \pm 1233,$$

where the uncertainty quoted is statistical. Furthermore, the ratio between the $B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$ and normalisation mode yields is measured

$$\frac{n_{B^0 \rightarrow D^{*-}K^+\pi^+\pi^-}}{n_{B^0 \rightarrow D^{*-}3\pi}} = (3.5 \pm 1.5)\%$$

This is lower than the corresponding ratio of branching fractions

$$\frac{\mathcal{B}(B^0 \rightarrow D^{*-}K^+\pi^+\pi^-)}{\mathcal{B}(B^0 \rightarrow D^{*-}3\pi)} = \frac{0.047 \pm 0.004\%}{0.721 \pm 0.029\%} = (6.5 \pm 0.6)\%,$$

which is due to the fact that only a subset of $B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$ decays are misidentified as $B^0 \rightarrow D^{*-}3\pi$ due to the application of pion PID requirements to the 3π system. All pions from the 3π candidate are required to have a difference in log-likelihood between the kaon and pion mass hypotheses, $\Delta_{LL}(K - \pi)$, below 10. The measured yield ratio suggests a $K \rightarrow \pi$ misidentification rate of $(54 \pm 24)\%$, which is consistent with the expectation from LHCb PID calibration data for the $\Delta_{LL}(K - \pi) < 10$ cut.

The results of the normalisation fit can be improved by increasing the signal purity of the data sample, which is explored in Section 4.

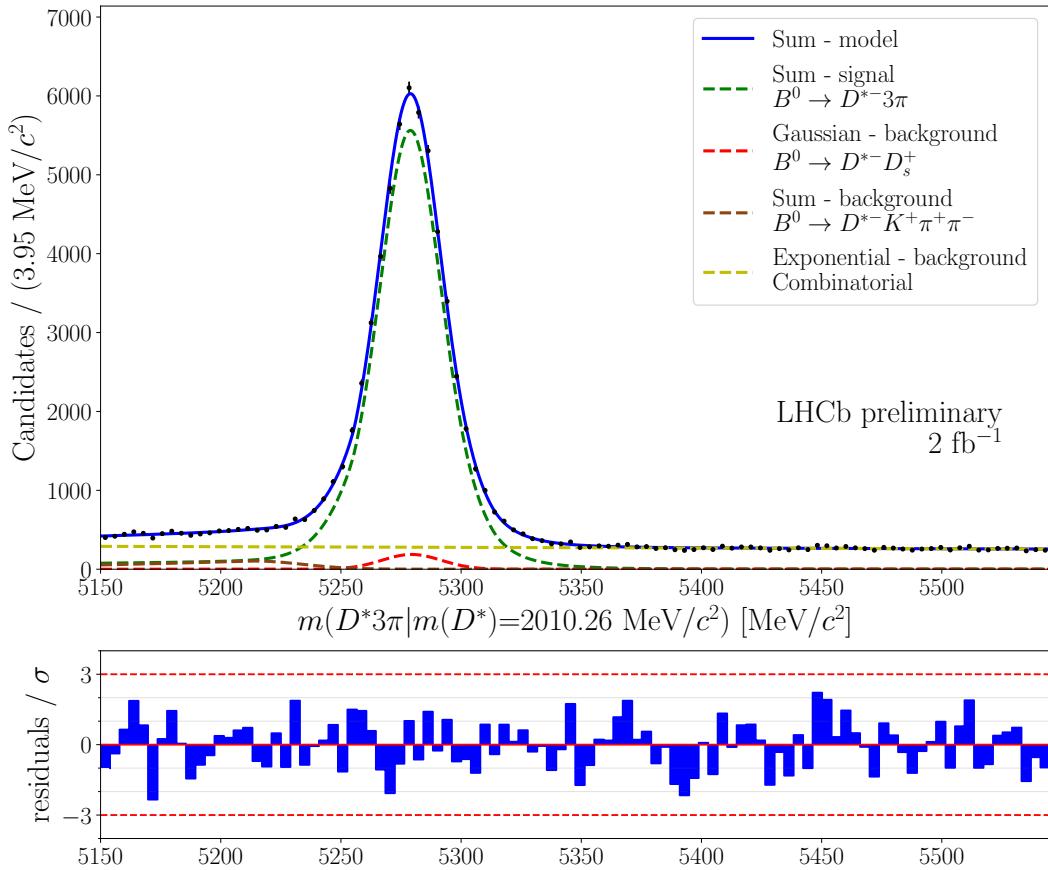


Figure 16. Invariant-mass fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in 2015 and 2016 LHCb data.

Component	Variable	Fitted value
$B^0 \rightarrow D^{*-}3\pi$	$n_{B^0 \rightarrow D^{*-}3\pi}$	$54,484 \pm 1233$
	μ_S	5278.90 ± 0.09
	$\sigma_{L,S}$	11.1 ± 0.3
	$\sigma_{R,S}$	18.9 ± 0.6
	$\alpha_{L,S}$	2.03 ± 0.11
	$\alpha_{R,S}$	-1.78 ± 0.13
	$n_{L,S}$	0.44 ± 0.13
	$f_{\frac{L}{R},S}$	0.461 ± 0.041
$B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$	$\frac{n_{B^0 \rightarrow D^{*-}K^+\pi^+\pi^-}}{n_{B^0 \rightarrow D^{*-}3\pi}}$	0.035 ± 0.015
Combinatorial	$n_{B,c}$	$27,254 \pm 654$
	$\lambda_{B,c}$	$(-3.2 \pm 1.9) \times 10^{-4}$

Table 5. Result of the invariant-mass fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in 2015 and 2016 LHCb data.

4. Multivariate analysis

The purpose of the multivariate analysis (MVA) is to classify signal and combinatorial background events in order to reduce the latter. This is achieved by training a Gradient Boosted Decision Tree (BDT) classifier using the scikit-learn library [18]. An illustration of the usefulness of the MVA is provided in annexe 7.1.3.

4.1 Training

A Gradient BDT with 400 trees and a learning rate equal to 0.1 is trained by comparing a set of discriminating variables in the signal and combinatorial background samples. The signal sample used is the MC $B^0 \rightarrow D^{*-} 3\pi$ sample. The background sample is taken from data in the high invariant-mass range between 5400 and 5570 MeV/ c^2 , as shown in Figure 17. This region is extremely pure combinatorial background, as confirmed by the fit in Section 3.5; at 5400 MeV/ c^2 , the ratio between the the $B^0 \rightarrow D^{*-} 3\pi$ and combinatorial components is about 1.2%.

In order to avoid applying the BDT to events it has been trained on, the dataset is divided at random into two equal subsamples, A and B . The BDT A (B) trained on dataset A (B) is applied to dataset B (A), which enables all of the data to be used in subsequent analysis. In particular, this allows the high-mass region used in the BDT training to be included in invariant-mass fits. Within each subsample A and B , further

50/50 splits are performed in order to create training and test samples for internal validation.

The set of discriminating variables used is listed in Table 6. These variables are chosen to maximise performance (i.e. the area under the ROC curve of the test sample), and minimise over-training (i.e. the training and test samples have similar BDT output distributions). This is achieved by selecting variables that are sufficiently different in the $B^0 \rightarrow D^{*-} 3\pi$ MC signal sample and the data combinatorial background sample.

Moreover, as the MC is used as a proxy of the $B^0 \rightarrow D^{*-} 3\pi$ signal peak, the chosen discriminating variables in MC must be distributed like in a pure $B^0 \rightarrow D^{*-} 3\pi$ sample. Assuming that the $B^0 \rightarrow D^{*-} 3\pi$ and $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$ have a similar behaviour in terms of how they compare to combinatorial background, the chosen variables have been selected so that they agree well between the $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$ weighted data and the MC dataset.

The distributions of the 24 discriminating variables are shown in annexe in 7.2.1. Definitions of the variables can be found in annexe, section 7.1.1. The most discriminating variables are the cosine of the $\ln(1 - x)$ of the DIRA angle w.r.t. the primary vertex and the impact parameter w.r.t. the primary vertex of the $D^* 3\pi$ system. The twelve most discriminating variables of the training of the two BDTs A and B are shown in section 7.2.2, in Table 9 and Table 10.

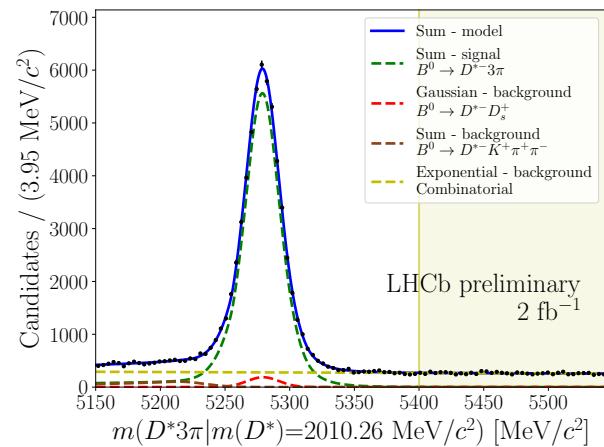


Figure 17. $m(D^* 3\pi)$ distribution in 2015 and 2016 LHCb data. The red lines indicate the region used as a combinatorial background sample for the BDT training. The edge effect visible at high mass is due to a prior selection cut applied to the data sample.

Particle	Variable
$D^{*}3\pi$	Transverse momentum p_T
	$\ln(x)$ of the χ^2 of the flight distance w.r.t. the own primary vertex (PV)
	χ^2 of the end vertex (EV) per d.o.f.
	$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV
D^*	Impact parameter w.r.t. the PV
	$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV
	Impact parameter w.r.t. the PV
	χ^2 of the impact parameter w.r.t. the PV
D^0	Transverse momentum p_T
	$\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV
	Impact parameter w.r.t. the PV
	$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV
3π	Transverse momentum p_T
	$\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV
	χ^2 of the EV per d.o.f.
	$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV
π from D^* K from D^0 π of D^0	Impact parameter w.r.t. the PV
	Transverse momentum p_T
	$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV

Table 6. Variables used in the BDT training.

4.2 Testing

After training of the two BDTs, ROC curves are evaluated using the test samples as shown in [Figure 18](#). The two ROC area under the curve values are 0.98, indicating that the BDT is highly separating for signal and combinatorial background. When applied to data, the BDT returns a variable that is referred to as *BDT* in this report. The more likely the event is to be signal (background), the higher (lower) the *BDT* value. A check for over-training is performed by comparing the BDT outputs of the testing and training samples in signal and background; as shown in [Figure 19](#), there is no evidence of any significant over-training.

In order to confirm this observation, Kolmogorov–Smirnov (KS) tests [13] (using the python library *scipy.stats* [24]) are performed between the BDT outputs in the test and training samples, considering the signal and background samples separately. The low values of the KS statistics and the high p-values of the KS tests in [Table 7](#) corroborate the observation that the BDTs are not significantly over-trained.

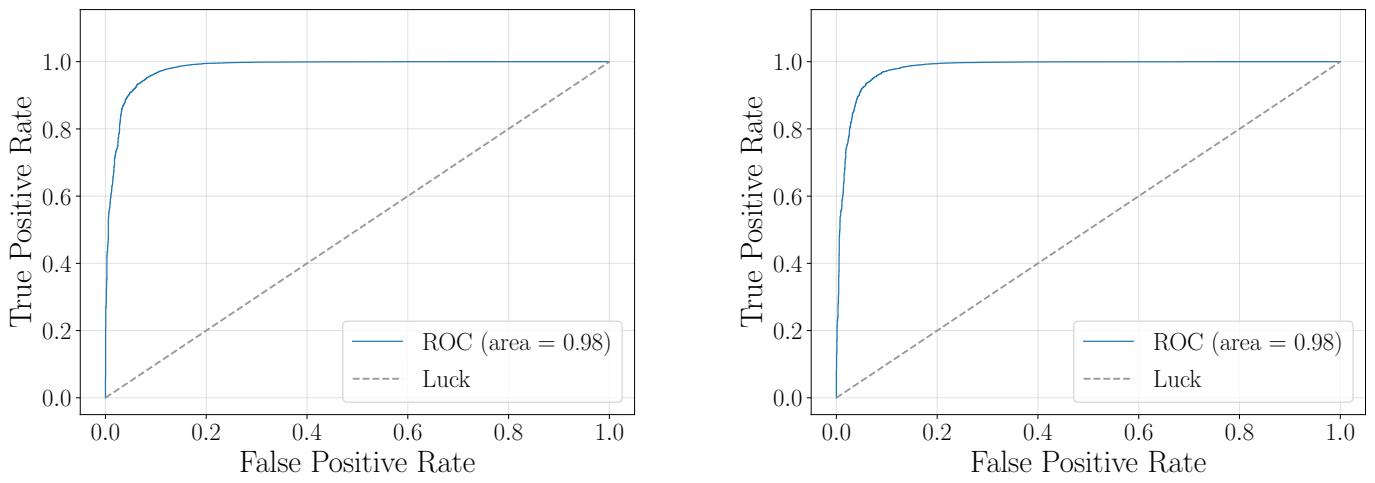
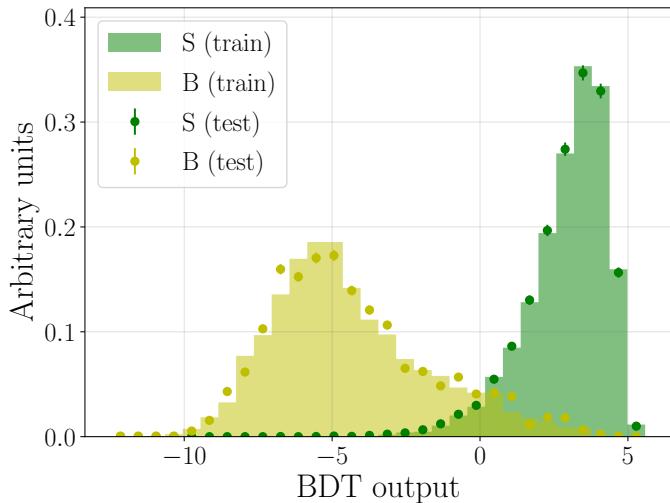
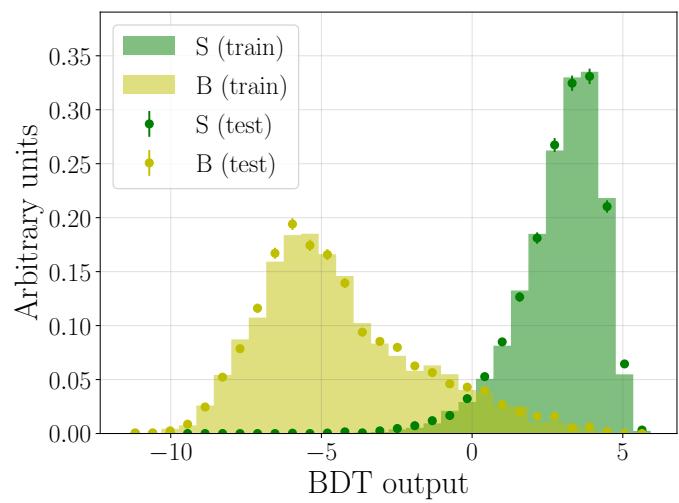


Figure 18. ROC curves for the two BDTs.



(a) Over-training plot for the BDT A .



(b) Over-training plot for the BDT B .

Figure 19. Over-training plots for the two BDTs, which show the BDT distribution in signal (green) and background (red) in both the training and test samples.

BDT	A		B		
	Training and test sample of the	Signal	Background	Signal	Background
KS score		0.012	0.022	0.0079	0.012
p-value		0.34	0.51	0.88	0.98

Table 7. KS scores and p-values of the KS test performed between the BDT outputs in the training and test samples, for the BDTs A and B .

4.3 Application to the data

As shown in [Figure 20](#), the BDT output distribution in data exhibits two bumps. The first, which appears around -5 , corresponds to combinatorial background while the second at higher values corresponds to the signal peak. This is confirmed by the two-dimensional histogram of $(BDT, m(D^*3\pi))$ shown in [Figure 22](#), where the signal contribution around $m(B^0)$ peaks at higher BDT values while the combinatorial background is evenly distributed across $m(D^*3\pi)$ at lower BDT values. Moreover, as shown in [Figure 21](#), the data with low BDT values contains very little signal. As such, a cut to retain data with a high BDT value will allow a substantial proportion of combinatorial background to be removed while retaining most of the signal.

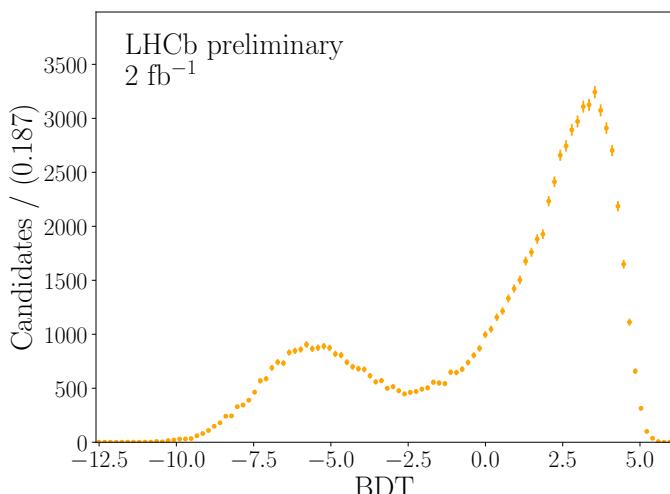


Figure 20. BDT output distribution in data in the $5150 - 5570$ MeV/ c^2 region of the $m(D^*)$ -constrained $m(D^*3\pi)$ invariant mass.

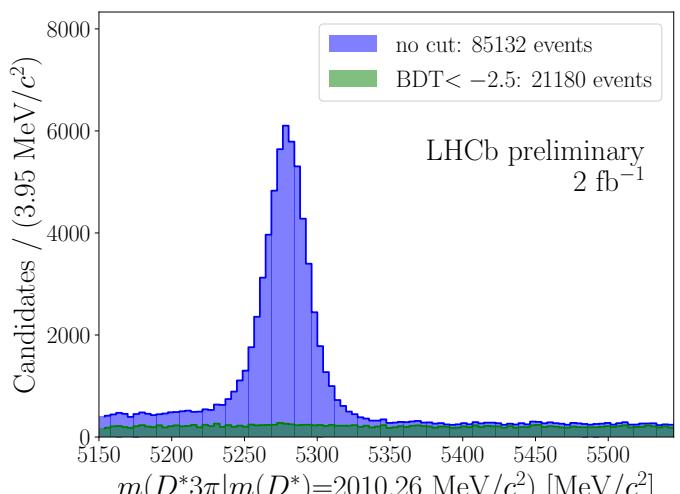


Figure 21. Data $m(D^*3\pi)$ distribution without any BDT cut applied (blue) and with a $BDT < -2.5$. cut (green).

Furthermore, the signal peak of the two-dimensional histogram of ($BDT, m(D^*3\pi)$) in data (in [Figure 22](#)) and in MC (in [Figure 23](#)) are entirely consistent.

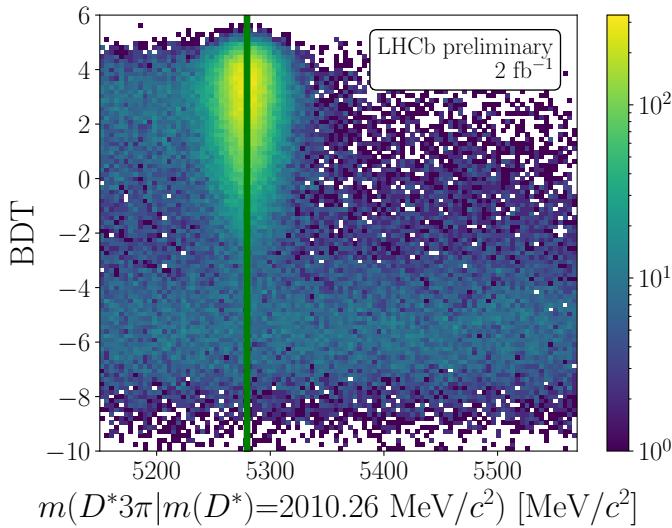


Figure 22. Two-dimensional distribution of ($BDT, m(D^*3\pi)$) in data.

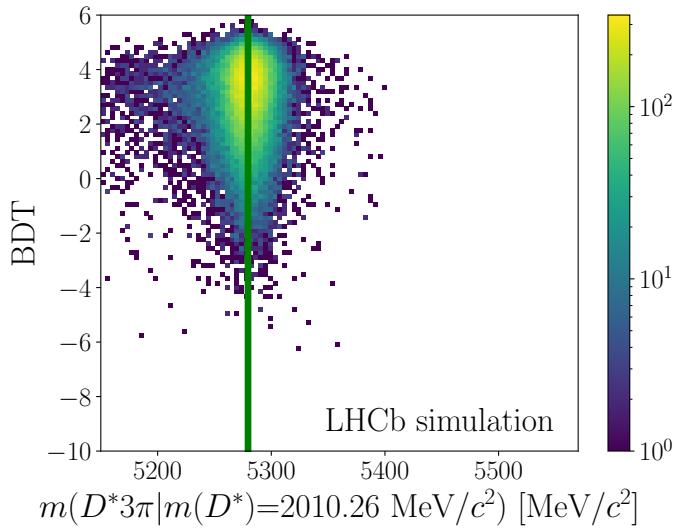


Figure 23. Two-dimensional distribution of ($BDT, m(D^*3\pi)$) in MC.

4.4 BDT cut optimisation

An optimisation is performed to determine the BDT cut that maximises the relative uncertainty $\frac{\Delta S}{S}$ on the normalisation mode yield $S = n_{B^0 \rightarrow D^{*-}3\pi}$, where ΔS denotes the absolute statistical uncertainty on S . To this end, a cut of $BDT > c$ is applied to the data and the fit procedure described above performed. This procedure is repeated for $c \in \{-4; -3.75; \dots, 1.75\}$, and the relative uncertainty on the normalisation mode yield evaluated for each BDT cut. In this procedure, some shape parameters are fixed to the values found in the full data fit (Section 3.5), namely the tail parameters α_L , α_R , and n_L , but also the ratio $\frac{n_{B^0 \rightarrow D^{*-}K^+\pi^+\pi^-}}{n_{B^0 \rightarrow D^{*-}3\pi}}$. In fixing the yield ratio, it is assumed that the variations in BDT efficiency for the $B^0 \rightarrow D^{*-}K^+\pi^+\pi^-$ and $B^0 \rightarrow D^{*-}3\pi$ decays are the same, which is sound given that the modes differ only in terms of having a kaon produced in the B^0 decay rather than a pion. Moreover, fixing the tail parameters in data is sound because the tails of the signal peak in MC are independent of the BDT cut applied to it, as shown in [Figure 24](#).

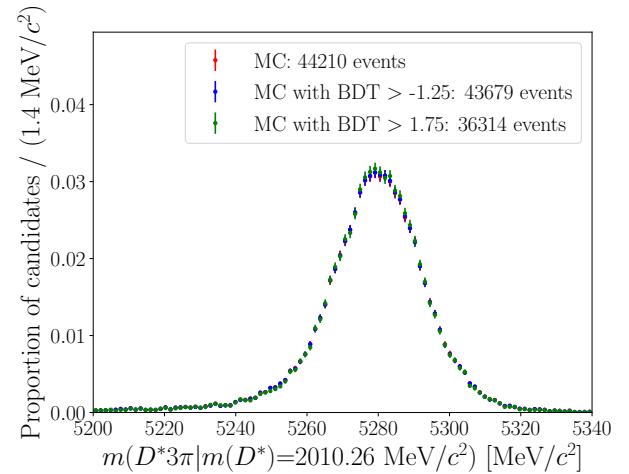


Figure 24. $m(D^*)$ -constrained $D^*3\pi$ invariant mass in MC where (1) no BDT cut is applied, (2) the optimal BDT cut $BDT > -1.25$ is applied and (3) the maximal BDT cut during optimisation $BDT > 1.75$ is applied.

The result of this procedure is shown in [Figure 25](#), where $\frac{\Delta S}{S}$ is shown as a function of BDT cut. The cut that minimises the relative uncertainty is found to be $BDT > -1.25$. As a consistency check, the ROC curve for the BDT cut optimisation is plotted in [Figure 26](#), where the signal efficiency and background rejection values are calculated by comparing the fitted yields at a given BDT cut to the value measured with the $BDT > -4$ cut applied. The optimal cut sits at the apex of the ROC curve, as expected.

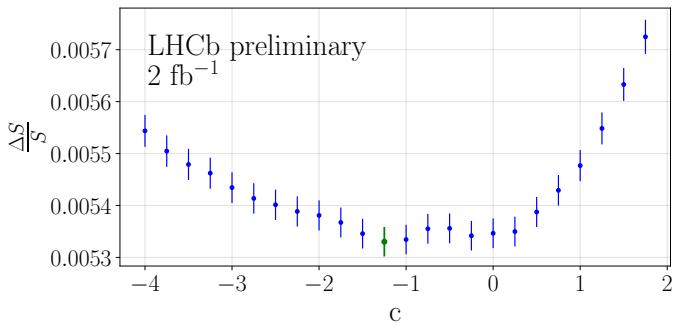


Figure 25. $\frac{\Delta S}{S}$ measured in the data fit as a function of c , where $BDT > c$ is the BDT cut applied. The optimal point is coloured green.

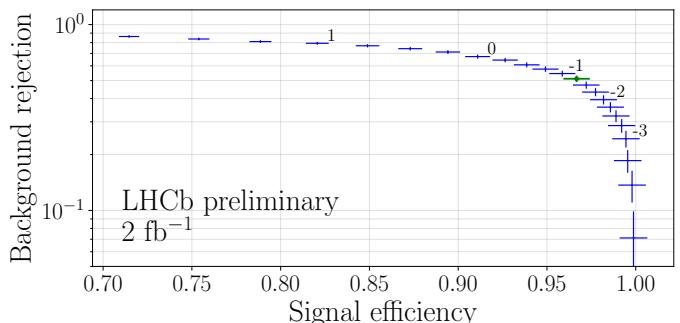


Figure 26. ROC curve for the BDT cut, which is constructed using signal and background yields measured in the data fit at each BDT cut. The labels on the upper right of some points indicate the BDT cut value, c . The optimal point is coloured green.

To trace the ROC curve, the number of combinatorial background candidates under the signal peak must be determined. This is done by integrating the combinatorial background PDF in the region $[\mu_S - 3\sigma_S, \mu_S + 3\sigma_S]$ around the normalisation peak, where $\mu_S \approx 5278.9$ MeV/ c^2 is the measured normalisation peak mean and $\sigma_S = f_{\bar{R},S} \times \sigma_{L,S} + (1 - f_{\bar{R},S}) \times \sigma_{R,S} \approx 15.3$ MeV/ c^2 is the arithmetic mean of the two normalisation CB PDF widths.

5. Optimised signal yield with systematic uncertainties

With the optimal $BDT > -1.25$ cut applied to data, the signal yield is measured to be

$$n_{B^0 \rightarrow D^{*-} 3\pi} = 52,322 \pm 279.$$

The result of the fit is shown in [Figure 27](#) and [Table 8](#). The result of the invariant mass fit to the $m(3\pi)$, the associated weights and the result of the fit to the weighted $m(D^*3\pi)$ spectrum with the optimal $BDT > -1.25$ cut applied to data are shown in annexe [7.2.3](#). The relative uncertainty $\frac{\Delta S}{S}$ is found to decrease from 2.3% to 0.53%. However, this is mainly due to the fact that the tail parameters and the ratio $\frac{n_{B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-}}{n_{B^0 \rightarrow D^{*-} 3\pi}} \approx 3.5\%$ have been fixed. The systematic uncertainty due to the use of these fixed parameters must therefore be evaluated.

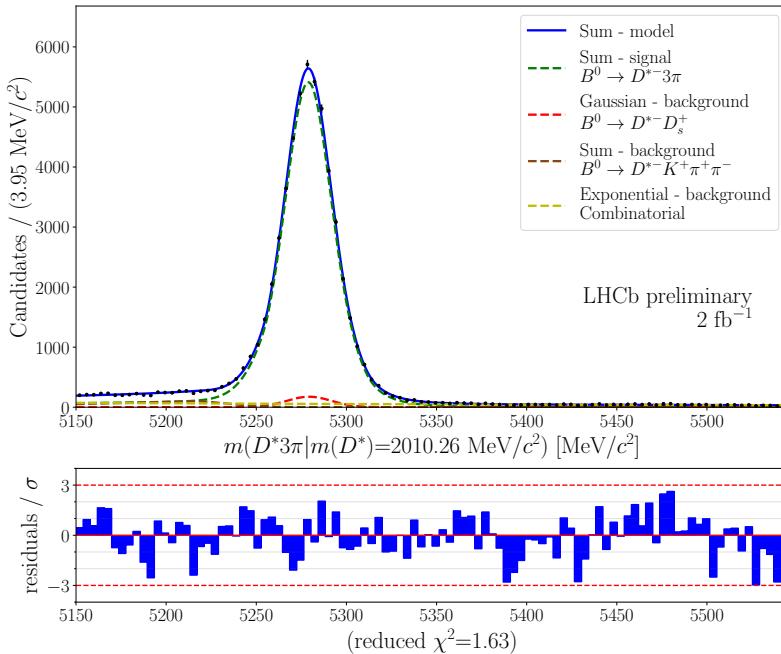


Figure 27. Invariant-mass fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of 2015 and 2016 LHCb data, where the optimal $BDT > -1.25$ cut is applied.

To determine the systematic uncertainty on the normalisation yield, the parameters $\alpha_{L,S}$, $\alpha_{R,S}$, $n_{L,S}$, and $n_{R,S} = n_{R,MC}$ are drawn randomly from Gaussian distributions with means equal to the nominal parameter values, and standard deviations equal to their uncertainty. If the drawn value has a different sign from the nominal value, the variable is drawn again. This prevents non-physical values of the $n_{L,S}$ and $n_{R,S}$ tail parameters, which by definition cannot be negative. Then, a fit to the $m(D^*3\pi)$ distribution in data with $BDT > -1.25$ is performed with the randomly drawn parameters fixed, and a new normalisation yield measured. This procedure is repeated about 1000 times, to build up a distribution of normalisation yield values under the systematic variation.

In a second phase, the same procedure is carried out with the ten parameters of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ component, namely the nine shape parameters from the RapidSim sample fit (see [Table 3](#)), and the yield ratio $\frac{n_{B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-}}{n_{B^0 \rightarrow D^{*-} 3\pi}}$. This procedure is applied a third time with the three parameters $n_{D^* D_s}$, $\mu_{D^* D_s}$ and $\sigma_{D^* D_s}$ of the $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$ contribution. As a crosscheck, the

Component	Variable	Fitted value
$B^0 \rightarrow D^{*-} 3\pi$	$n_{B^0 \rightarrow D^{*-} 3\pi}$	$52,322 \pm 279$
	$\sigma_{L,S}$	11.0 ± 0.2
	$\sigma_{R,S}$	18.1 ± 0.2
	$f_{\frac{L}{R},S}$	0.426 ± 0.021
	μ_S	5278.87 ± 0.08
Combinatorial	$n_{B,c}$	5002 ± 181
	$\lambda_{B,c}$	$(-1.93 \pm 0.27) \times 10^{-3}$

Table 8. Result of the invariant-mass fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution in 2015 and 2016 LHCb data, where the optimal $BDT > -1.25$ cut is applied.

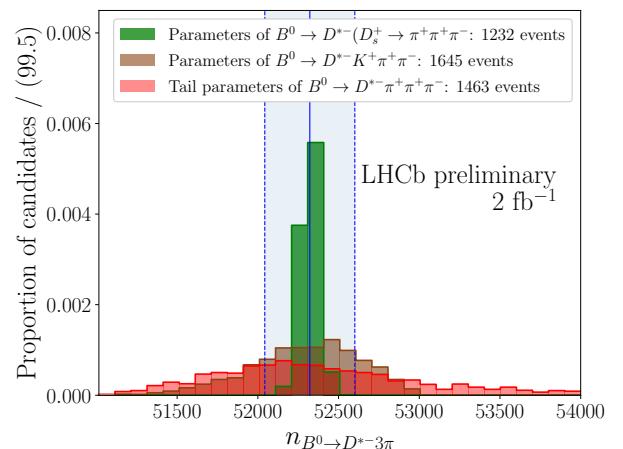


Figure 28. $n_{B^0 \rightarrow D^{*-} 3\pi}$ distributions obtained by varying the tail parameters of the $B^0 \rightarrow D^{*-} 3\pi$ component (in green) and the parameters of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ component (in brown). The blue region indicates the measured central value and statistical uncertainty of the normalisation mode yield, $n_{B^0 \rightarrow D^{*-} 3\pi} = 52,322 \pm 279$ (stat.).

distributions of the randomly drawn parameters are shown in annexe in section 7.2.4; they are found to be Gaussian-like.

The three signal yield distributions obtained by this procedure are shown in Figure 28. The standard deviations of these distributions are taken to represent the systematic uncertainties on the normalisation yield, and are found to be $\Delta_{\text{syst}} a(n_{B^0 \rightarrow D^{*-} 3\pi})_S = 806$ for the fixed signal shape parameters, $\Delta_{\text{syst}}(n_{B^0 \rightarrow D^{*-} 3\pi})_{B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-} = 339$ for the fixed $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ parameters and $\Delta_{\text{syst}}(n_{B^0 \rightarrow D^{*-} 3\pi})_{B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)} = 53$ for the fixed $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$ parameters. The total systematic uncertainty is obtained by summing these uncertainties in quadrature, where $\Delta_{\text{syst}}(n_{B^0 \rightarrow D^{*-} 3\pi}) = 876$ is found. The final normalisation mode yield measured in 2015 and 2016 data is thus

$$n_{B^0 \rightarrow D^{*-} 3\pi} = 52,322 \pm 279 \text{ (stat.)} \pm 876 \text{ (syst.)} = 52,322 \pm 919,$$

where the total relative uncertainty is 1.8%. It is lower than 2.3%, the relative statistical uncertainty of the non-optimised fit. This yield can be used as input to the calculation of $\mathcal{K}(D^*)$ as described in Section 1.2, which in turn will be used to evaluate $\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$ and thus $R(D^*)$. In future, the normalisation yield measurement performed in this work will be extended to use the full 9 fb^{-1} LHCb Run 1 + 2 dataset, and the offline selection cuts applied will be optimised for the hadronic $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ signal mode.

6. Conclusion

In order to determine the $R(D^*)$ ratio using the hadronic τ decay modes $\tau^+ \pi^+ \pi^+ \pi^- \bar{\nu}_\tau$ and $\tau^+ \pi^+ \pi^+ \pi^- \pi^0 \bar{\nu}_\tau$, the branching ratio of $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ can be measured using $B^0 \rightarrow D^{*-} 3\pi$ as a normalisation mode. In this work, the normalisation mode yield has been measured using 2015 and 2016 LHCb data, which will form part of the input to $R(D^*)$ in a future measurement.

A fit to the $m(D^{*3\pi}) - m(D^*) + m(D^*)_{PDG}$ distribution is performed, where $m(D^*)_{PDG}$ is the world-average value of the D^{*-} meson mass. In this fit, the normalisation mode yield corresponds to the yield of the signal PDF. Prior to the fit, the shape of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ background decay is determined using a RapidSim simulated sample. The s Weights associated to the $D_s \rightarrow 3\pi$ decay are computed using a fit to the 3π invariant mass spectrum around the $m(D_s)$ peak. The s Weighted $m(D^{*3\pi}) - m(D^*) + m(D^*)_{PDG}$ distribution is then fitted, from which the shape and yield of the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ background decay is determined. A fit to the $m(D^{*3\pi}) - m(D^*) + m(D^*)_{PDG}$ distribution of a fully-simulated LHCb MC sample of $B^0 \rightarrow D^{*-} 3\pi$ decays allows one of the tail parameters of the signal PDF to be fixed.

A multivariate analysis is performed using a gradient BDT in order to reject combinatorial background and improve the purity of the normalisation peak. Using the data fit, an optimisation procedure is carried out to determine the BDT cut which minimises the relative uncertainty on the normalisation mode yield.

Finally, the systematic uncertainty associated with the use of fixed terms in the normalisation mode fit is computed by varying the fixed terms in the fit within their prescribed uncertainties. The normalisation mode yield is measured to be

$$n_{B^0 \rightarrow D^{*-} 3\pi} = 52,322 \pm 279 \text{ (stat.)} \pm 876 \text{ (syst.)},$$

where the first uncertainty is statistical and the second is systematic. In future, this work will be extended to use the full 9 fb^{-1} LHCb Run 1 + 2 dataset, and an offline selection optimised for the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ mode will be used.

The use of the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ decay as a normalisation mode instead of the $B^0 \rightarrow D^{*-} 3\pi$ decay could also be explored, as a consistency check or an alternative approach. This mode is less favourable as a normalisation channel, however, because the $\sim 1\%$ $D_s^+ \rightarrow 3\pi$ sub-decay branching fraction decreases the statistical power by an order of magnitude compared to the $B^0 \rightarrow D^{*-} 3\pi$ decay.

The analysis carried out in this project will be extended to use the full LHCb dataset collected in 2011 and 2012 (Run 1) and between 2015 and 2018 (Run 2). Three BDTs dedicated to selecting $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decays will be used, where the performance of the BDTs will be evaluated using the fitted $B^0 \rightarrow D^{*-} 3\pi$ yields in data. Estimated $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ yields will be calculated from the normalisation mode yields using Eq. 2, allowing the BDT cuts to be optimised specifically for the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decay. With these optimised cuts applied to the $B^0 \rightarrow D^{*-} 3\pi$ data, a final $n_{B^0 \rightarrow D^{*-} 3\pi}$ value will be measured for use in a full Run 1 + Run 2 measurement of $R(D^*)$. This work will be carried out by the author as part of an upcoming Masters project at EPFL in Spring 2021.

Acknowledgment

I am extremely grateful for the constant and motivational support of Dr. Donal Hill. His assistance and dedication throughout the project was far beyond what I could expect. I also thank Prof. Olivier Schneider for offering me the opportunity to work on this project.

7. Annexe

7.1 Theory

Definition of geometrical and kinetic variables

The variables used in this report, especially as training variables of the BDT in section 4.1 are [22]:

- The **own primary vertex** (PV) of a particle is the p-p collision point it comes from (directly or indirectly via a parent particle).
- The **end vertex** (EV) of a particle is the point where it decays.
- The **origin vertex** (OV) of a particle is the the point where it is created, i.e., the EV of its parent particle.
- The **impact parameter** w.r.t. to the PV of a particle is the distance of its straight track (before the magnet) to the PV.
- The **transverse momentum** of a particle is the value of its 3-momentum projected to the direction orthogonal to the proton beam.
- The **DIRA angle** w.r.t. the PV of a particle is the angle between the straight line linking its PV and EV and the direction of its reconstructed 3-momentum.
- The **flight distance** w.r.t. the PV of a particle is the distance between its PV and its EV.

The vertices are measured by the vertex locator.

These physical quantities are represented in [Figure 29](#) in the case of the D^{*-} meson.

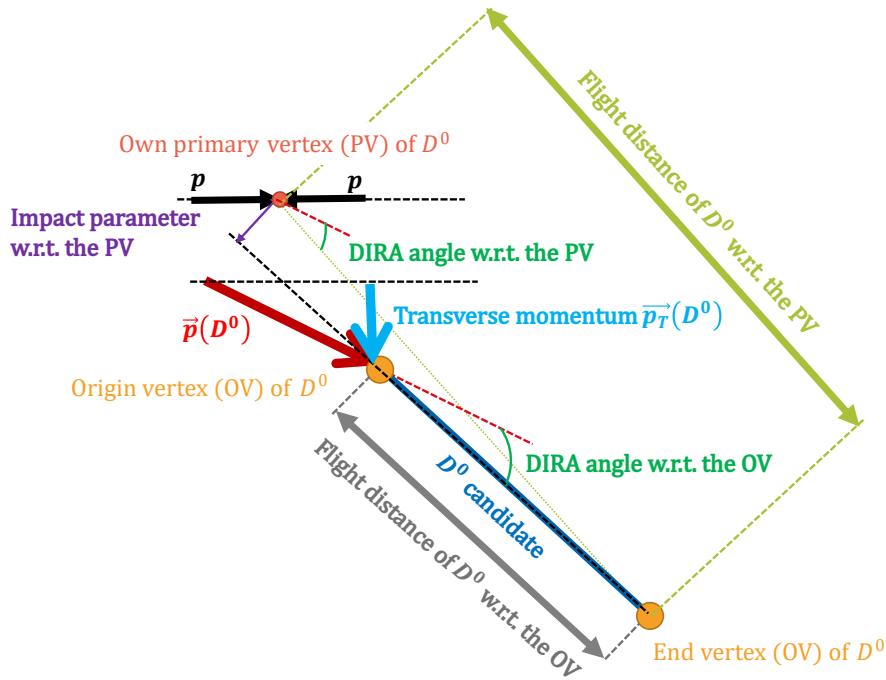


Figure 29. Representation of the 3-momentum, the transverse momentum \vec{p}_T , the Dirac angle θ , the primary vertex, the end vertex, the impact parameter and the flight distance of the D^0 meson.

s Weights

The s Weights are computed using the *hepstats* python library [15], which is based on a published paper [19].

In order to isolate the $B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$ contribution, modelled by a PDF denoted $p_{m(D^*3\pi), D_s^+}$ from the $D^{*-}3\pi$ invariant mass spectrum, modelled by $p_{m(D^*3\pi)}$, the mass of the 3π system is required to be in a ± 50 MeV/ c^2 window centered on the world-average invariant mass of the D_s meson. Additionally, weights, called s Weights, are applied to the $m(D^*3\pi)$ distribution. When assigned to all candidates, these weights are meant to project out the

candidates that does not contain a $D_s^+ \rightarrow 3\pi$ decay. This procedure requires that the variables $m(D^*3\pi)$ and $m(3\pi)$ are independent given that there contain or not a $D_s^+ \rightarrow 3\pi$ decay. In this section, the random variable $m(D^*3\pi)$ is denoted X to point out the fact that the $_s$ Weights do not depend not the choice of the variable X to which they are applied.

The $_s$ Weights $_s w(m_{3\pi})$ are computed using the data and the global PDF obtained from the fit to the $m(3\pi)$ spectrum around the invariant mass of the D_s meson, namely:

$$p_{m(3\pi)} = \frac{n_{D_s^+}}{n_{\text{tot}}} p_{m(3\pi)|D_s^+} + \frac{n_{\cancel{D}_s^+}}{n_{\text{tot}}} p_{m(3\pi)|\cancel{D}_s^+} = \sum_{k \in \{D_s^+, \cancel{D}_s^+\}} \frac{n_k}{n_{\text{tot}}} p_{m(3\pi)|k}, \quad (7)$$

where $p_{m(3\pi)|D_s^+}$ ($p_{m(3\pi)|\cancel{D}_s^+}$) is the (non-) $D_s^+ \rightarrow 3\pi$ conditional PDF and $n_{D_s^+}$ ($n_{\cancel{D}_s^+}$) is its yield. n_{tot} is the total number of candidates. Mathematically speaking, D_s^+ is interpreted as an event.

After computation of the $_s$ Weights, they are applied to every candidate c in the following way:

$$_s p_{X,m(3\pi)}(x[c], y[c]) = _s w(y[c]) \times p_{X,m(3\pi)}(x[c], y[c]), \quad (8)$$

where $x[c]$ ($y[c]$) is the observation of the random variable X ($m(3\pi)$) for the candidate c . $_s p_{X,m(3\pi)}$ denotes the $_s$ Weighted two-dimensional PDF of the variable $(X, m(3\pi))$ and $_s w(y[c])$ is the $_s$ Weight of the candidate c . The PDF of the variable X is obtained by simple integration over the variable $m(3\pi)$:

$$\forall x \in \mathbb{R}, \quad _s p_X(x) = \int_{\mathbb{R}} dy \quad _s w(y) p_{X,m(3\pi)}(x, y) \quad (9a)$$

$$= \int_{\mathbb{R}} dy \sum_{k \in \{D_s^+, \cancel{D}_s^+\}} \quad _s w(y) \frac{n_k}{n_{\text{tot}}} p_{X,m(3\pi)|k}(x, y) \quad (9b)$$

$$= \int_{\mathbb{R}} dy \sum_{k \in \{D_s^+, \cancel{D}_s^+\}} \quad _s w(y) \frac{n_k}{n_{\text{tot}}} p_{X|k}(x) p_{m(3\pi)|k}(y). \quad (9c)$$

To go from equation 9b to 9c, one assumes that the variables X and $m(3\pi)$ are conditionally independent, given D_s^+ or \cancel{D}_s^+ .

The $_s$ Weights, are defined by the fact that the $_s$ Weighted PDF $_s p_X$ must satisfy in average, in any interval I (e.g., bins of a histogram) of the X spectrum:

$$n_{\text{tot}} \int_I dx \quad _s p_X(x) = n_{D_s^+} \int_I dx \quad p_{X|D_s^+}(x). \quad (10)$$

Thanks to this requirement, assigning $_s$ Weights to the observations of the variable X allows to reconstruct the distribution of the variable X for the candidates that contain a $D_s^+ \rightarrow 3\pi$ decay.

The requirement is equivalent to demanding that the relation 10 is satisfied for any interval of size δx , centered in x , where $\delta x \rightarrow 0^+$. Concretely, it is required that:

$$\forall x \in \mathbb{R}, \quad n_{\text{tot}} \quad _s p_X(x) = n_{D_s^+} p_{X|D_s^+}(x). \quad (11)$$

This expression shows that the function $_s p_X(x)$ is not normalised. In order to find the $_s$ Weights, the left-hand side of the equation 11 is written explicitly using the equation 9c:

$$n_{\text{tot}} \quad _s p_X(x) = n_{\text{tot}} \int_{\mathbb{R}} dy \sum_{k \in \{D_s^+, \cancel{D}_s^+\}} \quad _s w(y) \frac{n_k}{n_{\text{tot}}} p_{X|k}(x) p_{m(3\pi)|k}(y) \quad (12a)$$

$$= \sum_{k \in \{D_s^+, \cancel{D}_s^+\}} p_{X|k}(x) n_k \int_{\mathbb{R}} dy \quad p_{m(3\pi)|k}(y) \quad _s w(y) \quad (12b)$$

Hence, the requirement 11 is verified if:

$$\forall k \in \{D_s^+, \cancel{D}_s^+\}, \quad \int_{\mathbb{R}} dy \quad p_{m(3\pi)|k}(y) \quad _s w(y) = \delta_{1k}, \quad (13)$$

where $\delta_{D_s^+, k}$ is the Kronecker symbol, equal to 1 if $k = D_s^+$ or 0 if $k = \bar{D}_s^*$. The $_s$ Weights should depend on $p_{m(3\pi)|D_s^+}$ and $p_{m(3\pi)|\bar{D}_s^*}$. Linear dependencies on these two PDFs are tried:

$$_s w(y) = \frac{\sum_{j \in \{D_s^+, \bar{D}_s^*\}} V_{D_s^+, j} \times p_{m(3\pi)|j}(y)}{n_{\text{tot}} p_{m(3\pi)}(y)}, \quad (14)$$

where $V_{D_s^+, \bullet} = \begin{pmatrix} V_{D_s^+, D_s^+} \\ V_{D_s^+, \bar{D}_s^*} \end{pmatrix}$ is a 2-dimensional vector yet to be defined. Equation 13 becomes:

$$\forall k \in \{D_s^+, \bar{D}_s^*\}, \sum_{j \in \{D_s^+, \bar{D}_s^*\}} V_{D_s^+, j} \int_{\mathbb{R}} dy \frac{p_{m(3\pi)|k}(y) p_{m(3\pi)|j}(y)}{n_{\text{tot}} p_{m(3\pi)}(y)} = \delta_{1k}. \quad (15)$$

$V_{D_s^+, \bullet}$ is promoted to a 2×2 matrix $V = \begin{pmatrix} V_{D_s^+, \bullet} & V_{\bar{D}_s^*, \bullet} \\ V_{D_s^+, \bar{D}_s^*} & V_{\bar{D}_s^*, \bar{D}_s^*} \end{pmatrix} = \begin{pmatrix} V_{D_s^+, D_s^+} & V_{D_s^+, \bar{D}_s^*} \\ V_{\bar{D}_s^*, D_s^*} & V_{\bar{D}_s^*, \bar{D}_s^*} \end{pmatrix}$. The updated requirement that must verify V is:

$$\forall l, k \in \{D_s^+, \bar{D}_s^*\}, \sum_{j \in \{D_s^+, \bar{D}_s^*\}} V_{l,j} \int_{\mathbb{R}} dy \frac{p_{m(3\pi)|k}(y) p_{m(3\pi)|j}(y)}{n_{\text{tot}} p_{m(3\pi)}(y)} = \delta_{l,k}. \quad (16)$$

It means that the matrix product of V and the 2×2 matrix $(\int_{\mathbb{R}} dy \frac{p_{m(3\pi)|k}(y) p_{m(3\pi)|j}(y)}{n_{\text{tot}} p_{m(3\pi)}(y)})_{(k,j) \in (\{D_s^+, \bar{D}_s^*\})^2}$ is required to be equal to the identity matrix.

Although the equation 16 with $k = \bar{D}_s^*$ is not used, this allows to define the matrix V and its sub-column $V_{D_s^+, \bullet}$ via the inverse matrix of V :

$$\forall k, j \in \{D_s^+, \bar{D}_s^*\}, V_{k,j}^{-1} = \int_{\mathbb{R}} dy \frac{p_{m(3\pi)|k}(y) p_{m(3\pi)|j}(y)}{n_{\text{tot}} p_{m(3\pi)}(y)}. \quad (17)$$

The matrix V defined thereby verifies the requirement 16. Moreover, if one wanted to project out the D_s^+ component, the $V_{\bar{D}_s^*, \bullet}$ column would be used.

In conclusion, the $_s$ Weights that allow to project out $D_s^+ \rightarrow 3\pi$ component of the $X = m(D^*3\pi)$ distribution are defined by:

$$_s w(y) = \boxed{\frac{\left(p_{m(3\pi)|D_s}(y) \quad p_{m(3\pi)|\bar{D}_s^*}(y) \right) \tilde{V}_{D_s^+, \bullet}}{n_{\text{tot}} p_{m(3\pi)}(y)}}, \quad (18)$$

where \tilde{V} is an estimation of the matrix V using the data:

$$\boxed{\tilde{V}_{i,j}^{-1} = \sum_c \frac{p_{m(3\pi)|k}(y[c]) p_{m(3\pi)|j}(y[c])}{n_{\text{tot}} p_{m(3\pi)}(y[c])}}. \quad (19)$$

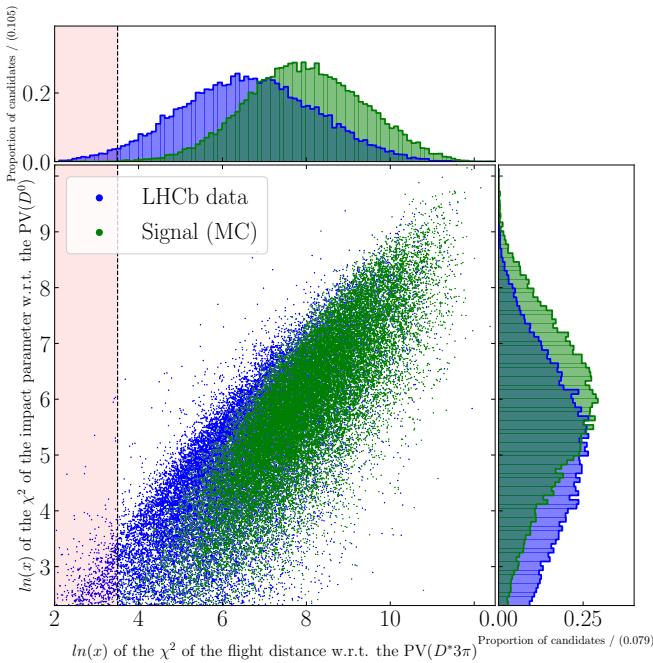
The sum is performed over the candidates c of the data.

Illustration of a multi-variate analysis

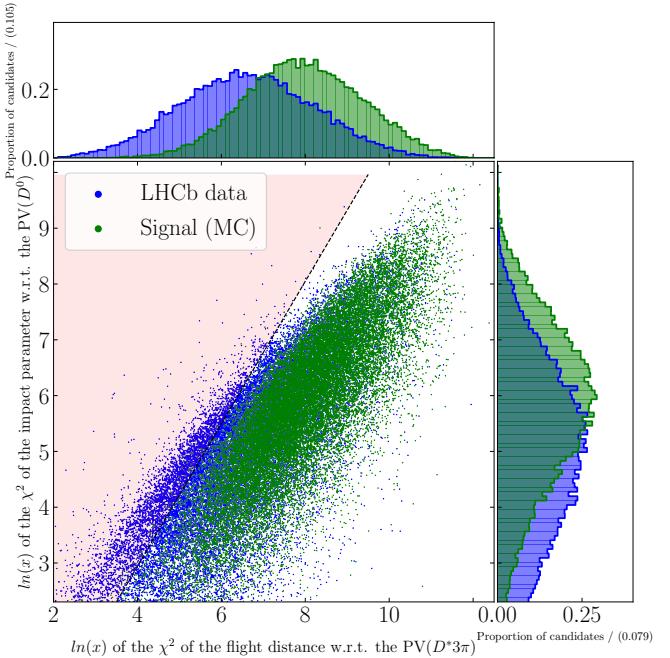
For brevity, in this section, X denotes the $\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV of the $D^*3\pi$ system and Y denotes the $\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV of the D^0 meson.

In order to pre-select $B^0 \rightarrow D^{*-}3\pi$ decays, preliminary cuts are usually applied to one-dimensional variables, such as the offline cuts described in section 2.1. A one-dimensional requirement can be chosen by comparing the LHCb data distribution to the MC distribution of a variable v and removing from the LHCb data the candidates with a variable v belonging to an interval poor in MC data. For instance, the $X > 3.5$ cut could be apply to the X variable, as shown in [Figure 30a](#).

A multi-variate analysis (e.g., the gradient BDT, used in section 4.1) allows to cut on multi-dimensional variables, where new discrepancies between the LHCb data and MC are visible, whereas they were not resolved in the one-dimensional approach. Indeed, [Figure 30b](#) illustrates a two-dimensional cut on (X, Y) , more efficient on genuine signal.



(a) The red region indicates the region that is removed after the $X < 3.5$ cut.



(b) Example of a two-dimensional cut applied to (X, Y) . The red region indicates the region that is removed after cut.

Figure 30. Two-dimensional scatter plot of (X, Y) in equal-sized samples of data (blue) and MC (green). The one-dimensional distributions of X (at the top) and of Y (on the right) are also plotted.

A more efficient approach is the training of a BDT using signal and background samples. The scatter plots of (X, Y) in these two training samples are shown in Figure 31. As expected, the suggested two-dimensional cut on (X, Y) would remove a substantial part of background while retaining most of the candidates belonging to the $B^0 \rightarrow D^{*-} 3\pi$ signal peak. After application of the BDT to the LHCb data, Figure 32 confirms that the optimal $BDT > -1.25$ cut removes most of the candidates that would have been excluded by applying the two-dimensional cut on (X, Y) .

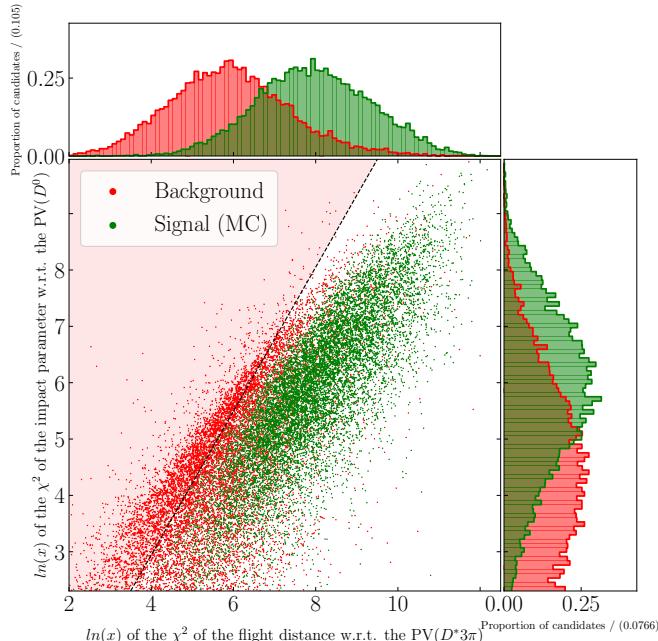


Figure 31. Two-dimensional scatter plot of (X, Y) in equal-sized samples of background (red) and MC (green). The red region indicates the two-dimensional cut of Figure 30b.

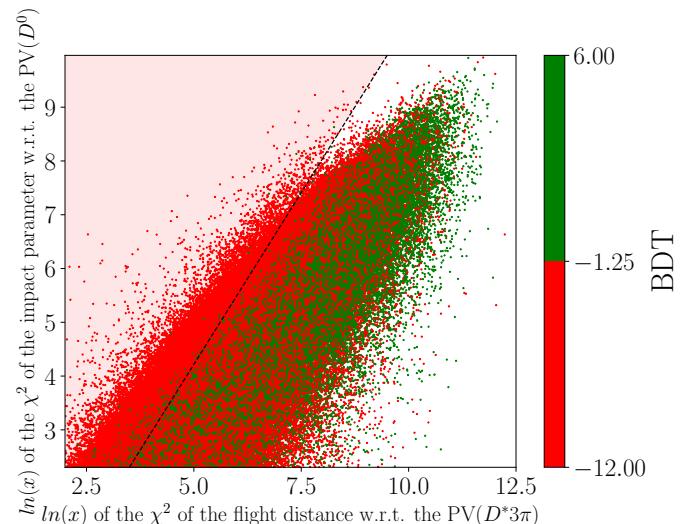


Figure 32. Two-dimensional scatter plot of (X, Y) in the LHCb data. The candidates that are retained after the optimal BDT cut $BDT > -1.25$ are coloured green, while the others are coloured red. The red region indicates the two-dimensional cut of Figure 30b.

Two-sample Kolmogorov-Smirnov test

The two-sample Kolmogorov-Smirnov test allows to check whether two sets of observations are drawn from the same distribution. It is implemented in `scipy.stats` [24].

Two collections of $M \in \mathbb{N}$ ($N \in \mathbb{N}$) independent and identically distributed real-valued random variables $X = (X_1, \dots, X_N)$ ($Y = (Y_1, \dots, Y_N)$) according to a probability distribution p_X (p_Y) are considered. In this report, p_X (p_Y) is the probability distribution of the BDT output of the training (test) sample. It is equivalent to consider the cumulative distribution function (CDF) F_X . The CDF can be estimated from an observation $x = (x_1, \dots, x_M)$ of (X_1, \dots, X_M) by the empirical CDF $\hat{F}_{X,M}[x_1, \dots, x_M] = \hat{F}_{X,M}[x]$:

$$\forall u \in \mathbb{R}, \hat{F}_{X,M}[x](u) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}_{]-\infty, u]}(x_i), \quad (20)$$

where $\mathbb{I}_{]-\infty, u]}$ is the indicator function of the subset $]-\infty, u]$ of \mathbb{R} . $\hat{F}_{X,M}[x](u)$ is the empirical rate of observations of values lower than u .

The two-sample Kolmogorov-Smirnov test is a non-parametric test where a null hypothesis $P_X = P_Y$ is confronted with a two-sided alternative hypothesis $P_X \neq P_Y$. The Kolmogorov-Smirnov statistic of the test is given by $K_{M,N}$:

$$K_{M,N}[X, Y] = \sqrt{\frac{MN}{M+N}} \sup_{u \in \mathbb{R}} |\hat{F}_{X,M}[X](u) - \hat{F}_{Y,N}[Y](u)|, \quad (21)$$

which should be *small* if the null hypothesis is verified. Assuming the null hypothesis, it can be proven that this statistic converges in law to a random variable \mathcal{K} following the Kolmogorov-Smirnov law, whose CDF is given by [20]:

$$\forall u \in \mathbb{R}, \lim_{M,N \rightarrow +\infty} \mathbb{P}(K_{M,N} \leq u) = 2 \sum_{k=1}^{+\infty} (-1)^{k+1} e^{-2k^2 u^2} = \mathbb{P}(\mathcal{K} \leq u) \mathbb{I}_{[0, +\infty[}(u), \quad (22)$$

where \mathbb{P} denotes the probability of the considered probability space. This expression can be used as an approximation for large samples, allowing to compute the p-value $\alpha^*[x, y]$ of the test:

$$\alpha^*[x, y] = \mathbb{P}(\mathcal{K} \geq K_{M,N}[x, y]),$$

which is, assuming the null hypothesis, the probability of observing a Kolmogorov-Smirnov score worse than $K_{M,N}[x, y]$, the score that has been observed. The p-value should be large if there is no statistically significant evidence against the null hypothesis. It can be compared to the restrictive 10% level; if $\alpha^*[x, y] > 10\%$, there is no evidence against the hypothesis that the two samples X and Y follow the same law.

7.2 Results

Distribution of the training variables in the signal and background samples of the BDTs

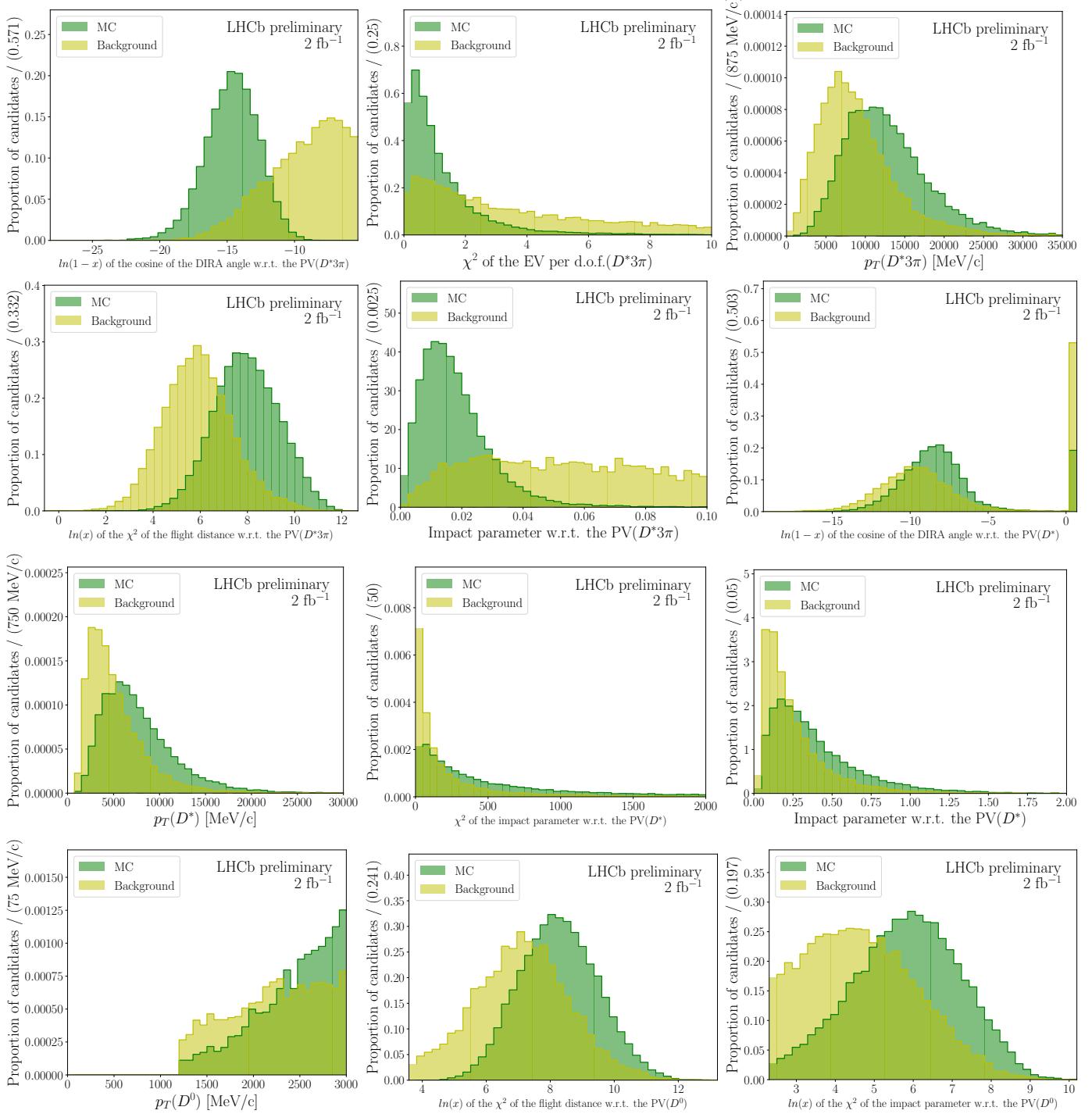


Figure 33. Distribution of the 24 variables used in the training of the BDT (in section 4.1), in the signal and background samples (first part).

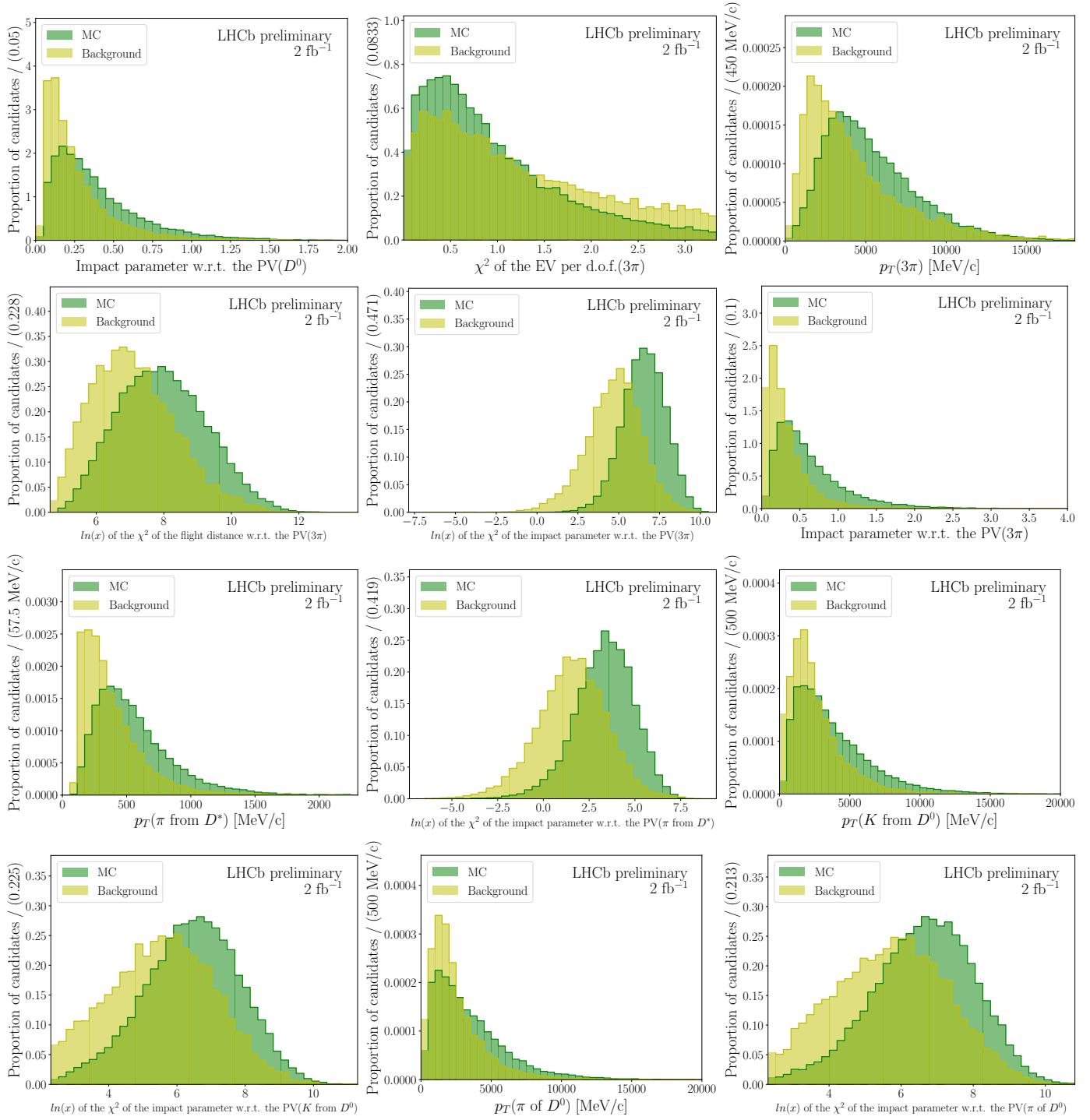


Figure 34. Distribution of the 24 variables used in the training of the BDT (in section 4.1), in the signal and background samples (second part).

Rankings by importance of the variables used in the BDT training

Variable	Importance
Impact parameter w.r.t. the PV($D^*3\pi$)	0.598
$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV($D^*3\pi$)	0.203
$\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV($D^*3\pi$)	0.071
χ^2 of the EV per d.o.f. ($D^*3\pi$)	0.070
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(π from D^*)	0.026
$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV(D^*)	0.007
$p_T(3\pi)$	0.006
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(3π)	0.005
$p_T(\pi$ from D^*)	0.004
$p_T(D^0)$	0.003
Impact parameter w.r.t. the PV(3π)	0.002
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(D^0)	0.002
χ^2 of the EV per d.o.f. (3π)	0.002
$\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV(3π)	0.001

Table 9. The twelve most discriminating variables used in the training of the BDT A , ranked by their importance.

Variable	Importance
Impact parameter w.r.t. the PV($D^*3\pi$)	0.485
$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV($D^*3\pi$)	0.318
χ^2 of the EV per d.o.f. ($D^*3\pi$)	0.065
$\ln(x)$ of the χ^2 of the flight distance w.r.t. the PV($D^*3\pi$)	0.047
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(π from D^*)	0.037
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(3π)	0.013
$\ln(1 - x)$ of the cosine of the DIRA angle w.r.t. the PV(D^*)	0.009
$p_T(3\pi)$	0.006
$p_T(D^0)$	0.005
$p_T(D^*)$	0.003
Impact parameter w.r.t. the PV(3π)	0.003
$p_T(\pi$ from D^*)	0.002
$\ln(x)$ of the χ^2 of the impact parameter w.r.t. the PV(D^0)	0.002
χ^2 of the EV per d.o.f. (3π)	0.002

Table 10. The twelve most discriminating variables used in the training of the BDT B , ranked by their importance.

Fits to the LHCb data with the optimal BDT > -1.25 cut

The result of the fit to the $m(3\pi)$ spectrum in the LHCb data where the optimal cut $BDT > -1.25$ is applied and the sWeights computed from this fit are shown in [Figure 35](#) and [Figure 36](#).

The sWeights are applied to the $m(D^*)$ -constrained $m(D^*3\pi)$ spectrum and the result of the fit is shown in [Figure 37](#) and [Table 11](#).

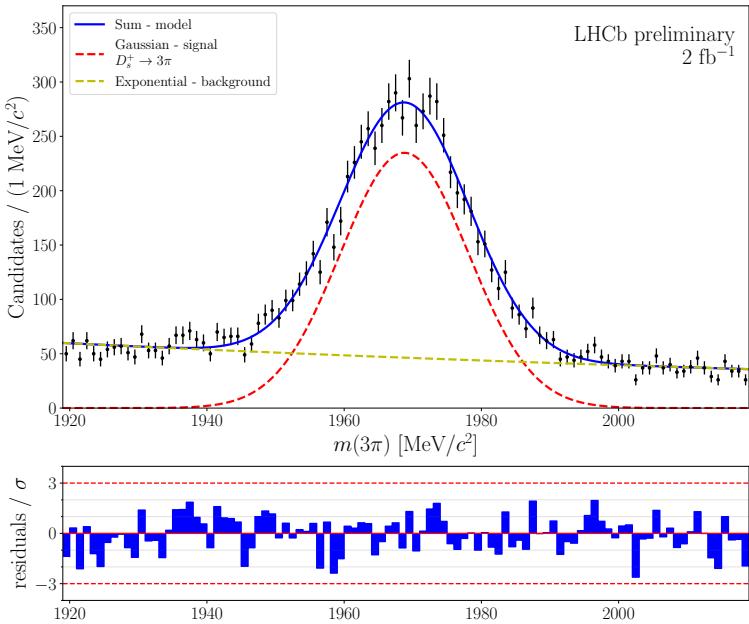


Figure 35. Fit to the $m(3\pi)$ distribution in the LHCb data where the optimal $BDT > -1.25$ cut is applied.

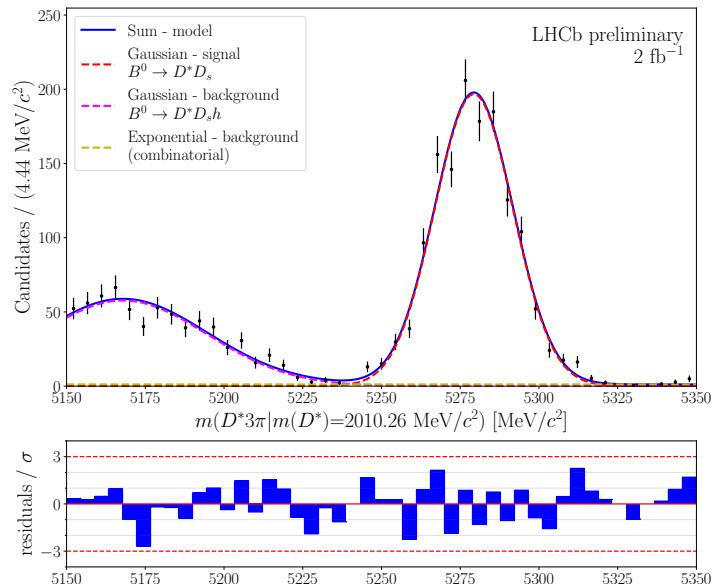


Figure 37. Fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of $_s$ Weighted LHCb data where the optimal $BDT > -1.25$ cut is applied. The $_s$ Weights assigned are derived from the $m(3\pi)$ fit shown in [Figure 35](#).

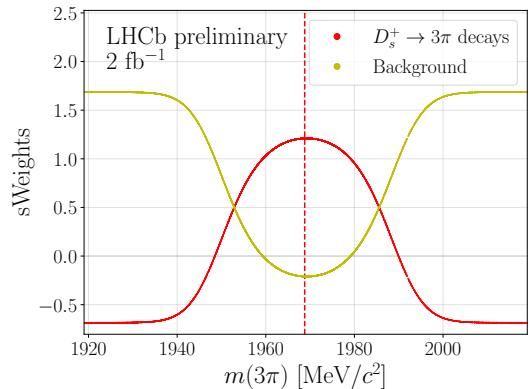


Figure 36. Signal and background $_s$ Weights computed from the fit to the $m(3\pi)$ distribution in the LHCb data where the optimal $BDT > -1.25$ cut is applied. The vertical dashed red line indicates the mass of the D_s^+ meson.

Component	Variable	Fitted value
$B^0 \rightarrow D^{*-}(D_s^+ \rightarrow 3\pi)$	$n_{D^*D_s}$	1381 ± 52
	$\sigma_{D^*D_s}$	12.45 ± 0.5
	$\mu_{D^*D_s}$	5279.3 ± 0.5
Partially reconstructed	$n_{B,D^*D_s h}$	26.7 ± 2.3
	$\sigma_{B,D^*D_s h}$	27 ± 2
	$\mu_{B,D^*D_s h}$	5168 ± 3
Combinatorial	$n_{B,c'}$	53 ± 83
	$\lambda_{B,c'}$	$(-0.0 \pm 4.2) \times 10^{-7}$

Table 11. Result of the fit to the $m(D^*)$ -constrained $m(D^*3\pi)$ distribution of $_s$ Weighted LHCb data where the optimal $BDT > -1.25$ cut is applied. The $_s$ Weights assigned are derived from the $m(3\pi)$ fit shown in [Figure 35](#).

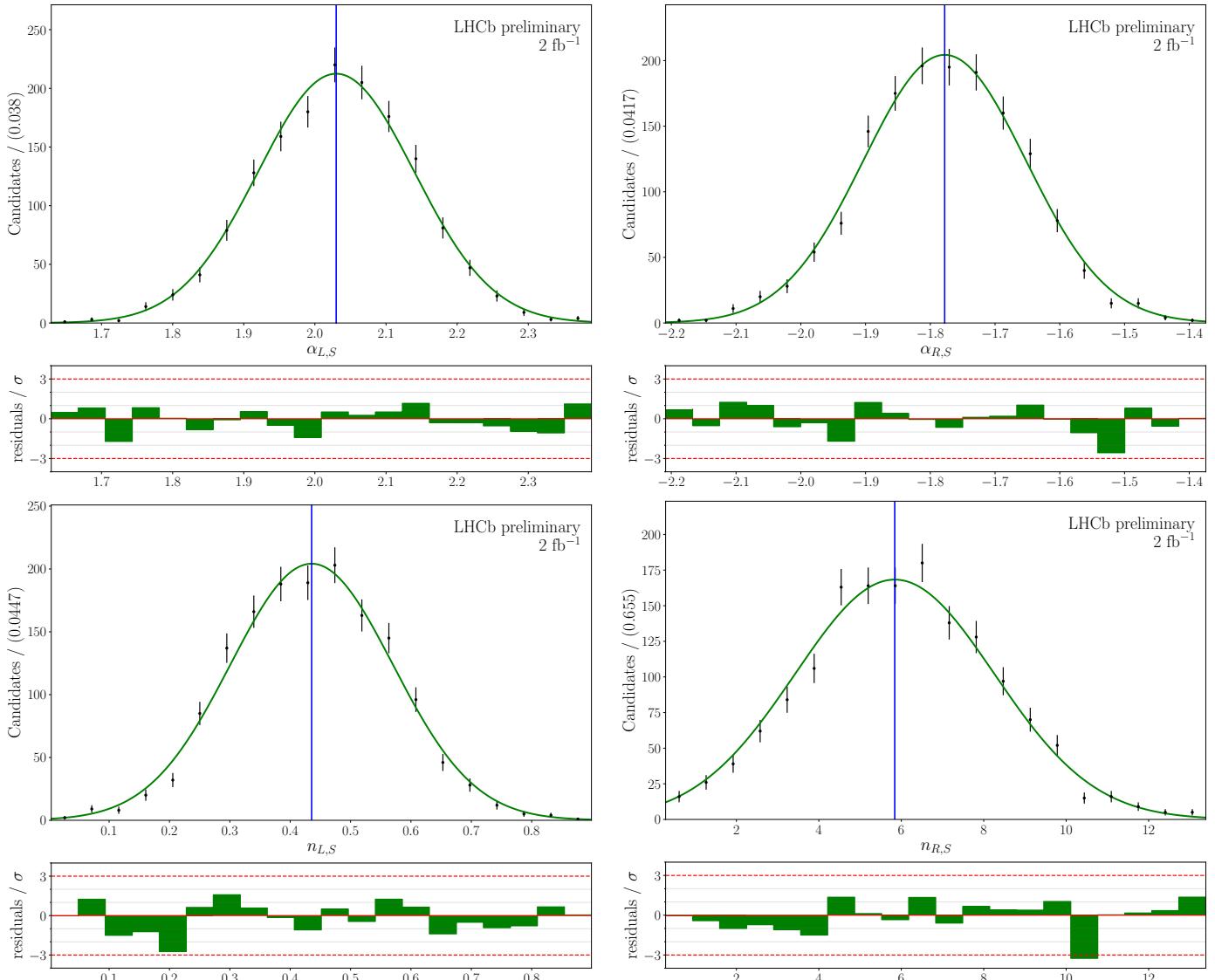
Distribution of the parameters fixed in the fit to $m(D^*3\pi)$ in the $BDT > -1.25$ LHCb data under systematic variations


Figure 38. $\alpha_{L,S}$, $\alpha_{R,S}$, $n_{L,S}$, $n_{R,S}$ distributions obtained by varying the tail parameters of the signal component in section 5 to find the associated systematic uncertainty on $n_{B^0 \rightarrow D^{*-} 3\pi}$. The vertical blue line indicates their nominal value. The curve is the Gaussian function from which the values of the parameters are drawn randomly.

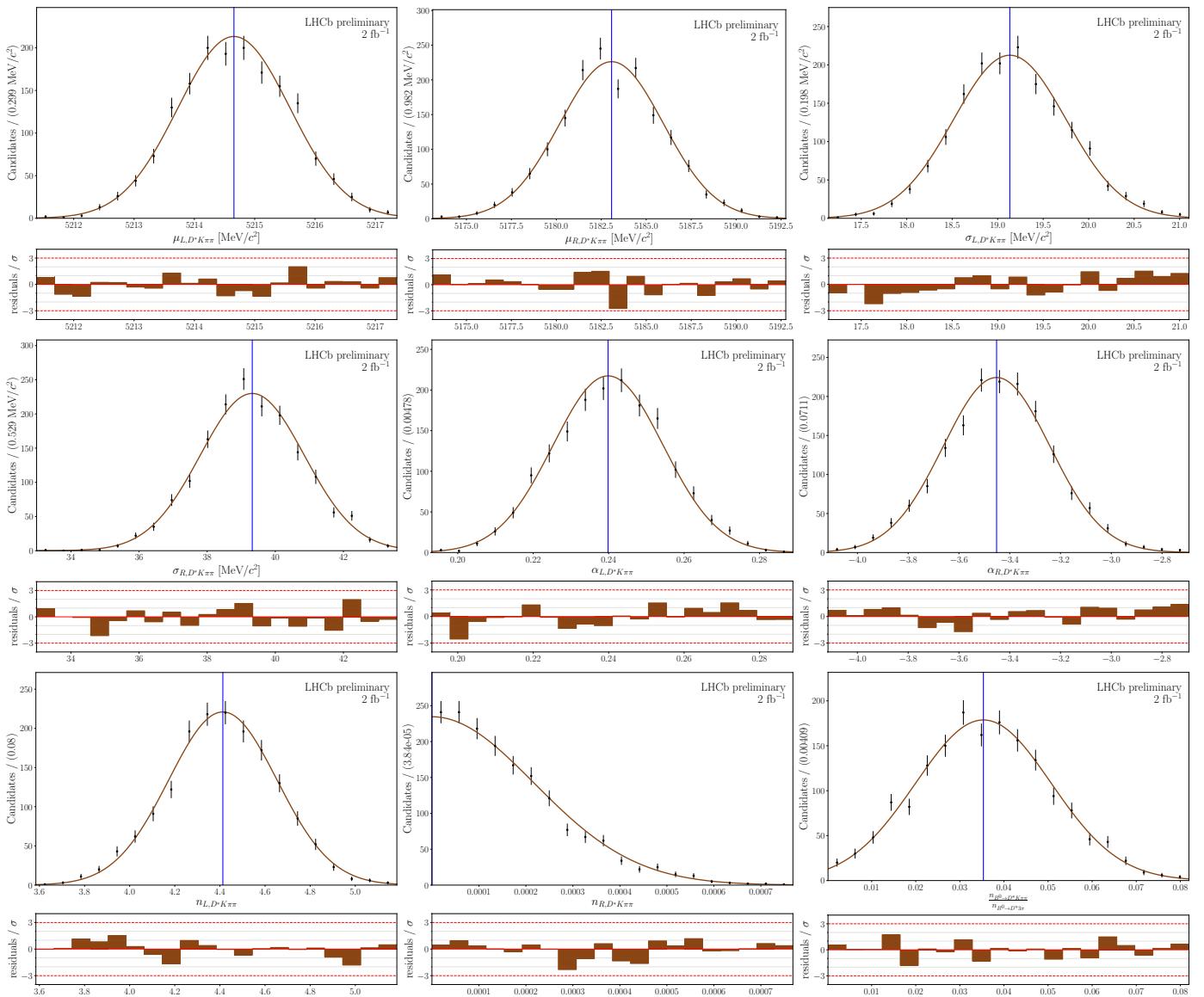


Figure 39. Distributions of the parameters of the $B^0 \rightarrow D^{*-} K^+ \pi^+ \pi^-$ component obtained by varying them to find the associated systematic uncertainty on $n_{B^0 \rightarrow D^{*-} 3\pi}$, as explained in section 5. The vertical blue line indicates their nominal value. The curve is the Gaussian function from which the values of the parameters are drawn randomly.

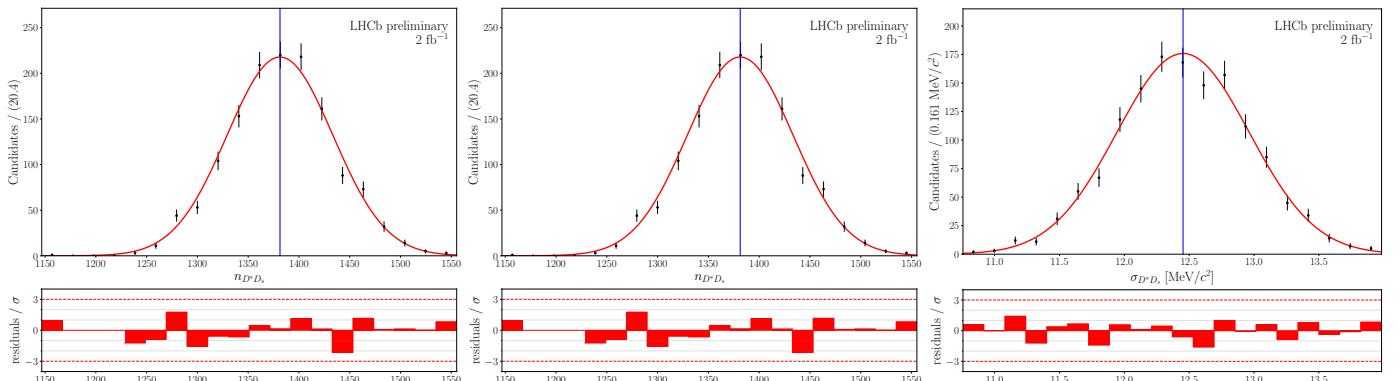


Figure 40. Distributions of the parameters of the $B^0 \rightarrow D^{*-} (D_s^+ \rightarrow 3\pi)$ component obtained by varying them to find the associated systematic uncertainty on $n_{B^0 \rightarrow D^{*-} 3\pi}$, as explained in section 5. The vertical blue line indicates their nominal value. The curve is the Gaussian function from which the values of the parameters are drawn randomly.

7.3 Python codes

The git directory of this project can be accessed from this link: <https://github.com/donalrinho/bd2dst3pi/tree/Anthony>. In particular, It contains:

- the folder *scripts* that contains python functions allowing for instance to load root files, plot 1D and 2D histograms; fit a distribution and plot the result of the fit; train, test and apply a BDT, etc.
- the folder *notebooks* that contains mainly jupyter notebooks, which use the functions in the folder *scripts* in order to analyse the data.

A brief summary of the notebooks and scripts is provided in *notebooks/Summary notebooks and data.ipynb*.

References

- [1] JOHANNES ALBRECHT AND. **The LHCb Trigger System: Present and Future.** *Journal of Physics: Conference Series*, **623**:012003, June 2015. Publisher: IOP Publishing. Available from: <https://doi.org/10.1088/1742-6596/623/1/012003>.
- [2] CERN. **The Large Hadron Collider**, dec 2020. Available from: <https://home.cern/science/accelerators/large-hadron-collider>.
- [3] LHCb COLLABORATION. **LHCb detector performance.** *International Journal of Modern Physics A*, **30**(07):1530022, March 2015. Publisher: World Scientific Publishing Co. Available from: <https://www.worldscientific.com/doi/abs/10.1142/S0217751X15300227>.
- [4] LHCb COLLABORATION. **LHCb: Trigger system**, dec 2020. Available from: <https://lhcb-public.web.cern.ch/en/DataCollection/Triggers2-en.html>.
- [5] LHCb COLLABORATION. **The LHCb trigger system**, dec 2020. Available from: <https://iopscience.iop.org/article/10.1088/1748-0221/9/09/C09015>.
- [6] LHCb COLLABORATION. **LHCb: VELO**, dec 2020. Available from: <https://lhcb-public.web.cern.ch/en/detector/VELO2-en.html>.
- [7] LHCb COLLABORATION, R AAIJ, ET AL. **Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau) / \mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)$.** *Physical Review Letters*, **115**(11):111803, sep 2015. arXiv: 1506.08614. Available from: <http://arxiv.org/abs/1506.08614>.
- [8] LHCb COLLABORATION, R. AAIJ, ET AL. **Measurement of the ratio of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ branching fractions using three-prong τ -lepton decays.** *Physical Review Letters*, **120**(17):171802, apr 2018. arXiv: 1708.08856. Available from: <http://arxiv.org/abs/1708.08856>.
- [9] G. A. COWAN, D. C. CRAIK, AND M. D. NEEDHAM. **RapidSim: An application for the fast simulation of heavy-quark hadron decays.** *Computer Physics Communications*, **214**:239–246, May 2017. Available from: <http://www.sciencedirect.com/science/article/pii/S0010465517300413>.
- [10] JONAS ESCHLE, ALBERT PUIG NAVARRO, RAFAEL SILVA COUTINHO, AND NICOLA SERRA. **zfit: Scalable pythonic fitting.** *SoftwareX*, **11**:100508, January 2020. Available from: <http://www.sciencedirect.com/science/article/pii/S2352711019303851>.
- [11] ANDREI GOLUTVIN. **LHCb: A question of asymmetry**, September 2008. Section: High-energy physics. Available from: <https://cerncourier.com/a/lhcba-question-of-asymmetry/>.
- [12] HFLAV. **Average of $R(D)$ and $R(D^{*+})$ for Spring 2019**, dec 2019. Available from: <https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/RDsDsstar/RDRDs.html>.
- [13] J. L. HODGES. **The significance probability of the smirnov two-sample test.** *Arkiv för Matematik*, **3**(5):469–486, Jan. 1958. Publisher: Institut Mittag-Leffler. Available from: <https://projecteuclid.org/euclid.afm/1485893310>.
- [14] F. JAMES AND M. ROOS. **Minuit - a system for function minimization and analysis of the parameter errors and correlations.** *Computer Physics Communications*, **10**(6):343–367, December 1975. Available from: <http://www.sciencedirect.com/science/article/pii/0010465575900399>.
- [15] MATTHIEU MARINANGELI, EDUARDO RODRIGUES, BRIAN POLLACK, AND JONAS ESCHLE. **scikit-hep/hepstats: 0.3.1**, November 2020. Available from: <https://zenodo.org/record/4185536>.
- [16] EDUARDO PICATOSTE OLLOQUI AND THE LHCb COLLABORATION. **LHCb Preshower(PS) and Scintillating Pad Detector (SPD): Commissioning, calibration, and monitoring.** *Journal of Physics: Conference Series*, **160**:012046, April 2009. Publisher: IOP Publishing. Available from: <https://doi.org/10.1088/1742-6596/160/1/012046>.
- [17] PARTICLE DATA GROUP, P. A. ZYLA, ET AL. **Review of Particle Physics.** *Progress of Theoretical and Experimental Physics*, **2020**(083C01), August 2020. Available from: <https://doi.org/10.1093/ptep/ptaa104>.
- [18] F. PEDREGOSA ET AL. **Scikit-learn: Machine Learning in Python.** *Journal of Machine Learning Research*, **12**:2825–2830, 2011.
- [19] M. PIVK. **sPlot: a quick introduction.** In *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pages 173–177. Imperial College Press, May 2006. Available from: https://www.worldscientific.com/doi/abs/10.1142/9781860948985_0036.
- [20] JOHN W. PRATT AND JEAN D. GIBBONS. **Kolmogorov-Smirnov Two-Sample Tests.** In JOHN W. PRATT AND JEAN D. GIBBONS, editors, *Concepts of Nonparametric Theory*, Springer Series in Statistics, pages 318–344. Springer, 1981. Available from: https://doi.org/10.1007/978-1-4612-5931-2_7.
- [21] TOMASZ SKWARNICKI. **A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances.** thesis, page 106, aug 1986. Available from: <https://inspirehep.net/literature/230779>.
- [22] KRZYSZTOF GRZEGORZ SOBCZAK. **Study of charmless three-body decays of neutral B mesons with the LHCb spectrometer.** PhD thesis, 2011. CERN-THESIS-2011-216, tel-00686906, 2011CLF22202.
- [23] MARÍA VIEITES DÍAZ. **Study of the $B^0 \rightarrow \rho(770)^0 K^*(892)^0$ decay with an amplitude analysis of $B^0 \rightarrow (\pi^+ \pi^-)(K^+ \pi^-)$ decays.** PhD thesis, 2019. Available from: <https://minerva.usc.es/xmlui/handle/10347/18670>.
- [24] PAULI VIRTANEN ET AL. **SciPy 1.0: fundamental algorithms for scientific computing in Python.** *Nature Methods*, **17**(3):261–272, Mar. 2020. Available from: <https://www.nature.com/articles/s41592-019-0686-2>.